



CIMAT

Centro de Investigación en Matemáticas, A.C.

---

# PROPUESTA DE UNA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS MEDIANTE UN ENFOQUE DE VIRTUALIZACIÓN DE DATOS PARA EL REPORTEO DE RENDIMIENTO EN PROGRAMAS DE VISITAS A DOMICILIO A NIÑOS EN EDAD TEMPRANA

**T E S I S**

Que para obtener el grado de  
**Maestro en Ingeniería de Software**

**Presenta**

Héctor Alessandro Milán Ortega

**Director de Tesis:**

Mtro. Carlos Abraham Carballo Monsivais

Autorización de la versión final



## Agradecimientos

A mis padres por todo el apoyo brindado a largo de mi vida y por enseñarme el verdadero valor de las cosas, gracias a ustedes he logrado superarme y ser la persona que soy, tanto en el ámbito personal como en el profesional.

A mi esposa Alejandra, gracias por tu tiempo y paciencia, por estar a mi lado en cada momento, por acompañarme en este difícil trayecto que emprendimos hace 3 años, por brindarme apoyo incondicional en todo momento y nunca apartarte de mi lado. Muchas gracias.

A mis amigos y compañeros por los momentos que compartimos, su compañerismo y apoyo, ustedes volvieron ameno cada día transcurrido en la Maestría.

Agradezco al CIMAT Unidad Zacatecas y a todo su personal por haberme brindado la oportunidad de continuar con mis estudios, por las su atención, conocimientos y enseñanzas a lo largo del programa académico.

Finalmente, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado para la realización de los estudios de maestría y el desarrollo del presente trabajo de tesis.

## Resumen

El reporte operativo es un enfoque muy empleado en agencias implementadoras de modelos de visitas a domicilio a niños en edad temprana en los Estados Unidos de América para generar informes de rendimiento organizacional. Sin embargo, se tienen problemas de tiempo de espera, desempeño e interferencia en las bases de datos operativas. Las herramientas comerciales de inteligencia de negocios ayudan en la mitigación de estos problemas, pero agregan un costo adicional, además de contar con un gran número de funcionalidades que finalmente no son empleadas, incrementando los riesgos de su implementación.

En este trabajo se propone la creación de una herramienta de Inteligencia de Negocios a la medida, capaz de aprovechar al máximo la información obtenida mediante la prestación de servicios de visitas a domicilio, ofreciendo una herramienta alineada a los objetivos de los programas de visitas a domicilio, con una funcionalidad de acuerdo con las necesidades de los usuarios contribuye a mejorar los problemas del reporte operativo. Nuestro enfoque de Virtualización de Datos acompañado de una táctica arquitectural de cacheo de datos tiene un impacto en la disminución de los tiempos de espera y la interferencia en las bases de datos operativas.

Los resultados validados por métodos estadísticos (ANOM, ANOVA, Kruskal-Wallis, Tukey) muestran una mejora en el tiempo empleado para la obtención de un informe de rendimiento (acceso a la información), la media del proceso actual tiene 5.044 segundos empleando el reporte operativo contra nuestra propuesta de una media de 1.187 segundos. Así como una mejora en la interferencia de las bases de datos, considerando lecturas a disco se obtuvo una reducción del 27.49% y escrituras a disco 23.57% respectivamente.

## Abstract

Operational reporting is an approach widely used by agencies that implement home visiting models for early childhood children in the United States of America for reporting organizational performance, resulting in problems such as waiting time, performance, and interference to operational databases. Commercial business intelligence tools are helpful in mitigating these problems, but implying an additional cost, in addition to having many functionalities that are ultimately not used, leading to failure in their implementation.

This work proposes the creation of a tailored business intelligence tool capable of making the most of the information obtained through the provision of home visiting services, offering a tool aligned to the objectives of home visiting programs, the functionality expected by the users according to their needs and the mitigation of the problems present in the operational reporting by proposing a Data Virtualization approach accompanied by a Data Cache architectural tactic to reduce waiting times and interference to operational databases.

The results show an improvement in the time used to obtain a performance report (access to information), going from an average of 5.044 seconds using operational reporting to 1.187 seconds. As well as a reduction in interference in the databases, going from an average of 13.46 disk reads and 98.1 disk writes per performance report using operational reporting to 3.70 disk reads and 23.12 disk writes per performance report using the tool proposed here.

## Índice

|  |     |
|--|-----|
| Agradecimientos .....                    | i   |
| Resumen.....                             | ii  |
| Abstract .....                           | iii |
| Introducción .....                       | 1   |
| Capítulo 1. Antecedentes .....           | 4   |
| 1.1.    Marco Teórico .....              | 4   |
| 1.1.1.    Visitas a domicilio .....      | 4   |
| 1.1.2.    Medición de rendimiento.....   | 6   |
| 1.1.3.    Inteligencia de Negocios ..... | 7   |
| 1.1.4.    Virtualización de datos .....  | 17  |
| 1.2.    Planteamiento del problema.....  | 20  |
| 1.3.    Objetivos .....                  | 22  |
| 1.3.1.    Objetivo general .....         | 22  |
| 1.3.2.    Objetivos específicos.....     | 22  |
| 1.4.    Hipótesis .....                  | 23  |
| 1.5.    Justificación .....              | 23  |
| 1.6.    Alcances y limitaciones .....    | 24  |
| Capítulo 2. Estado del arte .....        | 26  |

|   |   |    |
|---|---|----|
| 2.1.  | Revisión sistemática de la literatura .....                         | 26 |
| 2.1.1.  | Planificación de la revisión .....                                  | 27 |
| 2.1.2.  | Ejecución de la revisión .....                                      | 29 |
| 2.1.3.  | Resultados .....  | 32 |
| 2.2.  | Selección de tecnología.....  | 44 |
| 2.2.1.  | Lenguaje de programación .....                                      | 45 |
| 2.2.2.  | Framework de desarrollo.....  | 46 |
| 2.2.3.  | Base de datos .....   | 46 |
| 2.2.4.  | Arquitectura.....   | 46 |
| Capítulo 3. Metodología para el desarrollo del trabajo de investigación ..... |   | 49 |
| 3.1.  | Ejecución de la metodología para el desarrollo de la tesis .....    | 49 |
| 3.2.  | Métodos Estadísticos usados para validación de la herramienta ..... | 51 |
| 3.2.1.  | Prueba de Grubbs (Detección de Valores Atípicos) .....              | 53 |
| 3.2.2.  | Prueba de Anderson-Darling (Prueba de Normalidad).....              | 55 |
| 3.2.3.  | Prueba de Igualdad de Varianzas (Varianzas Homogéneas).....         | 56 |
| 3.2.4.  | Prueba de Análisis de Medias (Prueba ANOM).....                     | 59 |
| 3.2.5.  | Prueba ANOVA de un solo factor.....                                 | 62 |
| 3.2.6.  | Prueba de Suma de Rank Kruskal-Wallis para K poblaciones.....       | 67 |
| Capítulo 4. Propuesta de herramienta de inteligencia de negocios.....         |   | 69 |

|   |  |     |
|---|--|-----|
| 4.1.  | Desarrollo de la propuesta .....                                 | 69  |
| 4.1.1.  | Identificación de objetivos, preguntas y métricas a emplear..... | 70  |
| 4.1.2.  | Generación de modelo de la base de datos.....                    | 92  |
| 4.1.3.  | Generación de diseño arquitectural .....                         | 94  |
| 4.1.4.  | Desarrollo de la herramienta .....                               | 98  |
| Capítulo 5. Experimentación y Resultados..... |  | 116 |
| 5.1.  | Planificación del experimento controlado .....                   | 117 |
| 5.1.1.  | Participantes del experimento controlado .....                   | 118 |
| 5.1.2.  | Enfoques del experimento controlado .....                        | 119 |
| 5.1.3.  | Preparación del ambiente de pruebas experimental .....           | 121 |
| 5.1.4.  | Consideraciones sobre la muestra del experimento .....           | 121 |
| 5.1.5.  | Determinación de las Métricas del Experimento.....               | 124 |
| 5.2.  | Ejecución del experimento controlado.....                        | 125 |
| 5.3.  | Análisis e interpretación de resultados del experimento.....     | 126 |
| 5.3.1.  | Resultados del tiempo de respuesta.....                          | 127 |
| 5.3.2.  | Resultados de lecturas a disco .....                             | 132 |
| 5.3.3.  | Resultados de escrituras a disco .....                           | 138 |
| 5.3.4.  | Resultados del uso de procesador.....                            | 145 |
| 5.3.5.  | Resultados del uso de memoria .....                              | 150 |



|   |     |
|---|-----|
| Capítulo 6. Conclusiones y trabajo futuro ..... | 157 |
| 6.1.    Conclusiones .....                      | 157 |
| 6.2.    Contribuciones .....                    | 159 |
| 6.3.    Trabajo futuro .....                    | 159 |
| Referencias.....                                | 161 |

## Índice de Figuras

|   |    |
|---|----|
| Figura 1. Estructura general del programa MIECHV .....  | 5  |
| Figura 2. Sistema de inteligencia de negocios representado como una caja negra.....   | 8  |
| Figura 3. Transformaciones programadas mediante el uso de ETL.....  | 9  |
| Figura 4. Extracción de datos mediante el uso de ELT.....   | 10 |
| Figura 5. Arquitectura tradicional de inteligencia de negocios.....   | 14 |
| Figura 6. Implementación de un enfoque de virtualización de datos .....   | 17 |
| Figura 7. Fases y actividades realizadas durante la ejecución de la revisión sistemática de la literatura .....   | 26 |
| Figura 8. Resultados de la selección de estudios primarios.....   | 31 |
| Figura 9. Áreas abordadas en los estudios primarios de acuerdo con el número de menciones. ....   | 34 |
| Figura 10. 4 perspectivas del cuadro de mando integral.....   | 36 |
| Figura 11. Relación entre los beneficios directos e indirectos en la implementación de inteligencia de negocios para la gestión del rendimiento en una organización. .... | 44 |
| Figura 12. Lenguajes de programación más empleados al mes de abril de 2020 .....  | 45 |
| Figura 13. Metodología seguida para el desarrollo de la tesis .....   | 49 |
| Figura 14. Métodos Estadísticos empleados para validación de la herramienta.....  | 52 |
| Figura 15. Método empleado para el desarrollo de la herramienta propuesta .....   | 70 |
| Figura 16. Jerarquía del modelo GQM .....   | 73 |
| Figura 17. Notación empleada en un diagrama de flujo. ....  | 77 |
| Figura 18. Algoritmo empleado para obtener las métricas de la pregunta 1.....   | 78 |

|   |     |
|---|-----|
| Figura 19. Algoritmo empleado para obtener las métricas de la pregunta 2.....   | 80  |
| Figura 20. Algoritmo empleado para obtener las métricas de la pregunta 3.....   | 82  |
| Figura 21. Algoritmo empleado para obtener las métricas de la pregunta 4.....   | 85  |
| Figura 22. Algoritmo empleado para obtener las métricas de la pregunta 5.....   | 87  |
| Figura 23. Algoritmo empleado para obtener las métricas de la pregunta 6.....   | 90  |
| Figura 24. Modelo de la base de datos a emplear .....   | 93  |
| Figura 25. Niveles de abstracción del modelo C4. ....   | 95  |
| Figura 26. Diagrama de contexto para el sistema de inteligencia de negocios.....  | 96  |
| Figura 27. Diagrama de contenedores para el sistema de inteligencia de negocios. ....   | 96  |
| Figura 28. Diagrama de componentes para el sistema de inteligencia de negocios. ....  | 97  |
| Figura 29. Diagrama de código para el modelo de datos del sistema de inteligencia de negocios.<br>.....                                 | 98  |
| Figura 30. Ciclo de vida empleado para el desarrollo de la herramienta.....   | 99  |
| Figura 31. Estado actual de la herramienta empleada para la visualización de rendimiento<br>mediante el sistema Visit Tracker Web. .... | 100 |
| Figura 32. Inicio de sesión.....  | 101 |
| Figura 33. Filtros empleados por el panel de visualización.....   | 103 |
| Figura 34. Gráficas correspondientes al rendimiento obtenido por una agencia implementadora.<br>.....                                   | 104 |
| Figura 35. Visualización de la información empleada para el cálculo de una métrica. ....  | 105 |
| Figura 36. Configuración básica de una gráfica. ....  | 106 |
| Figura 37. Configuración de una métrica de rendimiento. ....  | 107 |

|  |     |
|--|-----|
| Figura 38. Estructura de una métrica de rendimiento. ....  | 108 |
| Figura 39. Diseño correspondiente a la primera métrica de rendimiento. ....  | 109 |
| Figura 40. Diseño correspondiente a la segunda métrica de rendimiento. ....  | 110 |
| Figura 41. Diseño correspondiente a la tercera métrica de rendimiento. ....  | 111 |
| Figura 42. Diseño correspondiente a la cuarta métrica de rendimiento. ....   | 112 |
| Figura 43. Diseño correspondiente a la quinta métrica de rendimiento. ....   | 113 |
| Figura 44. Diseño correspondiente a la sexta métrica de rendimiento. ....  | 114 |
| Figura 45. Vista general del panel de visualización de información. ....   | 115 |
| Figura 46. Fases del experimento controlado. ....  | 117 |
| Figura 47. Enfoque 1, reporte operativo sin inteligencia de negocios. ....   | 119 |
| Figura 48. Enfoque 2, arquitectura de almacén de datos centralizada. ....  | 120 |
| Figura 49. Procedimiento empleado para la ejecución del experimento controlado. ....   | 124 |
| Figura 50. Gráfica de valores atípicos sobre el tiempo empleado en segundos para la obtención de informes de rendimiento. ....     | 127 |
| Figura 51. Prueba de normalidad aplicada sobre el tiempo empleado en segundos para la obtención de un informe de rendimiento. .... | 128 |
| Figura 52. ANOM normal unidireccional del tiempo empleado en segundos para la obtención de un informe de rendimiento. ....         | 129 |
| Figura 53. Prueba de igualdad de varianzas sobre el tiempo empleado en segundos para la obtención de informes de rendimiento. .... | 129 |
| Figura 54. Análisis ANOVA sobre el tiempo empleado en segundos para la obtención de informes de rendimiento. ....                  | 130 |

|   |     |
|---|-----|
| Figura 55. Prueba de Kruskal-Wallis sobre el tiempo empleado en segundos para la obtención de informes de rendimiento. ....   | 130 |
| Figura 56. Comparaciones en parejas de Tukey sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.....                                     | 130 |
| Figura 57. Gráfica de intervalos sobre el tiempo empleado en segundos para la obtención de informes de rendimiento. ....  | 131 |
| Figura 58. Gráfica de valores atípicos sobre el número de lecturas a disco tras la obtención de un informe de rendimiento.....  | 132 |
| Figura 59. Gráfica de valores atípicos sobre el número de lecturas a disco tras la obtención de un informe de rendimiento (Tras la eliminación del valor atípico). .... | 133 |
| Figura 60. Prueba de normalidad aplicada sobre el total de lecturas a disco tras la obtención de un informe de rendimiento.....   | 134 |
| Figura 61. ANOM normal unidireccional del número de lecturas a disco tras la obtención de un informe de rendimiento.....  | 135 |
| Figura 62. Prueba de igualdad de varianzas sobre las lecturas a disco tras la obtención de informes de rendimiento. ....  | 136 |
| Figura 63. Análisis ANOVA sobre el total de lecturas a disco tras la obtención de informes de rendimiento. ....   | 136 |
| Figura 64. Prueba de Kruskal-Wallis sobre el total de lecturas a disco tras la obtención de informes de rendimiento. ....   | 136 |
| Figura 65. Comparaciones en parejas de Tukey sobre el total de lecturas a disco tras la obtención de informes de rendimiento.....                                       | 137 |

|  |     |
|--|-----|
| Figura 66. Gráfica de intervalos sobre el total de lecturas a disco tras la obtención de un informe de rendimiento. ....   | 138 |
| Figura 67. Gráfica de valores atípicos sobre el número de escrituras a disco tras la obtención de un informe de rendimiento. ....  | 139 |
| Figura 68. Gráfica de valores atípicos sobre el número de escrituras a disco tras la obtención de un informe de rendimiento (Tras la eliminación del valor atípico)..... | 140 |
| Figura 69. Prueba de normalidad aplicada sobre el total de escrituras a disco tras la obtención de un informe de rendimiento. ....                                       | 141 |
| Figura 70. ANOM normal unidireccional del número de escrituras a disco tras la obtención de un informe de rendimiento.....   | 142 |
| Figura 71. Prueba de igualdad de varianzas sobre las escrituras a disco tras la obtención de un informe de rendimiento.....  | 142 |
| Figura 72. Análisis ANOVA sobre el total de escrituras a disco tras la obtención de un informe de rendimiento. ....  | 143 |
| Figura 73. Prueba de Kruskal-Wallis sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.....   | 143 |
| Figura 74. Comparaciones en parejas de Tukey sobre el total de escrituras a disco tras la obtención de un informe de rendimiento. ....                                   | 143 |
| Figura 75. Gráfica de intervalos sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.....  | 145 |
| Figura 76. Gráfica de valores atípicos sobre el uso de procesador durante la obtención de un informe de rendimiento.....   | 146 |

|  |     |
|--|-----|
| Figura 77. Prueba de normalidad aplicada sobre el uso de procesador durante la obtención de un informe de rendimiento.....               | 146 |
| Figura 78. ANOM normal unidireccional del uso de procesador durante la obtención de un informe de rendimiento.....                       | 147 |
| Figura 79. Prueba de igualdad de varianzas sobre el uso de procesador durante la obtención de un informe de rendimiento.....             | 148 |
| Figura 80. Análisis ANOVA sobre el uso de procesador durante la obtención de un informe de rendimiento. ....                             | 148 |
| Figura 81. Prueba de Kruskal-Wallis sobre el uso de procesador durante la obtención de un informe de rendimiento.....                    | 148 |
| Figura 82. Comparaciones en parejas de Tukey sobre el uso de procesador durante la obtención de un informe de rendimiento. ....          | 149 |
| Figura 83. Gráfica de intervalos sobre el uso de procesador durante la obtención de un informe de rendimiento. ....                      | 150 |
| Figura 84. Gráfica de valores atípicos sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento. ....     | 151 |
| Figura 85. Prueba de normalidad aplicada sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento. ....   | 152 |
| Figura 86. ANOM normal unidireccional de la memoria disponible en MBytes durante la obtención de un informe de rendimiento. ....         | 153 |
| Figura 87. Prueba de igualdad de varianzas sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento. .... | 154 |

|  |     |
|--|-----|
| Figura 88. Análisis ANOVA sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.....                     | 154 |
| Figura 89. Prueba de Kruskal-Wallis sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento. ....          | 154 |
| Figura 90. Comparaciones en parejas de Tukey sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento. .... | 155 |
| Figura 91. Gráfica de intervalos sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento. ....             | 156 |



## Índice de Tablas

|  |    |
|--|----|
| Tabla 1. Modelos de prestación de servicios que cumplen con los criterios propuestos por el Departamento de Salud y Prestación de Servicios de los Estados Unidos de América ..... | 6  |
| Tabla 2. Arquitecturas de inteligencia de negocios más populares.....  | 11 |
| Tabla 3. Cadenas de búsqueda resultantes .....   | 29 |
| Tabla 4. Criterios de inclusión y exclusión empleados para la Revisión Sistemática de la Literatura.....   | 30 |
| Tabla 5. Procedimiento de selección de estudios primarios .....  | 30 |
| Tabla 6. Preguntas de evaluación de calidad aplicadas a los estudios primarios.....  | 32 |
| Tabla 7. Estudios primarios seleccionados para la Revisión Sistemática de la Literatura .....  | 32 |
| Tabla 8. Beneficios del uso de inteligencia de negocios para la gestión del rendimiento en una organización .....  | 43 |
| Tabla 9. Notación empleada en las fórmulas para el estadístico unilateral y bilateral. ....  | 54 |
| Tabla 10. Notación empleada en las fórmulas para los valores p para el estadístico de la prueba de Grubbs.....   | 55 |
| Tabla 11. Notación empleada en la fórmula para el estadístico de Anderson-Darling. ....  | 56 |
| Tabla 12. Notación empleada en la fórmula para la prueba de Bartlett.....  | 58 |
| Tabla 13. Notación empleada en la fórmula para la prueba de Levene. ....   | 59 |
| Tabla 14. Notación empleada en la fórmula para la media del ANOM.....  | 61 |
| Tabla 15. Notación empleada en la fórmula de la mediana principal ANOM.....  | 61 |
| Tabla 16. Notación empleada en la fórmula de desviación estándar ANOM. ....  | 61 |
| Tabla 17. Notación empleada en la fórmula de desviación estándar agrupada ANOM.....  | 62 |

|  |     |
|--|-----|
| Tabla 18. Notación empleada en la fórmula del modelo ajustado ANOVA. ....  | 63  |
| Tabla 19. Notación empleada en la fórmula de suma de los cuadrados del ANOVA. ....                               | 64  |
| Tabla 20. Notación empleada en la fórmula de Grados de Libertad del ANOVA. ....                                  | 65  |
| Tabla 21. Notación empleada en la fórmula de Cuadrados Medios del ANOVA. ....                                    | 65  |
| Tabla 22. Notación empleada en la fórmula del Estadístico F del ANOVA. ....                                      | 66  |
| Tabla 23. Notación empleada en la fórmula para Comparaciones Múltiples con el Método de Tukey. ....              | 66  |
| Tabla 24. Áreas de referencia establecidas por MIECHV. ....  | 71  |
| Tabla 25. Objetivo de medición identificado mediante el empleo de un enfoque GQM. ....                           | 74  |
| Tabla 26. Categorización de los datos para la pregunta 1. ....   | 79  |
| Tabla 27. Categorización de los datos para la pregunta 2. ....   | 81  |
| Tabla 28. Categorización de los datos para la pregunta 3. ....   | 83  |
| Tabla 29. Categorización de los datos para la pregunta 4. ....   | 86  |
| Tabla 30. Categorización de los datos para la pregunta 5. ....   | 88  |
| Tabla 31. Categorización de los datos para la pregunta 6. ....   | 91  |
| Tabla 32. Informes de rendimiento solicitados por Family Focus, Inc durante el año 2020.                         | 118 |
| Tabla 33. Características del ambiente de prueba. ....   | 121 |
| Tabla 34. Prerrequisitos de software. ....   | 121 |
| Tabla 35. Parámetros empleados para la obtención de los informes de rendimiento durante la experimentación. .... | 122 |
| Tabla 36. Métricas recolectadas tras la ejecución de un informe de rendimiento. ....                             | 124 |

|   |     |
|---|-----|
| Tabla 37. Estructura del conjunto de datos a emplear para el análisis e interpretación de resultados..... | 125 |
|---|-----|



## **Introducción**

Un programa de visitas a domicilio a niños en edad temprana es aquel que provee de soporte para mejorar la relación padre e hijo, contando con los siguientes objetivos (Health Resources & Services Administration, 2020):

- Mejorar la salud de la madre y el niño.
- Prevenir abuso y negligencia infantil.
- Alentar una crianza positiva.
- Promover el desarrollo del niño y la preparación para la escuela.

Estudios han examinado los beneficios que las visitas a domicilio tienen tanto en padres como en hijos, sugiriendo que aquellos que implementan este tipo de servicios tienden a tener mejores resultados en la obtención de los objetivos mencionados arriba (Elkan, R; et al, 2000) (Sweet & Appelbaum, 2004). La efectividad que presentan este tipo de servicios es medida periódicamente, respondiendo generalmente a tres preguntas:

- ¿Qué servicios fueron provistos?
- ¿A quién fueron provistos los servicios?
- ¿Se obtuvieron los resultados esperados?

Para responder las preguntas anteriores se requiere que un programa sea capaz de recolectar información histórica relacionada con la prestación de sus servicios, la cual será de ayuda para medir y mejorar la efectividad de estos. Por lo anterior, el uso de una herramienta de visualización de datos se vuelve un elemento crucial para un programa de visitas a domicilio a niños en edad temprana, facilitando la visualización y análisis de los resultados obtenidos, lo que eventualmente conducirá a la obtención de los resultados esperados. A pesar de esto, existe

una falta de herramientas que apoyen en el proceso de visualización e interpretación de los datos recolectados mediante la prestación de este tipo de servicios, resultando generalmente en reportes tabulares difíciles de interpretar, imposibilitando la comparación entre los resultados obtenidos y los resultados esperados, limitando la capacidad de toma de decisiones y por ende la mejora de un programa.

En este trabajo de investigación se presenta el desarrollo de una herramienta de Inteligencia de Negocios diseñada para dar soporte a la visualización de resultados obtenidos durante un periodo anual de prestación de servicios de visitas a domicilio. Esta herramienta pretende mejorar la legibilidad de los reportes presentados tradicionalmente, ser de ayuda para la gestión de rendimiento, definiendo éste como el grado en que una organización es capaz de brindar un servicio cumpliendo con los resultados esperados en un período determinado de tiempo. Para así funcionar como un elemento clave durante la toma de decisiones en un programa de visitas a domicilio a niños en edad temprana.

Este trabajo de investigación se compone de 7 capítulos, mismos que se detallan a continuación:

- **Capítulo 1. Antecedentes.** Presenta el marco teórico, mismo que es de ayuda para la comprensión de los términos empleados a través del trabajo de investigación. También se menciona la problemática a abordar, se presentan el objetivo general y los objetivos específicos, así como la justificación del trabajo de investigación.
- **Capítulo 2. Estado del arte.** Resume la situación actual del uso de herramientas de Inteligencia de Negocios para la gestión de rendimiento, junto

con las herramientas más famosas encontradas en el mercado y sus respectivas características.

- **Capítulo 3. Metodología para el trabajo de investigación.** Presenta las actividades realizadas para la generación de la herramienta de Inteligencia de Negocios propuesta. Incluyendo los métodos estadísticos aplicados en el Capítulo de Resultados.
- **Capítulo 4. Propuesta de la herramienta de inteligencia de negocios.** Describe cada uno de los elementos presentes en la herramienta de Inteligencia de Negocios propuesta, mostrando el diseño propuesto y la funcionalidad ofrecida.
- **Capítulo 5. Experimentación y resultados.** Presenta la planificación, ejecución y obtención de resultados del experimento controlado empleado para validar la herramienta de inteligencia de negocios.
- **Capítulo 6. Conclusiones y trabajo futuro.** Presenta las conclusiones, recomendaciones y trabajo futuro, lo anterior a partir del presente trabajo de investigación.

## **Capítulo 1. Antecedentes**

En este capítulo se presenta el marco teórico correspondiente a este trabajo de investigación, introduciendo los conceptos clave para la realización de este. Posteriormente se detalla el planteamiento del problema a abordar, los objetivos y su justificación.

### **1.1. Marco Teórico**

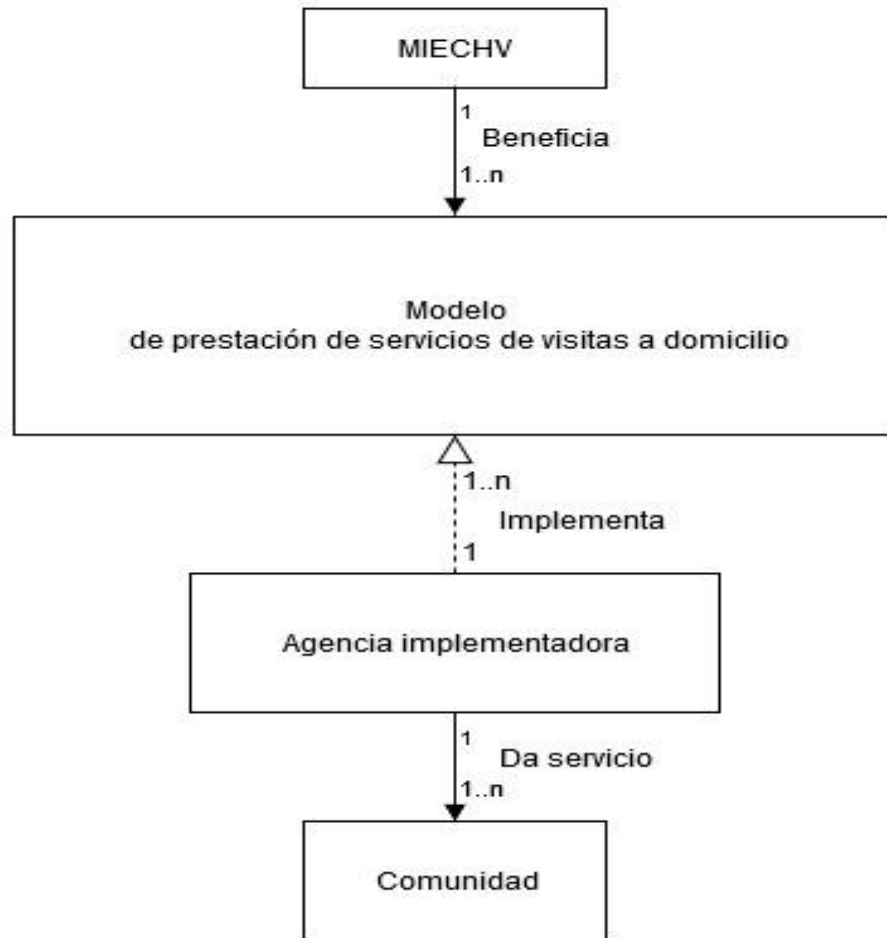
#### **1.1.1. Visitas a domicilio**

Visitas a Domicilio (Del inglés: Home Visiting) es un término empleado en los Estados Unidos de América para referirse a un tipo de servicio provisto de manera estructurada por un profesional entrenado. Comúnmente la prestación de servicios se realiza mediante visitas a domicilio periódicas a tutores con niños prenatales y/o niños desde el nacimiento hasta que alcanzan la edad para ingresar al kínder (regularmente a los 5 años), comúnmente a esta población de niños se le conoce como niños en edad temprana.

En el año 2010, es lanzado un programa federal llamado MIECHV (Del inglés: Maternal, Infant and Early Childhood Home Visiting), diseñado para mejorar las vidas de niños y familias en riesgo a través de visitas a domicilio con base en evidencia. MIECHV beneficia anualmente a estados y territorios dentro de los Estados Unidos de América, teniendo éstos la facilidad de contratar servicios a través de agencias locales implementadoras, las cuales prestarán servicio, atendiendo las necesidades específicas de una comunidad, enfocándose en mejorar el bienestar de las familias y niños dentro de la misma. Las agencias pueden elegir entre una variedad de modelos de prestación de servicios considerando las características y necesidades de la comunidad, siendo estos modelos evaluados periódicamente con el objetivo



de determinar su efectividad en la prestación de servicios. En la Figura 1 se presenta la estructura general del programa MIECHV.



*Figura 1. Estructura general del programa MIECHV*

En el año 2009 es lanzado HomVEE (Del inglés: Home Visiting Evidence of Effectiveness), conducido mediante la ejecución de una revisión de la literatura sobre los estudios presentes relacionados a las visitas a domicilio, y diseñado para proveer información sobre qué modelos de visitas a domicilio cuentan con la evidencia necesaria de efectividad de

acuerdo con lo propuesto por el Departamento de Salud y Servicios Humanos de los Estados Unidos de América. De acuerdo con (U.S. Department of Health & Human Services, 2020), los modelos de la Tabla 1 demuestran ser efectivos en cuanto a la prestación de servicios.

*Tabla 1. Modelos de prestación de servicios que cumplen con los criterios propuestos por el Departamento de Salud y Prestación de Servicios de los Estados Unidos de América*

| Identificador | Nombre del modelo  |
|---------------|--|
| 1             | Attachment and Biobehavioral Catch-Up (ABC) Intervention     |
| 2             | Child FIRST  |
| 3             | Health Families America                                      |
| 4             | Durham Connects/Family Connects                              |
| 5             | Home Instruction for Parents of Preschool Youngsters         |
| 6             | Early Head Start – Home-Based Option                         |
| 7             | Maternal Early Childhood Sustained Home Visiting Program     |
| 8             | Early Intervention Program for Adolescent Mothers            |
| 9             | Minding the Baby   |
| 10            | Early Start (New Zealand)                                    |
| 11            | Nurse-Family Partnership                                     |
| 12            | Family Check-Up for Children                                 |
| 13            | Parents as Teachers  |
| 14            | Family Spirit  |
| 15            | Play and Learning Strategies – Infant                        |
| 16            | Health Access Nurturing Development Services (HANDS) Program |
| 17            | SafeCare Augmented   |
| 18            | Healthy Beginnings   |
| 19            | Maternal Infant Health Program (MIHP)                        |

### **1.1.2. Medición de rendimiento**

Para el correcto entendimiento del significado que el término “Medición del rendimiento” tiene, es necesario definir el significado de rendimiento. De acuerdo con (Grüning, 2002) el rendimiento es la habilidad de una compañía para lograr sus objetivos. El diccionario Cambridge define rendimiento como el acto de hacer algo, así mismo menciona que el término rendimiento se refiere a que tan bien una actividad o trabajo es realizado. El

rendimiento puede ser medido a través de un conjunto de objetivos, un periodo de tiempo y un objeto de comparación (Hauber, 2003). Para determinar el grado en que un objetivo es logrado o no, se requiere entonces de una cuantificación en función del tiempo, y teniendo como base el objeto de comparación en sí. Es por esto que el término “Rendimiento” puede ser definido con base en el contexto en que se encuentre, una empresa dedicada al maquinado podría centrarse en el costo de un proceso, mientras que una agencia de automóviles fije sus objetivos en la venta de unidades.

Dado el contexto bajo el cual se presenta este trabajo de investigación, definiremos rendimiento como:

*“El grado en que una entidad es capaz de prestar un servicio cumpliendo con los resultados esperados por la misma y en un periodo de tiempo específico”.*

La medición de rendimientos es entonces el proceso mediante el cual una entidad cuantifica, analiza y comunica el rendimiento obtenido por la misma durante un periodo de tiempo específico, permitiendo así la comparación entre el resultado esperado y el resultado obtenido.

### **1.1.3. Inteligencia de Negocios**

La primera referencia al término Inteligencia de Negocios es presentada por (Luhn, 1958), definiendo “Negocios” como “una colección de actividades realizadas para cualquier propósito ya sea ciencia, tecnología, comercio, industria, leyes, gobierno, defensa, etc.” e “Inteligencia” como “la capacidad de comprender las interrelaciones de los hechos presentados de tal manera que guíen la acción hacia una meta deseada”. Definiendo un sistema de inteligencia de negocios como un sistema automático para difundir información a las diversas

secciones de cualquier organización, proporcionando medios para la difusión selectiva de información de acuerdo con sus requisitos.

Se puede definir Inteligencia de Negocios como un conjunto de metodologías, procesos, arquitecturas y tecnologías que transforman los datos sin procesar en información significativa y útil (Evelson, 2008).

Un sistema de inteligencia de negocios es una solución desarrollada para dar soporte y mejorar el proceso de toma de decisiones en una organización (Van der Lans, 2012).

Los usuarios de un sistema de inteligencia de negocios deberían verlo como una caja negra, empleando una interfaz de usuario para visualizar la información mediante una serie de paneles de información y pudiendo interactuar con los mismos a fin de obtener la información buscada mediante una serie de filtros o funcionalidades provistas por el sistema. Teniendo conocimiento del origen de los datos que visualizan, pero, sin conocer el formato de estos ni la manera en que son obtenidos (Figura 2).

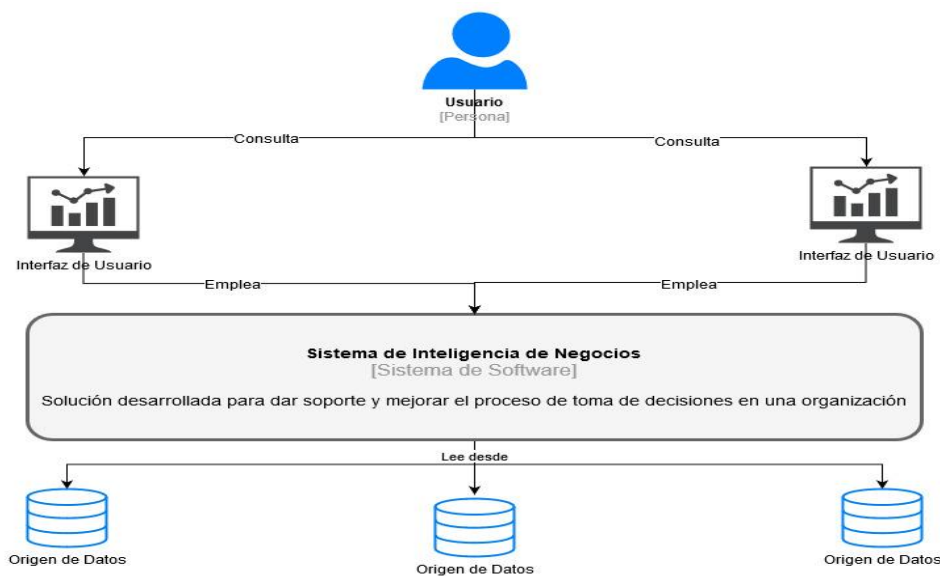


Figura 2. Sistema de inteligencia de negocios representado como una caja negra

### 1.1.3.1. ETL

ETL (Extracción Transformación Carga) es una técnica de integración de datos empleada principalmente en almacenes de datos. Mediante el uso de ETL la información puede ser copiada desde varios orígenes de datos, la integración de datos es realizada mediante un proceso de copiado desde los mismos, para posteriormente unir, transformar, limpiar y almacenar los datos obtenidos. Cada una de las operaciones a ser realizadas durante el proceso de extracción, transformación y carga es almacenada en un script, siendo éste ejecutado periódicamente, obteniendo únicamente la información nueva o modificada desde la última vez que se ejecutó el proceso, a este tipo de operaciones se les conoce como transformaciones programadas (Figura 3).



Figura 3. Transformaciones programadas mediante el uso de ETL

### 1.1.3.2. *ELT*

A diferencia de ETL, la técnica ELT (Extracción Carga Transformación) comienza con el proceso de extracción y carga de datos hacia un almacén de datos destino, una vez realizado esto, se ejecuta un segundo script encargado de realizar la transformación de los datos ya cargados y su posterior carga al mismo almacén de datos. Una característica principal de las herramientas que emplean ELT como técnica de extracción de datos, es que hacen uso de las características propias de las bases de datos para realizar el proceso de extracción y carga, delegando la mayor parte de las tareas a las mismas. En la Figura 4 se muestra el proceso básico empleado por herramientas ELT.

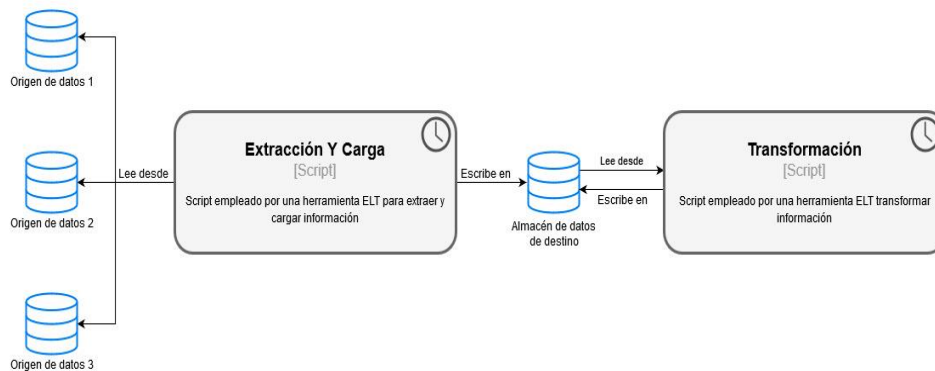


Figura 4. Extracción de datos mediante el uso de ELT

### 1.1.3.3. *Arquitectura de inteligencia de negocios*

Se define arquitectura de software como el conjunto de estructuras necesarias para razonar acerca del software en sí, comprendida por elementos de software, relaciones entre estos y sus propiedades (Bass, Clements, & Kazman, 2012). De acuerdo con (Van der Lans, 2012), una arquitectura de inteligencia de negocios es un conjunto de guías de diseño, descripciones y prescripciones de integración de datos empleadas en la construcción de una

solución de inteligencia de negocios. Cada arquitectura trata de hacer frente a diversos tipos de problemas relacionados principalmente a rendimiento, disponibilidad o consistencia de datos, empleando para esto diversos tipos de técnicas de extracción de datos. Aunque tradicionalmente a estas arquitecturas se les conoce como arquitecturas de almacén de datos (Del inglés: Data Warehouse Architecture), este término tiende a ser algo confuso, ya que algunas de las arquitecturas presentes en la actualidad no cuentan con un almacén de datos, es por esto que se emplea el término arquitectura de inteligencia de negocios.

Existen diversas maneras de desarrollar una solución de inteligencia de negocios, una organización tiene la libertad de seleccionar una solución conocida u optar por una nueva, una de las elecciones cruciales a realizar radica en la arquitectura de software a emplear. En la Tabla 2 se muestran las arquitecturas de inteligencia de negocios más populares de acuerdo con (Van der Lans, 2012).

*Tabla 2. Arquitecturas de inteligencia de negocios más populares*

| Nombre                                  | Descripción   |
|---|---|
| Data Warehouse Bus Architecture (DWBA)  | También conocido como el modelo Kimball, consiste en un área intermedia conocida como Staging Area y un conjunto de espacios de almacenamiento de datos asociados a áreas específicas, conocidos como Data Marts. |
| Corporate Information Factory           | También conocido como el modelo Inmon, consiste en un almacenamiento de datos operacional, un almacén de datos o Datawarehouse y, de ser necesario, un conjunto de Data Marts.                                    |
| Centralized Data Warehouse Architecture | La mayoría de los reportes son obtenidos mediante el uso de un almacén de datos centralizado.   |
| Federated Architecture                  | Todos los reportes y herramientas de análisis están ligados a una capa de federación de datos, la cual se encarga de obtener la   |

| Nombre                 | Descripción  |
|------------------------|--|
| Data Delivery Platform | Basada en Virtualización de Datos, una de sus principales características es que permite desacoplamiento entre los orígenes de datos y los consumidores de los mismos. Puede ser empleada como parte de otra arquitectura o tomar otra arquitectura como un componente más dentro de la misma. |

#### *1.1.3.4. Ventajas de los enfoques de inteligencia de negocios tradicionales*

Para el correcto funcionamiento de una herramienta de reporte se requiere de la consulta de información proveniente de un conjunto de datos recolectados de sistemas en producción, comúnmente almacenados en bases de datos. Estas bases de datos pueden ser consultadas directamente, pero esto conlleva una serie de problemas, entre los cuales destacan la consistencia de los datos y el desempeño obtenido al emplear un origen de datos que se encuentra ocupado constantemente por un sistema en funcionamiento. Para solventar estos problemas, los enfoques tradicionales de inteligencia de negocios hacen uso de almacenes de datos, diseñados específicamente para propósitos de reporte y analítica. Un almacén de datos (Del inglés: Data Warehouse) es un conjunto de datos integrados no volátiles, variantes en tiempo y orientados a temas o áreas de interés. El empleo de almacenes de datos para fines de reporte proporciona entonces las siguientes ventajas con respecto al uso directo de bases de datos de producción:

1. **Integración.** Los datos provenientes de diversos sistemas en producción son integrados en un solo almacén de datos, permitiendo así a una herramienta de reporte el uso de varios orígenes de datos.

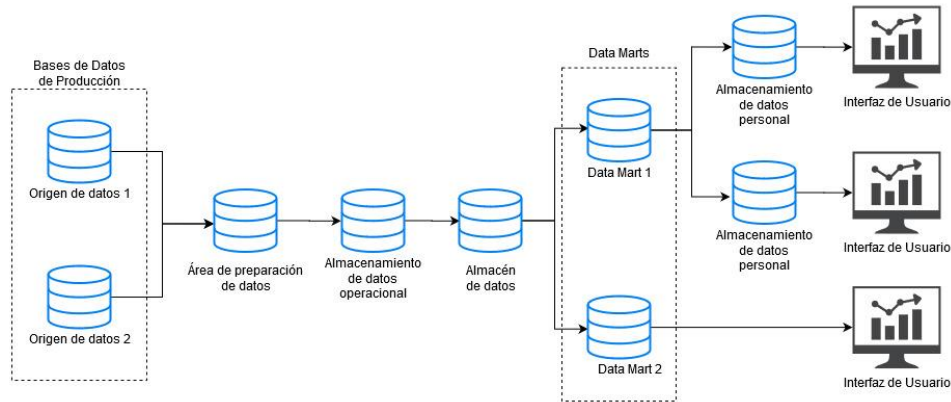


2. **Consistencia.** El empleo de scripts de transformación reduce la aparición de datos defectuosos mediante un proceso de limpieza de estos, ayudando a prevenir la aparición de inconsistencia en los datos almacenados tras ser extraídos y transformados.
3. **Desempeño.** Un almacén de datos puede ser diseñado y optimizado (mediante el uso de diversas técnicas como la de-normalización de datos) para fines de consulta de información, lo cual conlleva a mejorar en el desempeño de las consultas realizadas.
4. **Interferencia.** Debido a que las consultas realizadas por las herramientas de reporte se ejecutan en un almacén de datos, se reduce la interferencia a las bases de datos de producción. Acotándose únicamente al proceso de copiado de datos desde producción hacia un almacén de datos, lo cual comúnmente se realiza cuando los sistemas de producción presentan menos uso por parte de los usuarios de este.

#### ***1.1.3.5. Desventajas de los enfoques de inteligencia de negocios tradicionales***

Los enfoques tradicionales de inteligencia de negocios tienen como característica el empleo de múltiples bases de datos con propósitos distintos, como el uso de áreas de almacenamiento temporal, orígenes de datos operacionales, almacenes de datos, data marts y almacenes de datos personales, lo anterior con el objetivo de disminuir la carga en las bases de datos de origen, conocidas como bases de datos de producción. En la Figura 5 se presenta un

ejemplo de arquitectura de inteligencia de negocios empleando un enfoque tradicional, siendo muy notoria la gran cantidad de elementos presentes en la misma.



*Figura 5. Arquitectura tradicional de inteligencia de negocios*

Lo mencionado anteriormente tiene como resultado un impacto negativo en los siguientes aspectos:

- **Escalabilidad.** Debido a la gran cantidad de elementos presentes en una arquitectura con un enfoque tradicional, escalar la misma se vuelve un proceso complejo debido a los cambios que conlleva éste en cada uno de los elementos, lo anterior debido a que no existe un punto único de mantenimiento.
- **Mantenibilidad.** Debido al número de elementos presentes, mantener alguno de estos puede acarrear problemas en los elementos dependientes del mismo, siendo necesaria realizar una adecuación a todos y cada uno de ellos.
- **Disponibilidad.** Debido al proceso por el que la información debe de pasar antes de ser consultada por el usuario final, y al uso de scripts programados

periódicamente, el acceso a la información en tiempo real se vuelve una tarea prácticamente imposible.

De acuerdo con (Van der Lans, 2012), existen 6 desventajas asociadas a los enfoques de inteligencia de negocios tradicionales, la cuales son presentadas a continuación:

1. **Duplicación de datos.** Debido al número de elementos presentes, en donde los datos son almacenados, se produce una gran cantidad de datos duplicados. Los datos presentes en un almacén de datos personal son la copia de una parte de los datos presentes en data marts, que a su vez son una réplica de una parte de los datos presentes en el almacén de datos general, y así sucesivamente hasta llegar a las bases de datos de producción u orígenes de datos.
2. **Especificaciones de metadatos no compartidas.** La especificación de metadatos se refiere a la especificación de términos y relaciones entre las tablas presentes en una base de datos tradicional, estas especificaciones permiten a una herramienta de inteligencia de negocios volver las definiciones de reportes más independientes de los orígenes de datos. Debido al empleo de múltiples puntos finales de información y al uso de diversas herramientas por cada uno de ellos, el empleo de especificaciones se vuelve particular a cada uno de ellos, por lo cual no existe una definición compartida, provocando variaciones y por ende inconsistencia en los resultados entregados por cada reporte presente en la arquitectura en general.
3. **Flexibilidad limitada.** El uso de este tipo de enfoques dificulta el empleo de múltiples tecnologías, ligándose regularmente a un tipo de tecnologías en

concreto. Un ejemplo es el uso de bases de datos relacionales, una vez que se opta por este enfoque se vuelve muy complicada la migración hacia nuevos tipos de almacenamiento, como las bases de datos no relacionales.

4. **Decremento de la calidad de los datos.** El empleo de múltiples copias de datos conlleva a múltiples tareas de transformación de datos, lo cual a su vez implica un riesgo en la calidad de los datos. En otras palabras, entre más copias de datos existan en una arquitectura de inteligencia de negocios, más riesgo hay de que se presente un problema de inconsistencia de datos, conllevando a un decremento en la calidad de estos.
5. **Soporte limitado para reporte operativo.** Una de las características del reporte operativo es el uso de datos en tiempo real en el proceso de toma de decisiones. Con la existencia de múltiples copias y la ejecución periódica de scripts de transformación, el acceso a reportes que presenten información en tiempo real se vuelve una tarea imposible de realizar mediante un enfoque tradicional.
6. **Soporte limitado para el reporte de datos no estructurados y datos externos.** Los datos presentes en los almacenes de datos provienen comúnmente de bases de datos de sistemas en producción, raramente éstos provienen de fuentes no estructuradas o externas, cuando esta situación se presenta, la respuesta más común es la de estructurar estos datos primero para después ser almacenados, lo cual sugiere un esfuerzo adicional al no ofrecer la

flexibilidad para adaptarse a este tipo de escenarios, siendo necesaria la ejecución de una o más tareas adicionales antes de poder emplear estos datos.

#### 1.1.4. Virtualización de datos

La virtualización de datos es una tecnología que proporciona un único punto de vista de un conjunto heterogéneo de bases de datos y archivos, lo que permite que una aplicación recupere y manipule datos sin requerir detalles técnicos sobre los mismos, cómo se formatea la información en la fuente de origen o dónde está ubicada físicamente esta fuente. También se puede definir virtualización de datos como una forma de integración de datos para mostrarlos como un solo dato y permitiendo acceder a los mismos de una forma muy apegada a una consulta en tiempo real.

Para la correcta implementación de virtualización de datos se requiere el desarrollo de una capa intermedia que permita a una aplicación ocultar los detalles técnicos relacionados a los datos que le sean solicitados, de tal modo que cada parte interesada en la consulta de información acceda a la misma como si solo de una fuente de datos se tratara (Figura 6).

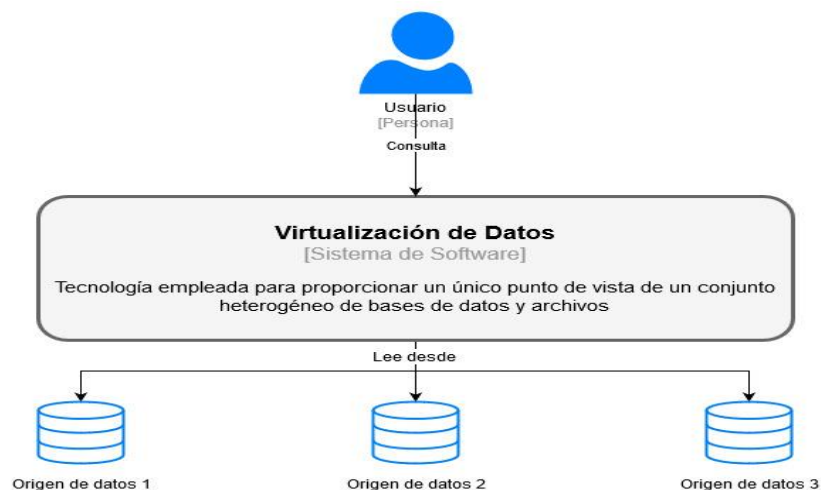


Figura 6. Implementación de un enfoque de virtualización de datos

#### ***1.1.4.1. Ventajas del uso de virtualización de datos***

De acuerdo con (Van der Lans, 2012), las ventajas que presenta el uso de virtualización de datos se agrupan en las siguientes categorías:

- 1. Desacoplamiento.** El empleo de una capa intermedia permite que los orígenes de datos a consultar estén desacoplados de las partes que los consultan, esto significa que las partes interesadas, por ejemplo, los usuarios de una aplicación no tienen conocimiento de la ubicación física ni de la tecnología empleada por los orígenes de datos a consultar. Adicionalmente, la adición o eliminación de orígenes de datos se vuelve más sencilla ya que esta se maneja únicamente en la capa intermedia, volviendo un sistema más flexible y facilitando la migración hacia nuevos orígenes de datos sin que las partes interesadas se vean afectadas por estos cambios.
- 2. Especificaciones de metadatos compartidas.** Las especificaciones de metadatos pueden ser almacenadas y gestionadas desde la capa intermedia, permitiendo que estas puedan ser compartidas con todas las partes interesadas en el uso de una herramienta o sistema de virtualización de datos. Lo anterior beneficia la consistencia de los reportes, vuelve más fácil la tarea de mantenimiento de especificaciones e incrementa la velocidad de desarrollo de reportes debido a que las especificaciones pueden ser reutilizadas.
- 3. Transformaciones bajo demanda.** Un sistema de virtualización de datos es capaz de transformar e integrar datos en el momento en que son accedidos, permitiendo así la consulta de datos operativos o en tiempo real.

Adicionalmente, debido a que existe la posibilidad de mezclar múltiples orígenes de datos, se pueden mezclar datos operativos provenientes de bases de datos de producción con datos históricos provenientes de almacenes de datos.

#### ***1.1.4.2. Desventajas del uso de virtualización de datos***

El empleo de virtualización de datos conlleva también ciertas desventajas, listadas a continuación:

1. **Capa extra de software.** El empleo de virtualización de datos implica el uso de una capa intermedia entre el origen de datos y la parte interesada en hacer uso de estos, lo cual requiere de software adicional, implicando la adición de tiempo de procesamiento debido al uso de esta capa adicional.
2. **Procesamiento repetido de transformaciones.** Con un enfoque de inteligencia de negocios tradicional, una vez que los datos han sido transformados mediante alguna de las técnicas mencionadas con anterioridad, estos son almacenados para posteriormente ser consultados sin la necesidad de pasar nuevamente por un proceso de transformación. Con el uso de virtualización de datos, debido a que las transformaciones se realizan en el momento en que los datos son requeridos, lo cual se conoce como transformaciones bajo demanda, estas mismas son realizadas repetidamente.
3. **Desempeño.** Un empleo de capas de software adicional en conjunto con un procesamiento repetido de transformaciones puede causar problemas de desempeño, motivo por el cual se considera que el empleo de virtualización de datos debe ser puesto a discusión teniendo en cuenta la cantidad de información

a procesar. Siendo esta la principal causa por la que este tipo de enfoques aún no es tan empleado, adicionalmente, esta preocupación se centra en el hecho de que el procesamiento y el uso de memoria volátil son costosos, lo cual conlleva a evaluar estos dos aspectos. Actualmente la capacidad de procesamiento de un equipo de cómputo es cada vez mayor, pronosticando que esta capacidad siga aumentando, del mismo modo el costo del uso de memoria volátil sigue disminuyendo, aplicándose incluso hardware de uso específico para el procesamiento y limpieza de datos.

## **1.2. Planteamiento del problema**

En la actualidad existen una serie de sistemas de seguimiento de casos empleados por las agencias implementadoras tanto para dar seguimiento a niños y sus familias como para reportar sus rendimientos a través de reportes prediseñados (Athena Software, 2020) (DataKeeper Technologies, 2020).

El reporte operativo sigue siendo un enfoque muy común empleado para la visualización del rendimiento organizacional, siendo este el caso de muchas agencias implementadoras de modelos de visitas a domicilio a niños en edad temprana en los Estados Unidos de América, en particular de los usuarios del sistema de seguimiento de caso Visit Tracker Web (DataKeeper Technologies, 2020). Estos enfoques realizan la consulta de información directamente sobre la base de datos empleada por los sistemas de producción, lo cual implica que un solo recurso sea compartido por dos o más entidades, resultando en problemas de desempeño e interferencia de datos.



Estos problemas se pueden mitigar mediante el empleo de una solución de inteligencia de negocios, al implementar una arquitectura que permita realizar la consulta de información sobre un origen de datos independiente al empleado por los sistemas en producción.

A pesar de esto, el empleo de herramientas comerciales de inteligencia de negocios para la medición del rendimiento organizacional en programas de visitas a domicilio implica un costo adicional para los usuarios de una agencia implementadora, ya que se requiere de adquisición de licencias de uso y de personal capacitado en el empleo de esta. Además de estar diseñadas para funcionar de una manera más general, contando con un gran número de funcionalidades que finalmente no son empleadas, lo cual tiene inconvenientes de usabilidad, generando una sensación de inutilidad hacia la herramienta y disminuyendo así el valor percibido del uso de esta, llevando al fracaso en la implementación de la herramienta de inteligencia de negocios.

Debido a los costos adicionales generados y a los problemas de usabilidad obtenidos a partir del uso de herramientas de inteligencia de negocios comerciales, surge la necesidad de la creación de una herramienta de Inteligencia de Negocios capaz de aprovechar al máximo la información obtenida mediante la prestación de servicios de visitas a domicilio, ofreciendo una herramienta alineada a los objetivos de los programas de visitas a domicilio, la funcionalidad esperada por los usuarios de acuerdo con sus necesidades y la mitigación de los problemas presentes al emplear un enfoque de reporte operativo.

### **1.3. Objetivos**

#### **1.3.1. Objetivo general**

Desarrollar una herramienta de inteligencia de negocios alineada a los objetivos del programa federal MIECHV, empleando un enfoque que facilite la visualización de los indicadores resultantes de la prestación de servicios de visitas a domicilio a niños en edad temprana y optimizando el uso de recursos computacionales empleados con respecto a un enfoque de reporte operativo tradicional.

#### **1.3.2. Objetivos específicos**

- Obtener el estado del arte respecto al uso de inteligencia de negocios para la gestión de rendimiento organizacional.
- Identificar arquitecturas de inteligencia de negocios existentes en el estado del arte.
- Comparar las arquitecturas de inteligencia de negocios identificadas para la selección de la arquitectura a emplear por la herramienta propuesta.
- Obtener el conjunto de funcionalidades necesarias para que un programa de visitas a domicilio sea capaz de medir su rendimiento.
- Realizar la propuesta de una herramienta de inteligencia de negocios para la gestión de rendimiento en programas de visitas a domicilio a niños en edad temprana.
- Validar la herramienta propuesta mediante un experimento controlado, usando métodos estadísticos para analizar los resultados.

#### **1.4. Hipótesis**

En comparación con un enfoque de reporte operativo para la obtención del rendimiento de una agencia implementadora de un modelo de prestación de servicios de visitas a domicilio a niños en edad temprana, un enfoque de inteligencia de negocios mediante Virtualización de Datos presenta beneficios al reducir la interferencia a las bases de datos de producción, el tiempo y los recursos computacionales empleados para la obtención de informes de rendimiento organizacional.

#### **1.5. Justificación**

De acuerdo con el Cuadrante mágico para plataformas de análisis e inteligencia de negocios presentado por (Gartner, 2020), Microsoft y Tableau se posicionan como los líderes en el mercado, seguidos por herramientas como Qlik y ThoughtSpot. A pesar de que todas las herramientas mencionadas brindan un gran conjunto de funcionalidades y actualmente son empleadas en todo el mundo, una gran cantidad de estas funcionalidades no son usadas por un programa de visitas a domicilio y aquellas funcionalidades que sí lo son, requieren de adaptación a las necesidades particulares de cada organización, requiriendo de personal capacitado en el uso de dichas herramientas, repercutiendo en tiempo, esfuerzo y dinero empleado.

De acuerdo con (Gartner, 2017), \$18.3 mil millones de dólares fueron los ingresos globales esperados para finales del año 2020 en el mercado de software de análisis e inteligencia de negocios, lo cual representó un crecimiento del 33% con respecto a los ingresos en el año 2016. Sin embargo, de acuerdo a un estudio realizado por (Rajnoha, R; et al, 2016) a 164 organizaciones de origen eslovaco, solo 7% de ellas hacían uso de inteligencia de negocios

dentro de sus operaciones diarias, 22% de ellas lo consideraron en un futuro cercano o a largo plazo, mientras que el 71% restante no consideraba el uso de inteligencia de negocios dentro de sus operaciones diarias, lo cual indica que menos de la tercera parte de dichas organizaciones hacen uso de inteligencia de negocios como parte fundamental de sus actividades diarias.

Este trabajo de investigación busca proveer de una herramienta de inteligencia de negocios capaz de ayudar a un programa de visitas a domicilio a niños en edad temprana a gestionar su rendimiento, ofreciendo las características necesarias para que dichos programas sean capaces de medir y obtener el rendimiento esperado tras la prestación anual de sus servicios, incluyendo impresión de reportes, desglose de información y uso en la nube. Lo anterior a través de un enfoque que permita la alineación de la funcionalidad proporcionada por la herramienta con los objetivos del programa.

### **1.6. Alcances y limitaciones**

El presente trabajo de investigación incluye el desarrollo de una solución de software resultante de la implementación de una herramienta de inteligencia de negocios a la agencia Family Focus, Inc. Implementadora del modelo de prestación de servicios de visitas a domicilio a niños en edad temprana HFA, presentado en la Tabla 1. El desarrollo de la herramienta se limita a las necesidades específicas requeridas por el programa federal MIECHV, diseñado por (Health Resources & Services Administration, 2020), Implementando la primera área de referencia estipulada para la medición del rendimiento, junto con sus 6 indicadores de rendimiento (Tabla 24).

La herramienta ofrece un panel de información compuesto por una serie de gráficos comparativos agrupados en áreas de referencia, cada área se compone por una serie de métricas representadas mediante dichos gráficos, permitiendo así a una agencia implementadora la visualización del rendimiento obtenido tras la prestación de servicios en un periodo de tiempo acotado, y de acuerdo a los resultados esperados por el programa federal MIECHV (Health Resources & Services Administration, 2020), presentados en el capítulo 4 de este trabajo.

Dado que la información empleada para la medición del rendimiento en la agencia implementadora anteriormente mencionada ha pasado previamente por un proceso de eliminación de información confidencial, se permite el libre uso de esta para los fines que convenga.

## Capítulo 2. Estado del arte

En este capítulo se presenta el estado del arte respecto al uso de herramientas de inteligencia de negocios para la gestión de rendimiento organizacional. Se analizan los trabajos encontrados y se presentan los métodos, herramientas y técnicas empleadas, por último, se presenta el beneficio que supone el uso de inteligencia de negocios en el rendimiento de una organización.

### 2.1. Revisión sistemática de la literatura

La Revisión Sistemática de la Literatura (Del inglés: (Systematic Literature Review), es un método propuesto por (Kitchenham, 2004) que permite identificar, evaluar, interpretar y sintetizar la evidencia disponible relevante a un tema o pregunta de investigación, reduciendo la posibilidad de sesgo en la búsqueda de información, ya que su ejecución se realiza de manera rigurosa e imparcial, aumentando así la posibilidad de obtener mejores resultados en la búsqueda de información. En la Figura 7 se muestran las fases y actividades que integran el método de la revisión sistemática de la literatura.

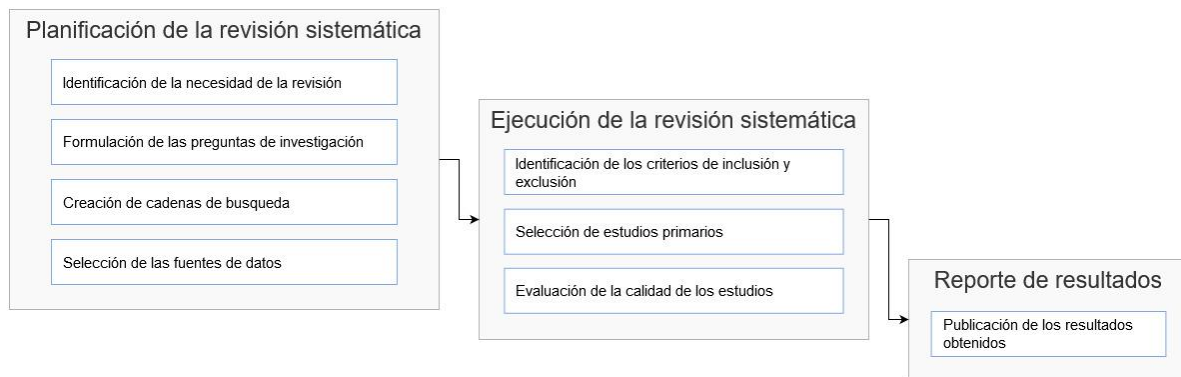


Figura 7. Fases y actividades realizadas durante la ejecución de la revisión sistemática de la literatura

### **2.1.1. Planificación de la revisión**

En la fase de planificación de la revisión sistemática se han de identificar la necesidad de esta, formular las preguntas de investigación a abordar, generar cadenas de búsqueda y seleccionar los orígenes de datos para la extracción de los estudios primarios.

#### ***2.1.1.1. Identificación de la necesidad***

Dado el rol crucial que tienen los datos y la visualización de estos en la gestión de rendimiento y mejora de un programa, el empleo de inteligencia de negocios se vuelve un enfoque a considerar para el reporte visual de la información. A pesar de esto, existen pocos o nulos precedentes sobre el uso de enfoques de inteligencia de negocios para la gestión de rendimiento en programas de visitas a domicilio a niños en edad temprana.

Dado el problema mencionado anteriormente, se vuelve necesario conocer los estudios existentes relacionados al uso de inteligencia de negocios para la medición de rendimiento, para así obtener un panorama general sobre el uso de métodos, técnicas o herramientas junto con los beneficios encontrados tras la implementación de las mismas, y de este modo obtener las bases que sirvan como fundamento para la propuesta de una herramienta de inteligencia de negocios útil para la medición de rendimientos en un programa de visitas a domicilio a niños en edad temprana.

#### ***2.1.1.2. Preguntas de investigación***

A continuación, se listan las preguntas de investigación a emplear para establecer el estado del arte respecto al uso de inteligencia de negocios para la gestión de rendimiento:

1. ¿Existe algún estudio sobre el uso de la inteligencia de negocios para la gestión de rendimiento en programas de visitas a domicilio a niños en edad temprana?

2. ¿Qué enfoques son empleados para la gestión de rendimiento mediante el uso de inteligencia de negocios?
3. ¿Cuáles son los factores de éxito en la implementación de inteligencia de negocios para la gestión rendimiento?
4. ¿Cuáles son los beneficios obtenidos del uso de inteligencia de negocios, en cuanto a rendimiento de un programa se refiere?

#### **2.1.1.3. Cadena de búsqueda**

Con la finalidad de facilitar la creación de la cadena de búsqueda, se realiza la identificación de las palabras clave resultantes de las preguntas de investigación establecidas, resultando las siguientes palabras clave:

1. Inteligencia de negocios
2. Rendimiento
3. Programa
4. Medición
5. Visitas a domicilio
6. Niños en edad temprana

Dado que el concepto visitas a domicilio está relacionado a familias con niños, se decide emplear los términos “Visitas a domicilio” y “Visitas a domicilio a niños en edad temprana”. La medición de rendimiento generalmente precede a la mejora de este, es por ello que se opta por emplear los términos “Medición de rendimiento” y “Mejora de rendimiento”. Por último, se realiza una combinación entre las palabras clave programa y rendimiento,



obteniendo los términos “Rendimiento de programa” y “Rendimiento corporativo”, agregándosele la palabra corporativo o corporación como sinónimo a la palabra programa.

Los términos mencionados se interrelacionan mediante el uso de conectores lógicos “AND” y “OR” y agrupación de estos, resultando en la Cadena de Búsqueda (CB) presentada en la Tabla 3, siendo esencial la existencia del término “Inteligencia de negocios”.

*Tabla 3. Cadenas de búsqueda resultantes*

| Identificador | Cadena de búsqueda  |
|---------------|---|
| CB1           | ("Business Intelligence") AND ("corporate performance" OR "program performance" OR "performance management" OR "performance measurement" OR "Early childhood home-visiting program" OR "Home-visiting program") |

#### **2.1.1.4. Orígenes de datos**

Para la selección de las fuentes u orígenes de datos se emplearon las más reconocidas en el área de la ingeniería de software, así como aquellas recomendadas por el asesor en este trabajo de investigación, resultando las siguientes:

- IEEE Xplore
- SpringerLink
- Elsevier Science
- Scopus

#### **2.1.2. Ejecución de la revisión**

La segunda fase de la revisión sistemática corresponde a la ejecución de la misma, en dicha fase han de definirse los criterios de inclusión y exclusión para la selección de los estudios primarios, así como la evaluación de la calidad de los mismos para la determinación de los estudios primarios a emplear para la extracción de información.

### 2.1.2.1. Criterios de inclusión y exclusión

En la Tabla 4 se muestran los criterios de inclusión y exclusión a emplear para realizar el filtrado de estudios provenientes de los orígenes de datos seleccionados.

Tabla 4. Criterios de inclusión y exclusión empleados para la Revisión Sistemática de la Literatura

| Criterio de inclusión  | Criterio de exclusión  |
|--|--|
| Las secciones título o Abstract contienen la palabra clave “Inteligencia de negocios” y al menos una de las palabras clave restantes en la cadena de búsqueda. | Ni el título ni el Abstract contienen la palabra clave “Inteligencia de negocios”.   |
| Estudios en Inglés o Español.  | Estudios en un idioma distinto a Inglés o Español.   |
| Estudios publicados entre el año 2014 y 2019.  | Estudios anteriores al año 2014.   |
| Estudios relacionados al empleo de inteligencia de negocios y su impacto en el rendimiento de un programa.   | Estudios que, por cuestión económica, legal o de cualquier índole estén fuera del límite de acceso de la institución o del estudiante.<br>Estudios duplicados. |

### 2.1.2.2. Selección de estudios primarios

Para la selección de los estudios primarios, de acuerdo con lo sugerido por (Kitchenham, 2004) se define un proceso de selección de estudios, comprendido por los pasos listados en la Tabla 5.

Tabla 5. Procedimiento de selección de estudios primarios

| Identificador | Descripción  |
|---------------|--|
| Paso 1        | Adaptación y ejecución de las cadenas de búsqueda en cada motor de búsqueda de las fuentes de datos seleccionadas.   |
| Paso 2        | Aplicar el primer criterio de inclusión para el filtrado de estudios mediante la lectura del título y el Abstract.   |
| Paso 3        | Aplicar el resto de los criterios de inclusión y exclusión, leyendo las secciones de Abstract, Introducción, Conclusiones y de ser necesario, el estudio completo. |
| Paso 4        | Selección de los estudios primarios.   |

En la Figura 8 se muestra la ejecución del proceso de selección de estudios primarios mediante la aplicación de los 4 pasos definidos para el mismo, resultando en la selección de 11 estudios primarios.

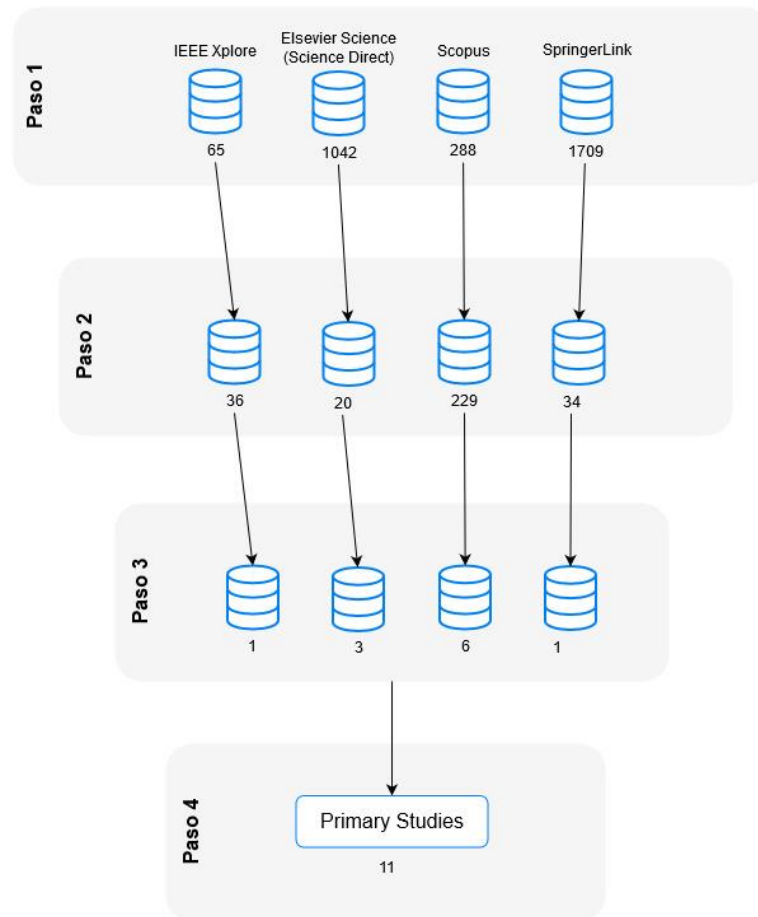


Figura 8. Resultados de la selección de estudios primarios

### 2.1.2.3. Evaluación de la calidad de los estudios

Mediante la evaluación de la calidad de los estudios primarios se garantiza que la información contenida en cada uno de ellos es relevante y aporta valor al trabajo de investigación a realizar. Para la evaluación de la calidad de los estudios se han de aplicar las

Preguntas de Evaluación de Calidad (PEC) listadas en la Tabla 6, eliminando aquellos estudios que no respondan a cuando menos una pregunta.

*Tabla 6. Preguntas de evaluación de calidad aplicadas a los estudios primarios*

| Identificador | Pregunta   |
|---------------|--|
| PEC1          | ¿El estudio contiene información acerca del uso de herramientas, métodos o técnicas de inteligencia de negocio aplicadas a la medición o gestión de rendimiento? |
| PEC2          | ¿El estudio explica cómo la inteligencia de negocios influye en el rendimiento de un programa?   |
| PEC3          | ¿El estudio presenta los factores de éxito que influyen en la correcta implementación de inteligencia de negocios?   |

Tras evaluar la calidad de los estudios provenientes del proceso de selección de estudios primarios se decide conservar todos estos, resultando en un total de 11 estudios empleados para la Revisión Sistemática de la Literatura.

### **2.1.3. Resultados**

A continuación, se muestran los resultados obtenidos de los estudios primarios (EP) seleccionados para la Revisión Sistemática de la Literatura, listados en la Tabla 7. Las secciones subsecuentes presentan los enfoques encontrados en la literatura con respecto a la implementación de inteligencia de negocios para la gestión de rendimientos, seguido por los factores de éxito y los beneficios encontrados tras la aplicación de un sistema de este tipo.

*Tabla 7. Estudios primarios seleccionados para la Revisión Sistemática de la Literatura*

| Identificador | Título  |
|---------------|---|
| EP1           | An empirical study of business intelligence impact on corporate performance management                      |
| EP2           | Business intelligence as a key information and knowledge tool for strategic business performance management |
| EP3           | Business intelligence for performance measurement: A case based analysis                                    |
| EP4           | Business intelligence systems and bank performance in Ghana: The balanced scorecard approach                |

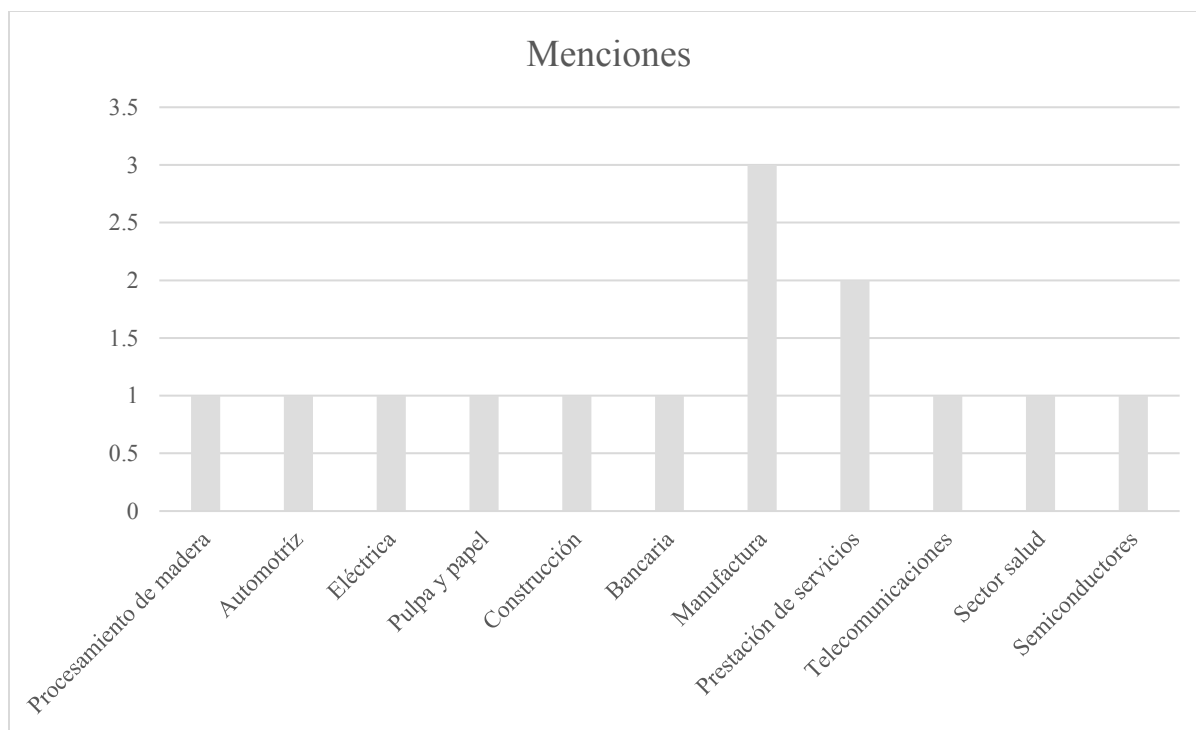
| Identificador | Título   |
|---------------|--|
| EP5           | Business intelligence systems use in performance measurement capabilities: Implications for enhanced competitive advantage                                   |
| EP6           | Evidence for Business Intelligence in Health Care: A Literature Review   |
| EP7           | Integration of business intelligence with corporate strategic management   |
| EP8           | Organizational improvisation and the reduced usefulness of performance measurement BI functionalities  |
| EP9           | The Impact of Business Intelligence on Corporate Performance Management  |
| EP10          | The Relation Between Using Business Intelligence Solutions and Corporate Performance   |
| EP11          | Using the balanced scorecard in assessing the impact of BI system usage on organizational performance: An empirical study of Taiwan's semiconductor industry |

### ***2.1.3.1. Inteligencia de negocios para la gestión de rendimiento en programas de visitas a domicilio***

Tras realizar una revisión de las áreas en las que los estudios primarios fueron enfocados, se detecta la aplicación de inteligencia de negocios en una amplia variedad de áreas, resaltando la mención de la industria de la manufactura y la prestación de servicios.

A pesar de lo anterior, no se encuentra evidencia del empleo de inteligencia de negocios en programas de visitas a domicilio, aunque destaca la aparición de un estudio que presenta los factores de éxito y beneficios encontrados tras la aplicación de inteligencia de negocios dentro del sector salud, correlacionado estrechamente con los sistemas de seguimiento de caso, empleados por los programas de visitas a domicilio para dar seguimiento a familias y niños atendidos.

La Figura 9 presenta el total de áreas abordadas, mediante la identificación de esta y el número de apariciones de acuerdo con lo mencionado en los estudios primarios.



*Figura 9. Áreas abordadas en los estudios primarios de acuerdo con el número de menciones.*

### **2.1.3.2. Enfoques de inteligencia de negocios para la gestión de rendimiento**

Relacionado a la pregunta de investigación número 2, de los 11 estudios primarios, 2 de ellos realizan la propuesta de marcos de trabajo para la implementación de inteligencia de negocios.

En (Vallurupalli, & Bose, 2018) se propone un marco de trabajo para la implementación de sistemas de gestión de rendimiento, llamado PMSIF (Del inglés: Performance Measurement System Implementation Framework), cuyo objetivo es facilitar el análisis de extremo a extremo de la implementación de un sistema de gestión de rendimiento en una organización. El marco propuesto es holístico y abarca todos los aspectos de la implementación de un sistema

de gestión de rendimiento (personas, procesos y tecnología), componiéndose de 6 etapas: iniciación, adopción, adaptación, aceptación, rutinización e infusión.

En (Alnoukari & Hanano, 2017) se propone un marco de trabajo, mediante el uso de características de inteligencia de negocios e inteligencia competitiva, para la creación de un análisis FODA y el desarrollo de objetivos de negocio mediante el empleo de cuadros de mando integral.

El empleo de cuadros de mando integral como parte de la implementación de inteligencia de negocios para la gestión de rendimiento es mencionado en 5 de los 11 estudios primarios. En (Kaplan & Norton, 1996) se acuñó el concepto de Cuadro de Mando Integral (BSC) como una herramienta de gestión que permite a las organizaciones medir las actividades realizadas en términos de visión y estrategia, proporcionando un panorama general del rendimiento de una empresa. Según los autores, una organización no puede ser evaluada únicamente mediante indicadores financieros, sino que existen otros indicadores no financieros igual de importantes, lo cual conlleva a la evaluación de una organización a través de indicadores financieros y no financieros, representados en las 4 perspectivas presentadas en la Figura 10.

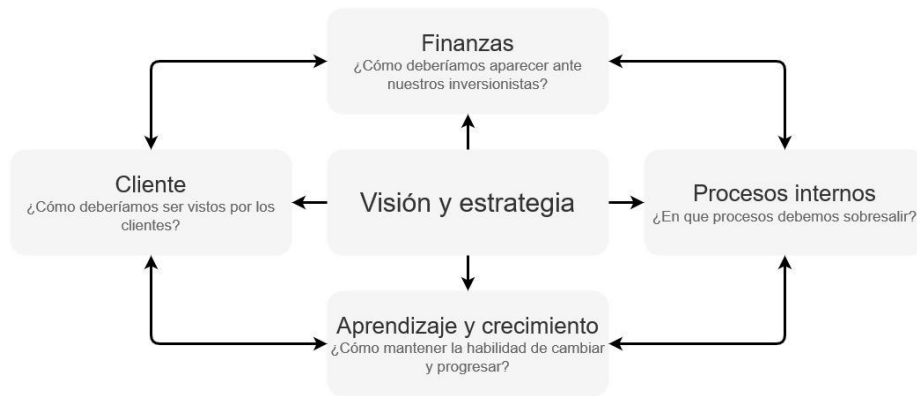


Figura 10. 4 perspectivas del cuadro de mando integral

### 2.1.3.3. Factores de éxito en la implementación de inteligencia de negocios para la gestión de rendimiento

A través del análisis de los estudios es posible identificar los factores de éxito para la implementación de inteligencia de negocios para la gestión de rendimiento, a continuación, se detallan los hallazgos realizados.

- *Funcionalidad y relevancia para las necesidades de la organización.* Funcionalidad de BI como gestión de paneles, desglose de datos, On-Line Analytical Processing (OLAP), data mining, entre otros, deben estar alineados con las necesidades de la organización. Este factor juega un papel importante en el éxito de BI en la organización; los sistemas heredados no tienen una forma estándar de informes e indicadores clave de rendimiento (KPI), por lo que causan problemas al revisar el desempeño y tomar acciones correctivas (Vallurupalli, & Bose, 2018).



Cuando se habla de la justificación del soporte de decisiones de BI, para reducir problemas como la rentabilidad o la eficiencia, BI debe estar dirigido por las necesidades de la organización, no por la tecnología (Thamir & Poulis, 2015). Si BI no alinea los objetivos de negocio de la organización o no satisface las necesidades de esta, como tomar decisiones bajo presión de tiempo o monitorear la competencia, toda la solución fallará.

- *Visión clara y caso de negocio bien establecido.* La organización requiere una visión sólida y bien establecida para garantizar que la decisión de adoptar y utilizar un sistema de BI sea fluida (Vallurupalli, & Bose, 2018; Loewen & Roudsari, 2017).

Se necesita una visión empresarial a largo plazo, basada principalmente en términos estratégicos y organizativos, para establecer un caso de negocio sólido, cumpliendo así los objetivos y necesidades del negocio; la posesión de un caso de negocio bien establecido, derivado de un análisis detallado de las necesidades de la organización, aumentará las posibilidades de obtener el apoyo de la alta dirección. Por último, un caso de negocio debe identificar los beneficios estratégicos propuestos, los recursos, los riesgos, los costos y el cronograma (Yeoh & Koronios, 2010; Zaiid, Grida, & Hussein, 2018)

- *Soporte de la alta dirección.* El apoyo de la alta dirección garantiza que la decisión de adoptar BI no sea anulada debido a intereses políticos y creados de las partes interesadas afectadas negativamente, de acuerdo con (Vallurupalli, &

Bose, 2018), cuando los beneficios fueron claros la alta dirección apoyó la adopción de BI, por lo que la decisión de adoptar BI fue fluida.

Un sistema de BI evoluciona de acuerdo con los requisitos empresariales cambiantes, por lo que requiere una financiación y una asignación de recursos coherentes directamente de la alta dirección para superar los problemas organizativos continuos (Yeoh & Koronios, 2010).

La alta dirección debe proporcionar los recursos necesarios, mano de obra calificada, los fondos de capital y todos los demás recursos, minimizando así la resistencia potencial causada por la estructura interna de la organización (Zaied, Grida, & Hussein, 2018).

- *Actividades de gestión de cambio.* Las actividades de gestión de cambios ayudan a la aceptación del sistema de BI, por ejemplo, mediante el uso de una combinación de medidas orientadas al usuario y medidas coercitivas (Vallurupalli, & Bose, 2018).

Los usuarios deben participar durante toda la implementación de BI. Una mejor participación de los usuarios en el proceso de cambio puede conducir a una mejor comunicación de sus necesidades, asegurando así una introducción exitosa de BI (Yeoh & Koronios, 2010).

La gestión del cambio se logra mediante la participación y la formación de los usuarios. De este modo, se reduce la resistencia al uso de la solución de BI recién implementada y ayuda a los usuarios a comprenderla mejor, haciéndola más eficiente y eficaz (Zaied, Grida, & Hussein, 2018).

- *Identificar las necesidades del usuario.* Comprender las necesidades de los usuarios ayuda a comprender la perspectiva de los usuarios y a diseñar un sistema adecuado para ellos (Vallurupalli, & Bose, 2018). Si los usuarios clave no están involucrados durante la implementación de BI, podría haber información valiosa que se puede pasar por alto. Las dimensiones de los datos, las reglas de negocio, los metadatos y el contexto de los datos que necesitan los usuarios deben incorporarse al sistema y validarse (Yeoh & Koronios, 2010).
- *Identificar a los líderes de proyecto.* Los líderes de proyecto ayudan a comprender mejor las necesidades y preocupaciones de los usuarios en las distintas etapas de la implementación de BI. Un líder de proyecto que tenga una excelente visión para los negocios podrá prever los desafíos organizacionales y cambiar de rumbo en consecuencia, viendo el sistema de BI desde una perspectiva organizacional, alineándose con los objetivos y necesidades de la organización (Yeoh & Koronios, 2010).

Los líderes del proyecto deben crear conciencia, proporcionar información, apoyo político y recursos materiales que afecten la adopción e implementación exitosa de BI, además, deben participar en actividades como la estandarización de datos, ingeniería de requisitos, análisis de calidad de datos y pruebas (Zaied, Grida, & Hussein, 2018).

- *Beneficios percibidos.* Una evaluación transparente de los beneficios que surgen del sistema de BI es importante para garantizar que la organización genere valor a partir de su uso (Vallurupalli, & Bose, 2018; Loewen & Roudsari, 2017).

Los beneficios percibidos por los usuarios, como la reducción de costos, el ahorro de tiempo y la mejora en la toma de decisiones, acelerarán la tasa de adopción de una implementación de BI (Zaied, Grida, & Hussein, 2018).

- *Calidad de los datos.* Muchos problemas relacionados con los datos dentro de los sistemas backend no se descubren hasta que los datos se vacían y consultan dentro del sistema de BI, lo que afecta la calidad de los informes gerenciales, y que a su vez influyen negativamente en los resultados de las decisiones (Yeoh & Koronios, 2010).

Más del 50 por ciento de los proyectos de BI fracasan debido a problemas de calidad de los datos; además, esto cuesta a EE. UU. más de \$ 600 mil millones de dólares anuales (Zaied, Grida, & Hussein, 2018).

Los datos deben ser de alta calidad; de lo contrario, la confiabilidad de las respuestas proporcionadas por BI podría verse comprometida, lo que generaría desconfianza de los usuarios y, finalmente, el fracaso de la solución (Vallurupalli, & Bose, 2018; Špičák, 2018; Loewen & Roudsari, 2017).

- *Utilidad percibida y facilidad de uso.* Los usuarios deben aprender a trabajar con una solución de BI, de no hacerlo, la solución podría no alcanzar su máximo potencial y, finalmente, fallar debido al escepticismo del usuario (Loewen & Roudsari, 2017; Špičák, 2018).

Las soluciones que son más fáciles de aprender se adoptan más rápidamente que aquellas que requieren que los adoptantes desarrollen nuevas habilidades de comprensión, por lo que los usuarios requieren una solución simple y estable

que satisfaga sus necesidades en el menor tiempo posible (Zaied, Grida, & Hussein, 2018).

- *Plataforma tecnológica subyacente.* Ser capaz de implementar y hacer uso con éxito de una solución de BI que demuestre ser eficaz en la implementación de factores de calidad como la disponibilidad, el rendimiento, la coherencia, la escalabilidad y la flexibilidad es importante para el éxito de la solución en sí (Loewen & Roudsari, 2017; Zaied, Grida, & Hussein, 2018).

La flexibilidad permitirá que una solución de BI interactúe con una variedad de fuentes de datos y herramientas analíticas, como otras soluciones de BI. La escalabilidad permitirá que una solución de BI se adapte a las necesidades empresariales cambiantes (Zaied, Grida, & Hussein, 2018).

- *Momento estratégico y flexibilidad organizacional.* Las competencias de improvisación organizacional implican que los gerentes utilicen solo formas mínimas de información de medición del desempeño, lo que reduce la utilidad de BI debido al momento estratégico y la flexibilidad organizacional, lo que hace que toda la solución falle (Peters, Wieder, & Sutton, 2018); lo que sugiere que la adopción de BI debe evaluarse en contextos donde prevalece la mejora organizacional.

#### ***2.1.3.4 Beneficios de la implementación de inteligencia de negocios en el rendimiento de una organización***

El desempeño de una organización se ve afectado positivamente por una serie de factores relacionados entre sí y no a través del empleo directo de inteligencia de negocios.

De acuerdo con (Richards, Yeoh, Chong, & Popovič, 2014), el grado en que se implementa efectivamente una solución de inteligencia de negocios influye directamente en la efectividad de la planificación y análisis de una organización.

A través del análisis, una solución de inteligencia de negocios tiene un impacto indirecto en la efectividad de los procesos operativos de una organización (Richards, Yeoh, Chong, & Popovič, 2014) (Hartl, Jacob, Mbep, Budree, & Fourie, 2016).

El empleo de una solución efectiva de inteligencia de negocios reduce el tiempo utilizado para la toma de decisiones (Rajnoha, R; et al, 2016) (Vallurupalli, & Bose, 2018) (Špičák, 2018), además de mejorar la calidad de la planificación estratégica y operativa, siendo entonces el empleo de un sistema de gestión del rendimiento estratégico respaldado por una solución de inteligencia de negocios la clave de la mejora del rendimiento en una organización (Rajnoha, R; et al, 2016).

El uso de una solución de inteligencia de negocios tiene un efecto positivo en el rendimiento financiero de una organización a través de sus procesos internos y del rendimiento de sus clientes (Hou, 2016) (Owusu, 2017). Por otra parte, el aprendizaje y crecimiento del personal tiene un efecto indirecto en el rendimiento financiero mediante la mejora de los procesos internos en una organización (Hou, 2016).

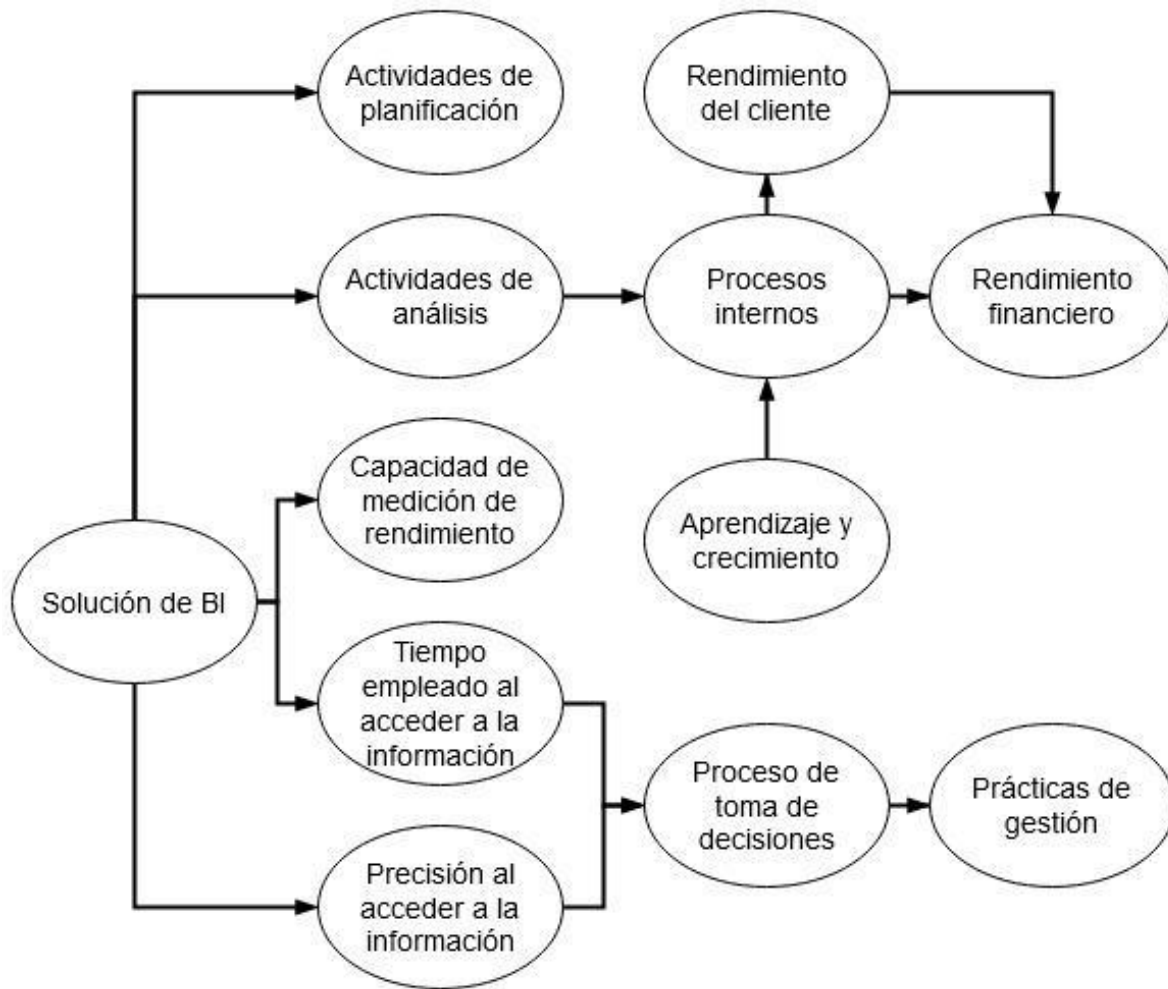
No se ha encontrado evidencia que indique que el tamaño de una organización impacta la efectividad de una solución de inteligencia de negocios (Richards, Yeoh, Chong, & Popovič, 2014), por otro lado, la utilidad de una solución de inteligencia de negocios es impactada negativamente por el momento estratégico y la flexibilidad organizacional, ya que la cantidad de información empleada para la medición de rendimiento tiende a reducirse (Peters, Wieder, & Sutton, 2018).

La Tabla 8 muestra los beneficios directos e indirectos resultantes de la implementación de una solución de inteligencia de negocios para la gestión del rendimiento de una organización.

*Tabla 8. Beneficios del uso de inteligencia de negocios para la gestión del rendimiento en una organización*

| Beneficios directos                         | Beneficios indirectos         |
|---|-------------------------------|
| Actividades de planificación                | Procesos internos             |
| Actividades de análisis                     | Prácticas de gestión          |
| Capacidad de medición de rendimiento        | Aprendizaje y crecimiento     |
| Tiempo empleado al acceder a la información | Proceso de toma de decisiones |
| Precisión al acceder a la información       | Rendimiento financiero        |
|   | Rendimiento del cliente       |

Dado lo anterior, los beneficios directos deben medirse como resultado del uso de la solución de BI, mientras que los beneficios indirectos deben medirse a través de una relación entre los otros beneficios presentes. La Figura 11 presenta la relación entre cada uno de los beneficios presentados.



*Figura 11. Relación entre los beneficios directos e indirectos en la implementación de inteligencia de negocios para la gestión del rendimiento en una organización.*

## 2.2. Selección de tecnología

En esta sección se presenta la tecnología disponible para la implementación de la propuesta de desarrollo de una herramienta de inteligencia de negocios, detallando la justificación de la elección del conjunto de elementos a emplear.



### 2.2.1. Lenguaje de programación

Un lenguaje de programación es un tipo de lenguaje compuesto por un conjunto de instrucciones, comandos y otros elementos léxicos y sintácticos, empleado para desarrollar software. La Figura 12 muestra los lenguajes de programación más populares dentro de la comunidad de desarrolladores de software al mes de abril de 2020 (TIOBE - The Software Quality Company, 2020).

| Apr 2020 | Apr 2019 | Change | Programming Language | Ratings      | Change        |
|----------|----------|--------|----------------------|--------------|---------------|
| 1        | 1        |        | Java                 | 16.73%       | +1.69%        |
| 2        | 2        |        | C                    | 16.72%       | +2.64%        |
| 3        | 4        | ▲      | Python               | 9.31%        | +1.15%        |
| 4        | 3        | ▼      | C++                  | 6.78%        | -2.06%        |
| 5        | 6        | ▲      | C#                   | 4.74%        | +1.23%        |
| <b>6</b> | <b>5</b> | ▼      | <b>Visual Basic</b>  | <b>4.72%</b> | <b>-1.07%</b> |
| 7        | 7        |        | JavaScript           | 2.38%        | -0.12%        |
| 8        | 9        | ▲      | PHP                  | 2.37%        | +0.13%        |
| 9        | 8        | ▼      | SQL                  | 2.17%        | -0.10%        |
| 10       | 16       | ▲▲     | R                    | 1.54%        | +0.35%        |
| 11       | 19       | ▲▲     | Swift                | 1.52%        | +0.54%        |
| 12       | 18       | ▲▲     | Go                   | 1.36%        | +0.35%        |
| 13       | 13       |        | Ruby                 | 1.25%        | -0.02%        |
| 14       | 10       | ▼▼     | Assembly language    | 1.16%        | -0.55%        |
| 15       | 22       | ▲▲     | PL/SQL               | 1.05%        | +0.26%        |
| 16       | 14       | ▼      | Perl                 | 0.97%        | -0.30%        |
| 17       | 11       | ▼▼     | Objective-C          | 0.94%        | -0.57%        |
| 18       | 12       | ▼▼     | MATLAB               | 0.93%        | -0.36%        |
| 19       | 17       | ▼      | Classic Visual Basic | 0.83%        | -0.23%        |
| 20       | 27       | ▲      | Scratch              | 0.77%        | +0.28%        |

*Figura 12. Lenguajes de programación más empleados al mes de abril de 2020*

Se puede observar que los 5 lenguajes de programación más populares son: Java, C, Python, C++ y C#. Los lenguajes de programación Java y C# tienen un respaldo por parte de

grandes corporaciones como Oracle y Microsoft. Mencionado lo anterior se procede a la selección del lenguaje de programación a emplear, optando por el lenguaje de programación C# debido al respaldo que tiene y a la compatibilidad con la tecnología de base de datos a emplear, mencionada posteriormente.

### **2.2.2. Framework de desarrollo**

Para que la herramienta de inteligencia de negocios a implementar pueda ser empleada por el personal de los programas de visitas a domicilio desde cualquier parte de los Estados Unidos de América, se ha de desarrollar una aplicación Web, motivo por el cual se opta por el empleo del framework de desarrollo Web ASP.NET MVC en conjunto con el lenguaje de programación C#.

### **2.2.3. Base de datos**

La base de datos a emplear para el desarrollo de la herramienta debe ser provista por (DataKeeper Technologies, 2020), siendo una base de datos relacional implementada en MS SQL Server y con un tamaño no mayor a 40 GB. Dicha base de datos contiene información de agencias implementadoras junto con la información del personal encargado de prestar servicios de visitas a domicilio, así como la información de las familias y niños a los que se les da servicio.

### **2.2.4. Arquitectura**

La selección de la arquitectura de software a emplear se basa en las arquitecturas de inteligencia de negocios más populares mencionadas en el capítulo 1. En cumplimiento con los factores de éxito encontrados en la literatura y las características de la base de datos a emplear, así como a limitantes de tiempo asociados principalmente a las necesidades propias

de la organización facilitadora de los datos a emplear, se han tomado los siguientes puntos como factores decisivos para la selección de la arquitectura de software:

- *Calidad de los datos.* Se requiere que la calidad de los datos a emplear no se vea afectada por el empleo de la herramienta, sin provocar alteraciones en los mismos y asegurando que la información presentada sea fidedigna.
- *Desacoplamiento.* Los usuarios que hagan uso de la herramienta de inteligencia de negocios no deben tener acceso directo a la base de datos de producción, evitando así poner en riesgo la integridad tanto de la base de datos como de los sistemas de producción empleados.
- *Reporteo operativo.* Se requiere que la información consultada sea tomada en tiempo real para de este modo no presentar información que pueda llegar a ser incorrecta en un periodo de tiempo determinado.
- *Mantenibilidad.* Debido a que se tiene contemplado seguir aumentando la funcionalidad ofrecida por la herramienta, esta misma debe de poder ser mantenida fácilmente por cualquier miembro de la organización.
- *Tiempo de desarrollo.* El tiempo demorado en el desarrollo de la herramienta es crítico para la organización ya que entre más tiempo pasa desde el inicio del desarrollo hasta el momento en que la herramienta es puesta a disposición de los usuarios, menos ventaja competitiva se tiene con respecto a las demás organizaciones proveedoras de este tipo de sistemas.
- *Desempeño.* La herramienta por desarrollar no deberá causar problemas de desempeño en los sistemas implementados actualmente por la organización,

siendo la principal preocupación el surgimiento de problemas de interbloqueo en las operaciones de lectura y escritura hacia la base de datos de producción.

Debido a los factores anteriores y al tamaño de la base de datos provista por (DataKeeper Technologies, 2020), la cual se considera pequeña al no sobrepasar los 10 GB de información, se considera que la arquitectura de Virtualización de Datos es la más adecuada debido a que permite el desacoplamiento de datos, así como reporteo operativo. Debido a que el número de elementos presentes en esta arquitectura es menor al presentado en arquitecturas tradicionales, se vuelve de interés en cuanto a mantenibilidad se refiere, adicionalmente el tiempo de desarrollo se disminuye por este factor.

El último factor por tomar en cuenta es el desempeño, mismo que comúnmente es solucionado mediante el uso de almacenes de datos en los enfoques tradicionales de inteligencia de negocios, pero, debido a que el uso de un enfoque de este tipo presenta más desventajas que ventajas con respecto a los factores considerados, la selección de este tipo de enfoques no es una opción. Para hacer frente a este factor mediante un enfoque de virtualización de datos, ha de ser empleada la técnica de cacheo de datos, disminuyendo de este modo la cantidad de lecturas realizadas hacia la base de datos de producción.

### Capítulo 3. Metodología para el desarrollo del trabajo de investigación

En este capítulo se describe la metodología empleada para el desarrollo de este trabajo de investigación. La Figura 13 muestra la serie de tareas que comprenden dicha metodología.



Figura 13. Metodología seguida para el desarrollo de la tesis

#### 3.1. Ejecución de la metodología para el desarrollo de la tesis

A continuación, se describen los pasos a realizar para la ejecución de la metodología para el desarrollo de la tesis.

- *Obtener el estado del arte respecto al uso de inteligencia de negocios para la gestión de rendimiento organizacional.* Con el objetivo de obtener el panorama general en cuanto al uso de inteligencia de negocios para la gestión de rendimiento organizacional enfocada en: 1) Inteligencia de negocios para la gestión de rendimiento en programas de visitas a domicilio, 2) enfoques de inteligencia de negocios para la gestión de rendimiento, 3) factores de éxito en la implementación de inteligencia de negocios para la gestión de rendimiento, 4) beneficios de la implementación de inteligencia de negocios en el

rendimiento de una organización, detallado todo esto en el capítulo 2 del trabajo de investigación.

- *Proponer un conjunto de funcionalidades necesarias para que un programa de visitas a domicilio sea capaz de medir su rendimiento.* Mediante el análisis de lo establecido por el Departamento de Salud y Servicios Humanos de los Estados Unidos de América para la demostración de la efectividad en la prestación de servicios por parte de un programa de visitas a domicilio a niños en edad temprana, y mediante el empleo de un enfoque de GQM (Del Inglés: Goal Question Metric), se realiza la propuesta de las funcionalidades necesarias para la gestión de rendimiento en la prestación de visitas a domicilio. La propuesta de funcionalidades se encuentra en el capítulo 4 de este trabajo de investigación.
- *Diseñar una herramienta de inteligencia de negocios mediante el conjunto de funcionalidades propuestas.* Posterior a la propuesta de funcionalidades se ha de diseñar la herramienta que incluya dichas funcionalidades, mediante el desarrollo arquitectural y detallado de la misma. El diseño de la herramienta se describe en el capítulo 4 de este trabajo de investigación.
- *Desarrollar la herramienta de inteligencia de negocios para la gestión de rendimiento en programas de visitas a domicilio a niños en edad temprana.* Mediante el diseño de la herramienta se ha de proceder al desarrollo de esta, para de este modo permitir la gestión y mejora de rendimiento organizacional

mediante el conjunto de funcionalidades propuesto e implementando el diseño arquitectural y detallado de la herramienta.

- *Validar la herramienta propuesta mediante un caso de estudio.* Por último, se ha de validar la herramienta propuesta mediante la ejecución de un experimento controlado.

### **3.2. Métodos Estadísticos usados para validación de la herramienta**

A continuación, se describen los métodos estadísticos seleccionados y utilizados para el análisis e interpretación de resultados del Capítulo 5, soportando además tanto las salidas gráficas como numéricas presentadas en el mismo. La Figura 14 muestra el procedimiento a seguir para la utilización de cada método estadístico.

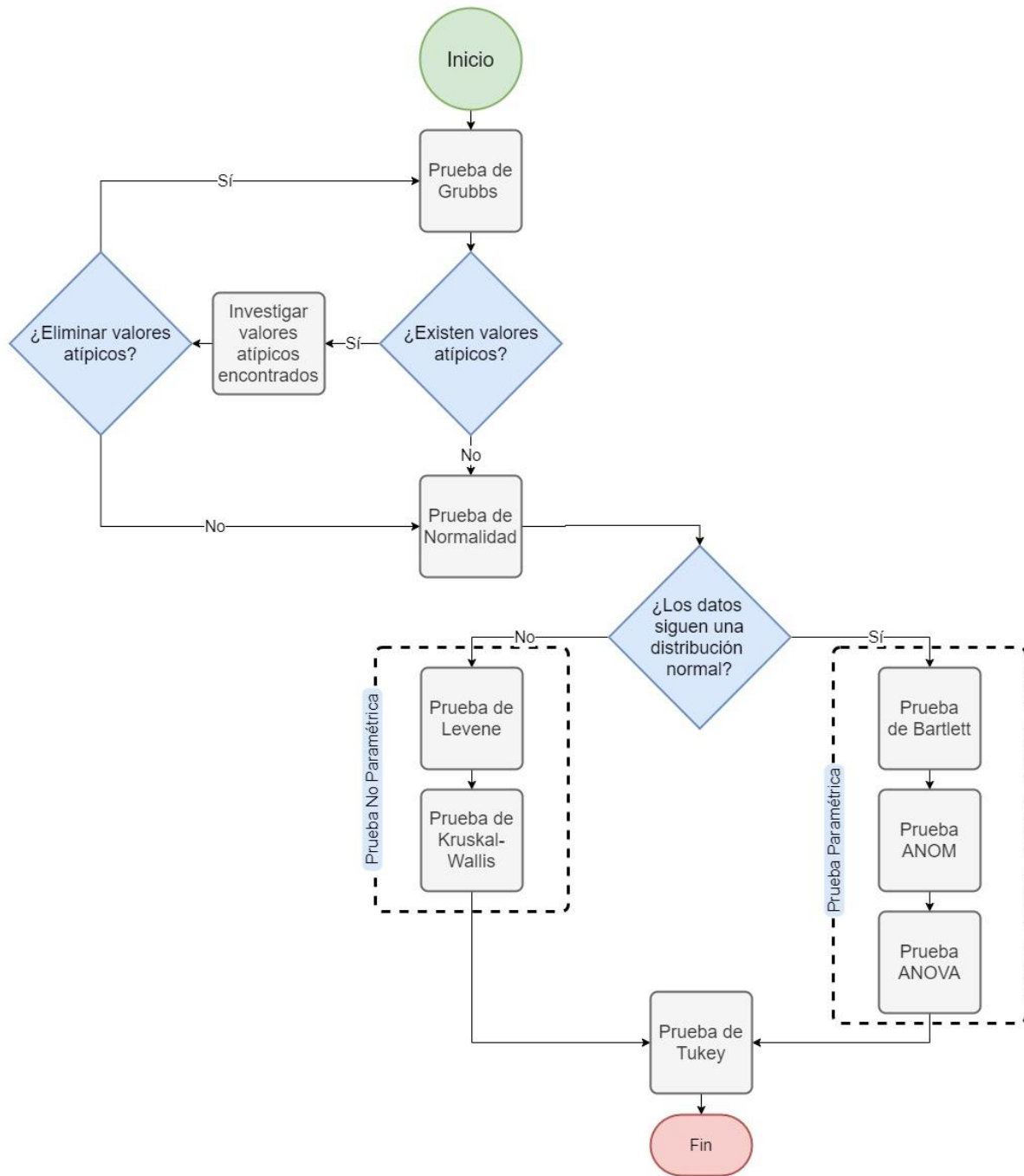


Figura 14. Métodos Estadísticos empleados para validación de la herramienta.



### 3.2.1. Prueba de Grubbs (Detección de Valores Atípicos)

Un valor atípico (Del inglés: Outlier) es un punto que no está de acuerdo con el conjunto de datos, o es un valor raro o una observación extrema, que se puede detectar por medio de una prueba de grubbs u otros métodos estadísticos.

#### **Objetivo de la Prueba de Grubbs**

Utilizar una prueba de valores atípicos para identificar un valor atípico individual en una muestra o variable de interés.

#### **Hipótesis para la Prueba de Grubbs**

Para una prueba de valores atípicos, las hipótesis son las siguientes:

- $H_0$ . Todos los valores de la muestra provienen de la misma población distribuida normalmente.
- $H_1$ . Uno de los valores de la muestra no proviene de la misma población distribuida normalmente.

#### **Métodos para la Prueba de Grubbs**

- **Fórmula para el estadístico unilateral para la Prueba de Grubbs**

Si se evalúa si el valor de datos más pequeño es un valor atípico, entonces el estadístico de prueba G viene dado por:

$$G = \frac{\bar{y} - y_1}{s}$$

Si se evalúa si el valor de datos más grande es un valor atípico, entonces G viene dado por:

$$G = \frac{y_n - \bar{y}}{s}$$

- **Fórmula para el estadístico bilateral para la Prueba de Grubbs**

Para una hipótesis bilateral, G viene dado por:

$$G = \text{máx.} \left\{ \frac{\bar{y} - y_1}{s}, \frac{y_n - \bar{y}}{s} \right\}$$

### Notación de la Prueba de Grubbs

Tabla 9. Notación empleada en las fórmulas para el estadístico unilateral y bilateral.

| Término   | Descripción                                |
|-----------|--|
| $\bar{y}$ | Media de la muestra                        |
| $y_i$     | El i-ésimo valor más pequeño en la muestra |
| s         | Desviación estándar de la muestra          |
| n         | Número de observaciones en la muestra      |

### Valores p para el estadístico de la prueba de Grubbs

- **Fórmula para una prueba unilateral del valor p de una prueba de Grubbs**

El valor p para una prueba unilateral es:

$$P \approx nP \left( T > \sqrt{\frac{n(n-2)G^2}{(n-1)^2 - nG^2}} \right)$$

- **Fórmula para una prueba bilateral del valor p de una prueba de Grubbs**

El valor p para una prueba bilateral es:

$$P \approx 2nP \left( T > \sqrt{\frac{n(n-2)G^2}{(n-1)^2 - nG^2}} \right)$$

- **Valores p exactos vs. aproximados del valor p de una prueba de Grubbs**

Si se cumple lo siguiente, entonces el valor p es exacto.

$$G \geq \sqrt{\frac{(n-1)(n-2)}{2n}}$$

De lo contrario, el valor p calculado representa un límite superior del valor p exacto.

Sin embargo, el límite superior es una aproximación muy adecuada al valor p exacto.

### Notación del valor p de una prueba de Grubbs

Tabla 10. Notación empleada en las fórmulas para los valores  $p$  para el estadístico de la prueba de Grubbs.

| Término | Descripción   |
|---------|---|
| G       | Estadístico de la prueba de Grubbs  |
| n       | Número de observaciones en la muestra   |
| T       | Variable aleatoria distribuida como una distribución t con $n - 2$ grados de libertad |

### 3.2.2. Prueba de Anderson-Darling (Prueba de Normalidad)

Una prueba de normalidad nos sirve para determinar si los datos siguen un comportamiento similar a una Distribución normal con parámetros media y varianza. Este es uno de los supuestos más importantes al usar métodos paramétricos. Existen métodos estadísticos como: Shapiro-Wilk, Kolmogorov-Smirnov y Anderson-Darling.

#### **Objetivo de una prueba Anderson-Darling**

Utilizar una prueba de normalidad para determinar si los datos no siguen una distribución normal.

#### **Hipótesis de una prueba Anderson-Darling**

Para una prueba de valores atípicos, las hipótesis son las siguientes:

- $H_0$ . Los datos siguen una distribución normal.
- $H_1$ . Los datos no siguen una distribución normal.

#### **Método de una prueba Anderson-Darling**

- **Fórmula para el estadístico de Anderson-Darling (A2)**

A2 mide el área entre la línea ajustada (que se basa en la distribución elegida) y la función de paso no paramétrica (que se basa en los puntos de la gráfica). El estadístico es una distancia.

$$A^2 = -N - \frac{1}{N} \sum (2i - 1) (\ln F(Y_i) + \ln(1 - F(Y_{N+1})))$$

### **Notación de una prueba Anderson-Darling**

Tabla 11. Notación empleada en la fórmula para el estadístico de Anderson-Darling.

| Término  | Descripción  |
|----------|--|
| $F(Y_i)$ | $\Phi \left( \frac{Y_i - \bar{x}}{s} \right)$ , que es la función de distribución acumulada de la distribución normal estándar |
| $Y_i$    | Datos ordenados  |

#### **• Valor p de una prueba Anderson-Darling**

Otra medida cuantitativa para notificar los resultados de la prueba de normalidad es el valor p. Un valor p pequeño es una indicación de que la hipótesis nula es falsa.

Si usted conoce  $A^2$ , puede calcular el valor p. Sea:

$$A'^2 = A^2 \times \left( 1 + \frac{0.75}{N} + \frac{2.25}{N^2} \right)$$

Dependiendo de  $A'^2$ , usted calculará p con las siguientes ecuaciones:

- Si  $13 > A'^2 > 0.600$ , entonces  $p = \exp(1.2937 - 5.709 * A'^2 + 0.0186(A'^2)^2)$
- Si  $0.600 > A'^2 > 0.340$ , entonces  $p = \exp(0.9177 - 4.279 * A'^2 - 1.38(A'^2)^2)$
- Si  $0.340 > A'^2 > 0.200$ , entonces  $p = 1 - \exp(-8.318 + 42.796 * A'^2 - 59.938(A'^2)^2)$
- Si  $A'^2 < 0.200$ , entonces  $p = 1 - \exp(-13.436 + 101.14 * A'^2 - 223.73(A'^2)^2)$

### **3.2.3. Prueba de Igualdad de Varianzas (Varianzas Homogéneas)**

Las varianzas homogéneas o igualdad de varianzas, es un método estadístico para comparar la variación entre dos o más grupos. Es un supuesto en una prueba ANOVA. Se tiene que determinar la normalidad de los datos para decidir el método a utilizar. Por ejemplo: si los

datos son normales se realiza una prueba de Bartlett, si los datos no son normales se realiza una prueba de Levene.

### **Objetivo de una prueba de igualdad de varianzas**

Utilizar la prueba de igualdad de varianzas para determinar si las varianzas o desviaciones estándar de dos o más grupos difieren.

### **Hipótesis de una prueba de igualdad de varianzas**

La prueba de hipótesis de varianzas iguales de un solo factor o variable es:

- $H_0$ . Las varianzas son iguales (varianzas homogéneas)
- $H_1$ . Las varianzas son diferentes (varianzas heterogéneas).

### **Limitaciones de una prueba de igualdad de varianzas**

Si los datos siguen una distribución normal o no normal el método utilizado cambia (Bartlett o Levene).

Existen dos métodos para evaluar pruebas de varianzas de más de dos grupos: Prueba de Bartlett y Prueba de Levene. La prueba de Bartlett es utilizada cuando los datos se comportan normales y la prueba de Levene es usada cuando los datos no son normales.

### **Métodos de una prueba de igualdad de varianzas**

#### **• Prueba de Bartlett**

El estadístico de prueba de Bartlett calcula el promedio aritmético ponderado y el promedio geométrico ponderado de cada varianza de muestra basada en los grados de libertad. Mientras mayor sea la diferencia en los promedios, es más probable que las varianzas de las muestras no sean iguales.  $B$  sigue una distribución  $\chi^2$  con  $k - 1$  grados de libertad. Si el valor  $p$  es menor que el nivel  $\alpha$ , se rechaza la hipótesis nula de que las varianzas son iguales.

- **Fórmula de una prueba de Bartlett**

$$B = \frac{(\sum v_i) \ln(\sum v_i S_i^2 / \sum v_i) - \sum v_i \ln S_i^2}{1 + \frac{\{\sum (1/v_i) - 1/\sum v_i\}}{\{3 (k - 1)\}}}$$

- **Notación de una prueba de Bartlett**

Tabla 12. Notación empleada en la fórmula para la prueba de Bartlett.

| Término | Descripción   |
|---------|---|
| $s^2$   | $\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 / (n_i - 1)$ |
| k       | Número de muestras                                  |
| $v_i$   | $n_i - 1$   |
| $n_i$   | Número de observaciones al i-ésimo nivel del factor |

- **Prueba de Levene**

El método de cálculo de la prueba de Levene es una modificación del procedimiento de Levene (Levene, 1960) desarrollado por Brown y Forsythe (1974). Este método considera las distancias de las observaciones con respecto a la mediana de la muestra en lugar de la media de la muestra. Al usar la mediana de la muestra en vez de la media de la muestra, esto hace que la prueba sea más sólida para las muestras más pequeñas y hace el procedimiento asintóticamente libre de distribución. Si el valor p es menor que el nivel  $\alpha$ , se rechaza la hipótesis nula de que las varianzas son iguales. (Levene, 1960 y (Brown y Forsythe, 1974).

- **Fórmula de una prueba de Levene**

$$S_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 / (n_i - 1)$$

- **Notación de una prueba de Levene**

Tabla 13. Notación empleada en la fórmula para la prueba de Levene.

| Término   | Descripción                                  |
|-----------|--|
| $V_{ij}$  | $ X_{ij} - \bar{X}_i $                       |
| $i$       | 1, ..., k                                    |
| $j$       | 1, ..., $n_i$                                |
| $X_i$     | Mediana                                      |
| $n_i$     | El número de observaciones en la muestra $i$ |
| $X_{ij}$  | La $i$ -ésima observación en la muestra $i$  |
| $\bar{X}$ | Media  |

### 3.2.4. Prueba de Análisis de Medias (Prueba ANOM)

El análisis de medias es una alternativa gráfica del ANOVA que prueba la igualdad de las medias poblacionales. La gráfica muestra cada una de las medias de los niveles de los factores, la media general y los límites de decisión (punto marcado en color rojo por minitab). Si un punto se encuentra fuera de los límites de decisión, existe evidencia de que la media de los niveles del factor representado por ese punto es significativamente diferente de la media general.

#### Objetivo de la prueba de análisis de medias

Utilizar el análisis de medias para determinar si la media de cada grupo difiere de la media general.

#### Hipótesis de la prueba de análisis de medias

La prueba de hipótesis ANOM de un solo factor es:

- $H_0$ . Todas las medias son iguales.
- $H_1$ . Al menos una media no es igual a las otras.

#### Limitaciones de la prueba de análisis de medias

Los datos pueden seguir una distribución normal, binomial o de Poisson. El análisis de medias es una alternativa gráfica al ANOVA. En este trabajo sólo se considera la distribución normal por el tipo de datos que tenemos (datos continuos).

### **Método de la prueba de análisis de medias**

El análisis de medias es un procedimiento para determinar si las medias del nivel de factor individual son diferentes de la media principal (la media de todas las observaciones en un factor). A continuación, se enumeran los pasos para calcular los resultados del ANOM para un modelo de un solo factor (una solo variable):

- Calcular la media en cada nivel del factor,  $\bar{y}_i$ . ( $i = 1, \dots, r$ ).
  - Calcular la media principal de todas las observaciones  $\bar{y} \dots$ .
  - Calcular  $s_p$ , un estimado de la desviación estándar de la observación.
  - Determinar el valor  $\alpha$ , que es el valor correspondiente al nivel de significancia elegido para la prueba y se utiliza en las líneas de decisión inferiores y superiores.
  - Calcular los límites de decisión inferiores (LDL) y superiores (UDL).
  - Graficar las medias de cada nivel de factor con líneas de referencia superiores e inferiores, y la línea central en la media principal.
- **Fórmula de la media del ANOM**

El promedio de las observaciones a cada nivel de factor. Se grafica la media para cada nivel de factor en la gráfica.

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$



- **Notación de la media del ANOM**

Tabla 14. Notación empleada en la fórmula para la media del ANOM.

| Término  | Descripción   |
|----------|---|
| $n_i$    | Número de observaciones a nivel del factor $i$ .                  |
| $y_{ij}$ | Valor de la $j$ -ésima observación al $i$ -ésimo nivel del factor |

- **Fórmula de la mediana principal (línea central) ANOM**

El promedio de todas las observaciones para todos los niveles del factor. Se utiliza la mediana principal como la línea central en la gráfica.

$$\bar{y}_{..} = \frac{y_{..}}{n_T}$$

- **Notación de la mediana principal (línea central) ANOM**

Tabla 15. Notación empleada en la fórmula de la mediana principal ANOM.

| Término  | Descripción                                       |
|----------|---|
| $y_{..}$ | La suma de todas las observaciones en la muestra. |
| $n_T$    | Número total de observaciones.                    |

- **Fórmula de desviación estándar (Desv.Est.) ANOM**

$$s_i = \sqrt{\frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1}}$$

- **Notación de desviación estándar (Desv.Est.) ANOM**

Tabla 16. Notación empleada en la fórmula de desviación estándar ANOM.

| Término  | Descripción   |
|----------|---|
| $y_{ij}$ | Valor de la $j$ -ésima observación al $i$ -ésimo nivel del factor |

| Término     | Descripción  |
|-------------|--|
| $\bar{y}_i$ | Media de observaciones al i-ésimo nivel del factor.  |
| $n_i$       | Número de observaciones al i-ésimo nivel del factor. |

- **Fórmula de desviación estándar agrupada ANOM**

Un estimado de la variación para todos los niveles del factor. La desviación estándar agrupada se utiliza para calcular los límites de decisión.

$$S_p = \sqrt{\frac{\sum_{i=1}^r [(n_i - 1) s_i^2]}{(n_T - r)}}$$

- **Notación de desviación estándar agrupada ANOM**

Tabla 17. Notación empleada en la fórmula de desviación estándar agrupada ANOM.

| Término | Descripción  |
|---------|--|
| $r$     | Número de niveles.                                   |
| $n_T$   | Número total de observaciones.                       |
| $n_i$   | Número de observaciones al i-ésimo nivel del factor. |
| $S_p^2$ | $\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 / (n_i - 1)$  |

### 3.2.5. Prueba ANOVA de un solo factor

#### Objetivo de una prueba ANOVA

Utilizar el ANOVA de un solo factor cuando tenga un factor categórico (discreto) y una respuesta continua y desee determinar si las medias de las poblaciones de dos o más grupos son diferentes.

#### Hipótesis de una prueba ANOVA

La prueba de hipótesis de un análisis de varianza de un solo factor es:

- $H_0$ . Todas las medias son iguales.

- $H_1$ . Al menos una media no es igual a las otras

### **Limitaciones de una prueba ANOVA**

Las K muestras de las poblaciones son independientes entre sí. Las poblaciones también deben estar distribuidas normalmente y tener la misma varianza.

### **Métodos de una prueba ANOVA**

- **Fórmula del Modelo Ajustado ANOVA**

El análisis de un solo factor del modelo de varianza se puede especificar de diferentes maneras. El modelo de medias de celda es:

$$y_{ij} = \mu_i + \epsilon_{ij}$$

Todas las observaciones para el nivel de factor tienen la misma expectativa,  $\mu_i$ . Puesto que  $\mu_i$  es una constante, todas las observaciones tienen la misma varianza, independientemente del nivel de factor. En el análisis de varianza, la estimación de los cuadrados mínimos se usa para ajustar el modelo y proporcionar estimaciones para los parámetros,  $\mu_i$ .

- **Notación del Modelo Ajustado ANOVA**

*Tabla 18. Notación empleada en la fórmula del modelo ajustado ANOVA.*

| Término         | Descripción  |
|-----------------|--|
| $\mu_i$         | Parámetros o la media real de observaciones al i-ésimo nivel del factor  |
| $\epsilon_{ij}$ | Error que está independiente y normalmente distribuido con una media de 0 y una varianza constante de $\sigma^2$ |

- **Fórmula de suma de los cuadrados (SC) del ANOVA**

La suma de las distancias al cuadrado. La SC Total es la variación total en los datos. El SC (Factor) es la desviación de la media del nivel de factor estimado alrededor de la media general. Esto se conoce también como la suma de los cuadrados entre los tratamientos. El error SC es la desviación de una observación desde su media de nivel de factor correspondiente. Esto se conoce también como error en los tratamientos.

Los cálculos son:

$$\text{Factor SC} = \sum n_i (\bar{y}_i - \bar{y}_{..})^2$$

$$\text{Factor SC} = \sum n_i (\bar{y}_i - \bar{y}_{..})^2$$

$$\text{SC Total} = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

- **Notación de la suma de los cuadrados (SC) del ANOVA**

Tabla 19. Notación empleada en la fórmula de suma de los cuadrados del ANOVA.

| Término        | Descripción  |
|----------------|--|
| $\bar{y}_i$    | Media de observaciones en el i-ésimo nivel del factor.       |
| $\bar{y}_{..}$ | Media de todas las observaciones.                            |
| $y_{ij}$       | Valor de la j-ésima observación al i-ésimo nivel del factor. |

- **Fórmula de Grados de Libertad (GL) del ANOVA**

Indica el número de elementos independientes en la suma de cuadrados. Los grados de libertad para cada componente del modelo son:

$$\text{(Factor) GL} = r - 1$$

$$\text{Error GL} = n_T - r$$

$$\text{Total} = n_T - 1$$

- **Notación de Grados de libertad (GL) del ANOVA**

Tabla 20. Notación empleada en la fórmula de Grados de Libertad del ANOVA.

| Término | Descripción                    |
|---------|--------------------------------|
| $n_T$   | Número total de observaciones. |
| $r$     | Número de niveles de factor.   |

- **Fórmula de Cuadrados Medios (CM) del ANOVA**

El cálculo del cuadrado medio del factor sigue:

$$\text{Factor CM} = \frac{\text{Factor SC}}{\text{Factor GL}}$$

El cálculo del cuadrado medio del error sigue:

$$\text{Error CM} = \frac{\text{Error SC}}{\text{Error GL}}$$

- **Notación de Cuadrados medios (CM) del ANOVA**

Tabla 21. Notación empleada en la fórmula de Cuadrados Medios del ANOVA.

| Término | Descripción            |
|---------|------------------------|
| CM      | Cuadrado medio.        |
| SC      | Suma de los cuadrados. |
| GL      | Grados de libertad.    |

- **Fórmula del Estadístico F del ANOVA**

Para estimar el estadístico F en el ANOVA de un solo factor se calcula:

$$F = \frac{\text{CM (Factor)}}{\text{CM (Error)}}$$

Los grados de libertad del numerador son  $r - 1$ . Los grados de libertad del denominador son  $n_T - r$ .

- **Notación del Estadístico F del ANOVA**

Tabla 22. Notación empleada en la fórmula del Estadístico F del ANOVA.

| Término | Descripción                    |
|---------|--------------------------------|
| $n_T$   | Número total de observaciones. |
| $r$     | Número de niveles del factor.  |

- **Fórmula para Comparaciones Múltiples con el Método de Tukey**

Los métodos de comparaciones múltiples permiten evaluar comparaciones de medias en pares para conformar grupo de medias iguales o diferentes. Existen varios métodos de comparación como: Tukey, Fisher, Dunnett, MCB de Hsu, Prueba de Welch y Games-Howell, entre otros. Para efectos de este trabajo, sólo consideramos el método de tukey para confirmar las diferencias entre medias observadas en el gráfico de ejes y resultado del ANOVA.

$$\bar{y}_i - \bar{y}_j \pm \frac{Q(1 - \alpha; r, n_T - r)}{\sqrt{2}} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Donde Q = percentil  $\alpha$  superior de la distribución de rango estudentizado con r y  $n_T - r$  grados de libertad.

Para encontrar la tasa de error individual a partir de la tasa de error simultánea, utilice la fórmula siguiente:

$$\alpha^* = 2[1 - P(t_u \leq \frac{1}{\sqrt{2}} Q_{1-\alpha, r, n_T - r})]$$

- **Notación Comparaciones Múltiples con el Método de Tukey**

Tabla 23. Notación empleada en la fórmula para Comparaciones Múltiples con el Método de Tukey.

| Término     | Descripción  |
|-------------|--|
| $\bar{y}_i$ | La media de la muestra para el i-ésimo nivel del factor. |
| $\bar{y}_j$ | La media de la muestra j-ésimo nivel del factor.         |
| $n_i$       | El número de observaciones en el nivel i.                |
| $n_j$       | El número de observaciones en el nivel j.                |

| Término    | Descripción   |
|------------|---|
| r          | El número de niveles.                                     |
| s          | La desviación estándar agrupada o raíz cuadrada (MSE).    |
| u          | Los grados de libertad para el error.                     |
| $\alpha$   | La probabilidad simultánea de cometer un error de tipo I. |
| $\alpha^*$ | La probabilidad individual de cometer un error de tipo I. |

- **Valor P del ANOVA (Significancia) con el Método de Tukey**

Se utiliza en las pruebas de hipótesis como ayuda para decidir si se puede rechazar o no una hipótesis nula. El valor p es la probabilidad de obtener una estadística de prueba que sea por lo menos tan extrema como el valor calculado real, si la hipótesis nula es verdadera. Un valor de corte comúnmente utilizado para el valor p calculado es 0.05. Por ejemplo, si el valor p de una estadística de prueba es menor que 0.05, se rechaza la hipótesis nula.

### **3.2.6. Prueba de Suma de Rank Kruskal-Wallis para K poblaciones**

Se utiliza una prueba de Kruskal-Wallis para determinar si las medias de dos o más grupos difieren. Los datos deben tener un factor categórico, una respuesta continua y los datos de todos los grupos deben tener distribución con una forma similar. Es utilizado cuando los datos no se distribuyen de una forma normal por lo que se requiere la aplicación de métodos no paramétricos.

#### **Objetivo de la prueba Kruskal-Wallis**

Validar si K muestras aleatorias podrían provenir de K poblaciones con la misma media.

#### **Hipótesis de la prueba Kruskal-Wallis**

Las hipótesis para la prueba no paramétrica Kruskal-Wallis se basa en las medianas:

- $H_0$ . Todas las medianas son iguales.
- $H_1$ . Al menos una mediana no es igual a las otras

### **Limitaciones de la prueba Kruskal-Wallis**

Cada tamaño de muestra debe ser de al menos 5 para que se utilice  $\chi^2$ , aunque no es necesario que los tamaños de muestra sean iguales. Las distribuciones de frecuencia K deben ser continuas.

### **Método de la prueba Kruskal-Wallis**

Las K muestras se combinan y organizan en orden de tamaño creciente y se les asigna un número de rango. Cuando se producen empates, se utiliza la media de los números de clasificación disponibles. Se calcula la suma de rango para cada una de las K muestras.

Sea  $R_j$  la suma de rangos de la j-ésima muestra,  $n_j$  el tamaño de la j-ésima muestra y N el tamaño de la muestra combinada. El estadístico de prueba es calculado por:

$$H = \left\{ \frac{12}{N(N+1)} \sum_{j=1}^K \frac{R_j^2}{n_j} \right\} - 3(N+1)$$

Esto sigue una distribución  $\chi^2$  con K - 1 grados de libertad. La hipótesis nula de medias iguales se rechaza cuando H excede el valor crítico. Los valores críticos de H para tamaños de muestra pequeños y K = 3, 4, 5 se dan en unas Tablas Kruskal-Wallis. Cuando se usa software estadístico solo se interpreta el valor p estimado.



## **Capítulo 4. Propuesta de herramienta de inteligencia de negocios**

En este capítulo se describe la propuesta de la herramienta, tomando en cuenta los factores de éxito y beneficios identificados en la revisión sistemática.

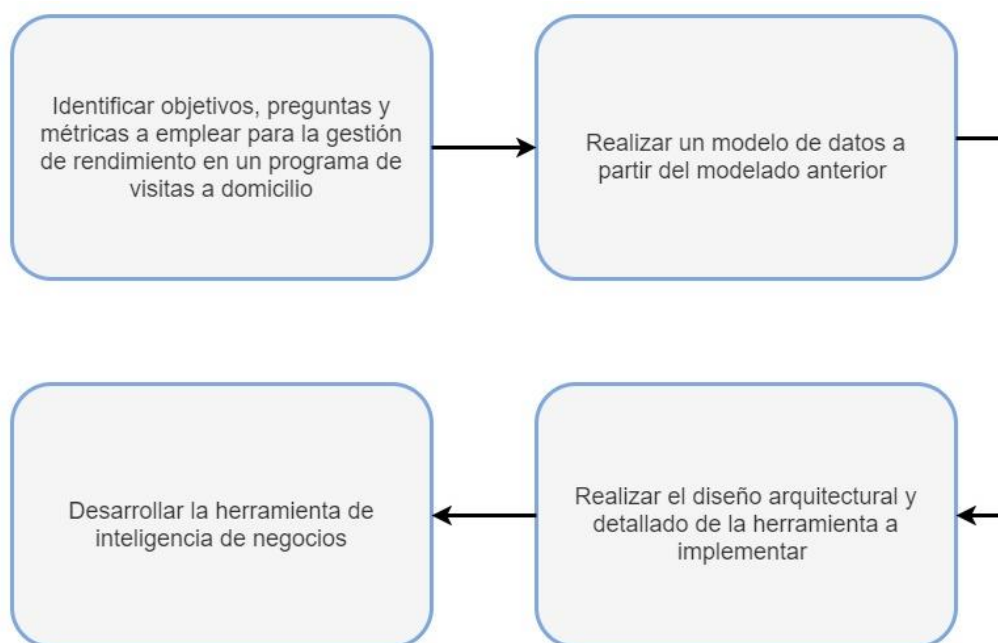
### **4.1. Desarrollo de la propuesta**

Para el correcto diseño de la herramienta de inteligencia de negocios, ha de emplearse un método que permita la identificación de los objetivos de un programa de visitas a domicilio, incluyendo las métricas a emplear junto con los indicadores asociados a cada una de estas, empleados para realizar una correcta gestión del rendimiento de dicho programa. Para la evaluación del rendimiento obtenido ha de realizarse una evaluación desde un punto de vista financiero y no financiero, para dar cumplimiento a lo anterior, la serie de métricas identificadas han de ser agrupadas en áreas, de tal manera que se pueda identificar correctamente el tipo de rendimiento obtenido tras la medición de cada una.

Tras la obtención de la arquitectura de software, junto con el diseño detallado y el modelo de datos a emplear, se procede al desarrollo de la herramienta de inteligencia de negocios.

Adicionalmente, los factores de éxito para la implementación de una herramienta de inteligencia de negocios han de tomarse en cuenta, lo anterior con el objetivo de implementar la herramienta propuesta.

En la Figura 15 se detallan los pasos a seguir por el método del desarrollo de la herramienta de inteligencia de negocios.



*Figura 15. Método empleado para el desarrollo de la herramienta propuesta*

#### **4.1.1. Identificación de objetivos, preguntas y métricas a emplear**

Para la correcta definición de los objetivos, preguntas y métricas a emplear para la gestión del rendimiento en un programa de visitas a domicilio ha de emplearse el programa MIECHV mencionado en el capítulo 1, lo anterior debido a que es un programa gestionado por el gobierno de los Estados Unidos de América a través de la Administración de Recursos y Servicios de Salud (HRSA), beneficiando anualmente a los modelos de prestación de servicio mencionados en la Tabla 1. HRSA requiere que los beneficiarios del Programa de Visitas a Domicilio MIECHV recopilen datos e informen sobre el desempeño de su programa a través de informes de desempeño anuales y trimestrales, y que cumplan con ciertos requisitos legales relacionados con la demostración de la mejora en al menos cuatro de las 6 áreas de referencia establecidas a través de una serie de indicadores de rendimiento (Tabla 24):

Tabla 24. Áreas de referencia establecidas por MIECHV.

| Área de referencia   | Indicador de rendimiento   |
|--|--|
| Mejoras en la salud materna, neonatal e infantil   | Porcentaje de bebés (entre las madres que se inscribieron antes de las 37 semanas de gestación) que nacen prematuros después de haberse inscrito.  |
|  | Porcentaje de bebés (entre las madres que se inscribieron antes del nacimiento) que fueron amamantados en cualquier cantidad a los 6 meses de edad.  |
|  | Porcentaje de tutores inscritos que se someten a pruebas de detección de depresión dentro de los 3 meses posteriores a la inscripción (para los que no están inscritos antes del nacimiento) o dentro de los 3 meses posteriores al parto (para los inscritos antes del nacimiento). |
|  | Porcentaje de niños inscritos que recibieron la última visita recomendada según el programa de la Academia Americana de Pediatría.   |
|  | Porcentaje de madres inscritas en visitas domiciliarias antes del nacimiento o dentro de los 30 días posteriores al parto que recibieron una visita posparto con un proveedor de atención médica dentro de las 8 semanas (56 días) posteriores al parto.                             |
|  | Porcentaje de tutores inscritos en visitas domiciliarias que informaron haber consumido tabaco o cigarrillos al momento de la inscripción y fueron derivados a asesoramiento o servicios para dejar de fumar dentro de los 3 meses posteriores a la inscripción.                     |
| Prevención de lesiones infantiles, abuso infantil, negligencia o maltrato y reducción de visitas a sala de emergencias | Porcentaje de bebés inscritos que siempre duermen boca arriba, sin compartir la cama ni ropa de cama blanda.   |
|  | Tasa de visitas al Departamento de Emergencias relacionadas con lesiones entre los niños inscritos.  |
|  | Porcentaje de niños inscritos con al menos 1 caso investigado de maltrato después de la inscripción dentro del período del informe.  |

| Área de referencia  | Indicador de rendimiento   |
|---|--|
| Mejoras en la preparación escolar y el rendimiento académico de los niños             | <p>Porcentaje de tutores inscritos que reciben una observación sobre interacción entre el tutor y el niño por parte del visitador a domicilio.</p> <p>Porcentaje de niños inscritos a los que durante una semana típica se les leía, contaban historias y / o cantaban canciones todos los días.</p> <p>Porcentaje de niños inscritos con una evaluación oportuna de retrasos en el desarrollo.</p> <p>Porcentaje de visitas a domicilio en las que se preguntó a los tutores si tenían alguna inquietud sobre el desarrollo, el comportamiento o el aprendizaje de su hijo.</p> |
| Reducción de la delincuencia o la violencia doméstica                                 | Porcentaje de tutores inscritos que son evaluados para detectar violencia doméstica.   |
| Mejoras en la autosuficiencia económica familiar                                      | <p>Porcentaje de tutores que se inscribieron sin un título de escuela secundaria o equivalente que posteriormente se inscribieron a este nivel de educación y mantuvieron una inscripción continua o completaron la escuela secundaria o equivalente.</p> <p>Porcentaje de tutores inscritos que tuvieron cobertura de seguro de salud continua durante al menos 6 meses consecutivos.</p>   |
| Mejoras en la coordinación y referencias para otros recursos y apoyos de la comunidad | <p>Porcentaje de cuidadores primarios referidos a servicios resultante de una prueba positiva de depresión.</p> <p>Porcentaje de niños inscritos con exámenes positivos de retrasos en el desarrollo que reciben servicios de manera oportuna</p> <p>Porcentaje de tutores inscritos con pruebas positivas de violencia doméstica que reciben información de referencia a los recursos existentes con respecto al tema.</p>  |

Tras obtenerse el conjunto de áreas de referencia, una serie de objetivos de medición, preguntas y métricas han de ser desarrollados, lo anterior mediante el empleo de un enfoque Goal Question Metric (GQM), propuesto por (Basili & Weiss, 1984).

GQM es un modelo de medición basado en la siguiente suposición: para que una organización pueda medir sus objetivos de manera intencionada, primero debe especificarlos para sí misma y sus proyectos, para posteriormente realizar una trazabilidad hacia los datos que los definirán operativamente mediante la refinación de una serie de preguntas; finalmente proporciona un marco para la recolección e interpretación de datos con respecto a los objetivos establecidos mediante una estructura jerárquica (Basili, Caldiera, & Rombach, 1994), presentada en la Figura 16.

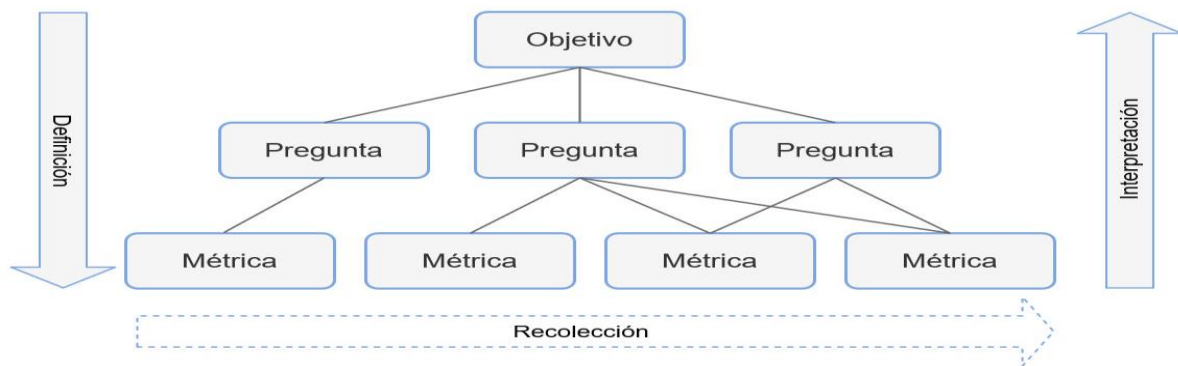


Figura 16. Jerarquía del modelo GQM

#### 4.1.1.1. Objetivos de medición

Una vez mencionadas las áreas de referencia empleadas para la medición del rendimiento, una serie de objetivos de medición han de ser creados por cada área.

Para este trabajo de investigación se propone la generación del objetivo de medición asociado a la primera área de referencia listada (Mejoras en la salud materna, neonatal e infantil), lo anterior debido a que es el área de que contiene el mayor número de indicadores

con respecto a las 5 áreas restantes y que parte de la información obtenida en la misma es utilizada a través de las demás áreas, como las familias, los niños inscritos y las visitas realizadas a estos.

Para la identificación del objetivo de medición han de emplearse los siguientes elementos propuestos por (Basili, Caldiera, & Rombach, 1994), mediante el empleo de un enfoque GQM:

- Propósito de la medición
- Problema por resolver mediante la medición
- Objeto relacionado a la medición
- Punto de vista de la medición
- Contexto en el que se realiza la medición

En la Tabla 25 se presentan los detalles del objetivo de medición identificado.

*Tabla 25. Objetivo de medición identificado mediante el empleo de un enfoque GQM.*

| Identificador | Elemento       | Descripción                                  |
|---------------|----------------|--|
| OM-01         | Propósito      | Mejorar                                      |
|               | Problema       | Salud  |
|               | Objeto         | Madres e hijos neonatales                    |
|               | Punto de vista | Usuario de un sistema de seguimiento de caso |
|               | Contexto       | Dentro de un programa de visitas a domicilio |

Dado lo anterior se procede con la definición del objetivo mediante el siguiente enunciado:

- **OM-01:** Analizar el nivel de salud de madres e hijos neonatales con el propósito de mejorarla, desde el punto de vista de un usuario de un sistema de seguimiento de caso y dentro de un programa de visitas a domicilio.

#### **4.1.1.2. Preguntas**

A continuación, se listan las preguntas generadas para la ayuda en el cumplimiento del objetivo de medición definido previamente, lo anterior haciendo uso de las medidas de rendimiento propuestas por (Health Resources & Services Administration).

Preguntas identificadas para dar cumplimiento al objetivo OM-01:

- **P-01.** De las madres embarazadas inscritas antes de la semana 37 de gestación, ¿Cuál es el porcentaje de niños inscritos nacidos antes de este tiempo?
- **P-02.** De los niños inscritos (entre las madres inscritas embarazadas) durante al menos 6 meses y con una edad entre 6 y 12 meses, ¿Cuántos fueron amamantados antes de cumplir 6 meses de edad?
- **P-03.** De las madres inscritas durante al menos 3 meses, ¿Cuántas son sometidas a pruebas de depresión dentro de los 3 meses posteriores a la inscripción (para aquellas no inscritas embarazadas) o dentro de los 3 meses posteriores al parto (para aquellas inscritas embarazadas)?
- **P-04.** De los niños inscritos, ¿Cuántos recibieron la última visita recomendada según el calendario de la Academia Americana de Pediatría (AAP)?
- **P-05.** ¿Cuál es el porcentaje de madres inscritas embarazadas o dentro de los 30 días posteriores al parto que recibieron una visita posparto dentro de las 8 semanas del parto?

- **P-06.** De los cuidadores primarios inscritos durante al menos 3 meses y que reportaron consumir tabaco o cigarrillos al inscribirse, ¿Cuántos fueron remitidos a servicios o asesoramiento para dejar de fumar dentro de los 3 meses posteriores a la inscripción?

#### **4.1.1.3. Métricas**

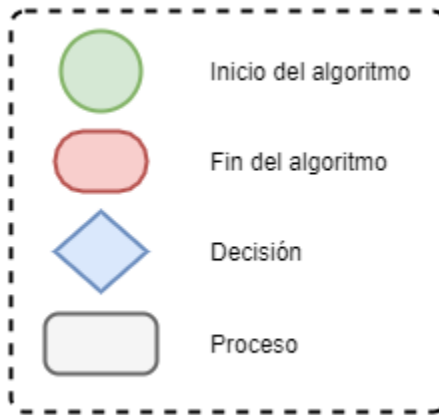
A continuación, se muestran las métricas a emplear para dar respuesta a cada pregunta generada y la lógica de negocios a seguir para realizar el cálculo del rendimiento obtenido.

Adicionalmente se proponen las siguientes dos métricas genéricas empleadas en cada pregunta identificada:

- **MG01.** Número de individuos (familias o niños) que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a una pregunta.
- **MG02.** Número de individuos (familias o niños) excluidos del cálculo del rendimiento asociado a una pregunta.

La representación del algoritmo empleado para obtener el valor correspondiente a cada métrica se realizará mediante el empleo de diagramas de flujo, utilizando la notación presentada en la Figura 17.





*Figura 17. Notación empleada en un diagrama de flujo.*

- *Pregunta 1*

Para realizar el cálculo del rendimiento asociado a la pregunta 1 se definen las siguientes métricas:

- **M01.** Número de nacimientos de madres inscritas embarazadas antes de la semana 37 de gestación.
- **M02.** Número de nacimientos prenatales (antes de la semana 37 de gestación) de madres inscritas embarazadas antes de la semana 37 de gestación.

La Figura 18 presenta la lógica a seguir para obtener el valor correspondiente a cada métrica para esta pregunta.

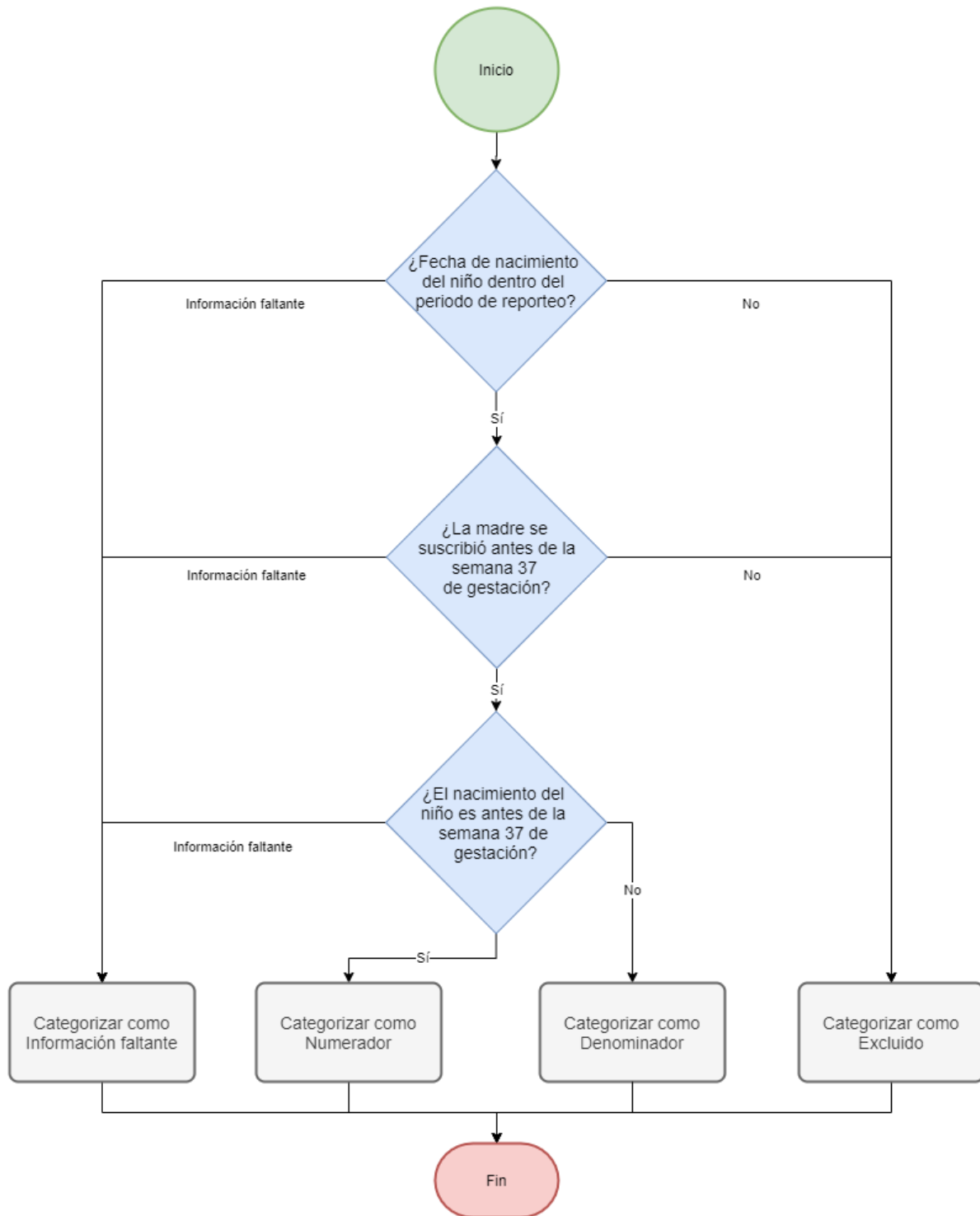


Figura 18. Algoritmo empleado para obtener las métricas de la pregunta 1.

Tras la ejecución del algoritmo indicado en el diagrama de flujo (Figura 18) se realiza una categorización de la información resultante, obteniendo 4 categorías (Tabla 26).

Tabla 26. Categorización de los datos para la pregunta 1.

| Categoría            | Métrica asociada | Descripción   |
|----------------------|------------------|---|
| Denominador          | M01              | Número de nacimientos de madres inscritas embarazadas antes de la semana 37 de gestación.   |
| Numerador            | M02              | Número de nacimientos prenatales (antes de la semana 37 de gestación) de madres inscritas embarazadas antes de la semana 37 de gestación. |
| Información faltante | MG01             | Número de niños que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a la pregunta 1.     |
| Excluido             | MG02             | Número de niños excluidos del cálculo del rendimiento asociado a pregunta 1.  |

Una vez obtenidos los valores correspondientes al numerador y denominador la siguiente fórmula ha de ser aplicada para la obtención del rendimiento:

$$\text{Indicador de Rendimiento 1} = \frac{M02}{M01 + M02} * 100\%$$

- *Pregunta 2*

Para realizar el cálculo del rendimiento asociado a la pregunta 2 se definen las siguientes métricas:

- **M03.** Número de niños activos entre 6 y 12 meses de edad inscritos prenatales en una agencia implementadora.

- M04.** Número de niños activos entre 6 y 12 meses de edad inscritos prenatales en una agencia implementadora y que fueron amamantados a los 6 meses de edad.

La Figura 19 presenta la lógica a seguir para obtener el valor correspondiente a cada métrica para esta pregunta.

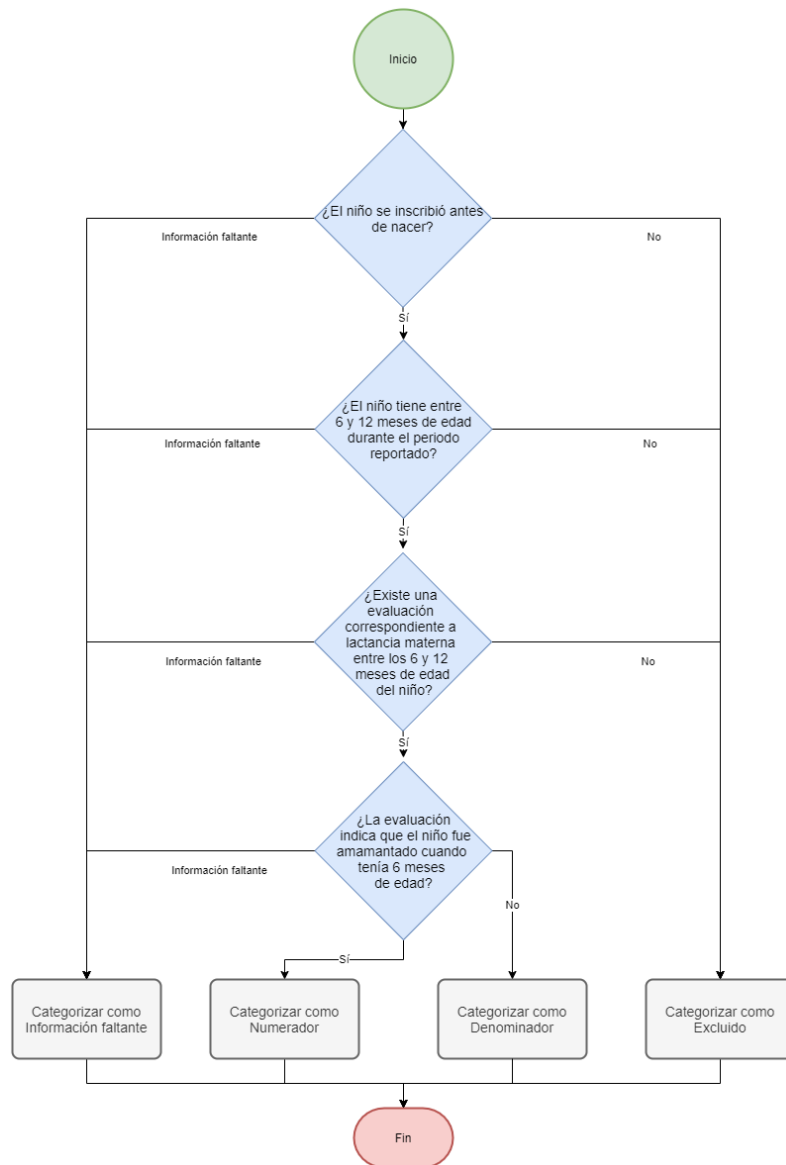


Figura 19. Algoritmo empleado para obtener las métricas de la pregunta 2.

Tras la ejecución del algoritmo indicado en el diagrama de flujo (Figura 19) se realiza una categorización de la información resultante, obteniendo 4 categorías (Tabla 27).

Tabla 27. Categorización de los datos para la pregunta 2.

| Categoría            | Métrica asociada | Descripción   |
|----------------------|------------------|---|
| Denominador          | M03              | Número de niños activos entre 6 y 12 meses de edad inscritos prenatales en una agencia implementadora.  |
| Numerador            | M04              | Número de niños activos entre 6 y 12 meses de edad inscritos prenatales en una agencia implementadora y que fueron amamantados a los 6 meses de edad. |
| Información faltante | MG01             | Número de niños que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a la pregunta 2.                 |
| Excluido             | MG02             | Número de niños excluidos del cálculo del rendimiento asociado a pregunta 2.  |

Una vez obtenidos los valores correspondientes al numerador y denominador la siguiente fórmula ha de ser aplicada para la obtención del rendimiento:

$$\text{Indicador de Rendimiento 2} = \frac{M04}{M03 + M04} * 100\%$$

- *Pregunta 3*

Para realizar el cálculo del rendimiento asociado a la pregunta 3 se definen las siguientes métricas:

- **M05.** Número de madres activas durante al menos 3 meses desde la fecha de inscripción o la fecha de parto para mujeres inscritas embarazadas.

- **M06.** Número de madres a las que se les aplicó una evaluación de depresión durante los primeros tres meses de actividad, comenzando por la fecha de inscripción o la fecha de parto para mujeres inscritas embarazadas.

La Figura 20 presenta la lógica a seguir para obtener el valor correspondiente a cada métrica para esta pregunta.

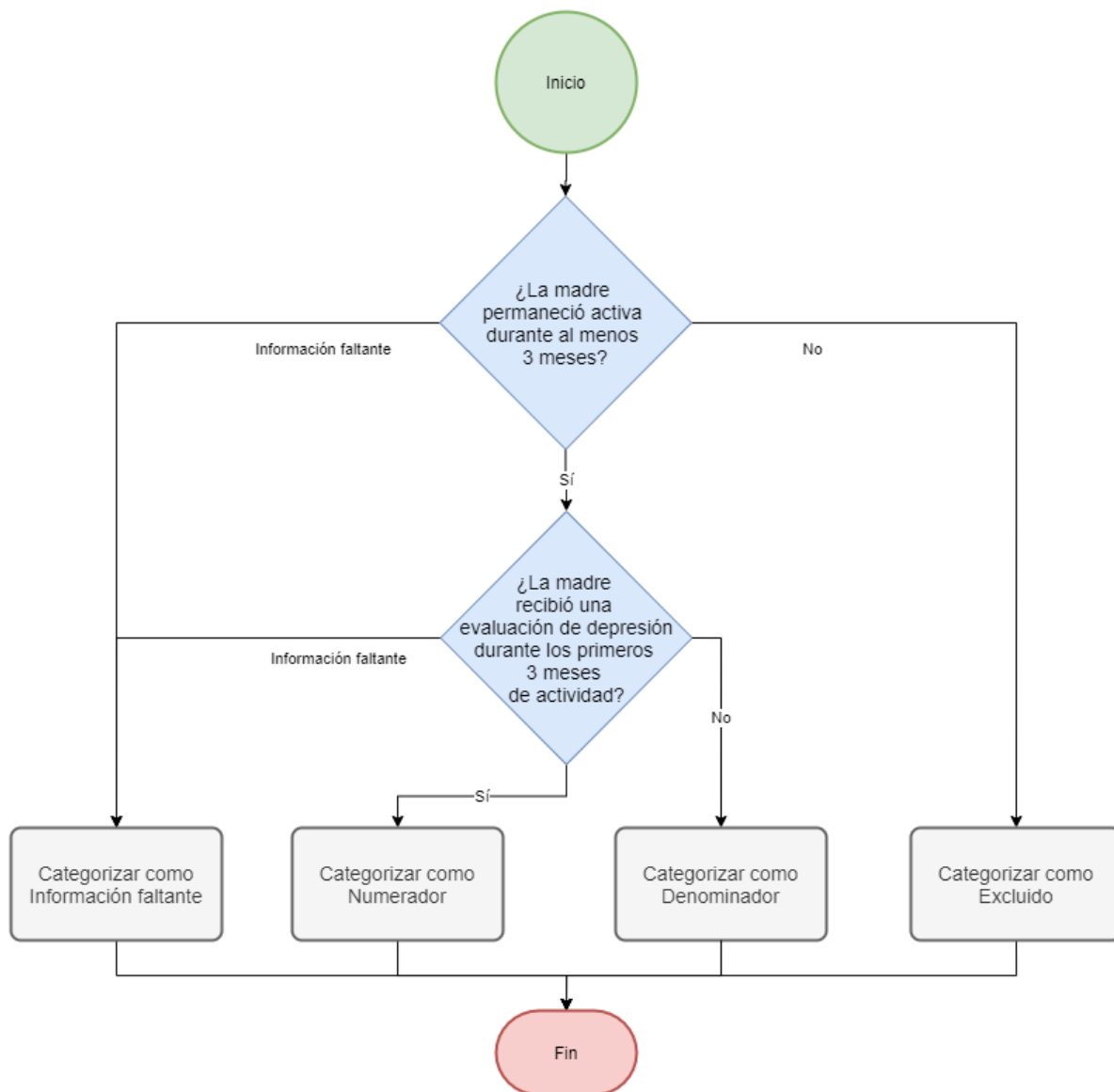


Figura 20. Algoritmo empleado para obtener las métricas de la pregunta 3.

Tras la ejecución del algoritmo indicado en el diagrama de flujo (Figura 20) se realiza una categorización de la información resultante, obteniendo 4 categorías (Tabla 28).

Tabla 28. Categorización de los datos para la pregunta 3.

| Categoría            | Métrica asociada | Descripción   |
|----------------------|------------------|---|
| Denominador          | M05              | Número de madres activas durante al menos 3 meses desde la fecha de inscripción o la fecha de parto para mujeres inscritas embarazadas.   |
| Numerador            | M06              | Número de madres a las que se les aplicó una evaluación de depresión durante los primeros tres meses de actividad, comenzando por la fecha de inscripción o la fecha de parto para mujeres inscritas embarazadas. |
| Información faltante | MG01             | Número de niños que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a la pregunta 3.   |
| Excluido             | MG02             | Número de niños excluidos del cálculo del rendimiento asociado a pregunta 3.  |

Una vez obtenidos los valores correspondientes al numerador y denominador la siguiente fórmula ha de ser aplicada para la obtención del rendimiento:

$$\text{Indicador de Rendimiento 3} = \frac{M06}{M05 + M06} * 100\%$$

- *Pregunta 4*

Para realizar el cálculo del rendimiento asociado a la pregunta 4 se definen las siguientes métricas:

- **M07.** Número de niños inscritos.
- **M08.** Número de niños inscritos que recibieron su última visita para seguimiento de desarrollo del niño basada en el calendario de la Academia Americana de Pediatría.

La Figura 21 presenta la lógica a seguir para obtener el valor correspondiente a cada métrica para esta pregunta.



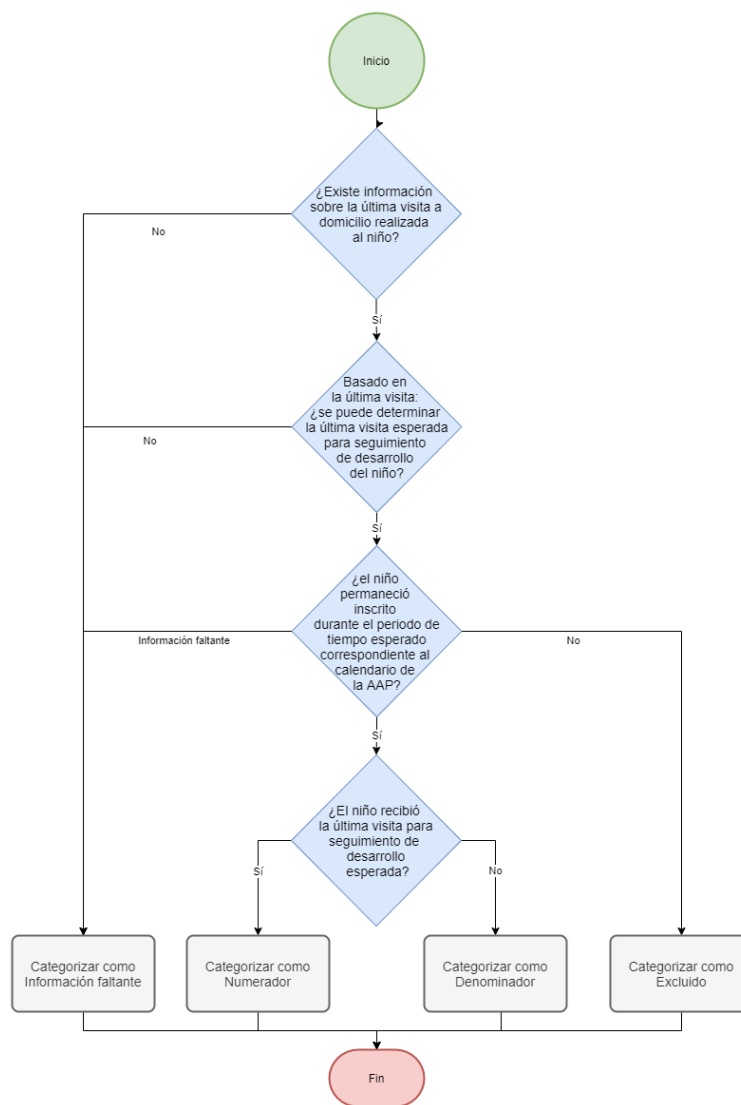


Figura 21. Algoritmo empleado para obtener las métricas de la pregunta 4.

Tras la ejecución del algoritmo indicado en el diagrama de flujo (Figura 21) se realiza una categorización de la información resultante, obteniendo 3 categorías (Tabla 29).

Tabla 29. Categorización de los datos para la pregunta 4.

| Categoría            | Métrica asociada | Descripción  |
|----------------------|------------------|--|
| Denominador          | M07              | Número de niños inscritos.   |
| Numerador            | M08              | Número de niños inscritos que recibieron su última visita para seguimiento de desarrollo del niño basada en el calendario de la Academia Americana de Pediatría. |
| Información faltante | MG01             | Número de niños que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a la pregunta 4.                            |

Una vez obtenidos los valores correspondientes al numerador y denominador la siguiente fórmula ha de ser aplicada para la obtención del rendimiento:

$$\text{Indicador de Rendimiento 4} = \frac{M08}{M07 + M08} * 100\%$$

- *Pregunta 5*

Para realizar el cálculo del rendimiento asociado a la pregunta 5 se definen las siguientes métricas:

- **M09.** Número de madres que se inscribieron antes del nacimiento del niño o dentro de los 30 días posteriores al parto y permanecieron inscritas durante al menos 8 semanas (56 días) después del parto.
- **M10.** Número de madres que se inscribieron antes del nacimiento del niño o dentro de los 30 días posteriores al parto que recibieron una visita posparto con un proveedor de atención médica dentro de las 8 semanas (56 días) posteriores al parto.

La Figura 22 presenta la lógica a seguir para obtener el valor correspondiente a cada métrica para esta pregunta.

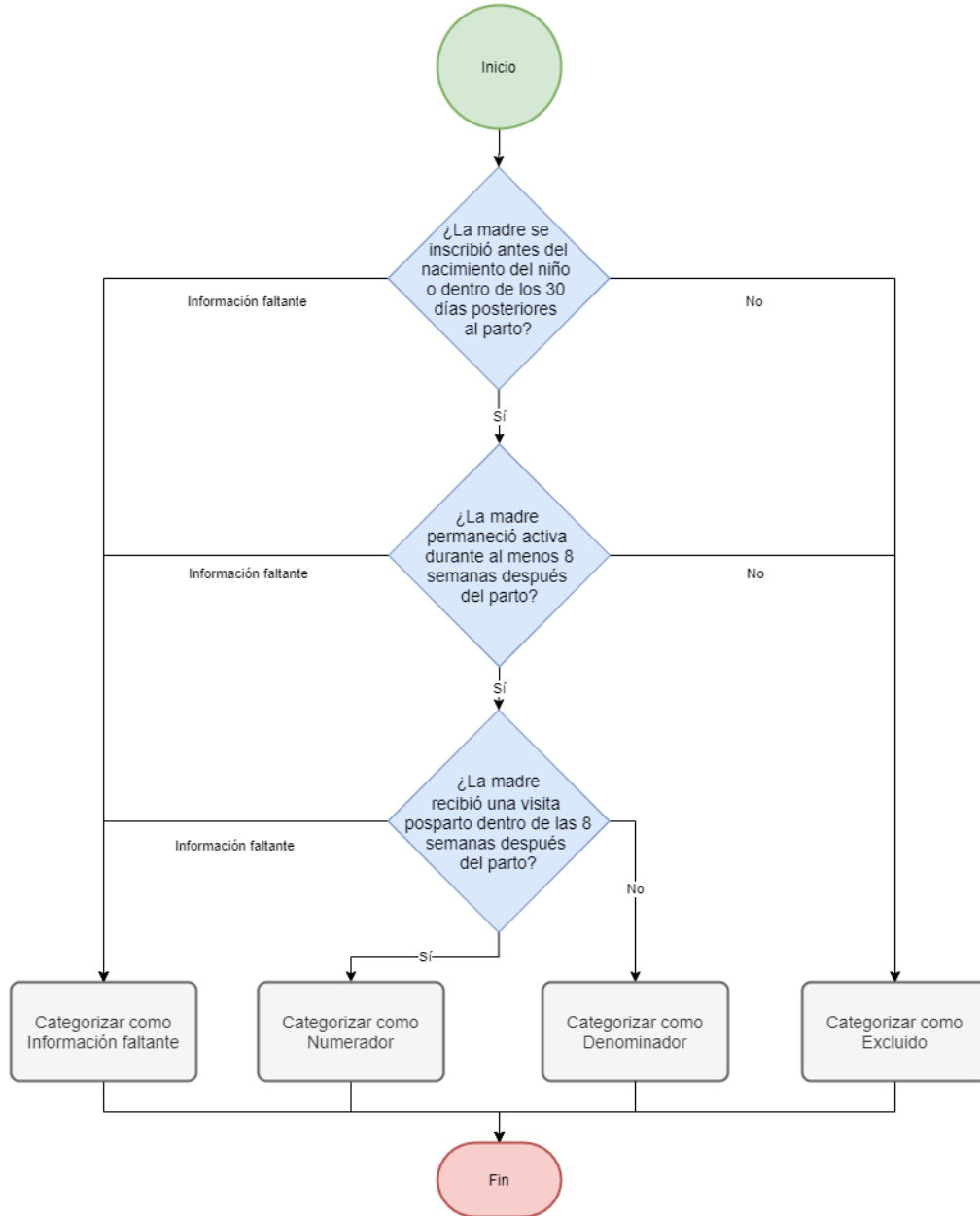


Figura 22. Algoritmo empleado para obtener las métricas de la pregunta 5.

Tras la ejecución del algoritmo indicado en el diagrama de flujo (Figura 22) se realiza una categorización de la información resultante, obteniendo 4 categorías (Tabla 30).

Tabla 30. Categorización de los datos para la pregunta 5.

| Categoría            | Métrica asociada | Descripción  |
|----------------------|------------------|--|
| Denominador          | M09              | Número de madres que se inscribieron antes del nacimiento del niño o dentro de los 30 días posteriores al parto y permanecieron inscritas durante al menos 8 semanas (56 días) después del parto.  |
| Numerador            | M10              | Número de madres que se inscribieron antes del nacimiento del niño o dentro de los 30 días posteriores al parto que recibieron una visita posparto con un proveedor de atención médica dentro de las 8 semanas (56 días) posteriores al parto. |
| Información faltante | MG01             | Número de niños que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a la pregunta 5.  |
| Excluido             | MG02             | Número de niños excluidos del cálculo del rendimiento asociado a pregunta 5.   |

Una vez obtenidos los valores correspondientes al numerador y denominador la siguiente fórmula ha de ser aplicada para la obtención del rendimiento:

$$\text{Indicador de Rendimiento 5} = \frac{M10}{M09 + M10} * 100\%$$

- *Pregunta 6*

Para realizar el cálculo del rendimiento asociado a la pregunta 6 se definen las siguientes métricas:

- **M11.** Número de tutores (padre o madre) inscritos que informaron haber consumido tabaco o cigarrillos en el momento de la inscripción y estuvieron inscritos durante al menos 3 meses.
- **M12.** Número de tutores (padre o madre) inscritos que informaron haber consumido tabaco o cigarrillos en el momento de la inscripción y fueron remitidos a servicios o asesoramiento para dejar de fumar dentro de los 3 meses posteriores a la inscripción.

La Figura 23 presenta la lógica a seguir para obtener el valor correspondiente a cada métrica para esta pregunta.

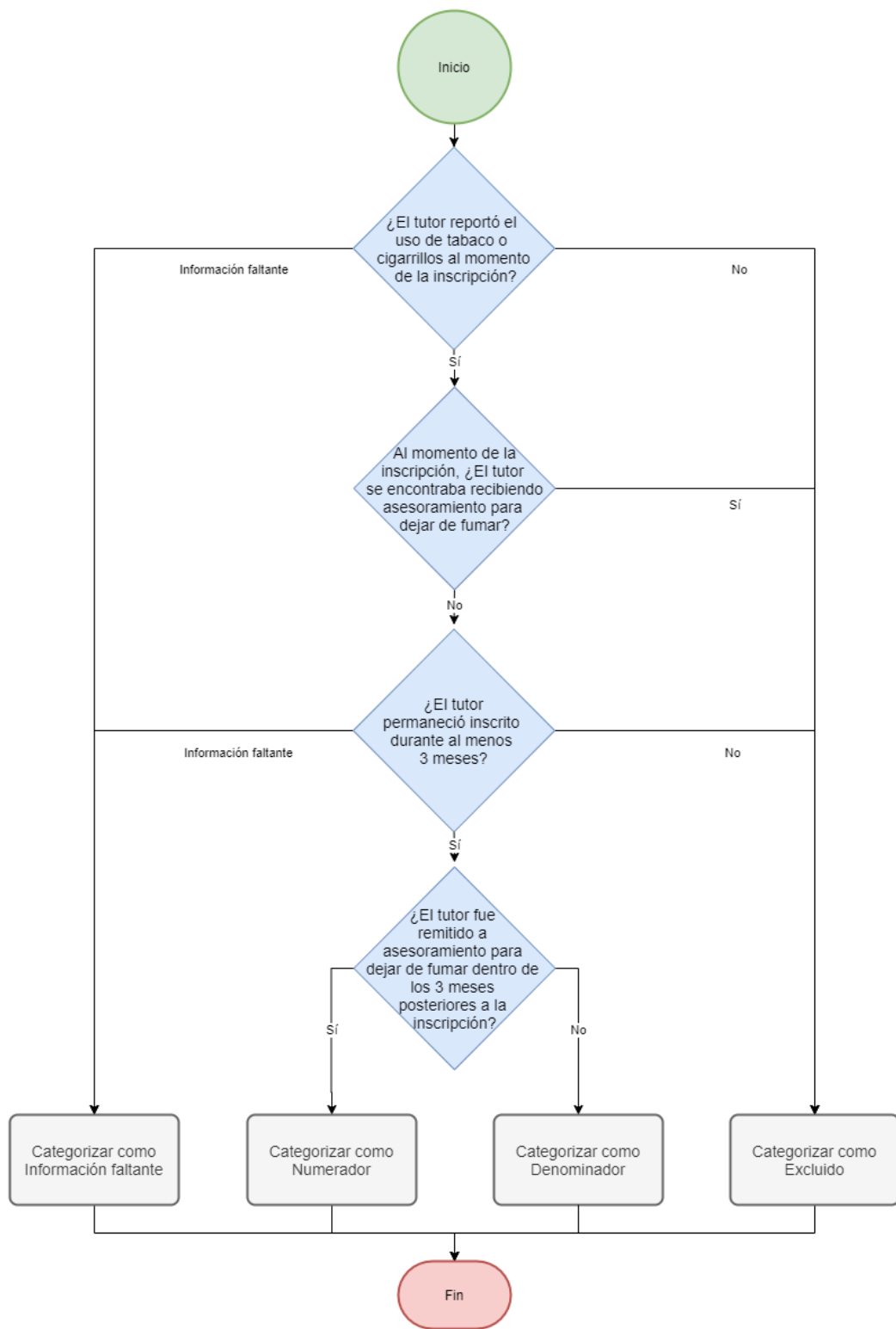


Figura 23. Algoritmo empleado para obtener las métricas de la pregunta 6.

Tras la ejecución del algoritmo indicado en el diagrama de flujo (Figura 23) se realiza una categorización de la información resultante, obteniendo 4 categorías (Tabla 31).

Tabla 31. Categorización de los datos para la pregunta 6.

| Categoría            | Métrica asociada | Descripción  |
|----------------------|------------------|--|
| Denominador          | M11              | Número de tutores (padre o madre) inscritos que informaron haber consumido tabaco o cigarrillos en el momento de la inscripción y estuvieron inscritos durante al menos 3 meses.   |
| Numerador            | M12              | Número de tutores (padre o madre) inscritos que informaron haber consumido tabaco o cigarrillos en el momento de la inscripción y fueron remitidos a servicios o asesoramiento para dejar de fumar dentro de los 3 meses posteriores a la inscripción. |
| Información faltante | MG01             | Número de niños que cuenten con alguna información faltante para la realización del cálculo del rendimiento asociado a la pregunta 6.  |
| Excluido             | MG02             | Número de niños excluidos del cálculo del rendimiento asociado a pregunta 6.   |

Una vez obtenidos los valores correspondientes al numerador y denominador la siguiente fórmula ha de ser aplicada para la obtención del rendimiento:

$$\text{Indicador de Rendimiento 6} = \frac{M12}{M11 + M12} * 100\%$$

#### **4.1.2. Generación de modelo de la base de datos**

Tras la obtención de los objetivos de medición, las preguntas a responder y las métricas asociadas a dichas preguntas, se genera un modelo de base de datos capaz de proveer la información necesaria para realizar la medición del rendimiento de acuerdo con lo planteado por (Health Resources & Services Administration), presentado en la Figura 24.



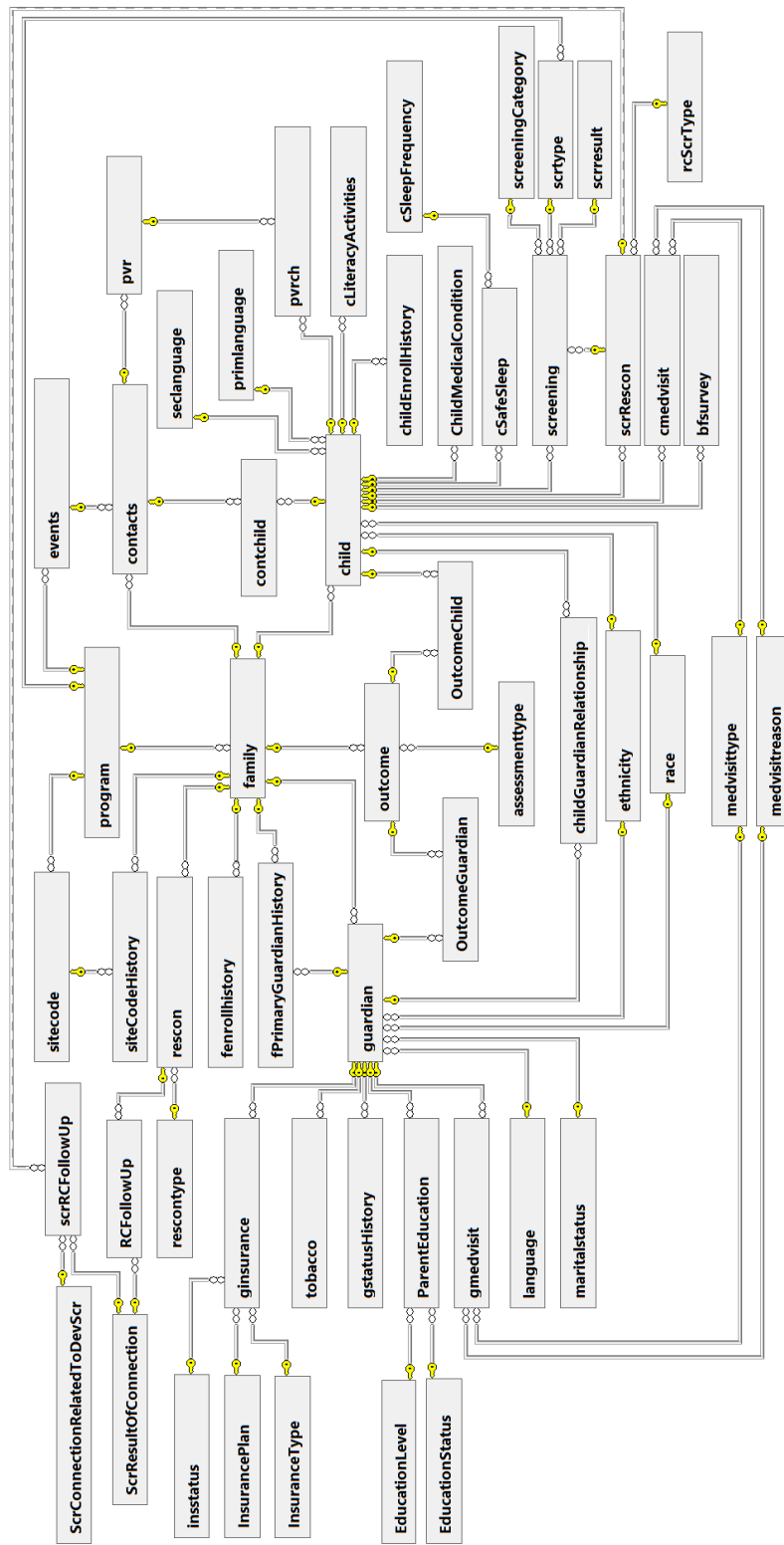


Figura 24. Modelo de la base de datos a emplear

### **4.1.3. Generación de diseño arquitectural**

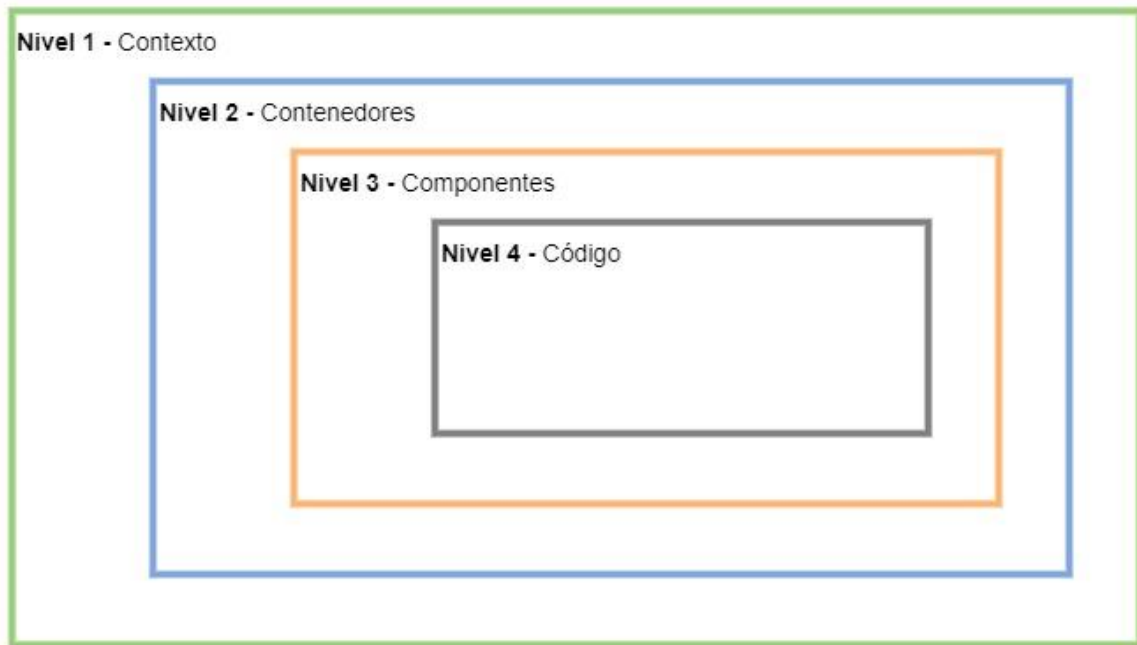
Tras obtener el modelo de datos se procede a la generación del diseño arquitectural, tomando en cuenta los puntos mencionados durante la selección tecnológica en el capítulo 2 de este trabajo de investigación.

Con el objetivo de asegurar que los datos obtenidos cumplan con el requisito de calidad ha de aplicarse un proceso de limpieza y filtrado de información al momento de consultar los paneles de información proporcionados por la herramienta de inteligencia de negocios. Una arquitectura en capas permitirá a los usuarios aplicar dicho proceso, adicionalmente beneficiará el desacoplamiento al permitir que el código necesario para la consulta de información de la fuente de datos se almacene en una capa independiente a aquella empleada para el proceso de limpieza y filtrado de datos.

Ya que se requiere de reporte operativo y que este no afecte el desempeño de la herramienta, se propone el empleo de cacheo de datos como táctica arquitectónica, mismo que ha de implementarse mediante el empleo de una base de datos en memoria a refrescarse periódicamente, sustituyendo el uso de almacenes de datos. Disminuyendo así el número de consultas realizadas a la base de datos de producción, manteniendo una sensación de reporte en tiempo real.

Para la correcta representación de la arquitectura a desarrollar ha de emplearse el modelo C4. El modelo C4 es un enfoque para diagramar la arquitectura de software, basado en abstracciones que reflejan cómo los arquitectos y desarrolladores de software piensan y construyen software (Brown, 2018). El modelo C4 está compuesto por 4 niveles de abstracción

que describen la estructura estática de un sistema de software (Figura 25), en términos de contenedores, componentes y código.



*Figura 25. Niveles de abstracción del modelo C4.*

#### **4.1.3.1. Contexto**

El diagrama de contexto presenta el sistema de software junto con los usuarios y sistemas con los que interactúa, mostrando la comunicación que existe entre cada elemento presente en este diagrama. La Figura 26 presenta el diagrama de contexto desarrollado para el sistema de inteligencia de negocios propuesto.



Figura 26. Diagrama de contexto para el sistema de inteligencia de negocios.

#### 4.1.3.2. Contenedores

Un contenedor es una unidad de despliegue o ejecución independiente, capaz de almacenar datos o ejecutar código. El sistema de inteligencia de negocios propuesto se compone de tres contenedores, incluyendo la aplicación Web empleada para la visualización del rendimiento obtenido por una agencia implementadora del programa federal MIECHV tras la prestación de servicios de visitas a domicilio (Figura 27).



Figura 27. Diagrama de contenedores para el sistema de inteligencia de negocios.

### 4.1.3.3. Componentes

Un diagrama de componentes muestra la forma en la que un contenedor se desarrolla a través de una serie de componentes de software y sus responsabilidades. Para la aplicación Web del sistema de inteligencia de negocios se proponen un total de 5 componentes, presentados en la Figura 28.

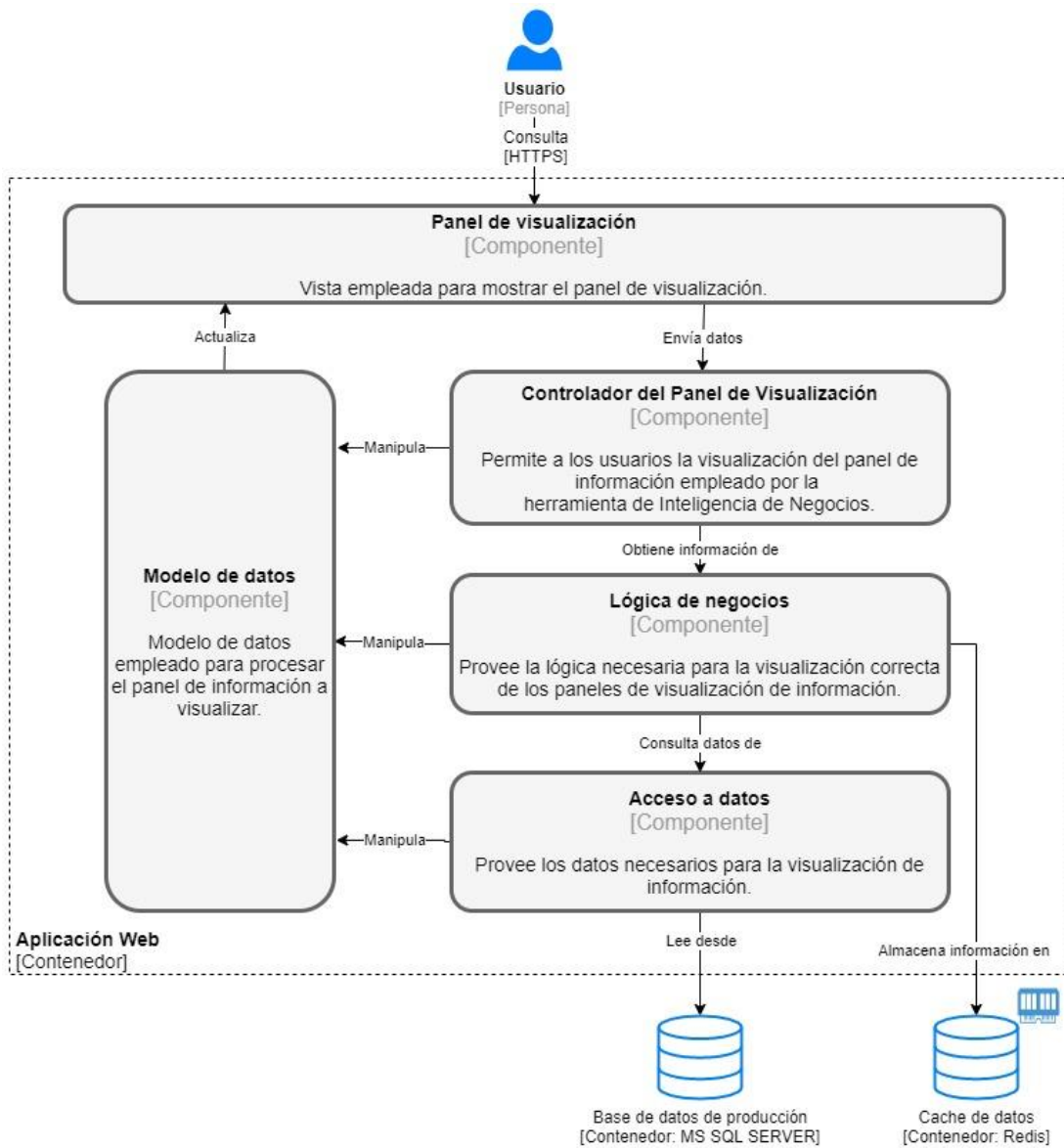


Figura 28. Diagrama de componentes para el sistema de inteligencia de negocios.

#### 4.1.3.4. Código

Un diagrama de código muestra un acercamiento a la implementación detallada de los componentes presentes en el sistema de software y es recomendado para los componentes más complejos o importantes en el mismo. Para el componente correspondiente al modelo de datos del sistema de inteligencia de negocios se propone el diagrama de código presentado en la Figura 29.

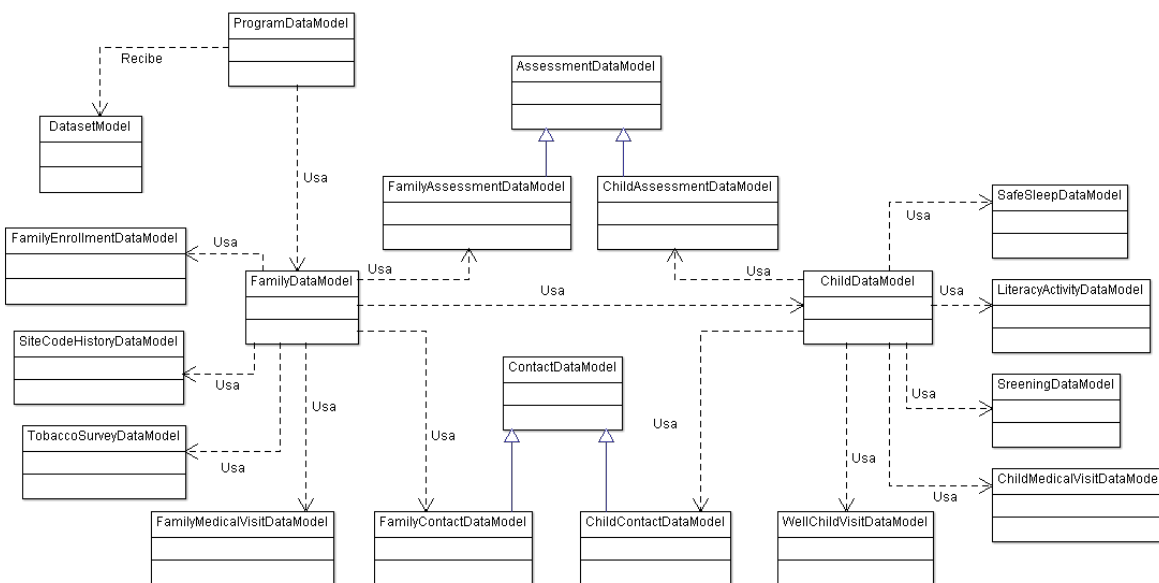
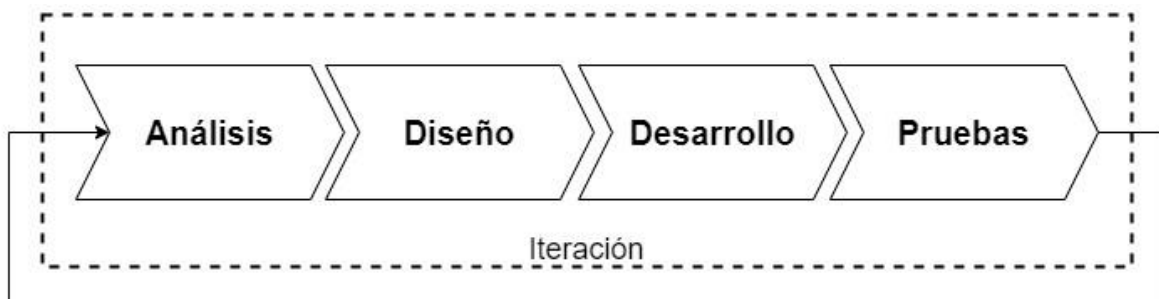


Figura 29. Diagrama de código para el modelo de datos del sistema de inteligencia de negocios.

#### 4.1.4. Desarrollo de la herramienta

A continuación, se presenta la herramienta desarrollada a partir de lo plasmado en las secciones anteriores, mediante una serie de figuras se demuestra la funcionalidad y uso de esta. También se detallan las especificaciones técnicas de la herramienta a fin de facilitar su posterior mantenimiento y despliegue.

Para el desarrollo de la herramienta y tomando en cuenta los factores de éxito y beneficios identificados en la revisión sistemática, se emplea una metodología de desarrollo de software tradicional basada en un ciclo de vida iterativo, compuesto por las siguientes actividades ejecutadas de forma secuencial. Cada iteración se ejecuta cuantas veces sea necesario hasta obtener el resultado esperado por la herramienta de inteligencia de negocios (Figura 30).



*Figura 30. Ciclo de vida empleado para el desarrollo de la herramienta.*

#### **4.1.4.1. Estado actual**

La Figura 31 muestra el estado actual de la herramienta empleada para la visualización de rendimiento mediante el sistema Visit Tracker Web (DataKeeper Technologies, 2020).

Observándose en la parte superior los filtros empleados para la visualización del rendimiento. Seguido de una serie de tablas, correspondientes al rendimiento obtenido por una agencia implementadora.

## MIECHV Form 2 IL

Program Name: Demo Parents As Teachers Program Code: 8888888 Date Range: 10/1/2020 - 5/31/2021 Site Codes: MIECHV(8888888)  
 Use the 1st MIECHV Start Date to calculate this report: N MIECHV Target Child only: Y Exclude families Exited or Inactive / On Hold due to COVID-19: N

Prepared 6/8/2021

Export all Constructs to CSV [Get CSV](#)

### CONSTRUCT 1: PRETERM BIRTH [export to CSV](#)

|        | Numerator | Denominator | Outcome (%) | Missing | Definitions   |
|--------|-----------|-------------|-------------|---------|---|
| Totals | 0         | 0           | 0           | 0       | <ul style="list-style-type: none"> <li><b>Numerator:</b> Children from the denominator that were born more than 21 days before their Due Date.</li> <li><b>Denominator:</b> Children born in the reporting period that were enrolled prenatally more than 21 days before their Due Date.</li> <li><b>Missing:</b> Children from the denominator whose Due Date is in the reporting period, no DOB, and are not exited.</li> </ul> |

### CONSTRUCT 2: BREASTFEEDING [export to CSV](#)

|        | Numerator | Denominator | Outcome (%) | Missing | Definitions  |
|--------|-----------|-------------|-------------|---------|--|
| Totals | 0         | 0           | 0           | 0       | <ul style="list-style-type: none"> <li><b>Numerator:</b> Children from the denominator that have a Breastfeeding measurement = Yes between 180 days or more from DOB OR a Breastfeeding measurement = Weaned where date weaned is 180 days or more from DOB.</li> <li><b>Denominator:</b> Children enrolled as a prenatal child AND are 180 days from DOB in date range AND have at least 1 PVR dated after the child reaches 180 days of age AND have at least one breastfeeding entry on or after DOB</li> <li><b>Missing:</b> Children that are 180 days past Due Date in the date range and without DOB information and have not been exited. AND children with no entry on or after DOB who are equal to 180 days from DOB in date range</li> </ul> |

### CONSTRUCT 3: DEPRESSION SCREENING [export to CSV](#)

|        | Numerator | Denominator | Outcome (%) | Missing | Definitions  |
|--------|-----------|-------------|-------------|---------|--|
| 1      | 0         | 0           | 0           | N/A     | <b>Guardians enrolled without a prenatal child</b> <ul style="list-style-type: none"> <li><b>Numerator:</b> Primary guardians reported in the denominator that have an EPDS within the 90 day enrollment period.</li> <li><b>Denominator:</b> Primary guardians enrolled for 90 days in the date range. These guardians did not enroll with a prenatal child and the family has at least 1 PVR after the 90 day enrollment period.</li> <li><b>Missing:</b> Not applicable.</li> </ul>   |
| 2      | 0         | 0           | 0           | 0       | <b>Guardians enrolled pregnant</b> <ul style="list-style-type: none"> <li><b>Numerator:</b> Primary guardians reported in the denominator that have an EPDS within their enrollment period up to the child reaching 90 days of age.</li> <li><b>Denominator:</b> Primary guardians that enrolled pregnant, gave birth, and the child is 90 days or older in the report date range. The family must have at least 1 PVR after the child reaches 90 days of age.</li> <li><b>Missing:</b> Primary guardians that enrolled pregnant and the prenatal child is 90 days past the Due Date with no DOB information and has not been exited.</li> </ul> |
| Totals | 0         | 0           | 0           | 0       |  |

Figura 31. Estado actual de la herramienta empleada para la visualización de rendimiento mediante el sistema Visit Tracker Web.

#### 4.1.4.2. Especificaciones técnicas

La herramienta desarrollada emplea las siguientes tecnologías para su funcionamiento:

- **Visual Studio Community 2015.** Entorno de desarrollo integrado.
- **IIS 7.** Servidor Web.
- **ASP.NET MVC.** Framework de desarrollo de aplicaciones Web mediante el patrón arquitectónico Modelo Vista Controlador.



- **C#.** Lenguaje de programación orientado a objetos.
- **Redis 6.2.** Almacén de estructura de datos en memoria de código abierto, empleado como base de datos, caché y bróker de mensajes.
- **MSSQL Server 2012.** Sistema gestor de base de datos relacional.

#### 4.1.4.3. *Funcionalidad y uso*

Por cuestiones de seguridad y para dar acceso a la información correcta de acuerdo con la agencia implementadora que se encuentre empleando la herramienta, el acceso al panel de información requiere que se proporcione un nombre de usuario y contraseña (Figura 32).

The image shows a login form for the 'Visit Tracker Dashboard'. At the top, there is a dark navigation bar with the text 'Visit Tracker Dashboard' on the left and 'Log in' on the right. Below this bar, the text 'Log in.' is centered. The form consists of two input fields: 'UserName' and 'Password'. Below the 'Password' field is a button labeled 'Log in'.

*Figura 32. Inicio de sesión.*

Una vez iniciada la sesión el usuario es redireccionado a la página principal, en donde se presenta el panel de visualización de información, compuesto de una serie de filtros empleados para obtener la información de rendimiento de una agencia implementadora. Una vez seleccionados los filtros es necesario presionar el botón azul de refrescado para la visualización de la información deseada (Figura 33). Los filtros empleados por una agencia implementadora son los siguientes:

- **Programa.** Dentro de la herramienta un programa representa a una agencia implementadora de un modelo de visitas a domicilio a niños en edad temprana.

Por defecto una agencia implementadora solo puede seleccionarse a sí misma, aunque existe la posibilidad de obtener la información de rendimiento correspondiente a varias agencias implementadoras mediante el empleo de un nombre de usuario y contraseña con acceso a todas las agencias, proporcionado en la sección de anexos.

- **Rango de fechas.** Representa el período de tiempo a emplear para la obtención del rendimiento de una agencia implementadora.
- **Código de sitio.** Una agencia implementadora emplea códigos de sitio para identificar el origen de los fondos recibidos para la atención de ciertas familias. Mediante el empleo de este filtro una agencia implementadora puede obtener la información de rendimiento correspondiente únicamente a las familias y niños apoyados mediante el empleo de un origen de fondeo específico.
- **Solo niños MIECHV.** Este filtro es empleado para filtrar solo aquellas familias y niños que son atendidos mediante el empleo de fondos provenientes del programa federal MIECHV (Health Resources & Services Administration, 2020).
- **Fecha empleada.** Este filtro es empleado para seleccionar que fecha ha de ser empleada como fecha de inicio de participación de una familia en una agencia implementadora. Existen dos variantes: 1) Fecha de suscripción, 2) Fecha en que comienza a recibir apoyo por parte de MIECHV.

- **Familias COVID-19.** Dada la contingencia actual se ha optado por excluir de los cálculos de rendimiento a aquellas familias que por alguna cuestión asociada a COVID-19 dejaron de recibir servicios ya sea temporal o permanentemente.

*Figura 33. Filtros empleados por el panel de visualización.*

Mediante la siguiente combinación de filtros es posible obtener la información de rendimiento de una agencia implementadora solo para aquellas familias que se encuentren recibiendo apoyo por parte de MIECHV:

- **Código de sitio:** MIECHV
- **Solo niños MIECHV:** Sí
- **Fecha empleada:** Fecha de inicio en MIECHV

Lo anterior con la finalidad de ayudar a la determinación de rendimiento y posterior reporte de resultados a las autoridades correspondientes por parte de una agencia implementadora.

Tras procesar la información necesaria para la visualización de la información a mostrar, una serie de gráficos son generados en el panel de información, correspondientes a la información de rendimiento obtenida (Figura 34).

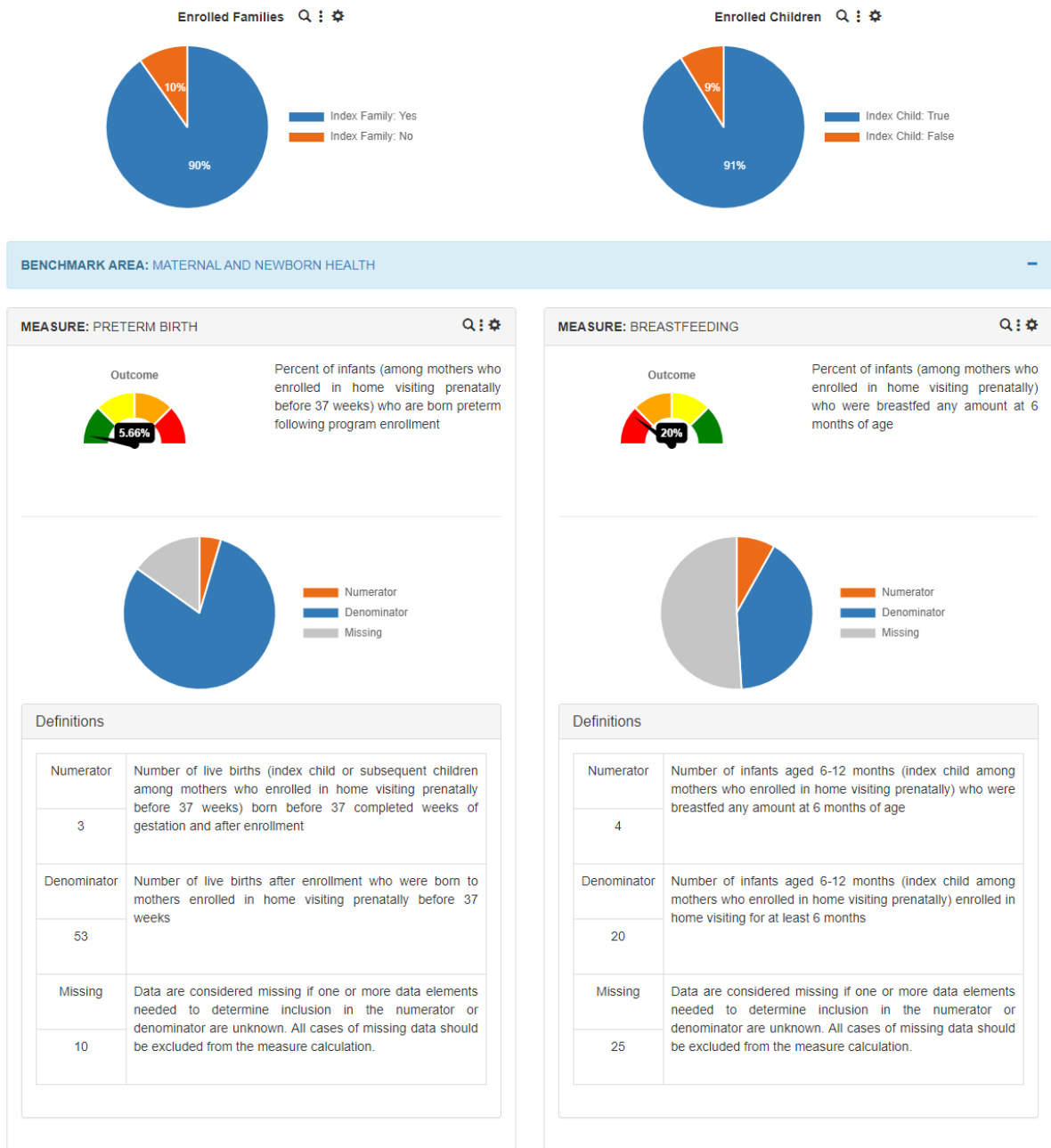


Figura 34. Gráficas correspondientes al rendimiento obtenido por una agencia implementadora.

Las primeras dos gráficas presentes en el panel de visualización de información corresponden a todas las familias y niños servidos actualmente por una agencia implementadora, realizando una categorización de quienes son tomados en cuenta para la

medición de rendimiento y quienes son ignorados de acuerdo con las indicaciones provistas por MIECHV para este reporte de rendimiento en específico (Figura 34).

Cada gráfica ofrece la posibilidad de visualizar los datos empleados para la obtención de los resultados correspondientes a cada indicador de rendimiento, mismos que se muestran en una ventana emergente ofreciendo la posibilidad de visualizar la información concreta de cada individuo, sea este niño o familia. Los datos mostrados en la ventana emergente incluyen información como el identificador único de cada individuo, la fecha en que comienza a participar en una agencia implementadora, código de sitio actual y la forma en que se categorizan los datos con respecto a la métrica que se consulte (Figura 35).

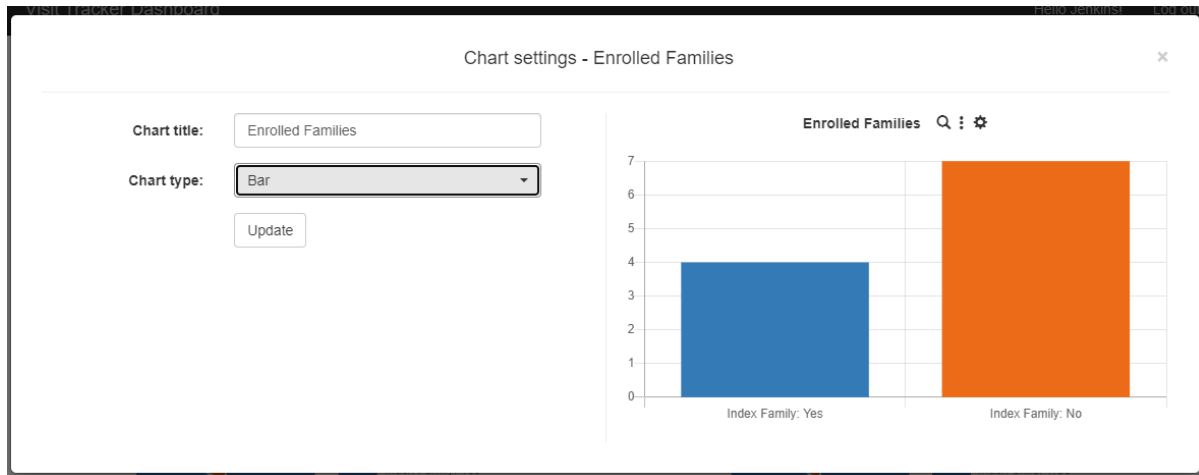
| Family ID | Family Name      | Index Family | Enroll Date | MIECHV Start Date | All-In-One Start Date | Date Used  | Exit Date  | Status | Due to COVID? | Site Code | Private Visit Records |
|-----------|------------------|--------------|-------------|-------------------|-----------------------|------------|------------|--------|---------------|-----------|-----------------------|
| 399989    | Johnson/Rains    | Yes          | 01/26/2017  |                   |                       | 01/26/2017 | 12/13/2019 | Exited |               | PI        | 3                     |
| 477033    | Callicutt        | Yes          | 03/19/2015  |                   |                       | 03/19/2015 | 06/30/2020 | Exited |               | PI        | 2                     |
| 610824    | Escobar-Gonzalez | Yes          | 09/19/2013  |                   |                       | 09/19/2013 |            | Active |               | PI        | 19                    |
| 611766    | Hussein          | Yes          | 01/09/2017  |                   |                       | 01/09/2017 | 12/11/2019 | Exited |               | PI        | 3                     |
| 641053    | Simmons          | Yes          | 02/12/2014  |                   |                       | 02/12/2014 |            | Active |               | PI        | 8                     |
| 680481    | Samano/Garcia    | Yes          | 04/15/2020  |                   |                       | 04/15/2020 |            | Active |               | PI        | 11                    |
| 687930    | Pinkston         | Yes          | 09/03/2014  |                   |                       | 09/03/2014 | 01/24/2020 | Exited |               | PI        | 7                     |
| 715150    | Serrano          | Yes          | 12/15/2014  |                   |                       | 12/15/2014 |            | Active |               | PI        | 16                    |
| 715162    | Byrd             | Yes          | 12/15/2014  |                   |                       | 12/15/2014 | 11/30/2019 | Exited |               | PI        | 5                     |
| 722228    | Trujillo         | Yes          | 01/22/2015  |                   |                       | 01/22/2015 |            | Active |               | PI        | 16                    |

Figura 35. Visualización de la información empleada para el cálculo de una métrica.

En la ventana emergente se ofrecen una serie de funcionalidades para facilitar la visualización de información a una agencia implementadora, incluyendo la posibilidad de ordenar los datos de manera ascendente o descendente con respecto a una columna en

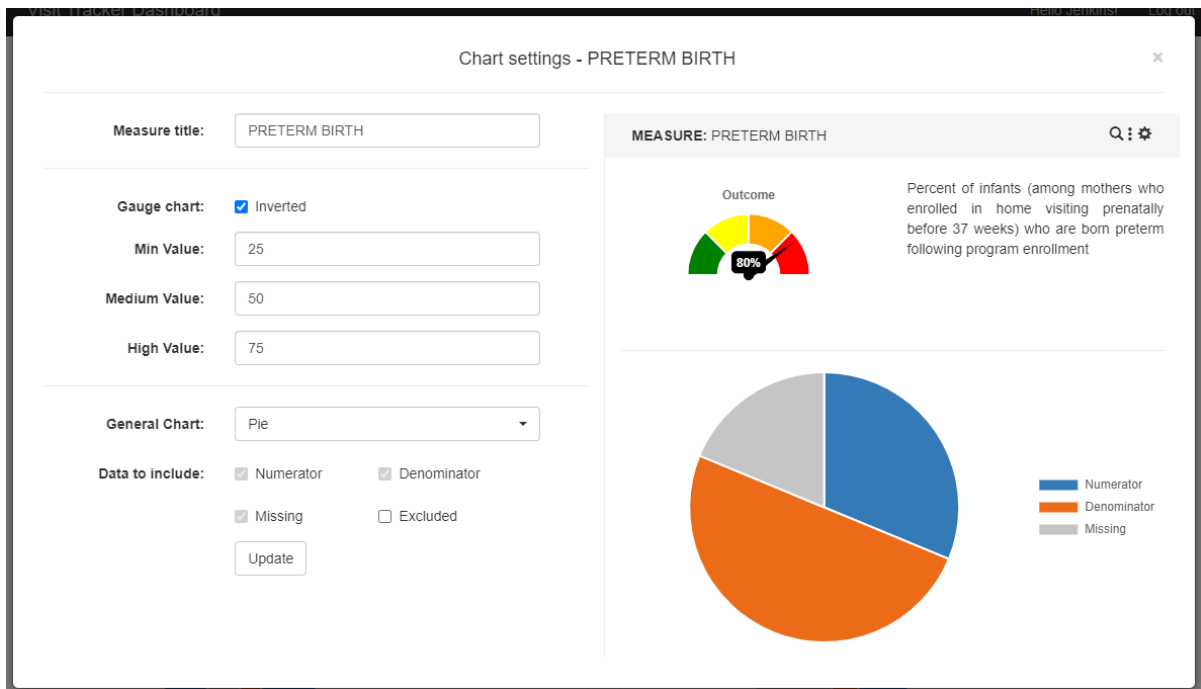
específico, paginación de la información a visualizar, búsqueda de información por columna y exportación de la información a un archivo CSV (Figura 35).

Cada gráfica presentada en el panel de información es configurable, permitiendo modificar el título de la gráfica y el tipo de visualización a emplear (Figura 36).



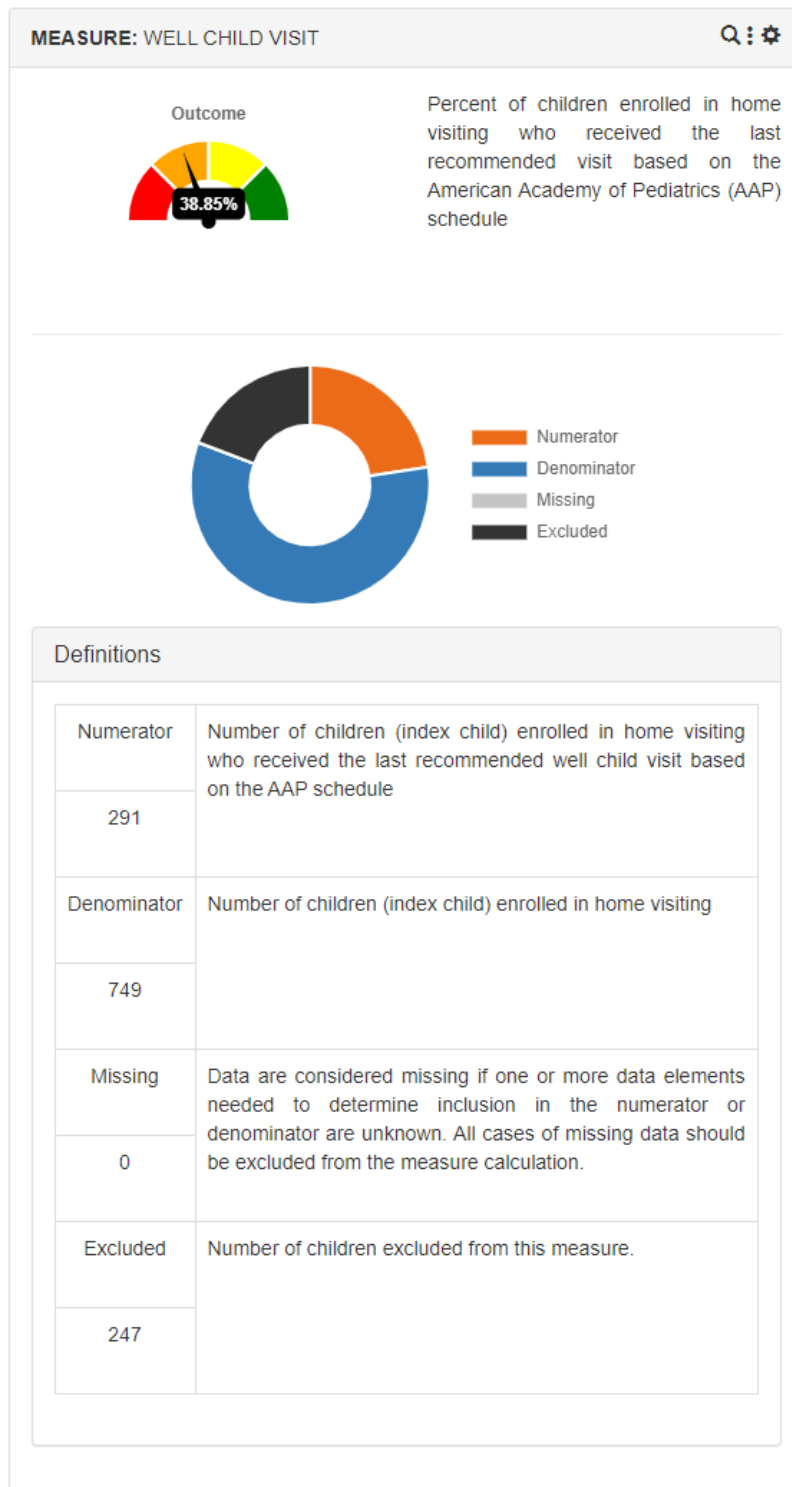
*Figura 36. Configuración básica de una gráfica.*

Para las gráficas correspondientes a cada métrica se ofrece la posibilidad de editar el título de esta, los valores máximos y mínimos del semáforo empleado para el despliegue del resultado de la medición de rendimiento, el tipo de visualización a emplear por los valores utilizados para el cálculo del rendimiento y la opción de incluir los valores excluidos para el cálculo de este (Figura 37).



*Figura 37. Configuración de una métrica de rendimiento.*

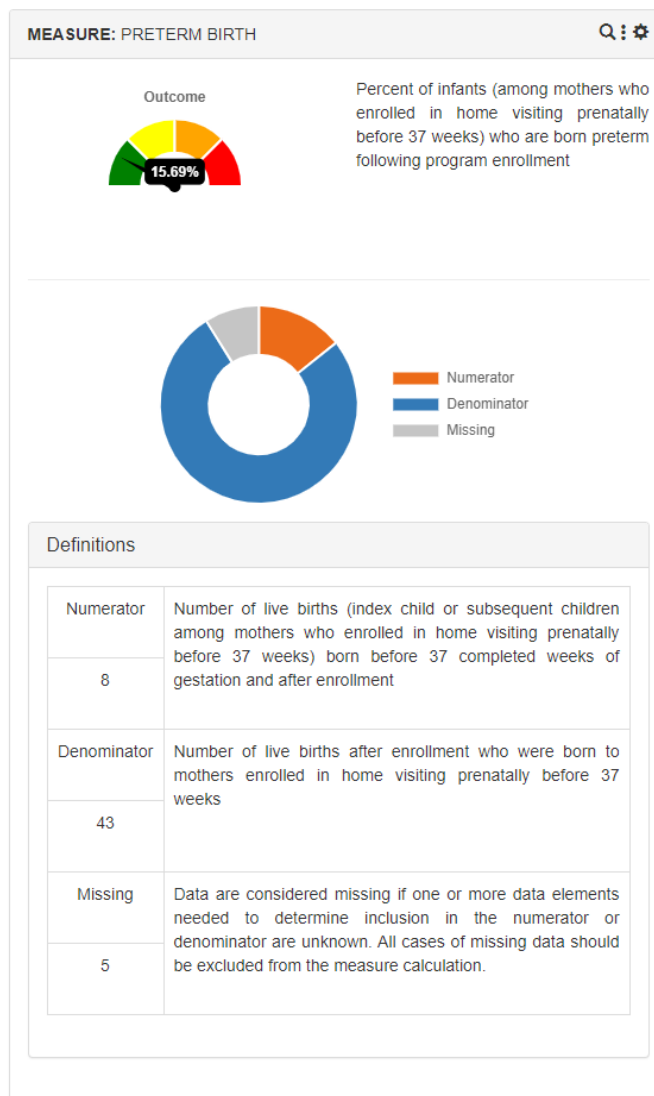
Cada métrica de rendimiento se compone de cuatro secciones, la primera sección corresponde al título, la segunda sección muestra el rendimiento obtenido junto con la descripción del cálculo correspondiente al mismo, la tercera sección muestra la información categorizada de acuerdo con lo propuesto para cada métrica y la cuarta sección muestra la definición para los datos categorizados como numerador, denominador, información perdida y excluido (Figura 38).



*Figura 38. Estructura de una métrica de rendimiento.*



La Figura 39 muestra el diseño de la primera métrica de rendimiento, correspondiente a la pregunta P-01, compuesta por el título, un botón para visualización de los datos empleados por esta métrica, un botón de configuración, el resultado de rendimiento 1, visualizado en la gráfica de semáforo presente, la descripción del indicador a la derecha del resultado, una gráfica correspondiente al total de datos categorizados en función del algoritmo correspondiente, y la definición de cada categoría.



*Figura 39. Diseño correspondiente a la primera métrica de rendimiento.*

La Figura 40 muestra el diseño de la segunda métrica de rendimiento, correspondiente a la pregunta P-02, compuesta por el título, un botón para visualización de los datos empleados por esta métrica, un botón de configuración, el resultado de rendimiento 2, visualizado en la gráfica de semáforo presente, la descripción del indicador a la derecha del resultado, una gráfica correspondiente al total de datos categorizados en función del algoritmo correspondiente, y la definición de cada categoría.

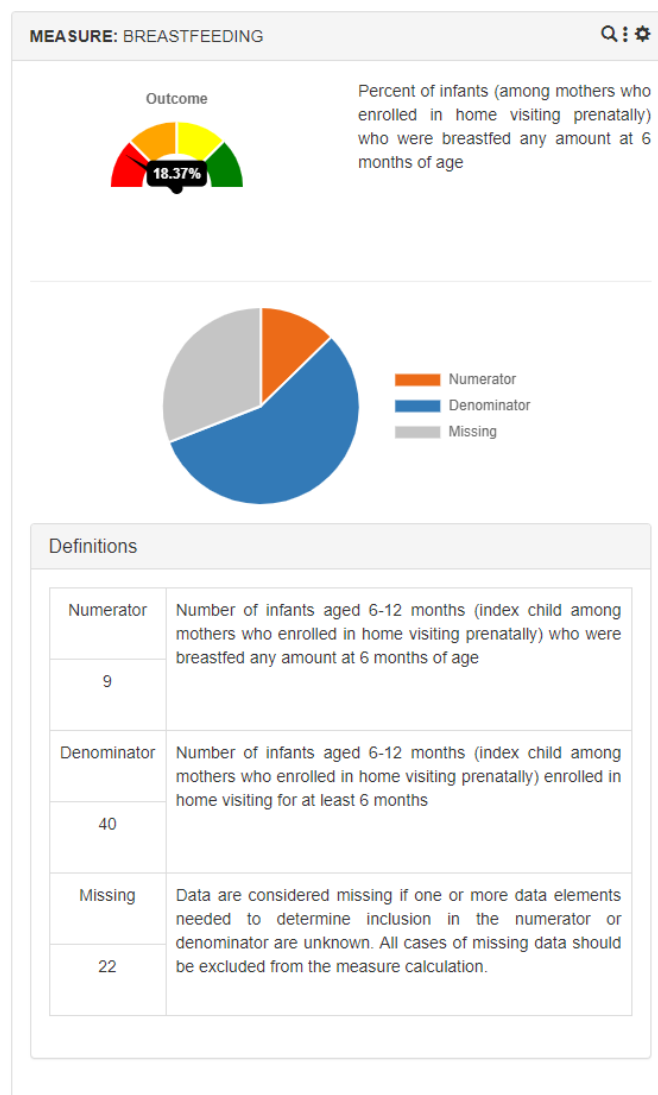


Figura 40. Diseño correspondiente a la segunda métrica de rendimiento.

La Figura 41 muestra el diseño de la tercera métrica de rendimiento, correspondiente a la pregunta P-03, compuesta por el título, un botón para visualización de los datos empleados por esta métrica, un botón de configuración, el resultado de rendimiento 3, visualizado en la gráfica de semáforo presente, la descripción del indicador a la derecha del resultado, una gráfica correspondiente al total de datos categorizados en función del algoritmo correspondiente, y la definición de cada categoría.

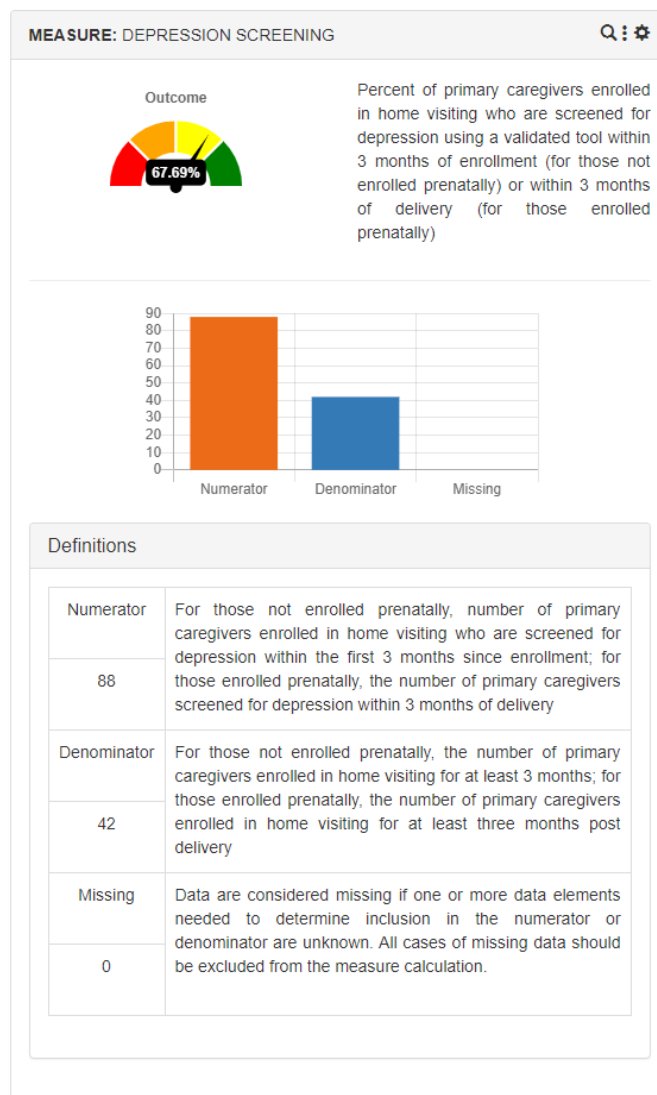
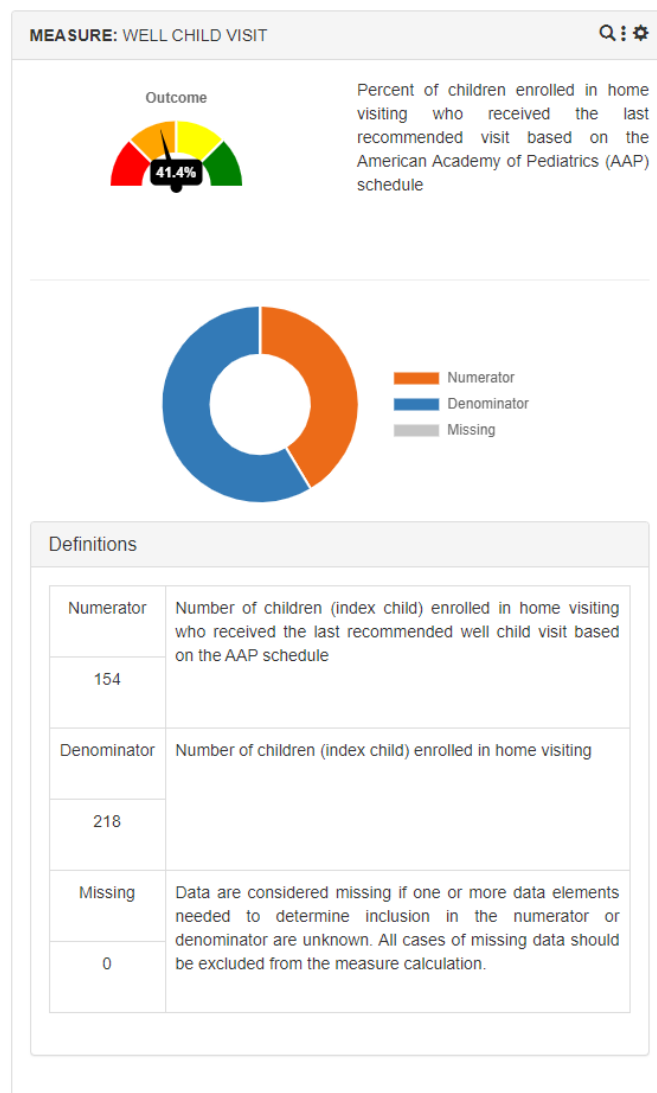


Figura 41. Diseño correspondiente a la tercera métrica de rendimiento.

La Figura 42 muestra el diseño de la cuarta métrica de rendimiento, correspondiente a la pregunta P-04, compuesta por el título, un botón para visualización de los datos empleados por esta métrica, un botón de configuración, el resultado de rendimiento 4, visualizado en la gráfica de semáforo presente, la descripción del indicador a la derecha del resultado, una gráfica correspondiente al total de datos categorizados en función del algoritmo correspondiente, y la definición de cada categoría.



*Figura 42. Diseño correspondiente a la cuarta métrica de rendimiento.*

La Figura 43 muestra el diseño de la quinta métrica de rendimiento, correspondiente a la pregunta P-05, compuesta por el título, un botón para visualización de los datos empleados por esta métrica, un botón de configuración, el resultado de rendimiento 5, visualizado en la gráfica de semáforo presente, la descripción del indicador a la derecha del resultado, una gráfica correspondiente al total de datos categorizados en función del algoritmo correspondiente, y la definición de cada categoría.

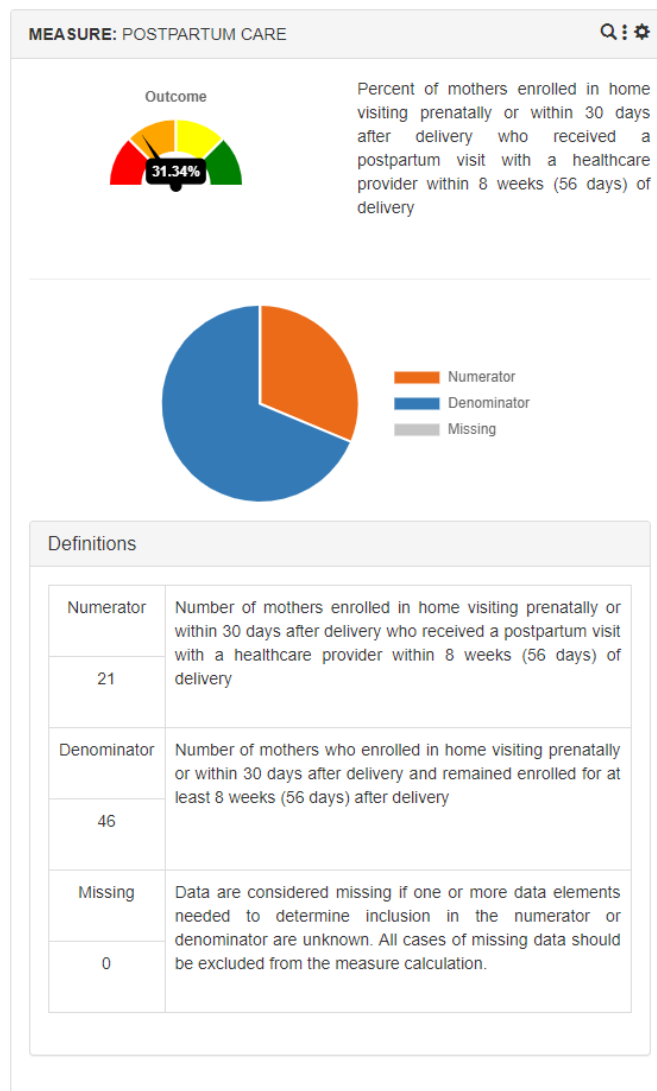


Figura 43. Diseño correspondiente a la quinta métrica de rendimiento.

La Figura 44 muestra el diseño de la sexta métrica de rendimiento, correspondiente a la pregunta P-06, compuesta por el título, un botón para visualización de los datos empleados por esta métrica, un botón de configuración, el resultado de rendimiento 6, visualizado en la gráfica de semáforo presente, la descripción del indicador a la derecha del resultado, una gráfica correspondiente al total de datos categorizados en función del algoritmo correspondiente, y la definición de cada categoría.

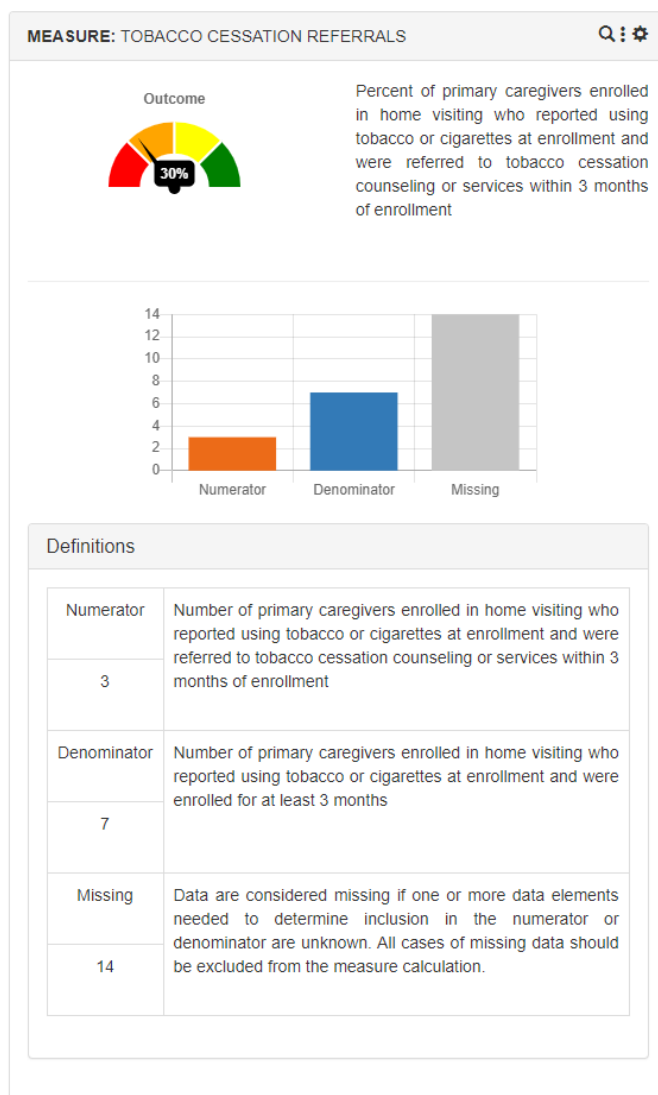


Figura 44. Diseño correspondiente a la sexta métrica de rendimiento.

La Figura 45 presenta la vista general del panel de visualización de información.

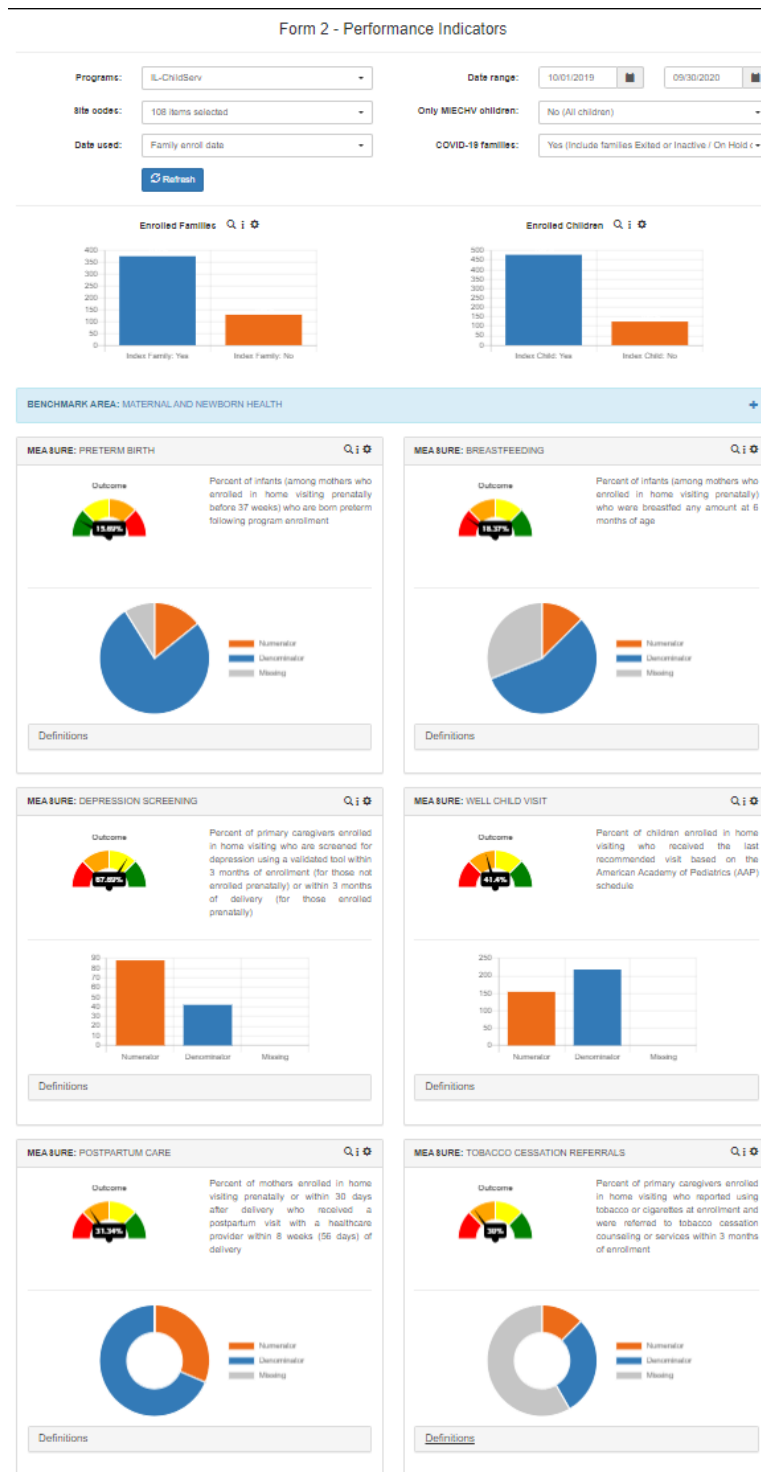


Figura 45. Vista general del panel de visualización de información.

## Capítulo 5. Experimentación y Resultados

De acuerdo con (Wohlin, Höst, & Henningsson, 2003), existen dos paradigmas principales de investigación:

- **Cualitativa.** Se enfoca en el estudio de objetos en su ambiente natural, intentando interpretar un fenómeno a través de las explicaciones aportadas sobre éste. Su enfoque principal es comprender, explicar, explorar, descubrir y esclarecer situaciones, sentimientos, percepciones, actitudes, valores, creencias y experiencias de un grupo de personas.
- **Cuantitativa.** Se cuantifica la relación o se realiza una comparación entre dos o más grupos identificando una relación causa-efecto. Son específicos, bien estructurados y su validez y confiabilidad han sido probados.

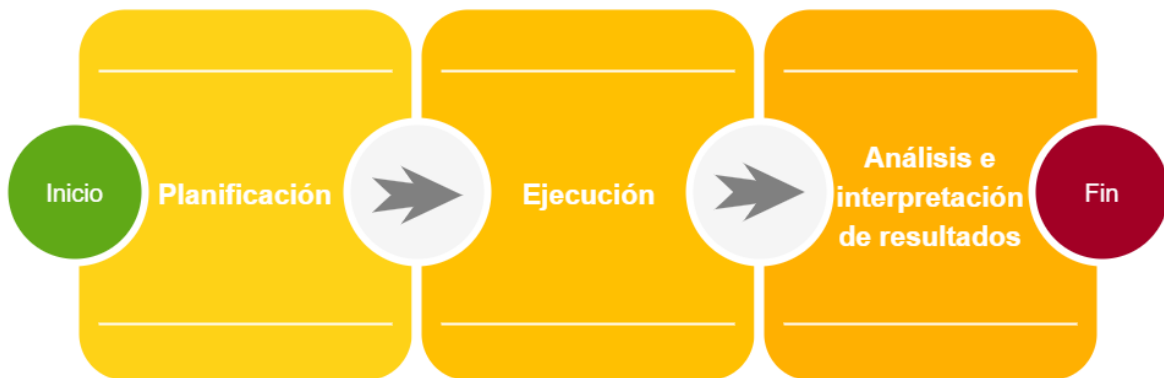
Adicionalmente, dependiendo del propósito de la evaluación, técnicas, métodos o herramientas y de las condiciones de la investigación, existen cuatro métodos empíricos principales, descritos a continuación (Wohlin, Höst, & Henningsson, 2003):

- **Experimento controlado.** Tiene un alcance limitado y la mayoría de las veces se ejecutan en un entorno controlado, su objetivo es el de manipular intencionalmente una o más variables independientes y analizar las consecuencias que esto tiene sobre una o más variables dependientes.
- **Estudio de caso.** Investiga un fenómeno dentro de una ambiente típico o real, busca establecer relaciones entre diferentes atributos o dar seguimiento a uno en específico, a diferencia de un experimento controlado el nivel de control empleado en un estudio de caso es menor.



- **Encuesta.** Busca recolectar información cualitativa y cuantitativa sobre una población específica a través de muestras representativas, los resultados son analizados para proporcionar conclusiones descriptivas y explicativas.
- **Análisis post-mortem.** Se realiza sobre hechos pasados, aunque no necesariamente concluidos, se relaciona tanto con un estudio de caso como con una encuesta y se puede realizar a través de la observación o mediante una serie de encuestas enfocadas al objeto de estudio a analizar.

Dado lo anterior, en este trabajo de investigación se utiliza un experimento controlado con la intención de saber cuál es el impacto del empleo de una arquitectura de inteligencia de negocios en el tiempo y los recursos computacionales que un usuario emplea para obtener un informe de rendimiento. De acuerdo con lo propuesto por (Wohlin, Höst, & Henningson, 2003), el experimento se compone por tres fases principales mostradas en la Figura 46.



*Figura 46. Fases del experimento controlado.*

### **5.1. Planificación del experimento controlado**

Este experimento consiste en el empleo de múltiples enfoques utilizados para el reporte de rendimiento organizacional, con el objetivo de obtener datos relacionados al

tiempo de respuesta de cada uno de ellos, los recursos computacionales empleados y posterior análisis e interpretación de los resultados obtenidos.

### 5.1.1. Participantes del experimento controlado

Para la selección de participantes en este experimento se utilizó una agencia implementadora de alguno de los modelos de visitas a domicilio avalados por el Departamento de Salud y Servicios Humanos de los Estados Unidos de América, los criterios para la selección de la agencia son los siguientes:

- Haber prestado servicios a familias durante el año 2020.
- Haber solicitado al menos un informe de rendimiento durante el año 2020.
- Haber empleado el sistema Visit Tracker (DataKeeper Technologies, 2020) para la obtención de cada uno de los informes de rendimiento.

Tras aplicar los criterios de selección se realizó un ordenamiento en función del total de familias servidas y el número de informes de rendimiento solicitados por agencia implementadora.

Tras realizar lo anterior se seleccionó la agencia implementadora con más informes solicitados, incluyendo a los empleados de la agencia que solicitaron dichos informes, resultando en un total de 4 empleados y 102 informes de rendimiento (Tabla 32).

*Tabla 32. Informes de rendimiento solicitados por Family Focus, Inc durante el año 2020.*

| Agencia implementadora | Empleado      | Informes de rendimiento solicitados |
|------------------------|---------------|-------------------------------------|
| Family Focus, Inc      | rosareal      | 4                                   |
| Family Focus, Inc      | RamirezM      | 78                                  |
| Family Focus, Inc      | StaceyMcK     | 18                                  |
| Family Focus, Inc      | MayraGonzalez | 2                                   |

### 5.1.2. Enfoques del experimento controlado

Para este trabajo de investigación se decidió emplear cuatro enfoques o arquitecturas de software utilizadas para la generación de informes de rendimiento. El objetivo de esta selección es el de realizar una comparativa entre tiempo y los recursos empleados para la generación de un informe de rendimiento por cada arquitectura de software.

El primer enfoque es una arquitectura compuesta por un sistema de reporte cuya información se obtiene directamente de la base de datos operativa, siendo este tipo de arquitecturas empleadas comúnmente en organizaciones que aún no adoptan inteligencia de negocios para la medición del rendimiento organizacional (Figura 47), además de ser el enfoque actual empleado por la agencia implementadora seleccionada para este experimento.



Figura 47. Enfoque 1, reporte operativo sin inteligencia de negocios.

El segundo y cuarto enfoque corresponden a una arquitectura de almacén de datos centralizada, siendo una de las arquitecturas de inteligencia de negocios más populares (Figura 48). Para el segundo enfoque se generó un almacén de datos a través de una serie de scripts, omitiendo la creación de índices y claves primarias provenientes de la base de datos de producción, con el objetivo de analizar el impacto que esto conlleva. Para el cuarto enfoque se generó un almacén de datos a través de una serie de scripts, respetando tanto los índices como las claves primarias provenientes de la base de datos de producción.

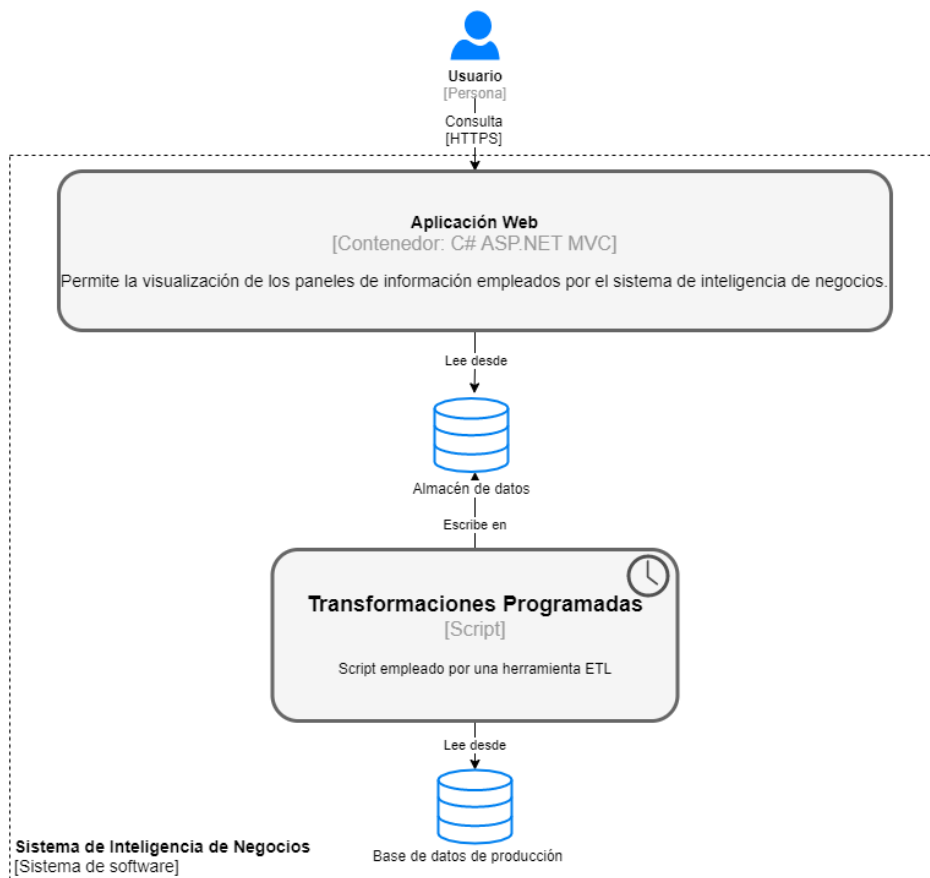


Figura 48. Enfoque 2, arquitectura de almacén de datos centralizada.

La tercera arquitectura empleada es la propuesta en este trabajo (Figura 28).

### 5.1.3. Preparación del ambiente de pruebas experimental

Previo a la ejecución del experimento controlado se preparó un ambiente de pruebas experimental empleado para la generación de informes de rendimiento. Las características del ambiente se muestran en la Tabla 33.

Tabla 33. Características del ambiente de prueba.

| Elemento          | Características                           |
|-------------------|---|
| Procesador        | Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz |
| Memoria RAM       | 16.0 GB DDR4 2400 MHz                     |
| Almacenamiento    | SSD 1 TB                                  |
| Sistema Operativo | Windows 10 Professional Edition           |

Adicionalmente, se realizó la instalación del software necesario para la implementación de los tres enfoques a emplear, mostrado en la Tabla 34.

Tabla 34. Prerrequisitos de software.

| Software           | Descripción  |
|--------------------|--|
| IIS 7              | Servidor Web   |
| Redis 6.2          | Base de datos en memoria.  |
| MSSQL Server 2012  | Base de datos de producción.   |
| .NET Framework 4.5 | Conjunto de estructuras y tecnologías que proporciona empleadas para el desarrollo y ejecución de software en sistemas operativos Windows. |

### 5.1.4. Consideraciones sobre la muestra del experimento

Tras la selección de la agencia implementadora a utilizar para la ejecución del experimento y tras haber seleccionado los cuatro enfoques a emplear se obtienen de manera aleatoria una serie de registros correspondientes a los informes de rendimiento solicitados por cada empleado durante el año 2020, incluyendo los parámetros con los que fueron solicitados.

Para la selección de los registros se aplicó un módulo 3 al identificador de cada informe de rendimiento solicitado, con lo cual, de un total de 102 informes de rendimiento se emplearon 29. Lo anterior para garantizar que cada uno de los registros seleccionados corresponda a un periodo distinto a lo largo del año 2020.

Con el objetivo de realizar una comparativa entre el tiempo tomado por cada enfoque para la generación de cada informe de rendimiento seleccionado a través de los 29 registros obtenidos, se duplicó la información de cada registro en los 4 grupos correspondientes a cada enfoque.

Por último, por cada registro obtenido se generaron nuevamente los informes de rendimiento correspondientes a los parámetros con los que fueron solicitados originalmente y en función del enfoque correspondiente, los parámetros empleados se presentan en la Tabla 35.

*Tabla 35. Parámetros empleados para la obtención de los informes de rendimiento durante la experimentación.*

| Nombre            | Descripción  |
|-------------------|--|
| Fecha inicial     | Fecha inicial del período de tiempo a emplear para la obtención del rendimiento de una agencia implementadora.   |
| Fecha final       | Fecha final del período de tiempo a emplear para la obtención del rendimiento de una agencia implementadora.   |
| Programa          | Dentro de la herramienta un programa representa a una agencia implementadora de un modelo de visitas a domicilio a niños en edad temprana. Para este experimento se seleccionó a la agencia Family Focus, Inc. |
| Códigos de sitio  | Códigos de sitio empleados por la agencia implementadora para identificar el origen de los fondos recibidos para la atención de ciertas familias.  |
| Solo niños MIECHV | Filtra solo aquellas familias y niños que son atendidos mediante el empleo de fondos provenientes del programa federal MIECHV  |

| Nombre            | Descripción   |
|-------------------|---|
|                   | (Health Resources & Services Administration, 2020).   |
| Fecha empleada    | Indica que fecha ha de ser empleada como fecha de inicio de participación de una familia en una agencia implementadora. Existen dos variantes: 1) Fecha de suscripción, 2) Fecha en que comienza a recibir apoyo por parte de MIECHV. |
| Familias COVID-19 | Indica la exclusión de los cálculos de rendimiento de aquellas familias que por alguna cuestión asociada a COVID-19 dejaron de recibir servicios ya sea temporal o permanentemente.   |

La Figura 49 muestra la representación visual del procedimiento a emplear.

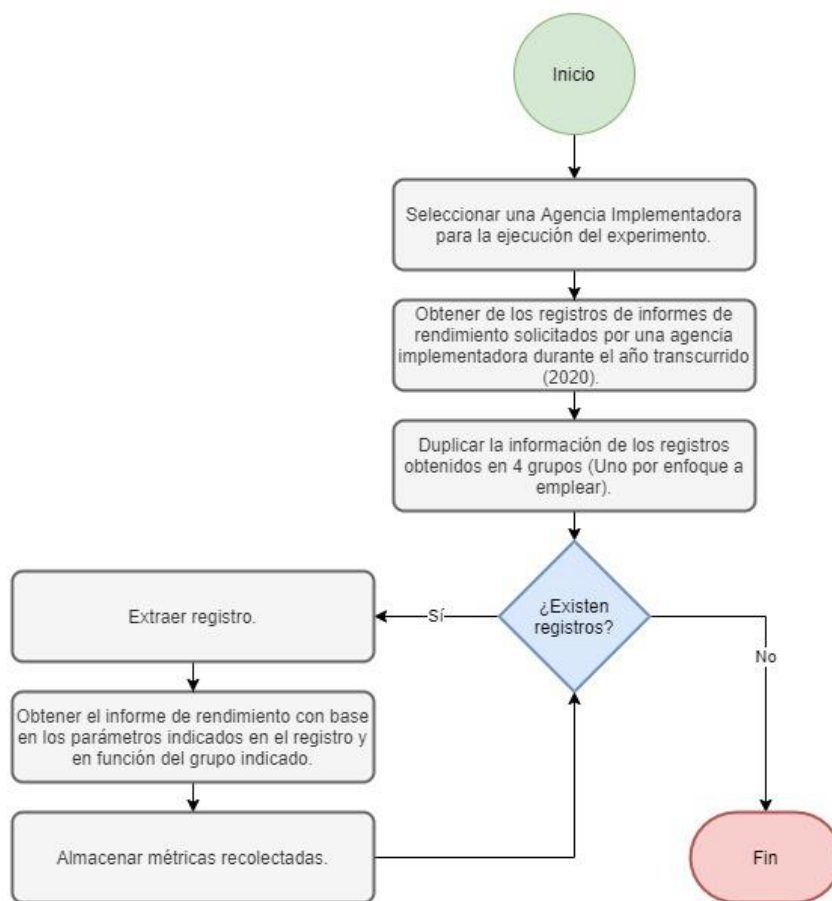


Figura 49. Procedimiento empleado para la ejecución del experimento controlado.

### 5.1.5. Determinación de las Métricas del Experimento

Tras procesar cada informe de rendimiento en función del enfoque correspondiente se almacenan una serie de métricas resultantes, incluyendo el tiempo tomado en obtener el informe, el número de lecturas y escrituras a disco, uso de procesador y la memoria disponible durante la obtención de éste. La Tabla 36 detalla las métricas recolectadas.

Tabla 36. Métricas recolectadas tras la ejecución de un informe de rendimiento.

| Métrica             | Descripción   | Método de cálculo                              |
|---------------------|---|--|
| Tiempo de respuesta | Tiempo tomado en obtener el informe de rendimiento. | Número de segundos por informe de rendimiento. |



| Métrica            | Descripción   | Método de cálculo  |
|--------------------|---|--|
| Lecturas a disco   | Lecturas realizadas a disco durante la obtención del informe de rendimiento.          | Número de lecturas por informe de rendimiento.           |
| Escrituras a disco | Escrituras realizadas a disco durante la obtención del informe de rendimiento.        | Número de escrituras por informe de rendimiento.         |
| Uso de procesador  | Porcentaje promedio del tiempo en el que el procesador se encuentra ocupado.          | Porcentaje de tiempo ocupado por informe de rendimiento. |
| Uso de memoria.    | Cantidad de memoria física, en Megabytes disponible inmediatamente para su asignación | Memoria física disponible por informe de rendimiento.    |

## 5.2. Ejecución del experimento controlado

Para la obtención de cada informe de rendimiento y recolección de métricas se empleó una planificación circular, enfocada en la obtención de un informe de rendimiento por cada grupo de manera alternada, comenzando por el primer grupo, hasta llegar al cuarto y comenzar nuevamente. De este modo se logró la obtención de 29 informes de rendimiento por grupo o enfoque, con un total de 116 informes de rendimiento nuevos.

Tras la obtención de todos los informes de rendimiento se obtuvieron y almacenaron las métricas correspondientes en un archivo separado por comas (CSV del Inglés: Comma Separated Value), correspondiente al conjunto de datos a emplear para el análisis e interpretación de resultados, cuya estructura se presenta en la Tabla 37.

*Tabla 37. Estructura del conjunto de datos a emplear para el análisis e interpretación de resultados.*

| Columna | Descripción   |
|---------|---|
| Grupo   | Enfoque empleado para la obtención del informe de rendimiento:<br>1. Reporteo Operativo |

| Columna             | Descripción  |
|---------------------|--|
|                     | 2. Arquitectura de almacén de datos centralizada sin emplear índices.<br>3. Arquitectura de Virtualización de datos.<br>4. Arquitectura de almacén de datos centralizada empleando índices |
| Tiempo de respuesta | Tiempo tomado en obtener el informe de rendimiento.  |
| Lecturas a disco    | Lecturas realizadas a disco durante la obtención del informe de rendimiento.   |
| Escrituras a disco  | Escrituras realizadas a disco durante la obtención del informe de rendimiento.   |
| Uso de procesador   | Porcentaje promedio del tiempo en el que el procesador se encuentra ocupado.   |
| Uso de memoria.     | Cantidad de memoria física, en megabytes, disponible inmediatamente para su asignación   |

### 5.3. Análisis e interpretación de resultados del experimento

En esta sección se presentan los resultados obtenidos tras la ejecución del experimento controlado, lo anterior mediante la visualización de la serie de análisis estadísticos realizados, incluyendo una prueba de valores atípicos, un análisis de medias, una prueba de varianzas, un análisis ANOVA y una prueba de Kruskal-Wallis, realizados todos mediante el software estadístico Minitab (Minitab, LLC., 2021). Los resultados son desglosados de la siguiente manera:

- Tiempo de respuesta por informe de rendimiento.
- Lecturas a disco por informe de rendimiento.
- Escrituras a disco por informe de rendimiento.
- Uso de procesador por informe de rendimiento.
- Uso de memoria por informe de rendimiento.

### 5.3.1. Resultados del tiempo de respuesta

La Figura 50 muestra los resultados de la prueba de valores atípicos correspondiente al tiempo empleado en segundos para la obtención de cada informe de rendimiento. Se puede observar que no existen valores atípicos por lo que se procede con el análisis de resultados.

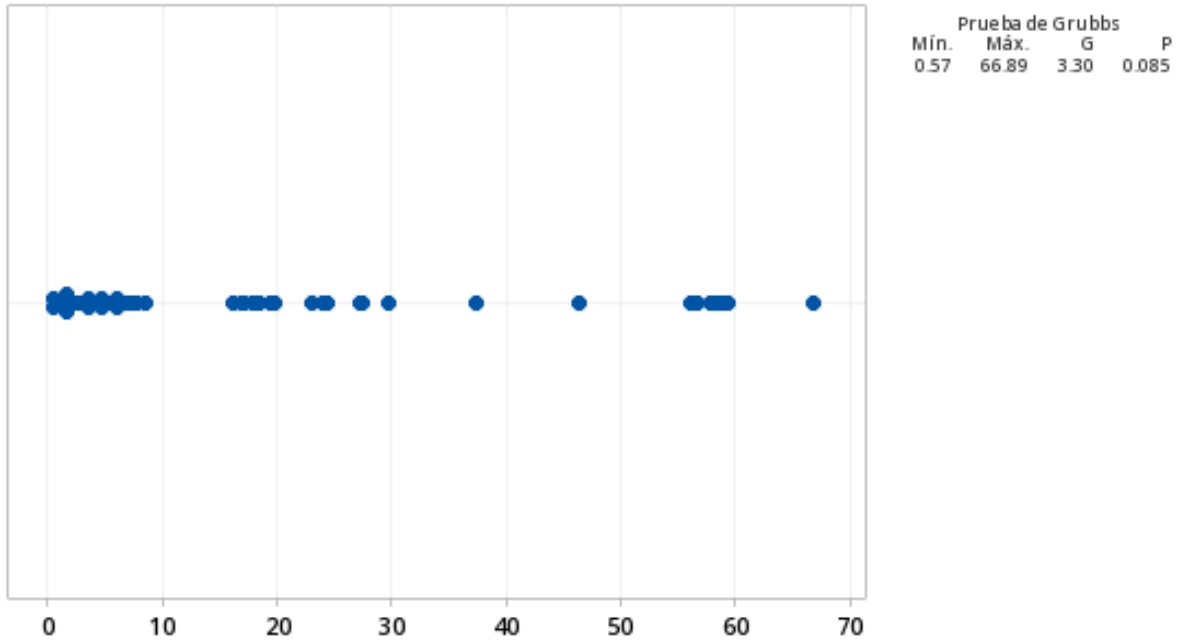


Figura 50. Gráfica de valores atípicos sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.

La Figura 51 muestra los resultados de una prueba de normalidad con un estadístico Anderson Darling aplicada al tiempo empleado en segundos para la obtención de cada informe de rendimiento, observándose un valor p estimado ( $<0.005$ ) menor a 0.05, por lo tanto, los datos no se distribuyen de manera normal, considerando un nivel de confianza del 95%.

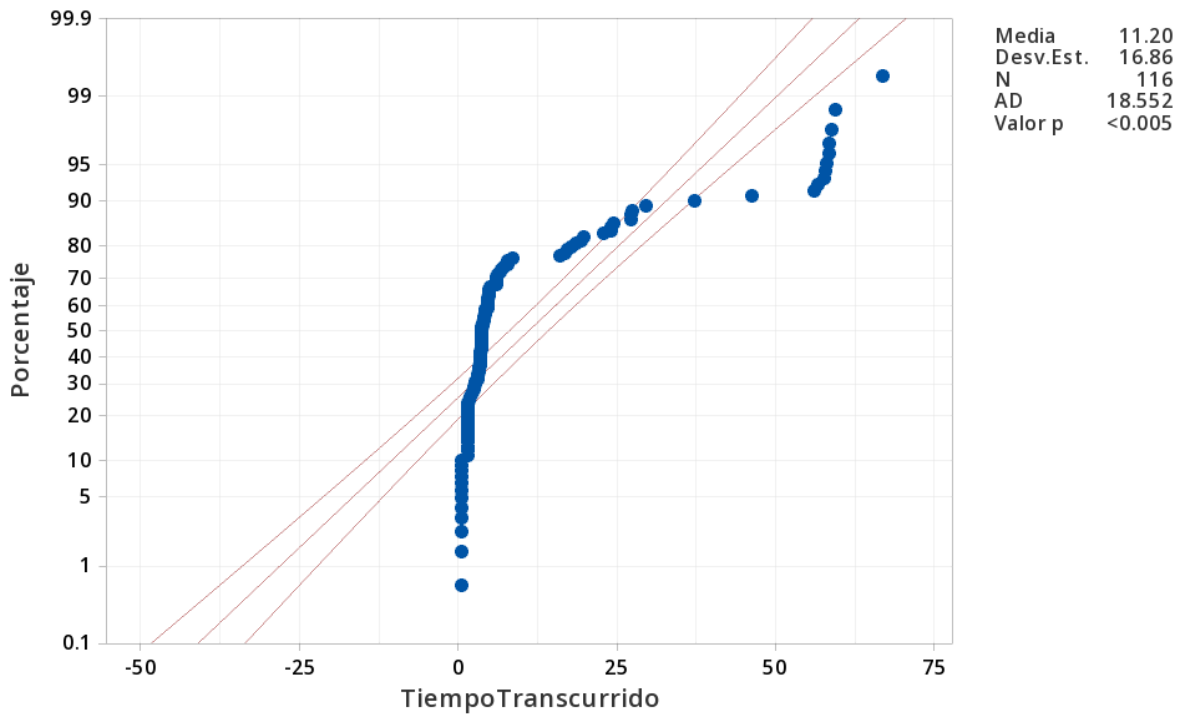


Figura 51. Prueba de normalidad aplicada sobre el tiempo empleado en segundos para la obtención de un informe de rendimiento.

El siguiente análisis realizado es un ANOM normal unidireccional del tiempo empleado en segundos para la obtención de un informe de rendimiento, presentado en la Figura 52. Se puede observar que el enfoque 2, correspondiente a una arquitectura de almacén de datos centralizado sin la presencia de índices ni claves primarias, presenta resultados por encima de la media, indicando que el empleo de este enfoque implica una repercusión negativa en cuanto a los tiempos de espera para la obtención de un informe de rendimiento, siendo mayores.

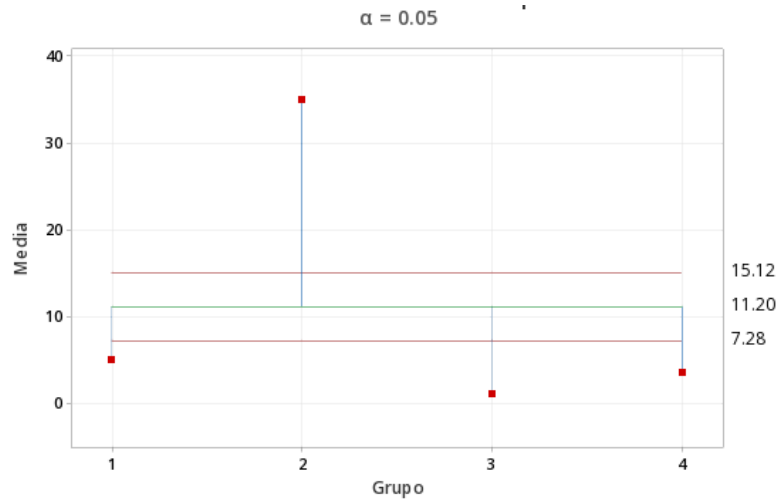


Figura 52. ANOM normal unidireccional del tiempo empleado en segundos para la obtención de un informe de rendimiento.

Posterior al análisis de medias se realizó una prueba de igualdad de varianzas (Figura 53), considerando que los datos son no normales, se interpreta la prueba de Levene, es decir, hay diferencias significativas entre las varianzas de los grupos ( $p = 0.000$  menor que  $\alpha = 0.05$ ). Adicionalmente, se detecta una sobreposición entre los intervalos correspondientes a los enfoques 1, 3 y 4. No siendo el caso para el enfoque 2, presentando un intervalo mayor y una desviación estándar significativamente diferente al de los otros enfoques.

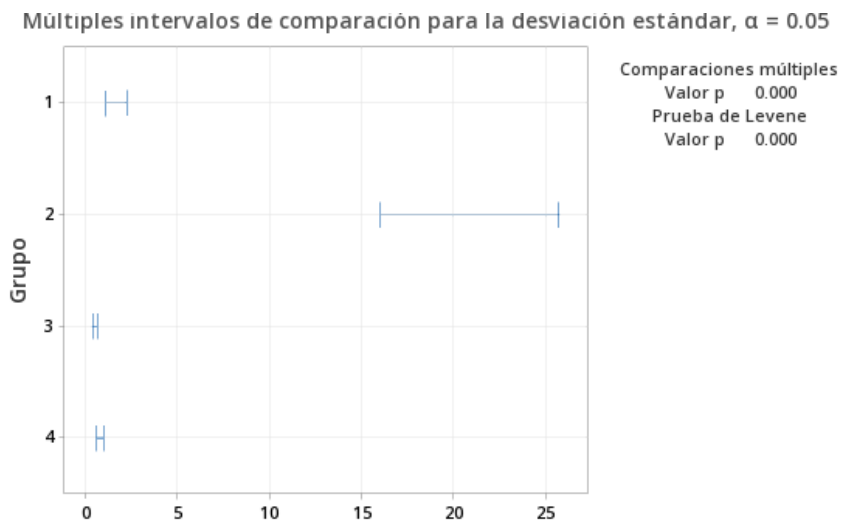


Figura 53. Prueba de igualdad de varianzas sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.

Una vez realizado el análisis de varianzas se procede con un análisis ANOVA (Figura 54) y una prueba de Kruskal-Wallis (Figura 55).

| Fuente | GL  | SC Ajust. | MC Ajust. | Valor F | Valor p |
|--------|-----|-----------|-----------|---------|---------|
| Grupo  | 3   | 22069     | 7356.22   | 77.46   | 0.000   |
| Error  | 112 | 10636     | 94.96     |         |         |
| Total  | 115 | 32705     |           |         |         |

Figura 54. Análisis ANOVA sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.

Hipótesis nula  $H_0$ : Todas las medianas son iguales  
 Hipótesis alterna  $H_1$ : Al menos una mediana es diferente

| Método                   | GL | Valor H | Valor p |
|--------------------------|----|---------|---------|
| No ajustado para empates | 3  | 95.78   | 0.000   |
| Ajustado para empates    | 3  | 95.79   | 0.000   |

Figura 55. Prueba de Kruskal-Wallis sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.

| Grupo | N  | Media  | Agrupación |
|-------|----|--------|------------|
| 2     | 29 | 34.97  | A          |
| 1     | 29 | 5.044  | B          |
| 4     | 29 | 3.607  | B          |
| 3     | 29 | 1.1874 | B          |

Figura 56. Comparaciones en parejas de Tukey sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.

Tras analizar los resultados obtenidos en las dos pruebas anteriores se observa que en ambas el valor p es menor que el nivel de significancia empleado, siendo el nivel de significancia 0.05 y el valor p para ambas pruebas menor a 0.01, lo cual indica que las

diferencias entre alguna de las medias son estadísticamente significativas. Por lo tanto, se rechaza una hipótesis nula, concluyendo que no todas las medias de población son iguales.

Adicionalmente, la prueba de Tukey (Figura 56) nos indica que el enfoque 2 forma un grupo individual, y los enfoques 1, 4, 3 forman un grupo similar, por lo que la media del enfoque 2 es muy diferente a los otros grupos.

La gráfica de intervalos presentada en la Figura 57 muestra la media y el intervalo de confianza para cada grupo, observándose una diferencia estadísticamente significativa entre el enfoque 2 y los enfoques restantes, siendo este enfoque el que más tiempo emplea para la obtención de un informe de rendimiento. El enfoque 3, correspondiente a la propuesta de este trabajo, presenta una media de muestra menor al de los otros enfoques, indicando que el tiempo empleado utilizando este enfoque puede ser menor, aun así, no se detecta una diferencia estadísticamente significativa con respecto al enfoque 1 y 4.

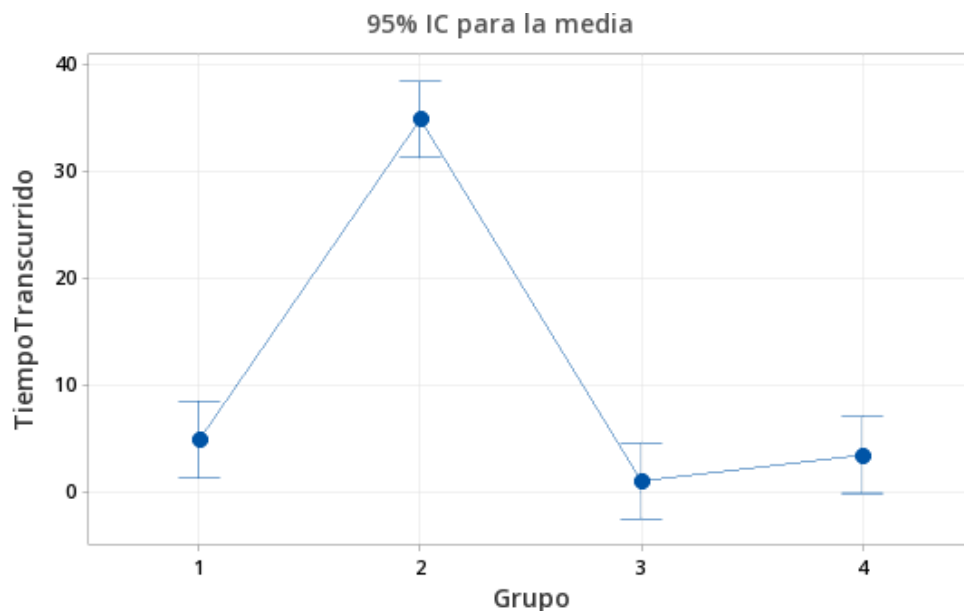


Figura 57. Gráfica de intervalos sobre el tiempo empleado en segundos para la obtención de informes de rendimiento.

### 5.3.2. Resultados de lecturas a disco

La Figura 58 muestra los resultados de la prueba de valores atípicos correspondiente al número de lecturas a disco tras la obtención de cada informe de rendimiento, observándose un valor atípico.

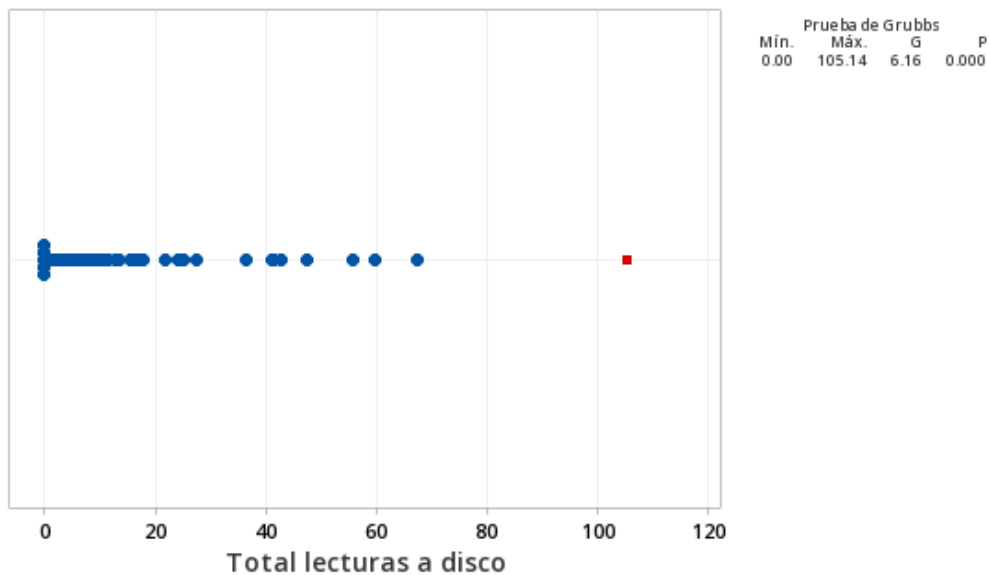


Figura 58. Gráfica de valores atípicos sobre el número de lecturas a disco tras la obtención de un informe de rendimiento.

Tras realizarse una revisión de los parámetros empleados para la obtención del informe de rendimiento correspondiente al valor atípico detectado se encontró un incremento en el rango de fechas empleado para su obtención, abarcando un periodo de informe de 1 año y 6 meses, lo cual conlleva a la obtención de un mayor número de datos a emplear para la obtención y despliegue de resultados. Dado lo anterior y debido a que el periodo de obtención de rendimiento organizacional en los demás informes no excede el periodo de 1 año en más de 1



mes, se decidió eliminar este registro con la finalidad de evitar que éste genere inconsistencias al momento de proceder con los métodos estadísticos aplicados.

Tras eliminar el valor atípico encontrado se ejecutó nuevamente una prueba de valores atípicos correspondiente al número total de escrituras a disco. La Figura 59 muestra los resultados obtenidos, detectando aun valores atípicos.

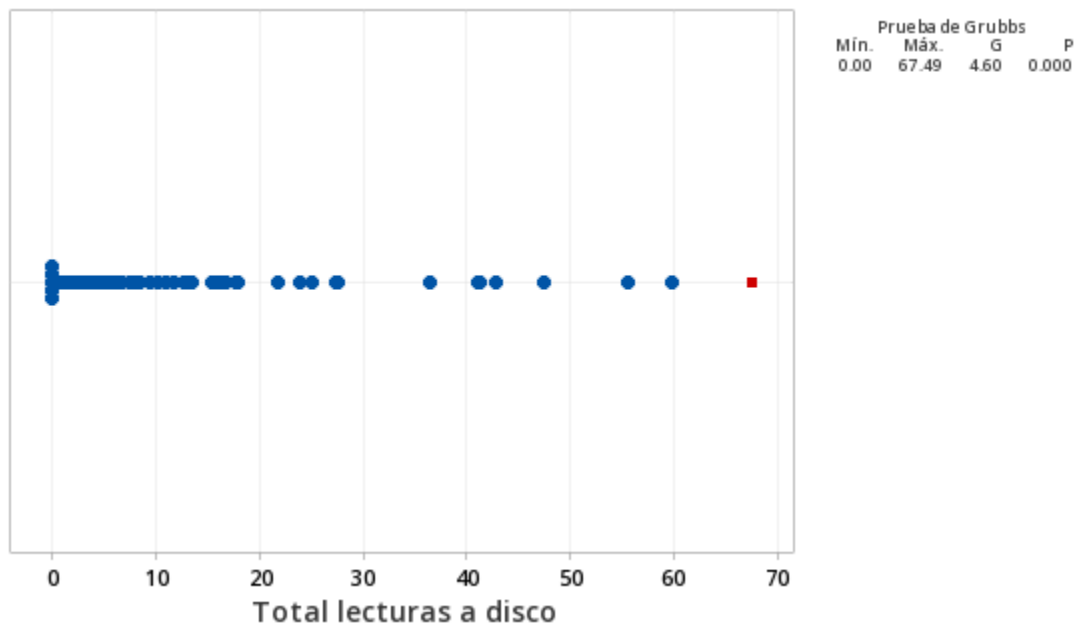


Figura 59. Gráfica de valores atípicos sobre el número de lecturas a disco tras la obtención de un informe de rendimiento (Tras la eliminación del valor atípico).

Como se puede observar, persisten valores atípicos a pesar de la eliminación del primero, siendo un total de 7. Debido a que la eliminación de los demás valores representaría una pérdida en la potencia estadística se decide conservarlos.

Tras analizar los parámetros con los que se obtuvieron los informes de rendimiento correspondientes a los valores atípicos restantes se encontró que 5 de ellos corresponden a los enfoques que emplean almacenes de datos. Dado que el problema se presenta en el número de

lecturas a disco, la falta de índices en el enfoque 2 provoca un aumento en estas. Para el enfoque 4, dado que hay presencia de índices, una revisión de estos es de ayuda en la disminución de lecturas a disco, sugiriendo que los índices, al provenir de la base de datos proporcionada para este trabajo, acarrearán problemas de diseño e implementación.

La Figura 60 muestra los resultados de una prueba de normalidad con un estadístico Anderson Darling aplicada al número de lecturas a disco tras la obtención de cada informe de rendimiento, observándose un valor p estimado ( $<0.005$ ) menor a 0.05, por lo tanto, los datos no se distribuyen de manera normal, considerando un nivel de confianza del 95%.

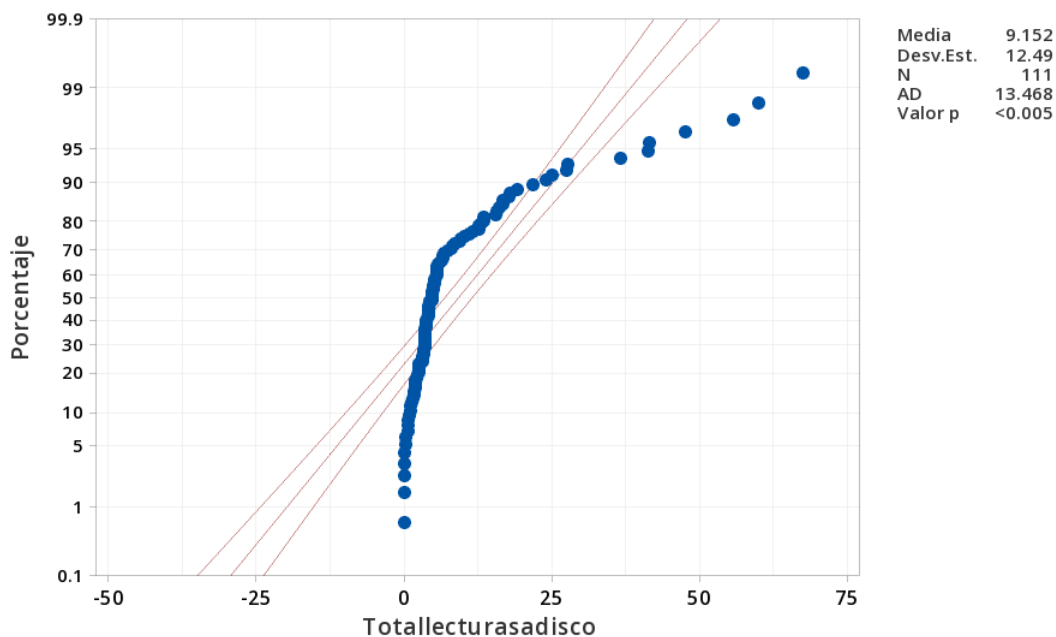
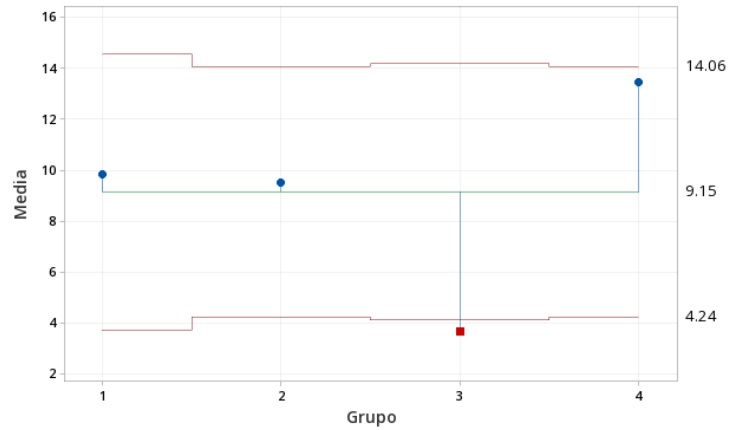


Figura 60. Prueba de normalidad aplicada sobre el total de lecturas a disco tras la obtención de un informe de rendimiento.

El siguiente análisis realizado es un ANOM normal unidireccional del total de escrituras disco tras la obtención de un informe de rendimiento, presentado en la Figura 61. Se puede observar que el enfoque 3, correspondiente a la propuesta de este trabajo, presenta

resultados por debajo de la media, indicando que el empleo de este enfoque implica una mejora en cuanto al número de lecturas a disco tras la obtención de un informe de rendimiento.



*Figura 61. ANOM normal unidireccional del número de lecturas a disco tras la obtención de un informe de rendimiento.*

Posterior al análisis de medias se realizó una prueba de igualdad de varianzas (Figura 62), considerando que los datos son no normales, se interpreta la prueba de Levene, es decir, hay diferencias significativas entre las varianzas de los grupos ( $p = 0.000$  menor que  $\alpha = 0.05$ ). Adicionalmente, se detecta una sobreposición entre los intervalos correspondientes a los enfoques 1 y 4. No siendo el caso para los enfoques 2 y 3, presentando una desviación estándar significativamente diferente al de los otros enfoques.

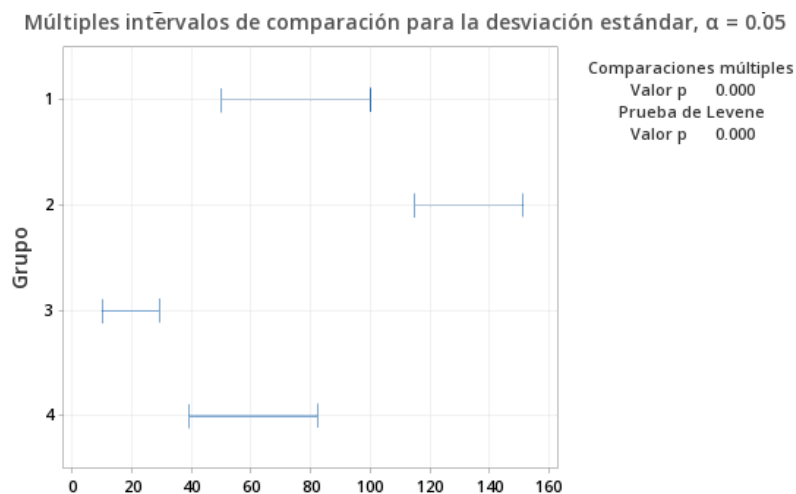


Figura 62. Prueba de igualdad de varianzas sobre las lecturas a disco tras la obtención de informes de rendimiento.

Una vez realizado el análisis de varianzas se procede con un análisis ANOVA (Figura 63) y una prueba de Kruskal-Wallis (Figura 64).

| Fuente | GL  | SC Ajust. | MC Ajust. | Valor F | Valor p |
|--------|-----|-----------|-----------|---------|---------|
| Grupo  | 3   | 1387      | 462.3     | 3.14    | 0.028   |
| Error  | 107 | 15769     | 147.4     |         |         |
| Total  | 110 | 17156     |           |         |         |

Figura 63. Análisis ANOVA sobre el total de lecturas a disco tras la obtención de informes de rendimiento.

| Método                   | GL | Valor H | Valor p |
|--------------------------|----|---------|---------|
| No ajustado para empates | 3  | 26.99   | 0.000   |
| Ajustado para empates    | 3  | 27.00   | 0.000   |

Figura 64. Prueba de Kruskal-Wallis sobre el total de lecturas a disco tras la obtención de informes de rendimiento.

| Grupo | N  | Media | Agrupación |
|-------|----|-------|------------|
| 4     | 29 | 13.46 | A          |
| 1     | 25 | 9.83  | A B        |
| 2     | 29 | 9.53  | A B        |
| 3     | 28 | 3.70  | B          |

*Figura 65. Comparaciones en parejas de Tukey sobre el total de lecturas a disco tras la obtención de informes de rendimiento.*

Tras analizar los resultados obtenidos en las dos pruebas anteriores se observa que en ambas el valor p es menor que el nivel de significancia empleado, siendo el nivel de significancia 0.05 y el valor p para ambas pruebas menor a 0.01, lo cual indica que las diferencias entre alguna de las medias son estadísticamente significativas. Por lo tanto, se rechaza una hipótesis nula, concluyendo que no todas las medias de población son iguales.

Adicionalmente, la prueba de Tukey (Figura 65) nos indica que los enfoques 1,2, 4 forman un grupo similar, mientras que los enfoques 1,2, 3 forman otro grupo similar, por lo que únicamente las medias entre los enfoques 3 y 4 son diferentes.

La gráfica de intervalos presentada en la Figura 66 muestra la media y el intervalo de confianza para cada grupo, no observándose una diferencia estadísticamente significativa entre ninguno de los enfoques. El enfoque 3, correspondiente a la propuesta de este trabajo, aunque no presenta una diferencia estadísticamente significativa con respecto a los otros, cuenta con una media de muestra menor a la de los otros enfoques, indicando que el total de lecturas a disco utilizando este enfoque es menor al de los otros enfoques, pero no de manera significativa.

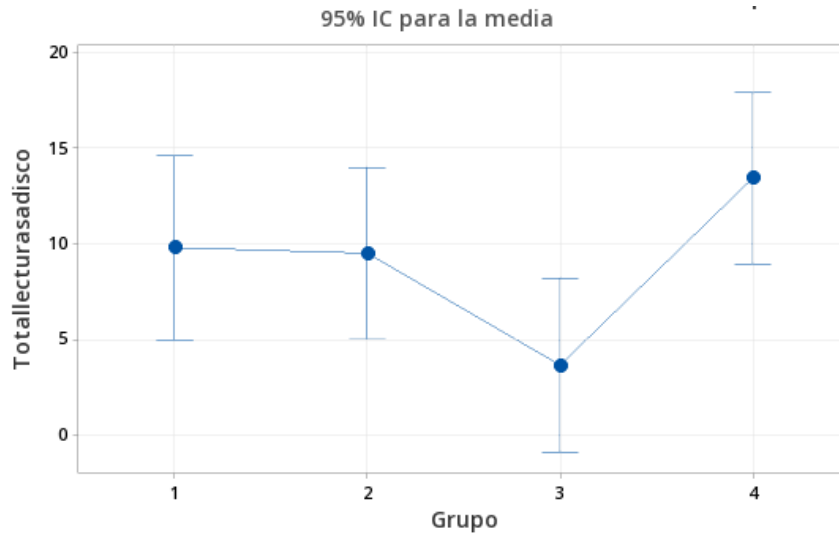
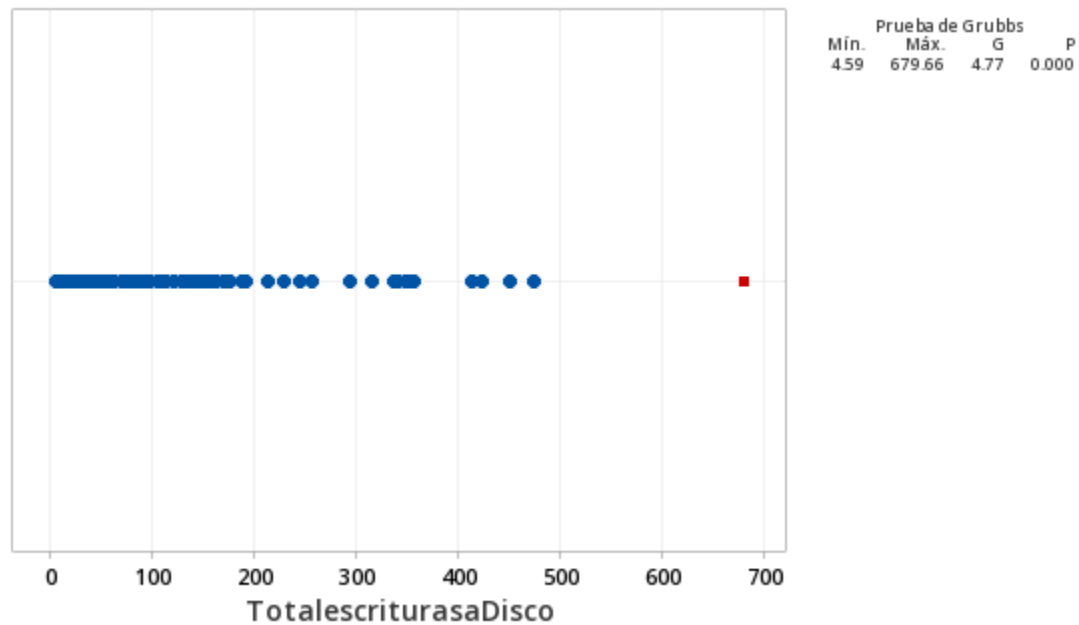


Figura 66. Gráfica de intervalos sobre el total de lecturas a disco tras la obtención de un informe de rendimiento.

### 5.3.3. Resultados de escrituras a disco

La Figura 67 muestra los resultados de la prueba de valores atípicos correspondiente al número de escrituras a disco tras la obtención de cada informe de rendimiento, observándose un valor atípico.



*Figura 67. Gráfica de valores atípicos sobre el número de escrituras a disco tras la obtención de un informe de rendimiento.*

Tras realizarse una revisión de los parámetros empleados para la obtención del informe de rendimiento correspondiente al valor atípico detectado se encontró un incremento en el rango de fechas empleado para su obtención, abarcando un periodo de informe de 1 año y 6 meses, lo cual conlleva a la obtención de un mayor número de datos a emplear para la obtención y despliegue de resultados. Dado lo anterior y debido a que el periodo de obtención de rendimiento organizacional en los demás informes no excede el periodo de 1 año en más de 1 mes, se decidió eliminar este registro con la finalidad de evitar que éste genere inconsistencias al momento de proceder con la serie de análisis estadísticos subsecuentes.

Tras eliminar el valor atípico encontrado se ejecutó nuevamente una prueba de valores atípicos correspondiente al número total de escrituras a disco. La Figura 68 muestra los

resultados obtenidos, se puede observar que no existen valores atípicos por lo que se procede con el análisis de resultados.

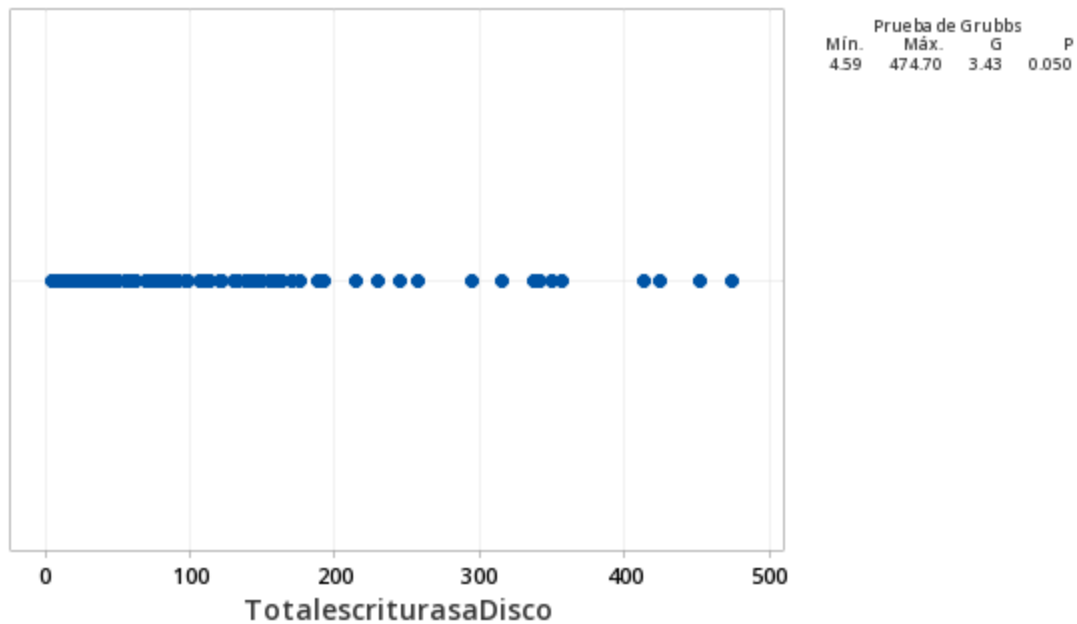


Figura 68. Gráfica de valores atípicos sobre el número de escrituras a disco tras la obtención de un informe de rendimiento (Tras la eliminación del valor atípico).

La Figura 69 muestra los resultados de una prueba de normalidad con un estadístico Anderson Darling aplicada al número de escrituras a disco tras la obtención de cada informe de rendimiento, observándose un valor p estimado ( $<0.005$ ) menor a 0.05, por lo tanto, los datos no se distribuyen de manera normal, considerando un nivel de confianza del 95%.



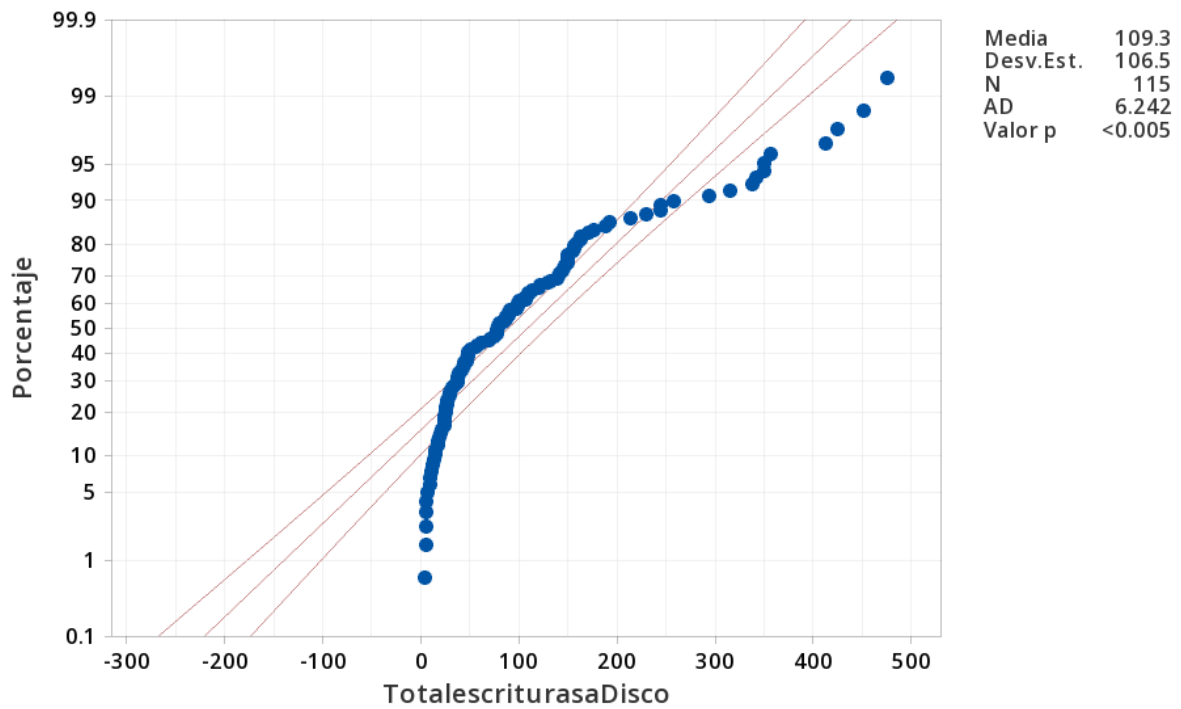


Figura 69. Prueba de normalidad aplicada sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.

El siguiente análisis realizado es un ANOM normal unidireccional del total de escrituras disco tras la obtención de un informe de rendimiento, presentado en la Figura 70. Se puede observar que el enfoque 2, correspondiente a una arquitectura de almacén de datos centralizado sin la presencia de índices ni claves primarias, presenta resultados por encima de la media, indicando que el empleo de este enfoque implica una repercusión negativa en cuanto al número de escrituras a disco tras la obtención de un informe de rendimiento. Adicionalmente, se observa que el enfoque 3, correspondiente a la arquitectura propuesta en este trabajo, presenta resultados por debajo de la media, lo cual indica que este enfoque reduce el número de escrituras a disco tras la obtención de un informe de rendimiento.

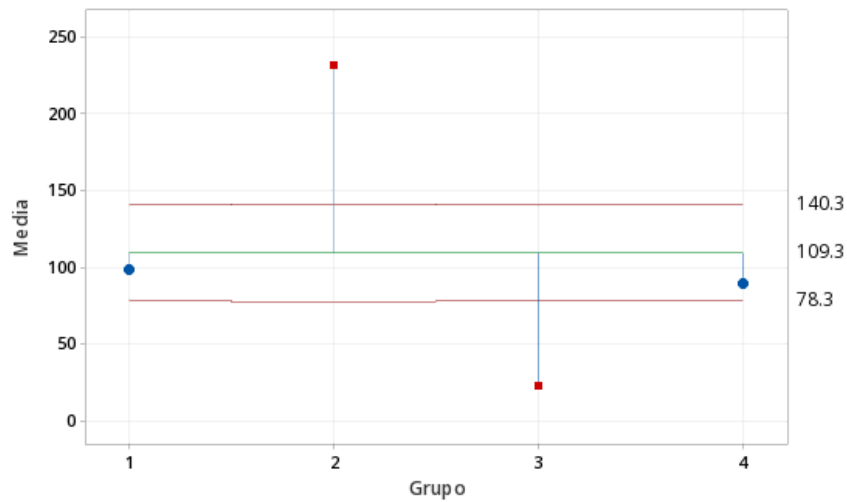


Figura 70. ANOM normal unidireccional del número de escrituras a disco tras la obtención de un informe de rendimiento.

Posterior al análisis de medias se realizó una prueba de igualdad de varianzas (Figura 71), considerando que los datos son no normales, se interpreta la prueba de Levene, es decir, no hay diferencias significativas entre las varianzas de los grupos ( $p = 0.252$  mayor que  $\alpha = 0.05$ ), detectándose una sobreposición entre los intervalos correspondientes a cada enfoque.

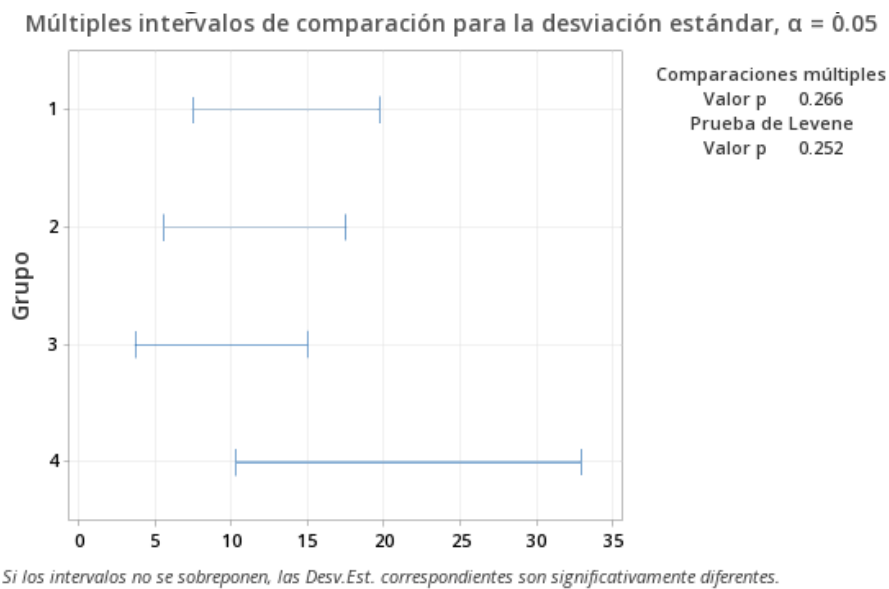


Figura 71. Prueba de igualdad de varianzas sobre las escrituras a disco tras la obtención de un informe de rendimiento.

Una vez realizado el análisis de varianzas se procede con un análisis ANOVA (Figura 72) y una prueba de Kruskal-Wallis (Figura 73).

| Fuente | GL  | SC Ajust. | MC Ajust. | Valor F | Valor p |
|--------|-----|-----------|-----------|---------|---------|
| Grupo  | 3   | 648007    | 216002    | 37.17   | 0.000   |
| Error  | 111 | 645111    | 5812      |         |         |
| Total  | 114 | 1293118   |           |         |         |

Figura 72. Análisis ANOVA sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.

Hipótesis nula  $H_0$ : Todas las medianas son iguales  
 Hipótesis alterna  $H_1$ : Al menos una mediana es diferente

| GL | Valor H | Valor p |
|----|---------|---------|
| 3  | 67.30   | 0.000   |

Figura 73. Prueba de Kruskal-Wallis sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.

| Grupo | N  | Media | Agrupación |
|-------|----|-------|------------|
| 2     | 28 | 231.3 | A          |
| 1     | 29 | 98.1  | B          |
| 4     | 29 | 89.0  | B          |
| 3     | 29 | 23.12 | C          |

Figura 74. Comparaciones en parejas de Tukey sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.

Tras analizar los resultados obtenidos en las dos pruebas anteriores se observa que en ambas el valor p es menor que el nivel de significancia empleado, siendo el nivel de significancia 0.05 y el valor p para ambas pruebas menor a 0.01, lo cual indica que las diferencias entre alguna de las medias son estadísticamente significativas. Por lo tanto, se rechaza una hipótesis nula, concluyendo que no todas las medias de población son iguales.

Adicionalmente, la prueba de Tukey (Figura 74) nos indica que el enfoque 2 y el enfoque 3 forman un grupo individual respectivamente, mientras que los enfoques 1 y 4 forman un grupo similar, por lo que únicamente las medias de los enfoques 1 y 4 son similares, mientras que la media del enfoque 3 es menor a las demás medias y la media del enfoque 4 es mayor.

La gráfica de intervalos presentada en la Figura 75 muestra la media y el intervalo de confianza para cada grupo, observándose una diferencia estadísticamente significativa entre el enfoque 2 y los enfoques restantes, siendo este enfoque el que más escrituras a disco realiza para la obtención de un informe de rendimiento. El enfoque 3, correspondiente a la propuesta de este trabajo, presenta una media de muestra menor a la de los otros enfoques, detectándose una diferencia estadísticamente significativa con respecto a los otros enfoques, indicando que el total de escrituras a disco utilizando este enfoque es significativamente menor al de los otros enfoques.

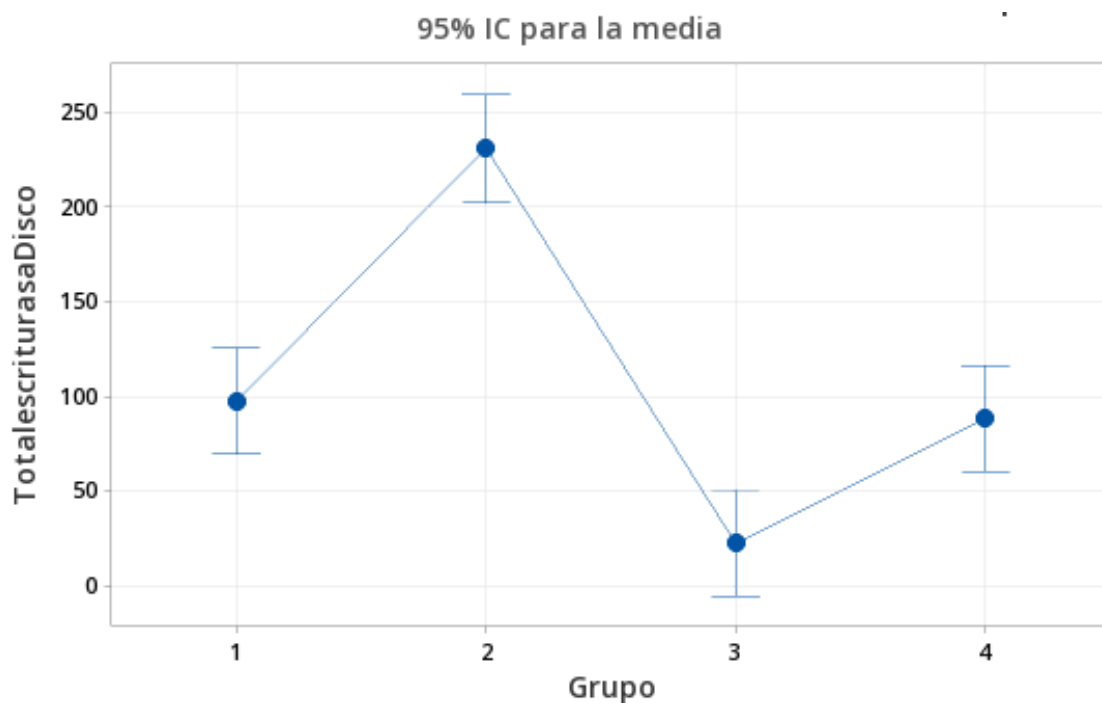


Figura 75. Gráfica de intervalos sobre el total de escrituras a disco tras la obtención de un informe de rendimiento.

#### 5.3.4. Resultados del uso de procesador

La Figura 76 muestra los resultados de la prueba de valores atípicos correspondiente al uso de procesador durante la obtención de cada informe de rendimiento. Se puede observar que no existen valores atípicos por lo que se procede con el análisis de resultados.

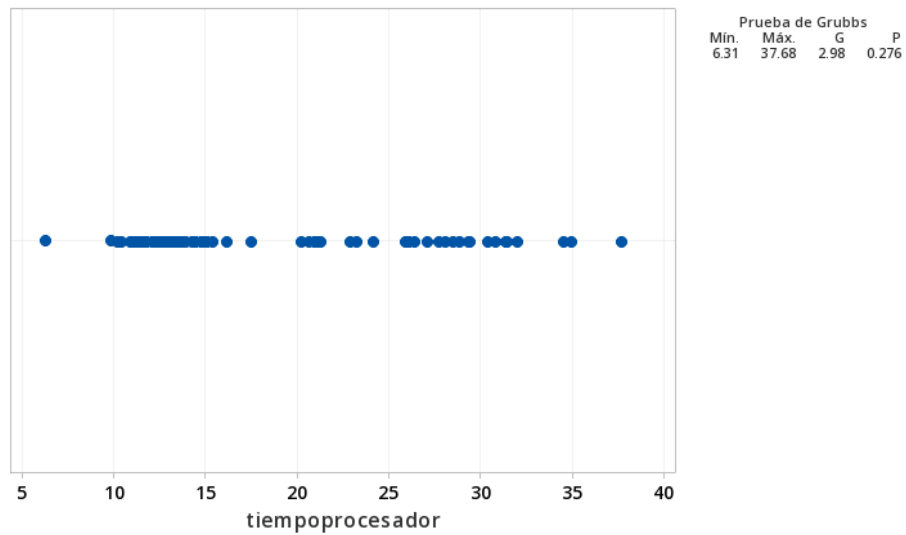


Figura 76. Gráfica de valores atípicos sobre el uso de procesador durante la obtención de un informe de rendimiento.

La Figura 77 muestra los resultados de una prueba de normalidad con un estadístico Anderson Darling aplicada al uso de procesador durante la obtención de cada informe de rendimiento, observándose un valor p estimado ( $<0.005$ ) menor a 0.05, por lo tanto, los datos no se distribuyen de manera normal, considerando un nivel de confianza del 95%.

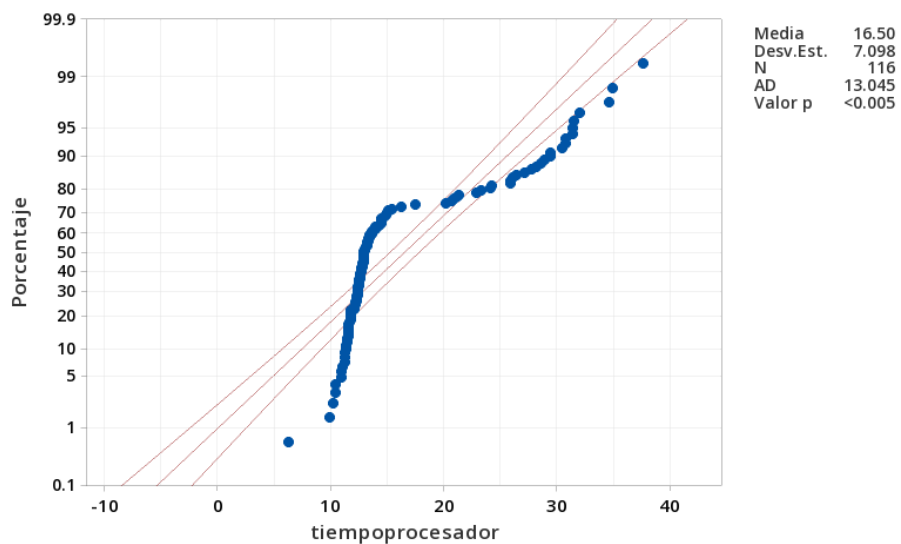


Figura 77. Prueba de normalidad aplicada sobre el uso de procesador durante la obtención de un informe de rendimiento.

El siguiente análisis realizado es un ANOM normal unidireccional del uso de procesador durante la obtención de un informe de rendimiento, presentado en la Figura 78. Se puede observar que el enfoque 3, propuesto en este trabajo, presenta resultados por encima de la media, indicando que el empleo de este enfoque implica una repercusión negativa en cuanto al porcentaje de uso de procesador durante la obtención de un informe de rendimiento, siendo mayor que al de los demás enfoques.

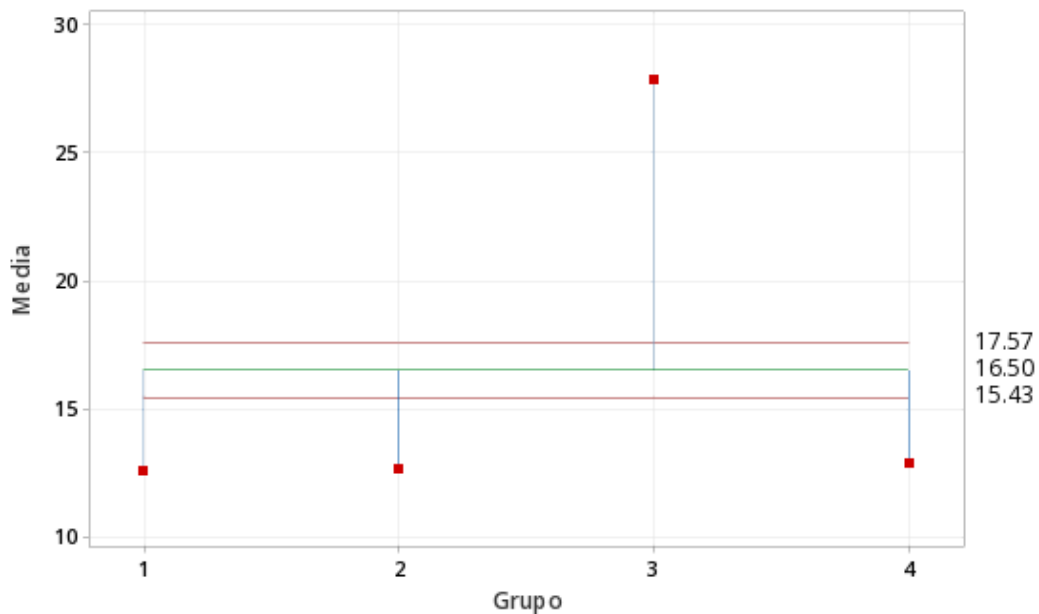


Figura 78. ANOM normal unidireccional del uso de procesador durante la obtención de un informe de rendimiento.

Posterior al análisis de medias se realizó una prueba de igualdad de varianzas (Figura 79), considerando que los datos son no normales, se interpreta la prueba de Levene, es decir, hay diferencias significativas entre las varianzas de los grupos ( $p = 0.000$  menor que  $\alpha = 0.05$ ). Adicionalmente, se detecta una sobreposición entre los intervalos correspondientes a los enfoques 1, 2 y 4. No siendo el caso para el enfoque 3, presentando una desviación estándar significativamente diferente al de los otros enfoques.

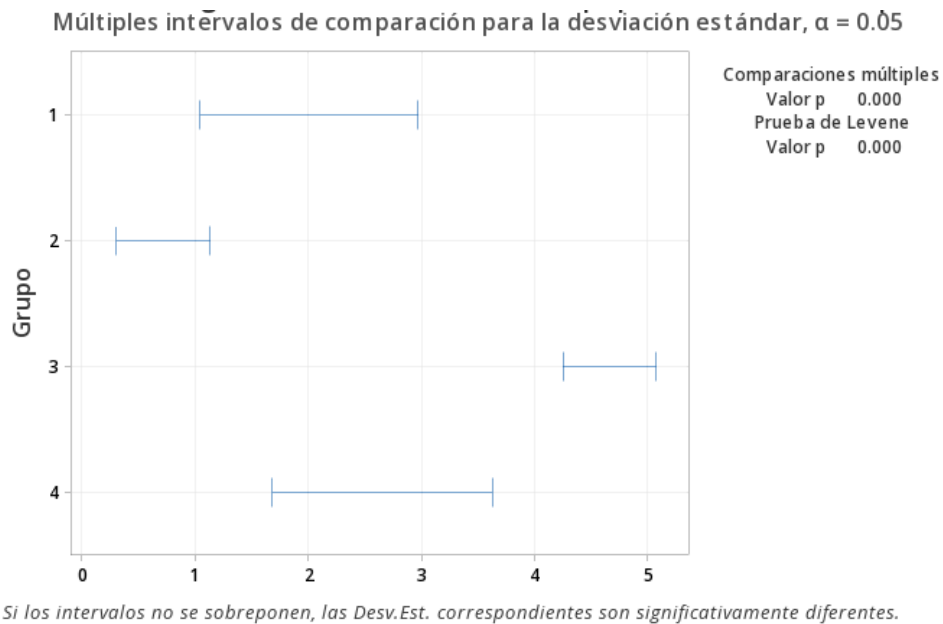


Figura 79. Prueba de igualdad de varianzas sobre el uso de procesador durante la obtención de un informe de rendimiento.

Una vez realizado el análisis de varianzas se procede con un análisis ANOVA (Figura 80) y una prueba de Kruskal-Wallis (Figura 81).

| Fuente | GL  | SC Ajust. | MC Ajust. | Valor F | Valor p |
|--------|-----|-----------|-----------|---------|---------|
| Grupo  | 3   | 4997.3    | 1665.78   | 234.44  | 0.000   |
| Error  | 112 | 795.8     | 7.11      |         |         |
| Total  | 115 | 5793.1    |           |         |         |

Figura 80. Análisis ANOVA sobre el uso de procesador durante la obtención de un informe de rendimiento.

Hipótesis nula  $H_0$ : Todas las medianas son iguales  
 Hipótesis alterna  $H_1$ : Al menos una mediana es diferente

| GL | Valor H | Valor p |
|----|---------|---------|
| 3  | 65.66   | 0.000   |

Figura 81. Prueba de Kruskal-Wallis sobre el uso de procesador durante la obtención de un informe de rendimiento.



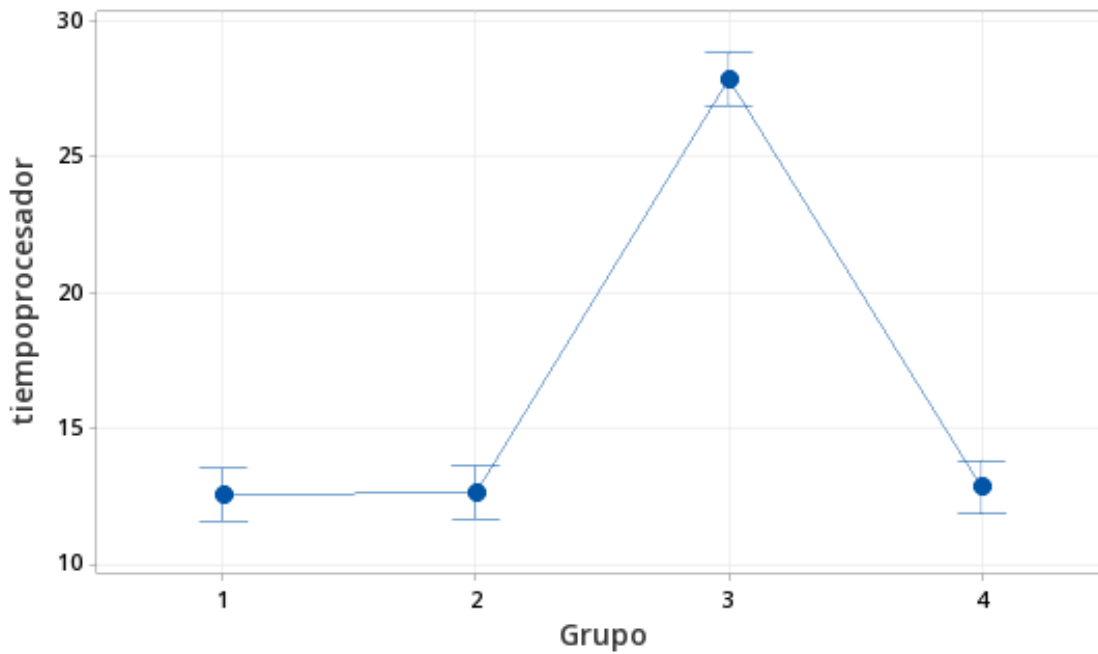
| Grupo | N  | Media  | Agrupación |
|-------|----|--------|------------|
| 3     | 29 | 27.867 | A          |
| 4     | 29 | 12.870 | B          |
| 2     | 29 | 12.662 | B          |
| 1     | 29 | 12.600 | B          |

*Figura 82. Comparaciones en parejas de Tukey sobre el uso de procesador durante la obtención de un informe de rendimiento.*

Tras analizar los resultados obtenidos en las dos pruebas anteriores se observa que en ambas el valor p es menor que el nivel de significancia empleado, siendo el nivel de significancia 0.05 y el valor p para ambas pruebas menor a 0.01, lo cual indica que las diferencias entre alguna de las medias son estadísticamente significativas. Por lo tanto, se rechaza una hipótesis nula, concluyendo que no todas las medias de población son iguales.

Adicionalmente, la prueba de Tukey (Figura 82) nos indica que el enfoque 3 forma un grupo individual, y los enfoques 1, 2, 4 forman un grupo similar, por lo que la media del enfoque 3 es muy diferente a los otros grupos.

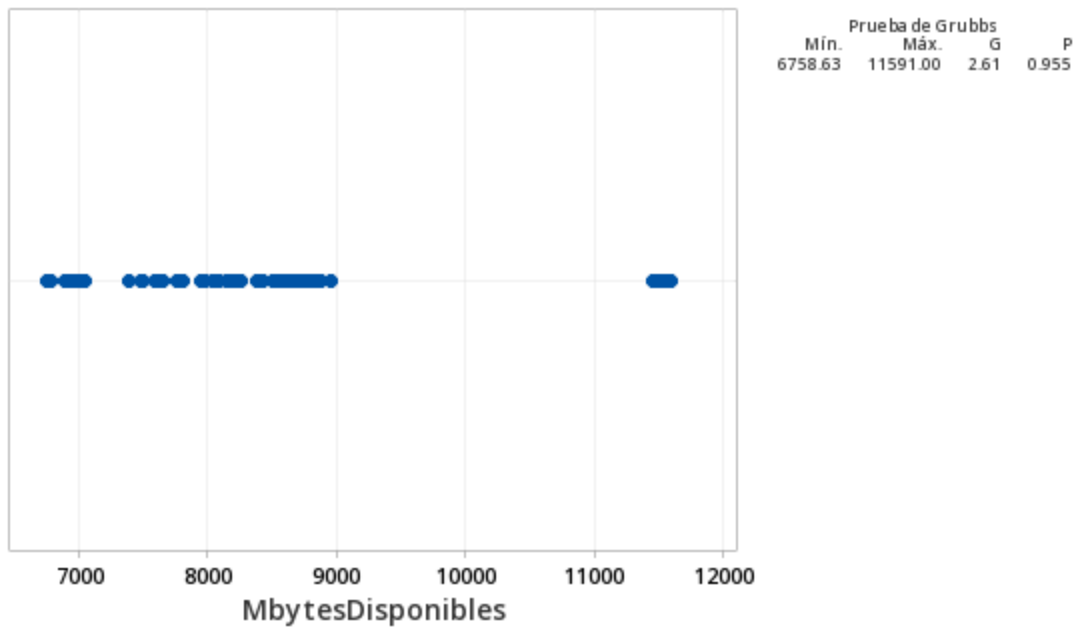
La gráfica de intervalos presentada en la Figura 83 muestra la media y el intervalo de confianza para cada grupo, observándose una diferencia estadísticamente significativa entre el enfoque 3 y los enfoques restantes, siendo este enfoque el que más tiempo de procesamiento emplea para la obtención de un informe de rendimiento. Adicionalmente, no se detecta una diferencia estadísticamente significativa entre los enfoques restantes.



*Figura 83. Gráfica de intervalos sobre el uso de procesador durante la obtención de un informe de rendimiento.*

### **5.3.5. Resultados del uso de memoria**

La Figura 84 muestra los resultados de la prueba de valores atípicos correspondiente a la memoria disponible en MBytes durante la obtención de cada informe de rendimiento. Se puede observar que no existen valores atípicos por lo que se procede con el análisis de resultados.



*Figura 84. Gráfica de valores atípicos sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.*

La Figura 85 muestra los resultados de una prueba de normalidad con un estadístico Anderson Darling aplicada al tiempo empleado en segundos para la obtención de cada informe de rendimiento, observándose un valor p estimado ( $<0.005$ ) menor a 0.05, por lo tanto, los datos no se distribuyen de manera normal, considerando un nivel de confianza del 95%.

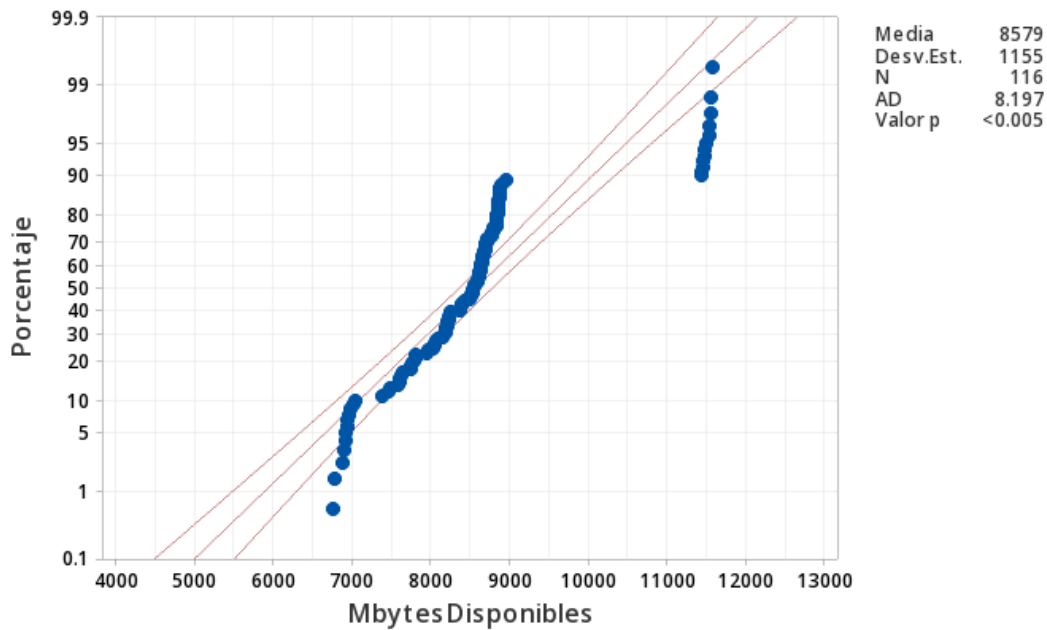
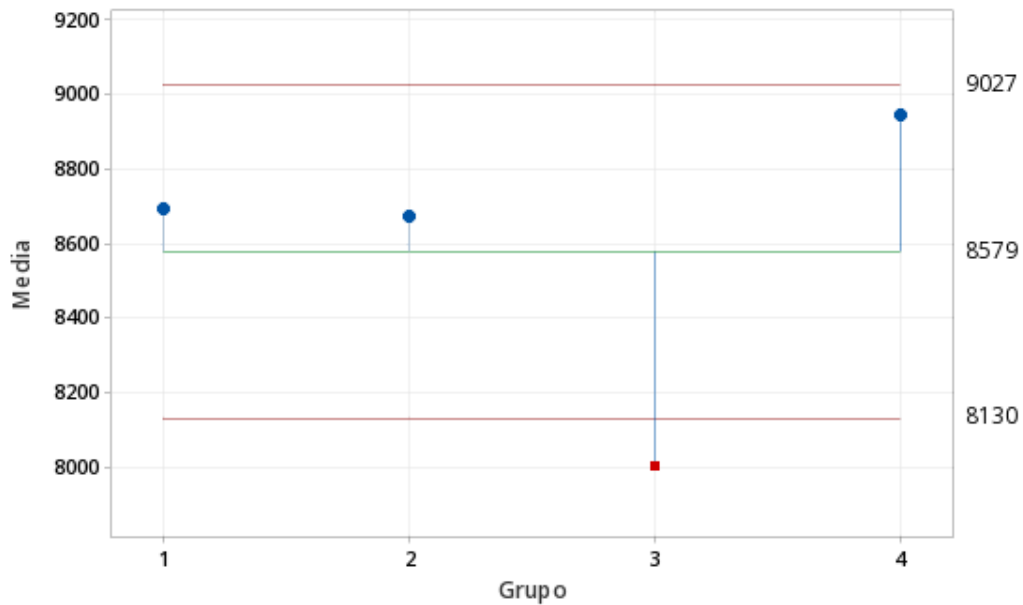


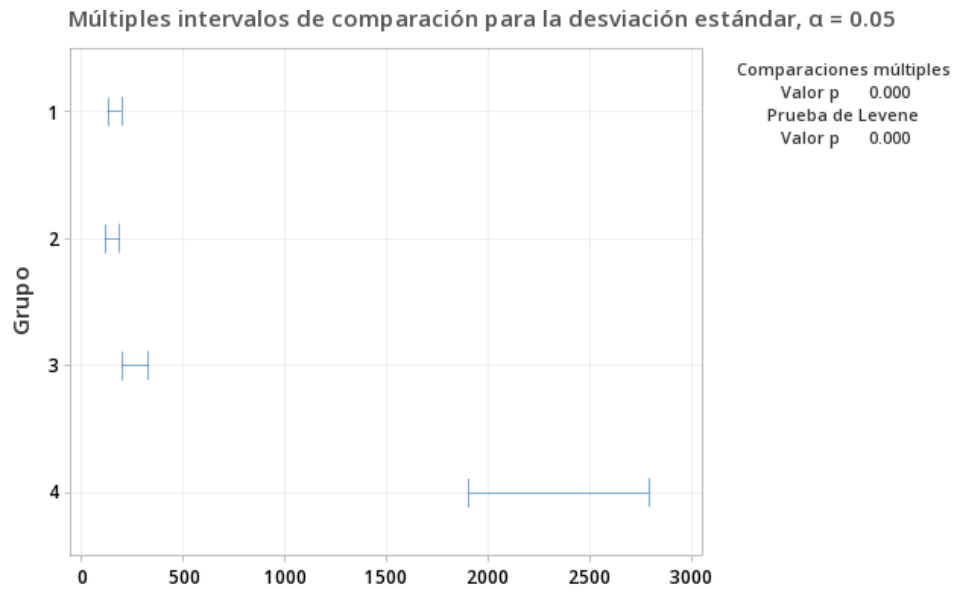
Figura 85. Prueba de normalidad aplicada sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.

El siguiente análisis realizado es un ANOM normal unidireccional del uso de procesador durante la obtención de un informe de rendimiento, presentado en la Figura 86. Se puede observar que el enfoque 3, propuesto en este trabajo, presenta resultados por debajo de la media, indicando en este indicador en particular, que el empleo de este enfoque implica una repercusión negativa en cuanto a la cantidad de memoria RAM (Del Inglés: Read Only Access) disponible durante la obtención de un informe de rendimiento.



*Figura 86. ANOM normal unidireccional de la memoria disponible en MBytes durante la obtención de un informe de rendimiento.*

Posterior al análisis de medias se realizó una prueba de igualdad de varianzas (Figura 87), considerando que los datos son no normales, se interpreta la prueba de Levene, es decir, hay diferencias significativas entre las varianzas de los grupos ( $p = 0.000$  menor que  $\alpha = 0.05$ ). Adicionalmente, se detecta una sobreposición entre los intervalos correspondientes a los enfoques 1, 2 y 3. No siendo el caso para el enfoque 4, presentando un intervalo mayor y una desviación estándar significativamente diferente al de los otros enfoques.



Si los intervalos no se superponen, las Desv.Est. correspondientes son significativamente diferentes.

Figura 87. Prueba de igualdad de varianzas sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.

Una vez realizado el análisis de varianzas se procede con un análisis ANOVA (Figura 88) y una prueba de Kruskal-Wallis (Figura 89).

| Fuente | GL  | SC Ajust. | MC Ajust. | Valor F | Valor p |
|--------|-----|-----------|-----------|---------|---------|
| Grupo  | 3   | 14203407  | 4734469   | 3.81    | 0.012   |
| Error  | 112 | 139212115 | 1242965   |         |         |
| Total  | 115 | 153415523 |           |         |         |

Figura 88. Análisis ANOVA sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.

Hipótesis nula  $H_0$ : Todas las medianas son iguales  
 Hipótesis alterna  $H_1$ : Al menos una mediana es diferente

| GL | Valor H | Valor p |
|----|---------|---------|
| 3  | 35.14   | 0.000   |

Figura 89. Prueba de Kruskal-Wallis sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.

| <u>Grupo</u> | <u>N</u> | <u>Media</u> | <u>Agrupación</u> |
|--------------|----------|--------------|-------------------|
| 4            | 29       | 8945         | A                 |
| 1            | 29       | 8695.0       | A B               |
| 2            | 29       | 8672.0       | A B               |
| 3            | 29       | 8001.6       | B                 |

*Figura 90. Comparaciones en parejas de Tukey sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.*

Tras analizar los resultados obtenidos en las dos pruebas anteriores se observa que en ambas el valor p es menor que el nivel de significancia empleado, siendo el nivel de significancia 0.05 y el valor p para en la prueba de Kruskal-Wallis menor a 0.01, lo cual indica que las diferencias entre alguna de las medias son estadísticamente significativas. Por lo tanto, se rechaza una hipótesis nula, concluyendo que no todas las medias de población son iguales.

Adicionalmente, la prueba de Tukey (Figura 90) nos indica que los enfoques 1,2, 4 forman un grupo similar, mientras que los enfoques 1,2, 3 forman otro grupo similar, por lo que únicamente las medias de los enfoques 3 y 4 son diferentes.

La gráfica de intervalos presentada en la Figura 91 muestra la media y el intervalo de confianza para cada grupo, observándose una diferencia entre el enfoque 3 y los enfoques restantes, no siendo estadísticamente significativa, siendo este enfoque el que más memoria RAM emplea durante la obtención de un informe de rendimiento.

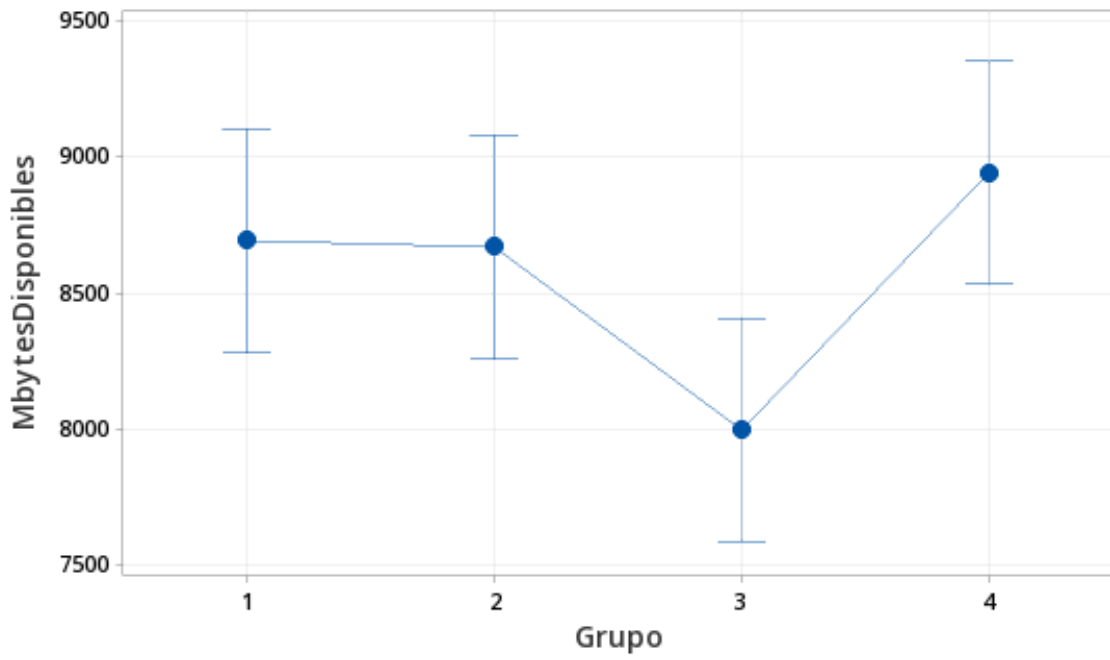


Figura 91. Gráfica de intervalos sobre la memoria disponible en MBytes durante la obtención de un informe de rendimiento.



## Capítulo 6. Conclusiones y trabajo futuro

### 6.1. Conclusiones

En la revisión sistemática de la literatura se resume que el uso de inteligencia de negocios beneficia el desempeño de una organización. Entre los beneficios encontrados se tienen mejoras en las actividades de planificación y análisis; procesos internos de una organización; en la capacidad de rendimiento; proceso de toma de decisiones; actividades de aprendizaje y crecimiento de los empleados; el rendimiento financiero y el rendimiento del cliente.

Se identificaron las siguientes arquitecturas en la literatura: Data Warehouse Bus Architecture, Corporate Information Factory, Centralized Data Warehouse Architecture, Federated Architecture, Data Delivery Platform, y Data Virtualization. La arquitectura seleccionada fue virtualización de datos considerando los siguientes criterios: calidad de los datos, desacoplamiento, reporte operativo, mantenibilidad, tiempo de desarrollo y desempeño.

El desarrollo de esta investigación resultó en la propuesta de una herramienta de inteligencia de negocios a emplear en programas de visitas a domicilio a niños en edad temprana. Esta herramienta será de ayuda para las agencias implementadoras del programa federal MIECHV, permitiéndoles realizar la medición del rendimiento obtenido tras la prestación de servicios a niños en edad temprana. La herramienta desarrollada ofrece la funcionalidad de: reporte operativo, selección del tipo de gráfica, configuración de criterios de aceptación y exportación de información. Conteniendo las siguientes características clave:

- Permite al usuario final el análisis de resultados obtenidos mediante un enfoque visual.

- Respetar lo estipulado por el programa federal MIECHV gestionado por HRSA.
- Visualización detallada de la información empleada para el cálculo de cada indicador de rendimiento.
- Exportación de la información empleada por cada indicador de rendimiento.

Se realizó un experimento controlado considerando cuatro grupos: 1) reporte de datos operativo; 2) arquitectura de almacén de datos centralizada sin empleo de índices 3) arquitectura de virtualización de datos y 4) arquitectura de almacén de datos centralizada mediante el empleo de índices. Se compararon y analizaron estadísticamente las siguientes variables: tiempo de respuesta, número de lecturas, número de escrituras a disco, uso de tiempo de procesamiento y uso de memoria.

El enfoque propuesto muestra mejores resultados en el tiempo de respuesta, el número de lecturas y el número de escrituras a disco. Cuando la información es precargada en una base de datos adicional, se reduce la interferencia a las bases de datos de producción, disminuyendo el número de lecturas y escrituras a disco tras la obtención de un informe de rendimiento. Así como la reducción del tiempo empleado tras la obtención de un informe de rendimiento en el tiempo de respuesta.

El uso de recursos computacionales en el enfoque de virtualización de datos no generó resultados favorables debido a que se observó un mayor uso de tiempo de procesamiento y de memoria RAM.

## **6.2. Contribuciones**

El presente trabajo de investigación aporta conocimiento nuevo respecto a los beneficios obtenidos del uso de una herramienta de inteligencia de negocios para la gestión de rendimiento dentro de programas de visitas a domicilio a niños en edad temprana.

Adicionalmente, se propone una solución de inteligencia de negocios tomando en cuenta los factores de éxito y beneficios encontrados en la literatura, centrándose en las necesidades específicas requeridas por el programa federal MIECHV, diseñado por (Health Resources & Services Administration, 2020).

Mediante la solución propuesta, los usuarios de una agencia implementadora del programa federal MIECHV, lograrán un mejor entendimiento del rendimiento por medio de las vistas gráficas (Dashboard), soportando mejores decisiones durante las visitas a domicilio a niños en edad temprana.

## **6.3. Trabajo futuro**

Como trabajo futuro se planea aumentar el conjunto de áreas de medición e indicadores ofrecidos por la herramienta de inteligencia de negocios, agregando al panel de visualización las siguientes 5 áreas de rendimiento y un total de 12 indicadores asociados a estas: 1) Prevención de lesiones infantiles, abuso infantil, negligencia o maltrato y reducción de visitas a sala de emergencias; 2) Mejoras en la preparación escolar y el rendimiento académico de los niños; 3) Reducción de la delincuencia o la violencia doméstica; 4) Mejoras en la autosuficiencia económica familiar y 5) Mejoras en la coordinación y referencias para otros recursos y apoyos de la comunidad.

En una segunda etapa, continuar con la implementación de los indicadores y sus áreas de medición, se plantea una prueba piloto directamente con los usuarios del sistema Visit Tracker Web. Dado que la agencia seleccionada para el experimento controlado pertenece al estado de Illinois en los Estados Unidos de América, la prueba piloto considera incorporar a las siguientes 3 agencias implementadoras en este estado: 1) Bright Start; 2) Healthy Start HFA y 3) Rockford Public Schools. Estas agencias implementadoras brindan servicio al mayor número de familias y niños. Adicionalmente se realizará una identificación, determinación y mitigación de riesgos del plan piloto.

En una tercera etapa se tiene contemplado el desarrollo de una herramienta de evaluación que permita medir los atributos de calidad no funcionales como: usabilidad, fiabilidad e interoperatividad directo con los usuarios de las agencias implementadoras. También deseamos validar los beneficios directos en la implementación de inteligencia de negocios considerando aspectos intangibles, considerando que el alcance este trabajo solo se limite a el tiempo empleado en acceder a la información y evaluar: 1) Actividades de planificación; 2) Actividades de análisis; 3) Capacidad de medición de rendimiento y 4) Precisión al acceder a la información.

Por último, se contempla publicar un artículo científico con los resultados obtenidos, en la publicación de HomVEE (U.S. Department of Health & Human Services, 2020).

## Referencias

- Kitchenham, B. (2004). Systematic reviews. *10th International Symposium on Software Metrics*.
- Alnoukari, M., & Hanano, A. (2017). Integration of business intelligence with corporate. *Journal of Intelligence Studies in Business*, 5-16.
- Athena Software. (2020). *Penelope [Web application software]*. Retrieved from <https://www.athenasoftware.net/>
- Basili, V. R., & Weiss, D. M. (1984). A methodology for collecting valid software engineering data. *IEEE Transactions on Software Engineering*, 1, 728-738.
- Basili, V. R., Caldiera, G., & Rombach, D. H. (1994). The Goal Question Metric Approach. *Encyclopedia of Software Engineering*, 2, 528-532.
- Bass, L., Clements, P., & Kazman, R. (2012). *Software Architecture in Practice: Software Architect Practice*. Addison-Wesley.
- Brown, S. (2018, 06 25). *The C4 model for visualising software architecture*. Retrieved from <https://c4model.com/>
- DataKeeper Technologies. (2020). *Visit Tracker [Web application software]*. Retrieved from <https://www.visittrackerweb.com/>
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics*. Boston, Massachusetts: Harvard Business Review Press.
- Elkan, R; et al. (2000). The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature. *Health Technology Assessment*, 4(13).

- Evelson, B. (2008). *Topic Overview: Business Intelligence*. Retrieved from Forrester Research:  
<https://www.forrester.com/report/Topic+Overview+Business+Intelligence/-/E-RES39218>
- Gartner. (2017, February 17). *Gartner Says Worldwide Business Intelligence and Analytics Market to Reach \$18.3 Billion in 2017*. Retrieved from Gartner:  
<https://www.gartner.com/en/newsroom/press-releases/2017-02-17-gartner-says-worldwide-business-intelligence-and-analytics-market-to-reach-18-billion-in-2017>
- Gartner. (2020, February 11). *Magic Quadrant for Analytics and Business Intelligence Platforms*. Retrieved 03 01, 2020, from Gartner:  
<https://www.gartner.com/doc/reprints?id=1-1YBTIWVR&ct=200211&st=sb>
- Grüning, M. (2002). *Performance-measurement-systeme – Messung und Steuerung von Unternehmensleistung*. Dissertation, Technische Universität Dresden.
- Hartl, K., Jacob, O., Mbep, F. L., Budree, A., & Fourie, L. (2016). The Impact of Business Intelligence on Corporate Performance Management. *49th Hawaii International Conference on System Sciences*. Koloa.
- Hauber, R. (2003). *Performance Measurement in der Forschung & Entwicklung*. Wiesbaden: Deutscher Universitäts-Verlag.
- Health Resources & Services Administration. (2020, May). *Data, Evaluation, & Continuous Quality Improvement*. (U.S. Department of Health and Human Services) Retrieved May 03, 2020, from <https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting/home-visiting-program-technical-assistance/performance-reporting-and-evaluation-resources>

- Health Resources & Services Administration. (2020, 1 20). *Home Visiting*. Retrieved from Maternal and Child Health Bureau: <https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting-overview>
- Health Resources & Services Administration. (n.d.). *MIECHV Performance Indicators and Systems Outcomes*. Retrieved 05 03, 2020, from <https://mchb.hrsa.gov/sites/default/files/mchb/MaternalChildHealthInitiatives/HomeVisiting/Federal-Home-Visiting-Program-Performance-Indicators-and-Systems-Outcomes-Summary.pdf>
- Hou, C.-K. (2016). Using the balanced scorecard in assessing the impact of BI system usage on organizational performance: An empirical study of Taiwan's semiconductor industry. *Information Development*, 32(5), 1545-1569.
- Kaplan, R. S., & Norton, D. P. (1996). *The Balanced ScoreCard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- Loewen, L., & Roudsari, A. (2017). Evidence for Business Intelligence in Health Care: A Literature Review. *Studies In Health Technology And Informatics*, 235, 579.
- Luhn, H. P. (1958). A Business Intelligence System. *IBM Journal*, 314-319.
- Minitab, LLC. (2021, 06 29). *Data Analysis Software | Statistical Software Package | Minitab*. Retrieved from Minitab: <https://www.minitab.com/>
- Owusu, A. (2017). Business intelligence systems and bank performance in Ghana: The balanced scorecard approach. *Cogent Business & Management*, 4(1).

- Peters, M. D., Wieder, B., & Sutton, S. G. (2018). Organizational improvisation and the reduced usefulness of performance measurement BI functionalities. *International Journal of Accounting Information Systems*, 29, 1-15.
- Rajnoha, R; et al. (2016). Business intelligence as a key information and knowledge tool for strategic business performance management. *Economics and Management*, 19(1), 183-203.
- Richards, G., Yeoh, W., Chong, A. Y.-L., & Popovič, A. (2014). AN EMPIRICAL STUDY OF BUSINESS INTELLIGENCE IMPACT ON CORPORATE PERFORMANCE MANAGEMENT. *PACIS 2014 Proceedings*, (p. 341).
- Špičák, D. (2018). The Relation Between Using Business Intelligence Solutions and Corporate Performance. *The Impact of Globalization on International Finance and Accounting* (pp. 395-401). Springer, Cham.
- Sweet, M., & Appelbaum, M. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, 75(5).
- Thamir, A., & Poulis, E. (2015). Business Intelligence Capabilities and Implementation Strategies. *International Journal of Global Business*, 8(1), 34-45.
- TIOBE - The Software Quality Company. (2020, April 1). *TIOBE Index for April 2020*. (TIOBE Software BV) Retrieved April 25, 2020, from <https://www.tiobe.com/tiobe-index/>



- U.S. Department of Health & Human Services. (2020, 01 23). *What is Home Visiting Evidence of Effectiveness?* Retrieved from Home Visiting Evidence of Effectiveness: <https://homvee.acf.hhs.gov/>
- Vallurupalli, V., & Bose, I. (2018). Business intelligence for performance measurement: A case based analysis. *Decision Support Systems*, 72-85.
- Van der Lans, R. F. (2012). *Data Virtualization for Business Intelligence Systems*. Elsevier Inc.
- Wohlin, C., Höst, M., & Henningson, K. (2003). Empirical Research Methods in Software Engineering. *Empirical Methods and Studies in Software Engineering*, 7-23.
- Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*, 50(3), 23-32.
- Zaied, A. N., Grida, M., & Hussein, G. S. (2018). Evaluation of critical success factors for business intelligence systems using fuzzy AHP. *Journal of Theoretical and Applied Information Technology*, 96(19), 6406-6422.