

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**



**FACULTAD DE INGENIERÍA**

**MAESTRÍA EN SISTEMAS DE INFORMACIÓN CON MENCIÓN EN DATA  
SCIENCE**

**“Diseño de una solución de inteligencia de negocios (BI) que apoye en la toma de  
decisiones en el sector de la construcción”**

**PROYECTO DE GRADUACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE MASTER EN  
SISTEMAS DE INFORMACIÓN CON MENCIÓN EN DATA SCIENCE**

**Sarango Rubio Álvaro Javier**

**Director**

**Ing. Rafael Melgarejo Heredia**

**Quito, 2023**

## **DEDICATORIA**

*Este proyecto se lo dedico a mis padres y hermanos, quienes sin su compañía esto no hubiese sido posible forman un pilar fundamental tanto en mi vida profesional como académica.*

*A mi abuelita Hortensia que me acompaña desde el cielo.*

**Álvaro Javier Sarango Rubio**

## **AGRADECIMIENTO**

*Quiero agradecer en primer lugar a dios por brindarme vida y salud, a mis padres y hermanos quien nunca han dejado de apoyarme.*

*Al Doctor Rafael Melgarejo por su acompañamiento, ayuda y paciencia en la elaboración de este proyecto.*

**Álvaro Javier Sarango Rubio**

# ÍNDICE GENERAL

DEDICATORIA .....	I
AGRADECIMIENTO .....	II
INDICE DE TABLAS .....	VIII
INDICE DE FIGURAS .....	IX
LISTAS DE ANEXOS .....	XI
Resumen.....	XII
Abstract.....	XIII
INTRODUCCIÓN .....	1
CAPÍTULO I .....	3
1. PROBLEMA DE LA INVESTIGACIÓN.....	3
1.1. JUSTIFICACIÓN .....	3
1.1.1. Contexto de la Industria de la construcción y las nuevas tendencias de manejo de información.....	4
1.2. PROBLEMA DE INVESTIGACIÓN.....	5
1.3. PREGUNTAS DIRECTRICES .....	6
1.3.1. Pregunta general .....	6
1.3.2. Preguntas específicas .....	6
1.4. OBJETIVOS .....	7
1.4.1. Objetivo general.....	7
1.4.2. Objetivos específicos .....	7
1.5. ALCANCE.....	7
1.6. METODOLOGÍA .....	8
1.7. Tipo de Investigación.....	8
CAPITULO II.....	9
2. BUSINESS INTELLIGENCE.....	9
2.1. Orígenes del Bussines Intelligence .....	9
2.2. BI y el sector de la construcción en Ecuador .....	10
2.3. Marco referencial .....	11
2.4. Marco Conceptual .....	13

2.4.1.	Conceptos básicos de la inteligencia de negocios .....	13
2.4.2.	Inteligencia de negocios.....	13
2.4.3.	Importancia de la Inteligencia de negocios.....	14
2.4.4.	Componentes de Bussines Intelligence.....	14
2.4.5.	Toma de decisiones (TDD).....	15
2.4.6.	Sistema de soporte a la decisión (DSS) .....	15
2.4.7.	Sistemas de información (SI).....	16
2.4.8.	Sistemas de Información Gerenciales (MIS) .....	17
2.4.9.	Cuadro de Mandos Integrales (CMI) .....	18
2.4.10.	Sistemas de business intelligence .....	18
2.5.	Metodología Ralph Kimball.....	19
2.5.1.	Ciclo de vida Ralph Kimball .....	21
2.5.2.	Extracción, Transformación y Carga (ETL) .....	22
2.5.3.	Data Warehouse (DW).....	23
2.5.4.	DataMart (DM) .....	23
2.5.5.	Esquema Estrella.....	23
2.5.6.	Esquema Copo de nieve.....	24
2.5.7.	Modelo Tabular.....	24
2.5.8.	Modelos Multidimensional .....	24
2.5.9.	SQL Server Management Studio (SMSS) .....	25
2.5.10.	SQL Server Integration Services (SSIS).....	25
2.5.11.	SQL Server Analysis Services (SSAS).....	26
2.6.	Métodos de agrupación .....	26
2.6.1.1.	Algoritmos de clasificación.....	26
2.6.1.	Tipos de algoritmos .....	27
2.6.1.1.	Algoritmos jerárquicos .....	27
2.6.1.2.	K-means.....	28
2.6.1.3.	DBSCAN.....	28
2.6.1.4.	Algoritmos de regresión .....	28
2.6.1.5.	Algoritmos de predicción .....	29
2.7.	Minería de datos .....	29

2.7.1.	Metodología Crisp-DM.....	30
2.7.1.1.	Fundamentos Básicos .....	30
2.8.	Elementos principales de la Metodología Crisp-DM.....	31
2.9.	Fases de la metodología Crisp-DM.....	31
2.9.1.	Comprensión del negocio .....	31
2.9.1.1.	Compresión de los datos.....	32
2.9.1.2.	Modelado.....	33
2.9.1.3.	Evaluación .....	34
2.9.1.4.	Despliegue .....	35
2.9.2.	Microsoft Power BI .....	35
2.9.3.	Microsoft Power BI Desktop.....	36
CAPITULO III.....		37
3.	METODOLOGÍA.....	37
3.1.	Flujo de proceso aplicado para el diseño de la solución de inteligencia de negocios.....	37
3.1.1.	Definición de requerimientos del negocio .....	39
3.2.	Diseño e Implementación de Subsistemas ETL.....	40
3.2.1.	Diseño de solución de Bussines Intelligence en Integration Services .....	40
3.2.2.	Selección de herramientas .....	40
3.2.3.	Estructura de datos.....	41
3.2.4.	Tabla de hechos y dimensiones .....	41
3.2.4.1.	Modelado Dimensional .....	41
3.2.4.2.	Tabla de hechos .....	42
3.2.4.3.	Tablas de dimensiones.....	42
3.2.4.4.	Contexto y universo.....	44
3.2.4.5.	Diseño Conceptual .....	45
3.2.4.6.	Diseño Lógico .....	46
3.2.4.7.	Diseño Físico.....	47
3.3.	Diseño e Implementación del ETL .....	48
3.3.1.	Diseño ETL.....	48
3.4.	Metodología de minería de datos .....	49

3.5.	Aplicación de la metodología Crisp-DM .....	50
3.5.1.	Comprensión del Negocio .....	50
3.5.2.	Determinar los objetivos del negocio .....	50
3.5.3.	Contexto.....	51
3.5.4.	Objetivo del Negocio .....	51
3.5.5.	Objetivos de la aplicación de inteligencia de negocios .....	51
3.5.6.	Recolección de datos iniciales .....	52
3.5.6.1.	Comprensión de los datos.....	52
3.5.6.2.	Descripción de los datos.....	53
3.5.6.3.	Exploración de los datos.....	53
3.5.6.4.	Verificar la calidad de los datos .....	56
3.5.7.	Preparación de los Datos.....	57
3.5.8.	Script SQL modelo de datos .....	57
3.5.8.1.	Script tabla de hechos.....	57
3.5.8.2.	Limpieza de los datos .....	58
3.5.8.3.	Integración de los datos .....	59
3.5.8.4.	Formato de los datos.....	60
3.5.9.	Modelado .....	60
3.5.9.1.	Selección de técnica de modelado.....	61
3.5.9.2.	Validación de técnicas de modelado .....	61
CAPÍTULO IV .....		62
4.	RESULTADOS .....	62
4.1.	Principales factores que inciden en el incremento de nuevas edificaciones.....	64
4.2.	Resultados de los grupos utilizando K medias.....	65
4.2.1.	Análisis y validación de resultados de K-medias .....	68
4.3.	Resultados de los grupos utilizando DBSCAN.....	72
4.3.1.	Análisis y validación de resultados de DBSCAN.....	73
4.3.2.	Comparativa de métricas de grupos K-medias – DBSCAN .....	74
4.4.	Modelo de Regresión lineal múltiple .....	75
4.4.1.	Selección de la técnica de la técnica de modelado .....	75

4.4.2.	Construcción del modelo .....	75
4.4.3.	Variables a utilizar .....	75
4.4.4.	Análisis y validación de los resultados de la estimación lineal .....	76
CAPÍTULO V .....		78
5.	CONCLUSIONES Y RECOMENDACIONES .....	78
5.1.	Conclusiones .....	78
5.2.	Recomendaciones.....	80
BIBLIOGRAFÍA .....		81
ANEXOS .....		87

## INDICE DE TABLAS

<b>Tabla 1.</b> Comparativo de motores de base de datos.....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 2.</b> Características de la metodología Kimball. ....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 3.</b> Detalle de requerimientos de negocio.....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 4.</b> Detalle de tabla de hechos de Edificaciones. ....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 5.</b> Detalle de tabla de dimensiones de Provincias .....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 6.</b> Detalle de tabla de dimensiones de Región .....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 7.</b> Detalle de tabla de dimensiones de Construcción. <b>¡Error! Marcador no definido.</b>	
<b>Tabla 8.</b> Detalle de tabla de dimensiones de Recursos de construcción... <b>¡Error! Marcador no definido.</b>	
<b>Tabla 9.</b> Detalle de tabla de dimensiones de Propósito de construcción. <b>¡Error! Marcador no definido.</b>	
<b>Tabla 10.</b> Detalle de tabla de dimensión del tipo de construcción..... <b>¡Error! Marcador no definido.</b>	
<b>Tabla 11.</b> Detalle de tabla de dimensión del tipo de propiedad de la edificación..... <b>¡Error! Marcador no definido.</b>	
<b>Tabla 12.</b> Detalle de tabla de dimensión de tiempo. ....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 13.</b> Principales aspectos de la metodología CRISP-DM..... <b>¡Error! Marcador no definido.</b>	
<b>Tabla 14.</b> Principales variables de la encuesta ESED.....	52
<b>Tabla 15.</b> Número de permisos de construcción por provincia .....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 16.</b> Periodicidad de ejecución de la encuesta ESED....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 17.</b> Evaluación del modelo K-MEDIAS .....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 18.</b> Detalle de grupos por provincia.....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 19.</b> Evaluación del modelo DBSCAN .....	<b>¡Error! Marcador no definido.</b>
<b>Tabla 20.</b> Evaluación de K-MEDIAS y DBSCAN.....	<b>¡Error! Marcador no definido.</b>

## INDICE DE FIGURAS

<b>Figura 1.</b> Componentes de Bussines Intelligence. ....	<b>¡Error! Marcador no definido.</b>
<b>Figura 2.</b> Toma de decisiones y sistemas de inteligencia de negocios.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 3.</b> Fases de los sistemas de información.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 4.</b> Fases de necesidades de información.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 5.</b> Ciclo de vida Kimball.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 6.</b> Enfoque de trabajo Kimball.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 7.</b> Estructura del proceso ETL.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 8.</b> Gráfico de algoritmo de clasificación.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 9.</b> Gráfico de algoritmo de regresión.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 10.</b> Proceso de minería de datos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 11.</b> Ciclo de vida de minería de datos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 12.</b> Fases de la comprensión del negocio.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 13.</b> Fases de entendimiento de los datos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 14.</b> Fases de modelado de los datos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 15.</b> Evaluación de la técnica de minería de datos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 16.</b> Flujo de servicios de Power BI.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 17.</b> Flujo de solución de inteligencia de negocios.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 18.</b> Estructura arquitectura técnica.....	39
<b>Figura 19.</b> Diseño conceptual.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 20.</b> Diseño lógico.....	45
<b>Figura 21.</b> Diseño físico.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 22.</b> Carga de datos entorno DSA.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 23.</b> Archivos de fuente de datos.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 24.</b> Permisos de construcción en las provincias económicamente más importantes. ...	<b>¡Error! Marcador no definido.</b>
<b>Figura 25.</b> Proporción de permisos de construcción por región.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 26.</b> Número de tipo de construcciones.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 27.</b> Script sql tabla de hechos edificaciones.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 28.</b> Errores en data inicial.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 29.</b> Data corregida.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 30.</b> Dashboard de resumen general.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 31.</b> Dashboard de distribución de numero de edificaciones por provincia	<b>¡Error! Marcador no definido.2</b>
<b>Figura 32.</b> Número de grupos recomendados.....	<b>¡Error! Marcador no definido.</b>

<b>Figura 33.</b> Gráfico de grupos de edificaciones.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 34.</b> Gráfico de grupos de edificaciones.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 35.</b> Dendograma .....	<b>¡Error! Marcador no definido.</b>
<b>Figura 36.</b> Dendograma .....	<b>¡Error! Marcador no definido.</b>
<b>Figura 37.</b> Calidad de clústeres .....	<b>¡Error! Marcador no definido.</b>
<b>Figura 38.</b> Optimización de clústeres.....	<b>¡Error! Marcador no definido.</b>
<b>Figura 39.</b> Variabilidad entre grupos .....	<b>¡Error! Marcador no definido.</b>
<b>Figura 40.</b> Gráfico valor óptimo de $\epsilon$ .....	71
<b>Figura 41.</b> Grupos generados mediante el algoritmo DBSCAN .....	71
<b>Figura 42.</b> Relación de variables a utilizar en el modelo .....	74
<b>Figura 43.</b> Proyección de edificaciones a construir .....	75

## LISTAS DE ANEXOS

<b>Anexo 1.</b> Variables y descripción ESED.....	85
<b>Anexo 2.</b> Diseño de ETL de carga de información.....	86
<b>Anexo 3.</b> Modelo Tabular Analysis Services.....	86
<b>Anexo 4.</b> Scripts sql dimensión Construcción.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 5.</b> Scripts sql dimensión Propiedad.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 6.</b> Scripts sql dimensión Propósito Construcción.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 7.</b> Scripts sql dimensión Propósito Provincias.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 8.</b> Scripts sql dimensión Recursos.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 9.</b> Scripts sql dimensión Región.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 10.</b> Scripts sql dimensión Tiempo.....	<b>¡Error! Marcador no definido.</b>
<b>Anexo 11.</b> Scripts sql dimensión Tipo Construcción.....	90
<b>Anexo 12.</b> Scripts sql dimensión Zona.....	90
<b>Anexo 13.</b> Distancia euclidiana.....	91
<b>Anexo 14.</b> Relaciones de distancias y principales variables.....	91
<b>Anexo 15.</b> Clústeres generados con k-medias.....	92
<b>Anexo 16.</b> Relación de variables a utilizar en el modelo.....	92
<b>Anexo 17.</b> Grupos generados y migrados a Power BI.....	93
<b>Anexo 18.</b> Resultados del modelo de regresión.....	93
<b>Anexo 19.</b> Resultados de diagnóstico de residuos del modelo.....	<b>¡Error! Marcador no definido.</b>

**TITULO:** “DISEÑO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS (BI) QUE APOYEN EN LA TOMA DE DECISIONES EN EL SECTOR DE LA CONSTRUCCIÓN”

**Autor:** Sarango Rubio Álvaro Javier

**Tutor:** Dr. Rafael Melgarejo Heredia

## **Resumen**

En la actualidad el análisis de las grandes cantidades de información marca un punto de inflexión en el sostenimiento y crecimiento tanto de empresas como de sectores productivos, el sector de la construcción representa para la economía ecuatoriana un eje de crecimiento significativo dado su injerencia en la generación de empleos, activación económica y atracción de inversión extranjera, contar con una herramienta que consolide información tanto histórica como actual representa una gran ventaja puesto que se puede potenciar el análisis no solo a nivel descriptivo sino también generar un valor agregado, la presente investigación tiene el objetivo del diseño y construcción de una solución de inteligencia de negocios que permita realizar un procesamiento y consolidación de datos utilizando la metodología CRISP- DM y técnicas de minería de datos para su posterior consumo a través de tableros de control utilizando la información de la encuesta ESED que proporciona estadísticas de edificaciones.

**PALABRAS CLAVE:** DATA MART/ DATA WAREHOUSE/ ETL/ CRISP-DM/ BUSSINES INTELLIGENCE/ MINERÍA DE DATOS/ DASHBOARDS

**TITLE:** “DESIGN OF A BUSINESS INTELLIGENCE (BI) SOLUTION TO SUPPORT DECISION-MAKING IN THE CONSTRUCTION SECTOR”

**Author:** Sarango Rubio Álvaro Javier

**Tutor:** Dr. Rafael Melgarejo Heredia

### **Abstract**

Currently, the analysis of large amounts of information marks a turning point in the sustainability and growth of both companies and productive sectors. The construction sector represents an axis of significant growth for the Ecuadorian economy given its interference in the generation of jobs, economic activation and attraction of foreign investment, having a tool that consolidates both historical and current information represents a great advantage since the analysis can be enhanced not only at a descriptive level but also generate added value, the This research has the objective of designing and building a business intelligence solution that allows data processing and consolidation using the CRISP-DM methodology and data mining techniques for subsequent consumption through control panels using the information from the ESED survey that provides building statistics.

**KEYWORDS:** DATA MART/ DATA WAREHOUSE/ ETL/ CRISP-DM/ BUSSINES INTELLIGENCE/ DASHBOARDS

## INTRODUCCIÓN

En la actualidad las distintas empresas y organizaciones buscan ser más competitivas y ampliar su nicho de mercado en base a diferentes estrategias de crecimiento empresarial mismas que se sustentan en la toma de decisiones que a su vez depende de la situación de los diferentes indicadores de gestión.

Ante este panorama la información con la que cuenta una empresa representa un activo muy importante ya que a partir del procesamiento de la misma se puede analizar y determinar el pasado presente y futuro de una organización. Es por esto que se vuelve indispensable contar con metodologías específicas que faciliten el tratamiento de los datos.

Dentro del contexto ecuatoriano poco a poco las distintas empresas y organizaciones han ido entendiendo el valor de los datos que se producen día a día, de modo que se ha venido adoptando las nuevas tendencias en lo que respecta al análisis de los datos con el objetivo de generar un punto de inflexión y obtener una ventaja en el mercado que marque la diferencia con la competencia.

Para esto el uso, desarrollo y construcción de herramientas de bussines intelligence que contribuyan al procesamiento y análisis dentro de una empresa constituyen una necesidad sustancial que en el corto plazo organice la información, en el mediano plazo consolide y gestione la data y en el largo plazo constituirá un sistema sólido de indicadores de medición para la toma de decisiones.

Dentro de la innovación la aplicación de bussines intelligence se vuelve indispensable si lo que se busca es la excelencia, crecimiento y expansión dentro de un determinado nicho de mercado.

La presente investigación consta de cuatro capítulos el primero detalla las características del plan de investigación, el segundo abordará la temática relacionada al marco conceptual, referencial y estado del arte, en el tercer capítulo se hará énfasis en detallar la metodología CRISP-DM, el tercer capítulo número tres se centrará en el desarrollo de la solución de business intelligence, finalmente el capítulo final se refiere a las conclusiones y recomendaciones.

# CAPÍTULO I

## 1. PROBLEMA DE LA INVESTIGACIÓN

### 1.1. JUSTIFICACIÓN

El sector de la construcción constituye para el Ecuador una de las actividades que a lo largo de su historia ha contribuido mayormente a la dinamización de la economía siendo un eje de priorización por su importancia en la generación de plazas de empleo.

“La construcción es, además, un indicador clave de la evolución y crecimiento de la economía; así, cuando la economía está en recesión, el sector de la construcción es uno de los más afectados y, por el contrario, en épocas de bonanza económica, la construcción se constituye en uno de los sectores más dinámicos y activos” (Flacso, 2011)

Es por tal motivo que analizar continuamente la evolución y progreso del mismo significa un punto de inflexión en futuras normas y políticas que pudieran desarrollarse e implementarse en pro de conseguir un avance sostenido. Siempre y cuando se cuente con información confiable, veraz y disponible en todo momento de manera que todas las decisiones que puedan tomarse sean con conocimiento de todos los posibles escenarios y se deje de lado las condiciones de incertidumbre.

La toma de decisiones dentro de una organización, empresa o sector productivo puede describirse como el pilar de crecimiento pues afecta a todos los niveles y departamentos con que cuenta ya sea a nivel administrativo u operativo, basándose en elegir los mejores caminos y alternativas que se ajusten a las necesidades y recursos.

Sumado a la problemática que existe al momento de analizar grandes cantidades de información a medida que las dimensiones de los mismos se van incrementando. Los modelos de

bussines intelligence proporcionan un enorme beneficio dado que su estructura logra centralizar todas sus fuentes de información y la mantiene actualizada permanentemente.

### **1.1.1. Contexto de la Industria de la construcción y las nuevas tendencias de manejo de información**

Dentro de la dinámica de una economía existen varios sectores estratégicos cuyo desempeño es determinante para el crecimiento económico reflejado a través de varios aspectos e indicadores como por ejemplo el PIB y la tasa de desempleo.

“Un sector clave para el crecimiento económico es el de la Construcción, éste se ha visto severamente afectado por las crisis económicas. La construcción refleja el comportamiento macroeconómico nacional, ya que presenta un desempeño similar al de la economía en su conjunto. Además de ser uno de los sectores que más efectos multiplicadores tiene en otras actividades tanto en empleo como en la producción, el valor agregado que genera y la aportación en la formación bruta de capital fijo” (Robles Rodriguez & Velásquez García, 2001)

En el contexto actual la mayor parte de empresas dentro del mercado ecuatoriano han decidido invertir en áreas propias de BI y de ciencia de datos pues el beneficio, así como productividad de las mismas muestran una retribución a nivel de resultados que marcan una pauta importante en la gestión y dirección de una determinada compañía.

Desde hace algunos años el desarrollo y aparición de nuevas herramientas relacionadas a la gestión de datos ha generado una revolución y una nueva visión respecto al valor que ofrece un buen manejo de información todo esto debido al incremento exponencial en el volumen de datos.

(Hernández Leal, Duque Méndez, & Moreno Cadavid, 2017) mencionan que el incremento del volumen de datos a través de diversos sistemas ha conllevado una necesidad y cambio de cultura en la optimización de modelos de almacenamiento y procesamiento de

grandes cantidades de información, para esto se han desarrollado varios ecosistemas y aplicaciones adaptados para cumplir con la construcción de distintos indicadores empresariales basados en analítica avanzada.

La investigación propuesta arroja un enorme beneficio y transformación en la forma actual en la que se gestionan y analizan los datos ya que permitirá obtener mayor precisión a la hora de determinar información como ingresos, gastos, operatividad y avance de los mismos que a su vez logrará mejorar la efectividad, optimización, planificación y distribución de los distintos recursos de producción.

El desarrollo e implementación de una solución de inteligencia de negocios lograra generar una cultura en mejora continua en la gestión y control de los distintos proyectos y por ende agilizará la toma de decisiones y optimizará los análisis futuros respecto a la gestión de un determinado proyecto.

## **1.2. PROBLEMA DE INVESTIGACIÓN**

Uno de los pilares principales en los que se basa en crecimiento económico ecuatoriano se centra en el comportamiento y desarrollo del sector de la construcción ya a que a partir del buen desempeño de este se pueden generar diferentes plazas de empleo además de lograr dinamizar la economía. Dentro de la industria de la construcción en Ecuador el manejo y gestión de proyectos es de vital importancia para el fortalecimiento y crecimiento de esta, bajo este ese contexto la manipulación y uso de la información constituye un factor clave en la toma de decisiones. Ante este escenario nace la necesidad de que las empresas que constituyen el sector de construcción. En la actualidad los datos y estadísticas de diferentes pro-

yectos de construcción a nivel nacional se encuentran distribuidos en varias fuentes de información lo cual ocasiona un cierto nivel de dificultad a la hora de analizar y medir el desempeño del sector e industria de la construcción, además que el incremento en el flujo de la información complica que se logre construir una base de datos consolidada y actualizada.

Dentro de la industria de la construcción en Ecuador el manejo y gestión de proyectos es de vital importancia para el fortalecimiento y crecimiento de la misma, bajo este ese contexto la manipulación y uso de la información constituye un factor clave en la toma de decisiones.

Ante este escenario nace la necesidad de que las empresas que constituyen el sector de construcción, generen una herramienta que les facilite el análisis de la información y por ende la toma de decisiones y de manera global una solución basada en la inteligencia de negocios que permita evaluar y medir la evolución del sector.

### **1.3. PREGUNTAS DIRECTRICES**

#### **1.3.1. Pregunta general**

¿Es posible desarrollar una solución de bussines intelligence para el proceso de consumo y consolidación de información del sector de la construcción?

#### **1.3.2. Preguntas específicas**

¿Cómo integrar y procesar la información perteneciente a las estadísticas de edificaciones ESED para generar grupos y proyecciones basadas en la aplicación de técnicas de minería de datos que permitan tomar decisiones a las empresas del sector de la construcción?

¿Cuáles son los principales factores que inciden en el incremento de nuevas edificaciones?

¿De qué manera se puede mejorar el consumo de información actual para un análisis cuantitativo, actualizado y consistente más eficiente para las empresas del sector de la construcción?

¿De qué manera se puede caracterizar la información para obtener tendencias en el comportamiento y desarrollo del sector de la construcción?

## **1.4. OBJETIVOS**

### **1.4.1. Objetivo general**

Desarrollar una solución de inteligencia de negocios para el consumo y consolidación de información del sector de la construcción.

### **1.4.2. Objetivos específicos**

- Generar información con valor agregado que contribuya al análisis de las Estadísticas de edificaciones ESED.
- Determinar los principales factores que inciden en el incremento en la construcción de nuevas edificaciones.
- Diseñar una arquitectura de bussines intelligence que facilite el consumo y consolidación de información para su posterior análisis.
- Aplicar algoritmos de minería de datos que permitan generar grupos y proyecciones en el sector de la construcción.

## **1.5. ALCANCE**

El presente proyecto de investigación pretende generar una solución basada en bussines intelligence y metodologías de minería de datos que responda las necesidades en la gestión

y consumo de datos, que permita contar con información actualizada y fiable en cualquier momento a través de tableros de control.

La información para el desarrollo de esta investigación fue recolectada por el INEC<sup>1</sup> mediante la encuesta de Estadísticas de Edificaciones ESED<sup>2</sup> que recoge datos sobre los potenciales nuevos proyectos inmobiliarios y nuevos permisos de construcción.

## **1.6. METODOLOGÍA**

La metodología a usarse para el desarrollo e implementación de la solución de inteligencia de negocios se sustentará en Kimball en conjunto con la metodología Crisp-DM de manera que se facilite la integración paulatina de las diferentes etapas en proyectos relacionados con minería de datos a través de técnica de minería de datos.

## **1.7. Tipo de Investigación**

La investigación propuesta tiene un enfoque descriptivo analítico en función a la aplicación, seguimiento y desarrollo de las fases de implementación en función de la metodología Kimball para la construcción de un data Warehouse, en conjunto con las etapas de CRISP-DM para la generación de la solución de bussines intelligence.

---

<sup>1</sup> Instituto Nacional de Estadísticas y Censos órgano encargado de generar estadísticas oficiales del Ecuador.

<sup>2</sup> Estadísticas de Edificaciones es una operación estadística cuya función principal es recolectar información sobre el ritmo de crecimientos de los potenciales proyectos inmobiliarios.

## CAPITULO II

### 2. BUSINESS INTELLIGENCE

En este capítulo se abarcar los elementos más importantes teóricamente respecto a la inteligencia de negocios.

La definición de inteligencia de negocios en la actualidad ha tomado gran relevancia dentro de las empresas y organizaciones debido al alto crecimiento de flujo de información y el valor de la misma a la hora de tomar decisiones.

#### 2.1. Orígenes del Bussines Intelligence

Las primeras impresiones sobre bussines intelligence se dieron a conocer en la década de 1960 un periodo en donde las organizaciones y sectores productivos se apoyaban mayormente en los sistemas tradicionales y el nivel de analítica de datos era limitado debidos a los recursos disponibles y a las diferentes fuentes de información.

Con el avance de los años y el desarrollo tecnológico cada vez más fueron apareciendo alternativas en temas relacionado al procesamiento de datos, lo cual facilitaba cada vez más las tareas operativas y el rol del analista iba cambiando orientándose a un enfoque más apegado al de determinar y encontrar factores de comportamiento y de predecir futuros escenarios.

“Para los años 80’s y los 90’s, se empezaron a gestionar grandes bases de datos con lo que permitía analizar grandes cantidades de información en un tiempo muy corto y real. En el nuevo milenio se da lo que llamaremos el CRM (Customer Relationship Management), concepto que detallaremos más adelante. Finalmente, para el 2005 entra el término objeto de estudio de este artículo, Business Intelligence (BI)” (Muñoz, Osorio, & Zuñiga, 2016)

El termino como tal aparece en el año 1958 a través de la publicación Bussines Intelligence System de Hans Peter Luhn por aquel entonces investigador de IBM en donde menciones las primeras definiciones y aproximaciones de aplicabilidad.

Para el año 1969 ya con el concepto de base de datos más desarrollado se comienzan a incorporar a las distintas empresas

## **2.2. BI y el sector de la construcción en Ecuador**

El sector de la construcción en Ecuador representa un eje fundamental de inversión, desarrollo y crecimiento para la economía del país.

Dentro de este contexto el uso y manejo de la distinta información es de vital importancia para el fortalecimiento del sector y considerarla como una herramienta de mejora y crecimiento continuo en base a un proceso adecuado de toma de decisiones.

En la actualidad en Ecuador y después del periodo marcado por la pandemia uno de los sectores productivos que ha logrado reactivarse con mucho éxito es el de la construcción generando muchos proyectos y plazas de empleo que contribuyendo al crecimiento económico del país.

Dentro del mercado productivo del país existe varias empresas vinculadas al ámbito de producción, construcción y manufactura que a través de los años han ido creciendo y participando en varios proyectos dentro de la economía ecuatoriana. Por tal motivo se ha producido un aumento considerable en la cantidad de información que manejan las distintas organizaciones.

Según datos del INEC el sector de la construcción registro datos negativos durante el periodo más intenso de la pandemia.

### **2.3. Marco referencial**

Una primera investigación realizada por (Sanchez, 2012) muestra la efectividad, optimización y recursividad de los procesos y fortalecimiento de toma de decisiones mientras se desarrolla la implementación de una solución basada en inteligencia de negocios, la cual permitió tener un panorama mucho más claro respecto a la información relacionada con el cliente. De manera que se logró adoptar e integrar un sistema CRM que logra centralizar todos los movimientos y cambios que se realizan en tiempo real para posteriormente ser analizados en una herramienta de visualización mediante indicadores que evalúan el desempeño y gestión realizada.

En la investigación realizada por (Navarrete Carrasco, 2002) se busca determinar los principales factores en la estructura de bussines intelligence y el impacto de su aplicación e implementación de manera que se puedan identificar puntos clave y moldearlos con el fin de construir una ventaja competitiva, de manera que se vea reflejado en beneficios en varios ámbitos para la organización y que repercuta sustancialmente en la reducción de costos e incremento de ingresos y generación de valor en departamentos clave de una empresa.

La tesis de grado de (Chavez Huapaya & Contreras Ochoa, 2018) sobre la aplicación de herramientas de inteligencia de negocio en una empresa del sector de comercialización y servicios, ante la falta de una estructura que soporte el volumen de datos que se genera y ante la dificultad de procesar información ante esto se propone el desarrollo e implementación de una solución de bussines intelligence en donde se pretende darle un orden a la información y se agilice la construcción de reportes para ello se hizo uso de programas de gestión de datos SQL Server y de visualización Power BI. Posterior a su puesta a producción se logró reducir

el tiempo en la construcción de indicadores de gestión y por lo tanto se produjo una mejora en la toma de decisiones.

(Aspiazu Prado, 2019) en su trabajo de titulación muestra la mejora de procesos y optimización de tiempos en la generación de KPI de gestión, donde la mayor contribución se vio reflejada una disminución del 30% en el procesamiento de información.

El proyecto de investigación de (Delgado Palacios, 2014) realiza un análisis profundo de todas las implicaciones y ventajas de la aplicación de soluciones de bussines intelligence y su repercusión en el manejo y gestión de los distintos proyectos basado en las mejores prácticas a través del uso de distintos modelos y procedimientos.

El trabajo de maestría realizado por (Mavesoy Murcia, 2018) se menciona y demuestran los grandes beneficios de utilizar y generar metodologías de analítica de datos agiles basado en la aplicación de CRISP-DM generando como producto final tableros de control, orientado a satisfacer las necesidades de información del cliente interno.

En la investigación realizada por (Ortega Encina, 2013) se realiza una aplicación de la metodología CRISP-DM en donde cada etapa se adaptan a las necesidades de información requerida y se conjunta con el desarrollo de una solución de bussines intelligence, todo esto se refleja en la construcción de distintos indicadores de performance en relación a la efectividad y eficiencia de los distintos procesos.

El proyecto de investigación de (Riscos Ramos, 2022) se hace énfasis en la aplicación de herramientas de bussines intelligence en empresas pequeñas y medianas de sectores productivos enmarcado en metodologías agiles de minería de datos, obteniendo al final resultados

muy prometedores y positivos a través de distintos tableros de control e indicadores en función de inteligencia de negocios y modelos predictivos mejorando la identificación de oportunidades de crecimiento.

## **2.4. Marco Conceptual**

### **2.4.1. Conceptos básicos de la inteligencia de negocios**

Dentro de los principales términos que podemos encontrar tanto teóricamente como conceptualmente dentro de la inteligencia de negocios son esenciales para el éxito de la implementación de un determinado proyecto de BI a continuación, detallaremos los más importantes.

### **2.4.2. Inteligencia de negocios**

La inteligencia de negocios se define como el conjunto de herramientas enfocados a los recursos administrativos en donde las distintas organizaciones buscan obtener el máximo beneficio de la información con la que cuenta con el objetivo de lograr obtener una ventaja competitiva que pueda marcar la diferencia en su crecimiento. (Muñoz, Osorio, & Zuñiga, 2016)

En la actualidad diversas organizaciones se refieren a la inteligencia de negocios como el mejor camino para lograr un desarrollo sustentable.

“El objetivo básico de la Business Intelligence es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones” (Cano, 2007)

Ante este escenario es de vital importancia en primera instancia definir el termino de BI para esto vamos a citar los siguientes conceptos. La inteligencia de negocios guarda una estrecha relación con la optimización de recursos mediante el desarrollo de estrategias y la aplicación en un determinado campo.

“Se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.” (Conesa Caralt & Curto Díaz, 2010)

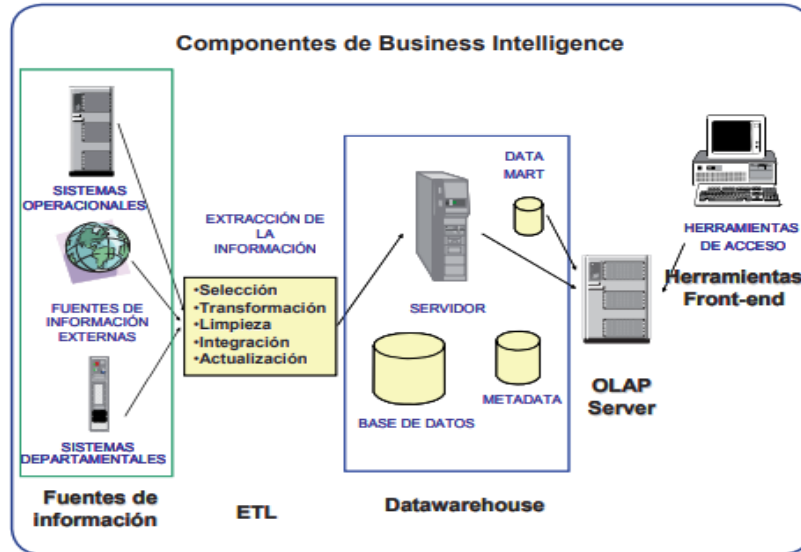
### **2.4.3. Importancia de la Inteligencia de negocios**

En el contexto actual todo lo relacionado a la aplicación y desarrollo de soluciones de inteligencia de negocios marcan la diferencia entre el crecimiento o estancamiento de una determinada empresa, de igual forma el saber correctamente explotar la información con que se cuenta arroja un valor agregado sobre la toma de decisiones respecto a presentes y futuros escenarios.

### **2.4.4. Componentes de Bussines Intelligence**

En la concepción de un sistema bussines intelligence se tienen que tomar en consideración diferentes elementos cuyo principal objetivo es optimizar y efectivizar de manera que la información que consume el usuario final tenga un alto nivel de confiabilidad.

(Cano, 2007) menciona que los componentes relacionados con Bussines Intelligence



**Figura 1.** Componentes de Bussines Intelligence  
Fuente: (Cano, 2007)

#### 2.4.5. Toma de decisiones (TDD)

La elección de una decisión puede definirse como la elección de las mejores opciones en base a un análisis previo (Umanzor, 2011) menciona que la toma de decisiones se basa en un proceso de aprendizaje natural o estructurado.

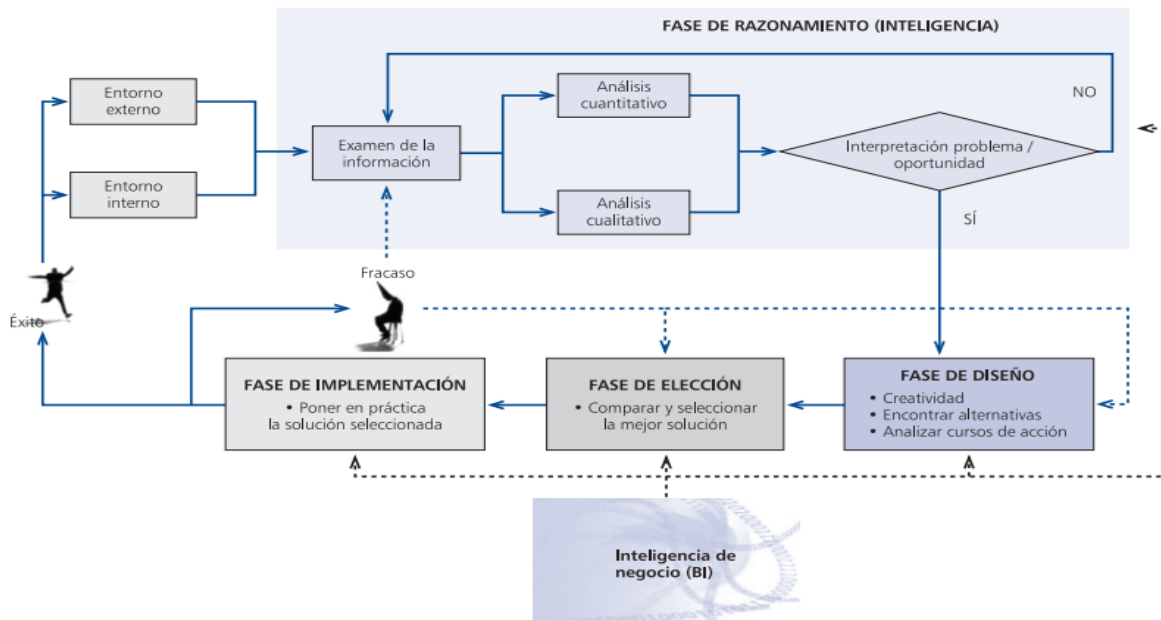
#### 2.4.6. Sistema de soporte a la decisión (DSS)

Los sistemas de soporte a la decisión representan la base y sustento con las que los diferentes niveles de una empresa se apoyan para la toman las decisiones.

“Un Sistema de Soporte a la Decisión (DSS) es una herramienta de Business Intelligence enfocada al análisis de los datos de una organización” (Vásquez Castrillon & Sucerquia Osorio, 2011)

(Bonilla Botia & Briceño Díaz, 2006) definen a un sistema de soporte a la decisión esta como el conjunto de y subsistema de administración de la base de datos, del directorio de datos y la facilidad para acceder a los mismos.

#### Toma de decisiones



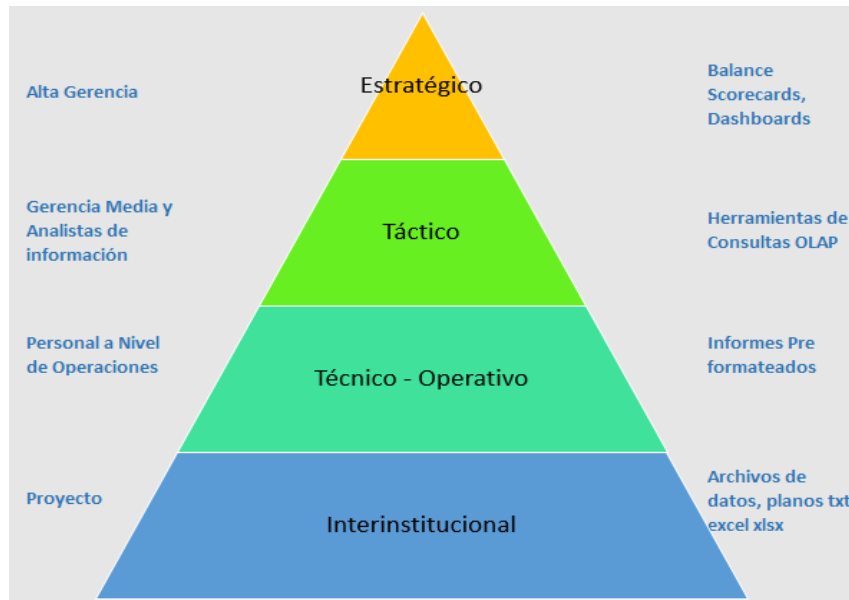
**Figura 2.** Toma de decisiones y sistemas de inteligencia de negocios  
Fuente: (Salgueiro Roldán, Cepeda Carrión, & Galán González, 2012)

#### 2.4.7. Sistemas de información (SI)

(Prieto & Martínez, 2004) menciona que un sistema de información como una herramienta fundamental y de gran ayuda en la toma de decisiones, de manera que se pueden generar una mejora en la eficiencia en lo correspondiente a la efectividad gerencial y a nivel profesional elementos claves en la generación de una estrategia de negocio.

“Un sistema de información SI puede ser cualquier combinación organizada de personas, hardware, software, redes de comunicación y recursos de información que almacene, recupere, transforme y disemine la información en una organización” (O’Brien & Marakas, 2004)

### Fases de sistemas de información



**Figura 3.** Fases de los sistemas de información  
Elaboración: Autor

#### 2.4.8. Sistemas de Información Gerenciales (MIS)

Son todos aquellos reportes que muestran información de acuerdo a un determinado periodo de tiempo con el objetivo mantener un control y gestión en la toma de decisiones en función al análisis de los distintos indicadores.

(Laudon & Laudon, 2012) menciona que un sistema de información gerencial es un conjunto de componentes interrelacionados cuya principal función es recolectar, procesar, almacenar y distribuir información que a su vez contribuyan en el proceso de toma de decisiones.

“Los sistemas de información contienen información sobre personas, lugares y cosas importantes dentro de la organización, o en el entorno que la rodea. Por información nos referimos a los datos que se han modelado en una forma significativa y útil para los seres humanos” (Laudon & Laudon, 2012)

#### **2.4.9. Cuadro de Mandos Integrales (CMI)**

Un cuadro de mando integral funge como un instrumento ayudo al control y análisis de los distintos indicadores de manera que se pueda establecer un panorama general de la situación de una empresa.

(Kaplan & Norton, 1997) mencionan que centra en la expansión de una o varias unidades de negocio y a generación de valor de las mismas hacia sus clientes actuales y futuros además de lograr optimizar las capacidades, personal y recursos en el mediano y largo plazo.

“El Cuadro de Mando Integral pone énfasis en que los indicadores financieros y no financieros deben formar parte del sistema de información para empleados en todos los niveles de la organización. Los empleados de primera línea han de comprender las consecuencias financieras de sus decisiones y acciones: los altos ejecutivos deben comprender los inductores del éxito financiero a largo plazo” (Kaplan & Norton, 1997)

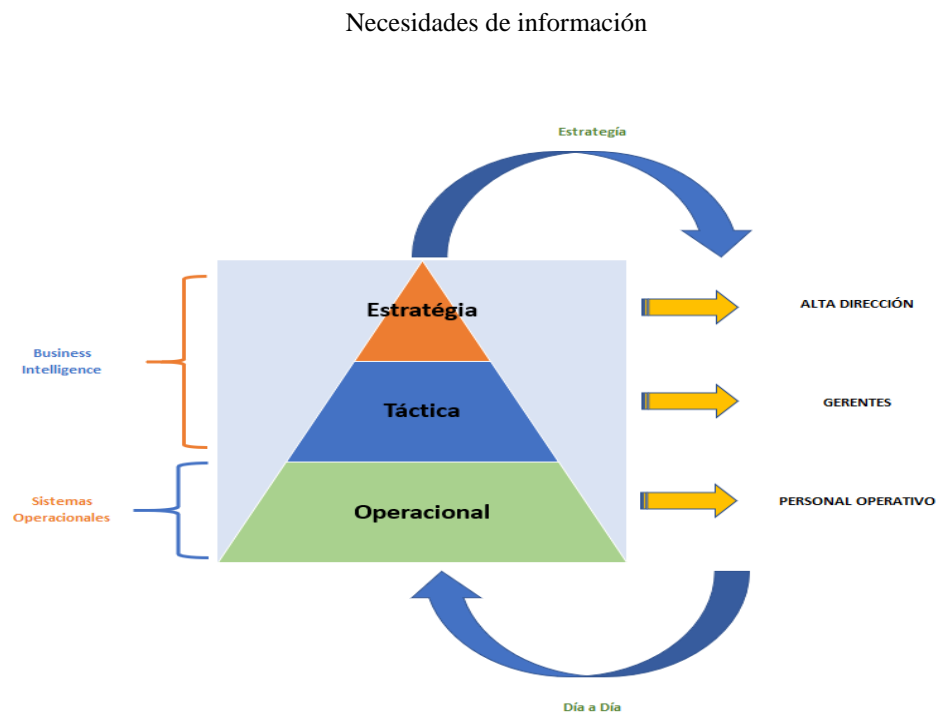
#### **2.4.10. Sistemas de business intelligence**

Los sistemas de business intelligence generalmente se basan en la generación de automatizaciones de procesos operativos manuales a través de una adecuada centralización de información de manera que su acceso sea relativamente fácil.

(Salgueiro Roldán, Cepeda Carrión, & Galán González, 2012) describen a los sistemas de business intelligence como un marco conceptual en el que se conjuntan diferentes sistemas desarrollados de procesamiento de información y la generación de valor de la misma.

“Teniendo en consideración este marco conceptual, se puede afirmar que las aplicaciones que se construyen sobre un sistema BI tienen como propósito auxiliar en las distintas tareas que componen el proceso de toma de decisiones. De este modo, en la fase de razonamiento o inteligencia

se puede contar, entre otras aplicaciones, con cuadros de mando o dashboards, así como herramientas de minería de datos que capaciten a la organización para la identificación de problemas u oportunidades.” (Salgueiro Roldán, Cepeda Carrión, & Galán González, 2012)



**Figura 4.** Fases de necesidades de información  
Elaboración: Autor

## 2.5. Metodología Ralph Kimball

La metodología seleccionada para el desarrollo y construcción de la solución de bussines intelligence es Kimball dada la estructura y necesidad de los datos.

En ese contexto Ralph Kimball propone un enfoque de construcción basado en los siguientes parámetros.

- Planteamiento del problema
- Definición de requerimientos del giro de negocio

- Definición y diseño de la arquitectura tecnológica
- Definición y diseño del modelo dimensional
- Diseño lógico y físico
- Preparación de los datos
- Especificaciones y desarrollo de indicadores de analítica
- Implementación
- Mantenimiento

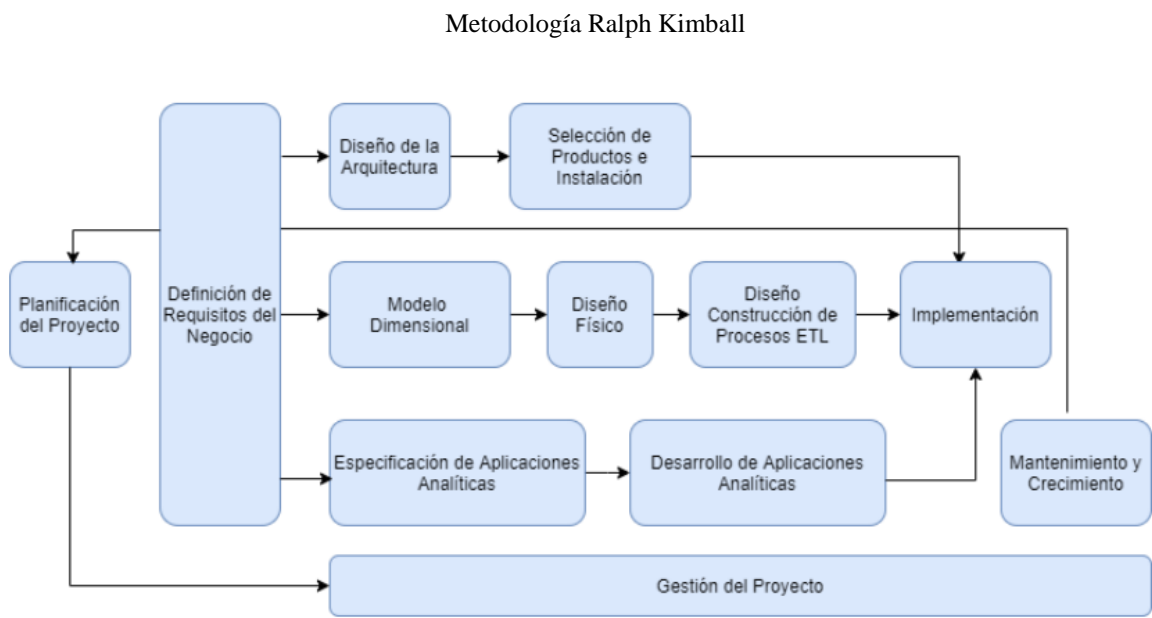
De igual forma Kimball fundamenta su aplicación en cuatro principios fundamentales que detallamos a continuación.

- **Centrarse en el negocio.** - Indispensable centrarse en los principales aspectos referentes a los requerimientos del negocio y generar un análisis que cuente con un valor agregado mediante la planificación y aplicación final.
- **Infraestructura adecuada.** - Se debe generar una estructura confiable de fácil uso y de rendimiento optimizado que permita obtener resultados confiables y alineados a las necesidades del negocio.
- **Aumentos incrementales.** - Se debe considerar que pueden existir cambios e inclusión de nuevos aspectos dentro de la infraestructura construida de manera que sea fácil de modificar en un plazo de tiempo corto y que no afecte a la información actual es decir que la carga de nueva data sea de forma incremental.
- **Solución completa.** - Una vez construida e implementada la solución esta debe asegurar todos los recursos necesarios para que los usuarios puedan obtener un valor

agregado sobre el análisis de datos que debe ir de la mano con la respectiva capacitación de uso y respectiva documentación para consultas.

### 2.5.1. Ciclo de vida Ralph Kimball

Dentro de las fases de construcción de una solución de bussines intelligence es de suma importancia seguir el ciclo de vida propuesto por Ralph Kimball ya que su diseño y seguimiento permite desarrollar en paralelo todas las actividades en la creación de un data Warehouse.



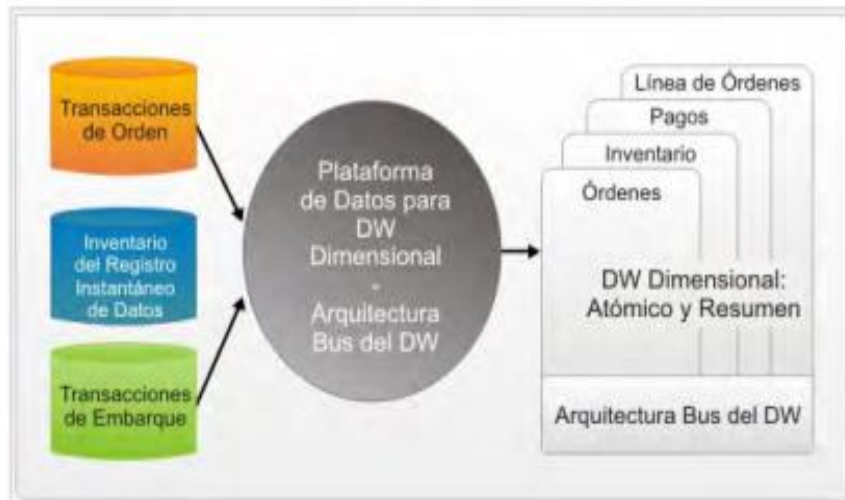
**Figura 5.** Ciclo de vida Kimball  
Elaboración: (Forero Castañeda & Sánchez Garcia, 2021)

Como mencionan (Gerrero García & Rodas Silva, 2022) Kimball permite desarrollar un determinado proyecto en función a una base empírica dado su gran versatilidad que permite una construcción de un determinado Data Warehouse de forma escalonada e integral alineados a los requerimientos del negocio.

“El enfoque de Kimball es una metodología de trabajo Bootom-up se caracteriza por unir diferentes DataMarts que se encuentren estructurados, a su vez crear DataMart y luego ir agregando

otros que comparten las dimensiones que estén establecidas u otras nuevas.” (Gerrero García & Rodas Silva, 2022)

### Enfoque Kimball



**Figura 6.** Enfoque de trabajo Kimball  
Elaboración: (Gerrero García & Rodas Silva, 2022)

### 2.5.2. Extracción, Transformación y Carga (ETL)

Un proceso ETL se define como un proceso que se encarga de extraer, transformar y cargar un determinado conjunto de datos desde un origen hacia un destino en específico tomando en cuenta las necesidades del giro del negocio de una empresa u organización.

“El proceso ETL es una parte de todo proceso de integración de datos. Su función tiene gran relevancia, ya que completa el resultado de desarrollo de aplicaciones y sistemas imprimiendo la cohesión necesaria” (PowerData, 2017)

### Proceso ETL



**Figura 7.** Estructura del proceso ETL  
Elaboración: Autor

### 2.5.3. Data Warehouse (DW)

Dentro de una definición simple un data Warehouse es un almacén de datos que agrupa y centraliza todos los data mart construidos dentro de una organización de manera que se agilite y facilite la obtención de información.

“En su nivel más elemental un Data Warehouse (Almacén de datos) es un área de transferencia de información para soporte de decisiones. Un Data Warehouse recopila datos provenientes de diversas aplicaciones en los sistemas corporativos en actividad; integra los datos en un modelo lógico de áreas comerciales temáticas; almacena la información de modo tal que sea accesible y comprensible para los tomadores de “decisiones no técnicos”, provee la información a los tomadores de decisiones en todo el ámbito corporativo a través de distintas herramientas de consulta y generación de informes” (Jaime, Matto, & Schuster, 1999)

### 2.5.4. DataMart (DM)

(Ramos, 2016) define a un data mart como una copia de las distintas transacciones en forma estructura que cumple con las condiciones para realizar consultas y análisis previamente diseñado con las premisas establecidas.

(Silva Peñafiel, Zapata Yáñez, Morales Guamán, & Toaquiza Padilla, 2019) definen a un data mart como una estructura de datos en donde toda la información contenida está diseñada con el objetivo de favorecer el análisis y divulgación eficiente de los datos.

### 2.5.5. Esquema Estrella

El esquema estrella es un modelo de datos tradicional cuyo principal beneficio es proporcionar una comprensión de relación más simple y fácil de entender para el usuario.

#### **2.5.6. Esquema Copo de nieve**

El esquema copo de nieve se caracteriza principalmente por la construcción de diferentes dimensiones que se relacionan directamente a la tabla de hechos.

#### **2.5.7. Modelo Tabular**

Dentro del modelamiento de datos un modelo tabular ofrece una gran ventaja dentro del análisis de datos puesto que su concepción garantiza recursividad en relación y construcción de un esquema de base de datos.

(Torres Sánchez, Espinosa Cervantes, Simon Cuevas, García Hernández, & Simon Cuevas, 2015) mencionan que un modelo tabular mantiene una estructura en base de datos se fundamenta en las tablas y la relación de las mismas de manera que se generan capacidades de autoservicio en la exploración de información.

#### **2.5.8. Modelos Multidimensional**

Un modelo multidimensional se basa en una estructura que agrupa una o varias medidas cuantitativas relacionadas a un determinado estudio o giro de negocio que permite realizar funciones de analítica profunda debido a la granularidad obtenida a través de la construcción de varias dimensiones.

(Cedeño Trujillo, 2006) menciona que un modelo multidimensional conlleva ciertas particularidades y procesos de una determinada empresa dividiéndolos en mediciones numéricas

en torno a una tabla de hechos que se basa en un control de registros en función a una fecha y temporalidad.

### 2.5.9. SQL Server Management Studio (SMSS)

SQL server es un software que permite gestionar grandes cantidades de datos a través la ejecución de diferentes sentencias de lenguaje SQL.

(Microsoft, 2023) menciona que SQL server funge como un entorno integrado que permite administrar cualquier infraestructura de base de datos de manera que es posible implementar, monitorear y actualizar distintos componentes a nivel de información que utilizan aplicaciones que ayudan a generar consultas y scripts.

**Tabla 1.** Comparativo de motores de base de datos

BASE DE DATOS	ORACLE	POSTGRESSQL	SQL SERVER
<b>DESARROLLADOR</b>	Oracle Corporation	PostgreSQL Global Development Group	Microsoft
<b>LICENCIA</b>	Privada	Libre	Privada
<b>VENTAJAS</b>	Modelo de objeto relacional	Sistema multiplafaroma	Escalabilidad, estabilidad y seguridad
	El uso de particiones mejora la eficiencia de replicación	Escabilidad y confiabilidad	Utiliza comandos DDL y DML
<b>DESVENTAJAS</b>	Precio elevado por licencia	Deficiencia en la velocidad de respuesta	Su soporte se limita a Windows

**Fuente:** (Microsoft, Microsoft SQL Server Notas técnicas del producto, 2017)

Elaboración: Autor

### 2.5.10. SQL Server Integration Services (SSIS)

Integration services es un software que permite extracción, transformación y carga de datos mediante la aplicación de diversos componentes, así como la aplicación y definición de diferentes parámetros de control.

(Microsoft, 2023) menciona que Analysis services facilita la creación de soluciones empresariales específicamente en la transformación e integración de datos de manera que se puedan resolver problemas complejos en el almacenamiento, limpieza, minería y administración de datos.

### **2.5.11. SQL Server Analysis Services (SSAS)**

Analysis Services es un servicio desarrollado por Microsoft que cumple el objetivo de creación de modelos tabulares y multidimensionales que permite generar conexiones directas a distintas instancia y fuentes de datos.

“Analysis Services es un motor de datos analíticos (VertiPaq) que se utiliza en el soporte de decisiones y análisis de negocios. Proporciona capacidades de modelo de datos semánticos de nivel empresarial para inteligencia empresarial (BI), análisis de datos y aplicaciones de generación de informes como Power BI, Excel, Reporting Services y otras herramientas de visualización de datos” (Microsoft, 2023)

## **2.6. Métodos de agrupación**

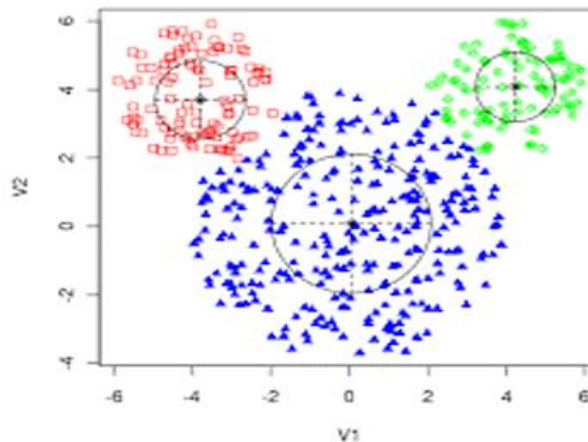
Son técnicas y herramientas que principalmente nos ayudan a lograr determinar grupos en función a diferentes características o variables de manera que se puedan ser consistentes entre si a continuación se detallaran los más relevantes.

### **2.6.1.1. Algoritmos de clasificación**

Utilizado para fraccionar un grupo de información en conjuntos de manera que cada elemento se integre o ubique lo más próximo en base a una probabilidad a un determinado conjunto. (Salazar Torres, 2021)

(Sandoval, 2018) menciona que básicamente se centra en elegir uno o varios patrones que permitan la clasificación y comparación de distintos datos en función de la construcción de diferentes grupos, mediante el cálculo interno de similitudes que generen homogeneidad entre si y heterogéneo entre grupos.

Gráficos de clasificación



**Figura 8.** Gráfico de algoritmo de clasificación  
Fuente: (Sandoval, 2018)

## 2.6.1. Tipos de algoritmos

### 2.6.1.1. Algoritmos jerárquicos

Los algoritmos jerárquicos parte de la premisa de lograr formar grupos o clústeres de manera que se minimicen las distancias entre los distintos elementos que integran un grupo en función a sus similitudes.

### **2.6.1.2. K-means**

El método de k-means consiste en la generación de un proceso iterativo que permita que la asignación de cada elemento a un determinado grupo o clúster, es especialmente efectivo cuando se cuenta con un gran número de observaciones.

“El algoritmo k-means obtiene el conjunto de datos D y el parámetro k como entrada, y luego divide un conjunto de datos D de n objetos en k grupos. Esta partición depende de la medida de similitud, de modo que el resultado obtenido tendrá una similitud dentro del grupo tan alta pero la similitud entre los grupos será baja. La similitud de los conglomerados se mide utilizando el valor medio de los objetos en un conglomerado, que puede mostrarse como la media del conglomerado.” (Sathya, Anu, & Divyadharshini , 2017)

### **2.6.1.3. DBSCAN**

DBSCAN es un algoritmo desarrollado en 1996 que base su aplicación en la densidad de cada uno de los objetos de manera que la asignación de cada elemento de un determinado conjunto de datos se acopla a la solución de forma más óptima.

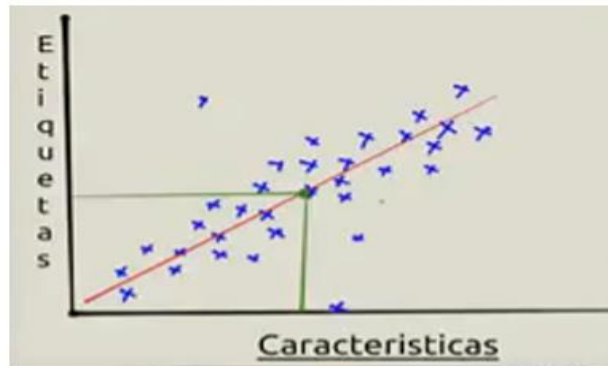
“DBSCAN no sólo puede adaptarse al conjunto de muestras no convexas, sino que también tiene un muy buen grado de ajuste al conjunto de muestras convexo. La clave del algoritmo de agrupamiento DBSCAN es la determinación de dos parámetros Eps y MinPts, donde Eps representa el umbral de distancia de vecindad, y MinPts representa el valor crítico del número de muestras en el vecindario.” (Jianfeng, y otros, 2022)

### **2.6.1.4. Algoritmos de regresión**

Los algoritmos de regresión son técnicas y métodos que mediante su aplicación nos permiten realizar estimaciones y predicciones en función a la utilización de distintas variables tanto dependientes como independientes.

(Cormen , Leiserson, Rivest, & Clifford, 2022) mencionan que un algoritmo puede considerarse como un conjunto de instrucciones correctamente definidas y finitas que permiten realizar una actividad mediante la ejecución de una serie de pasos hasta lograr la obtención de un determinado resultado.

## Algoritmo de regresión



**Figura 9.** Gráfico de algoritmo de regresión

Fuente: (Sandoval, 2018)

“Es el proceso que intenta determinar los valores de una o varias variables, a partir de un conjunto de datos. La predicción de valores continuos puede planificarse por las técnicas estadísticas de regresión.” (Salazar Torres, 2021)

### 2.6.1.5. Algoritmos de predicción

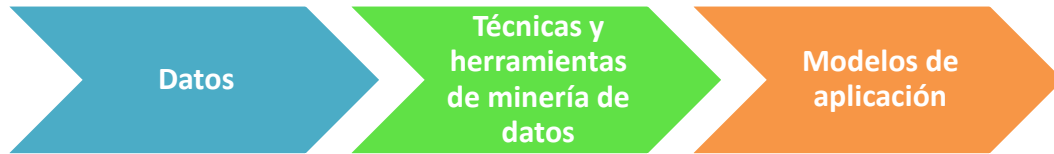
## 2.7. Minería de datos

(Riquelme, Ruiz, & Gilbert, 2006) mencionan que la minería de datos busca darle un sentido adecuado a la exploración de los datos y explosión de la información que en la actualidad puede ser almacenada.

De manera general se puede definir a la minería de datos como el conjunto de herramientas y métodos cuyo objetivo es darle un valor agregado al procesamiento y análisis de la información y que repercuta de forma positiva en la toma de decisiones.

“La minería de datos, es un conjunto de técnicas agrupadas con el fin de crear mecanismos adecuados de dirección, entre ellas puede citarse la estadística, el reconocimiento de patrones, la clasificación y la predicción” (Rodríguez Suárez & Díaz Amador, 2009)

## Proceso de minería de datos



**Figura 10.** Proceso de minería de datos

Elaboración: Autor

### **2.7.1. Metodología Crisp-DM**

Es una metodología que sirve de guía en el desarrollo y construcción de diferentes proyectos en función a la aplicación de una o varias técnicas de minería de datos,

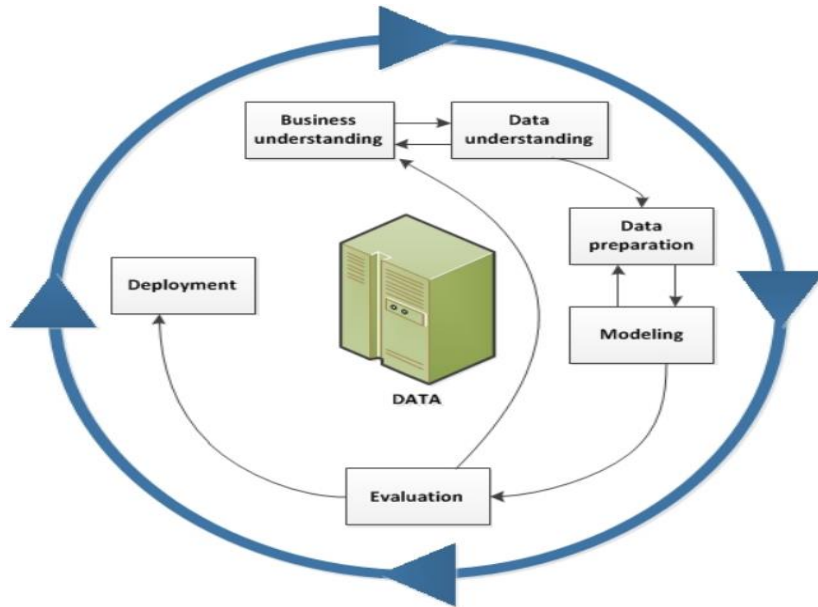
(Gironés Roig, 2013) menciona que la metodología CRISP-DM que se basa práctica y en la experiencia de los diferentes analistas en función a la organización de procesos y actividades.

#### **2.7.1.1. Fundamentos Básicos**

CRISP-DM (Cross Industry Standard Process for Data Mining) Proceso Estándar Industrial Híbrido para la Minería de Datos.

La metodología Crisp-DM se centra en técnicas y estándares de minería de datos a través de distintas fases cada una de ellas con un objetivo y una misión específico encaminados al éxito del proceso de un determinado proyecto.

Minería de datos



**Figura 11.** Ciclo de vida de minería de datos  
Elaboración: (IBM, 2021)

## 2.8. Elementos principales de la Metodología Crisp-DM

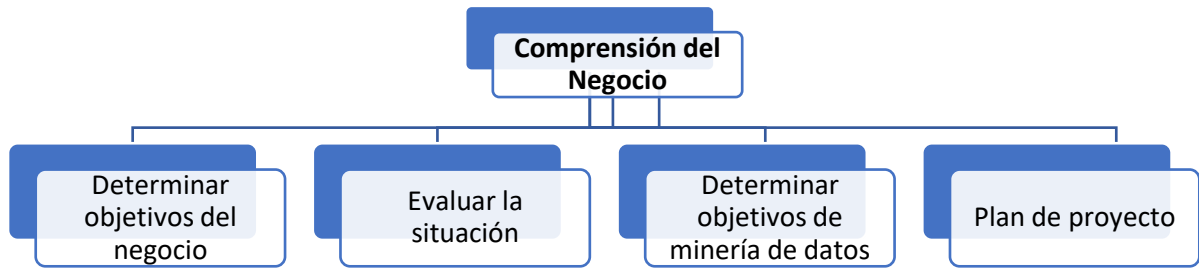
Dentro de la metodología Crisp-DM es fundamental definir un entorno de trabajo enmarcado en normas y flujos de trabajo dentro de los distintos procesos de manera que sirvan de apoyo en la realización y conclusión de un proyecto o investigación.

“Un proceso se puede ver como una colaboración entre roles para alcanzar cierta meta u objetivo. Para dirigir su representación, podemos restringir la orden en la cual las actividades deben ser, o pueden ser ejecutadas. También es necesario definir la "forma" del proceso en un determinado tiempo, y la estructura del ciclo de vida en términos de fases y de iteraciones” (Menéndez Domínguez & Castellanos Bolaños, 2015)

## 2.9. Fases de la metodología Crisp-DM

### 2.9.1. Comprensión del negocio

Comprensión del negocio

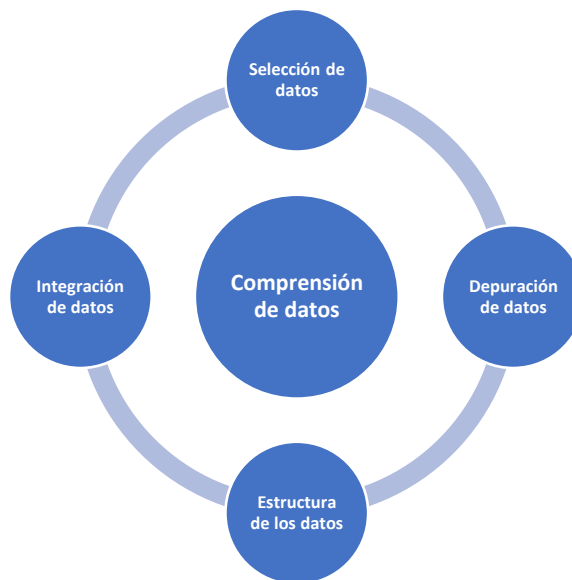


**Figura 12.** Fases de la comprensión del negocio  
 Fuente: (Chapman, y otros, 1999)

### 2.9.1.1. Comprensión de los datos

Una vez establecido todas las aristas en torno al negocio la siguiente etapa dentro de la metodología CRISP-DM se sustenta en realizar un análisis a fondo a nivel técnico de la estructura y calidad de los datos, mientras que a nivel situacional realizar una exploración descriptiva de la situación actual.

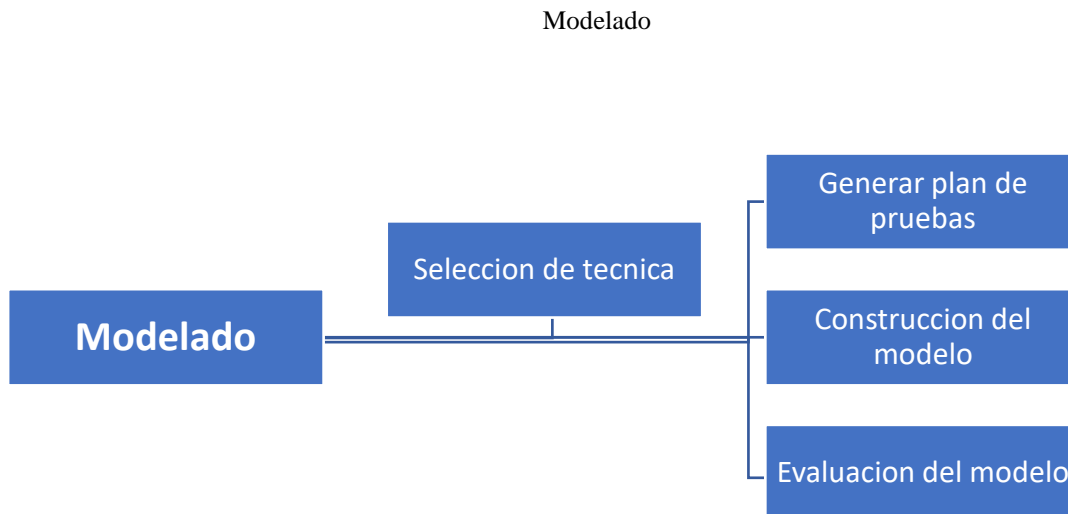
#### Comprensión de los datos



**Figura 13.** Fases de entendimiento de los datos  
Fuente: (Salcedo Parra, Galeano, & Rodriguez, 2010)

### 2.9.1.2. Modelado

En esta fase se procederá a elegir una técnica de minería de datos que mejor se adecue y alinee a la problemática, estructura de datos y objetivos planteados en función a test de pruebas, de manera que se logren cumplir todos los estándares y criterios de éxitos en la propuesta y construcción de un modelo de bussines intelligence.



**Figura 14.** Fases de entendimiento de los datos  
Fuente: (Salcedo Parra, Galeano, & Rodriguez, 2010)

### **2.9.1.3. Evaluación**

En esta fase es indispensable tener claro el panorama completo de la problemática de estudio y en función a esto seleccionar la técnica más adecuada que nos ayude a cumplir los objetivos planteados.

(Chapman, y otros, 2000) mencionan que todos los métodos y técnicas son aplicables dependiendo a un análisis previo en donde se deje en constancia tanto los beneficios como limitaciones direccionando todo esto a un uso adecuado de los recursos disponibles.

Evaluación



**Figura 15.** Evaluación de la técnica de minería de datos  
Fuente: (Chapman, y otros, 2000)

#### **2.9.1.4. Despliegue**

La fase final de la metodología CRISP-DM consiste en la implementación final de la solución de minería de datos en función a los recursos disponibles.

(Corona Organiche, Jiménez Alfaro, & Cortés Barrera, 2020) mencionan que para esta fase es indispensable establecer estrategias de ejecución, así como un monitoreo y mantenimiento de los modelos desarrollados.

Además de establecer un informe que detalle todo lo concerniente a la aplicación en base a las diferentes experiencias y análisis final de los puntos positivos y negativos.

#### **2.9.2. Microsoft Power BI**

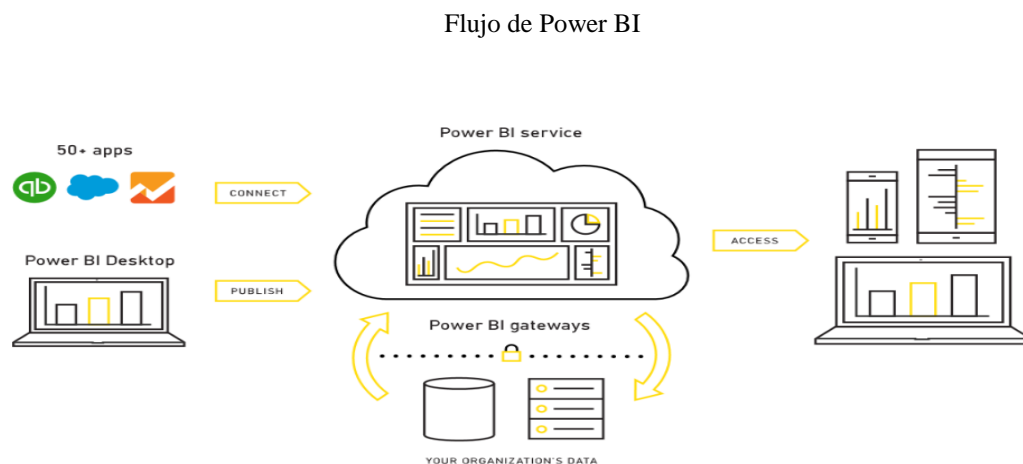
Una de las principales herramientas con respecto a la visualización y análisis de indicadores es Power Bi desarrollada por Microsoft con el objetivo de facilitar la construcción de reportería a nivel administrativo y gerencial.

(Microsoft, 2023) define a Power Bi como una plataforma unificada y escalable de Business Intelligence mediante la creación y publicación de aplicaciones y objetos virtuales recomendada para empresas que manejan grandes cantidades de datos.

### 2.9.3. Microsoft Power BI Desktop

Power BI Desktop es un software de uso libre que permite generar reporteria en función de diferentes objetos visuales permitiendo conexiones directas desde diferentes fuentes de datos.

(Microsoft, 2023) menciona a Power BI Desktop como desarrollador de conocimientos profundos y prácticos dentro de una amplia variedad de escenarios mediante el uso de herramientas de modelado de datos y el uso de funciones y creación de medidas.

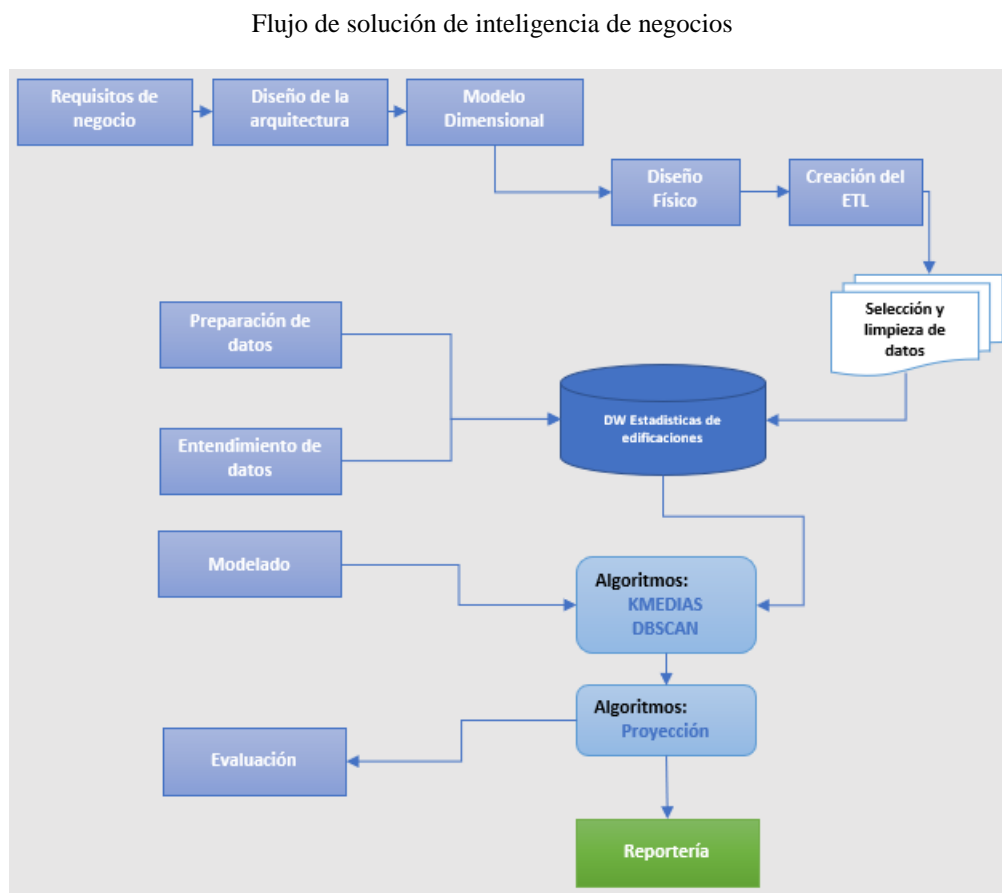


**Figura 16.** Flujo de servicios de Power BI  
Fuente: (Microsoft, Power BI: analítica de negocios para obtener información de valor, 2016)

## CAPITULO III

### 3. METODOLOGÍA

#### 3.1. Flujo de proceso aplicado para el diseño de la solución de inteligencia de negocios.



**Figura 17.** Flujo de solución de inteligencia de negocios

La figura 17 muestra el flujo de desarrollo a manera general de la solución de inteligencia de negocios, en la primera etapa se detalla la construcción y diseño del Data Warehouse

además de la creación y posterior ejecución del proceso ETL que alimentara al almacén de datos con la información de estadísticas de edificaciones.

En la segunda etapa se detallan las fases de ejecución de la metodología CRISP-DM en donde se desarrollan y aplicaran los algoritmos de clusterización y la validación de los mismos, a su vez se realizará y validara la proyección lineal. Finalmente, toda la información procesada se reflejará dentro de un reporte de Power BI.

Como se propuso en la metodología se utilizará Kimball para el desarrollo y construcción de la solución de bussines intelligence puesto que ofrece grandes beneficios y ventajas en su utilización

**Tabla 2.** Características de la metodología Kimball

	CARACTERISTICAS	DETALLE
<b>METODOLOGÍA KIMBALL</b>	Enfoque en el negocio	Se inicia con una identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio.
	Infraestructura de información	Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
	Realizar entregas en incrementos significativos	Creación de DataWarehouse en incrementos entregables en plazos cortos. Hay que usa el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos.
	Solución completa	Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible.

**Fuente:** (Gerrero García & Rodas Silva, 2022)

Elaboración: Autor

Por las características y objetivos planteados en la elaboración de este proyecto de investigación se ha optado por elegir la metodología Kimball.

### 3.1.1. Definición de requerimientos del negocio

Actualmente toda la información y estadísticas referente al sector de construcción son generadas y almacenadas por el Instituto Nacional de Estadística y Censos (INEC) que es una entidad pública encargada de la recolección de datos y el procesamiento de los mismos mediante la generación de encuestas de diferentes aspectos sociales y económicos

Para el objetivo de nuestra información nos enfocaremos en la información específica correspondiente a la construcción y la manera en que se puede agilizar y facilitar el acceso a dicha información mediante tableros de control o dashboards.

**Tabla 3.** Detalle de requerimientos de negocio

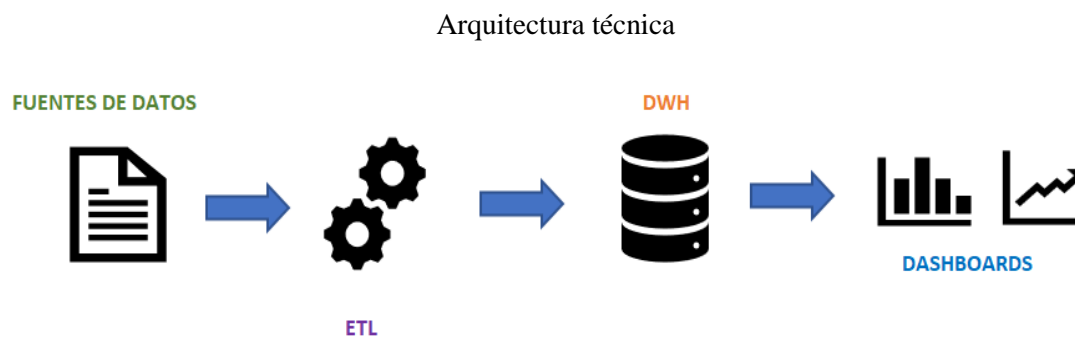
TEMA	REQUERIMIENTO	PROCESO DE NEGOCIO	OBSERVACION
<b>PLANIFICACIÓN DE LAS ACTIVIDADES PRINCIPALES</b>	Analizar información histórica en conjunto del sector de la construcción	Registro de datos recolectados mediante encuestas	A nivel de provincia y región
	Generar indicadores de desempeño y evolución del sector de la construcción	Generación de comparativas semestrales y anuales	A nivel de provincia y región
	Analizar el número de permisos de construcción	Clasificación por área (urbana, rural)	A nivel de provincia y región
	Analizar el motivo de la construcción	Generar clasificación del motivo de construcción, evolución y crecimiento	A nivel de provincia y región

Elaboración: Autor

## 3.2. Diseño e Implementación de Subsistemas ETL

### 3.2.1. Diseño de solución de Bussines Intelligence en Integration Services

Como parte del desarrollo integral de la solución de inteligencia de negocios se decidió utilizar Integration Services (SSIS) puesto que ofrece muchas ventajas y es compatible con la base de datos.



**Figura 18.** Estructura arquitectura técnica  
Elaboración: Autor

### 3.2.2. Selección de herramientas

- **Base de datos**

**SQL Server.** - Almacena los datos provenientes de la ejecución del proceso ETL.

- **Modelado de datos**

**Power Designer.** – Permite generar un modelo entidad relación lógico, físico y conceptual para generar la base de datos.

- **ETL**

**SSIS.** - Integration services la herramienta creada por Microsoft permite la construcción y ejecución de los procesos ETL.

- **Reporteria**

**Power BI.** - Herramienta de reporteria creada por Microsoft que se alimenta de la data almacenada en los data Warehouse y data mart

### **3.2.3. Estructura de datos**

En primera instancia para evitar generar complejidad en el desarrollo y construcción de la solución de bussines intelligence nos basaremos en la utilización del modelo estrella dado que la posterior migración de información se simplifica y se optimiza tiempos de ejecución.

Para la construcción y consumo de información utilizaremos como fuente primaria la encuesta ESED<sup>3</sup> generada por el INEC en donde se registran datos concernientes a permisos de construcción autorizados por los Gobiernos Autónomos Descentralizados Municipales.

En el Anexo numero 1 podemos observar el detalle de información y el tipo de información recolectada.

### **3.2.4. Tabla de hechos y dimensiones**

Una parte fundamental en la construcción de un DataMart consiste en la correcta definición y construcción de la tabla de hechos y dimensiones a continuación detallaremos a profundidad cada una de ellas y la utilidad e importancia de las mismas.

#### **3.2.4.1. Modelado Dimensional**

En lo referente al modelo dimensional las principales tareas se centrarán en detallar la estructura de las diferentes variables, así como los diferentes estándares de esquemas dentro de la base de datos.

---

<sup>3</sup> Estadísticas de edificaciones

Además, que a través del modelado se lograra tener una visión mucho más amplia del modelo entidad relación que se pretende utilizar en pro del cumplimiento de los objetivos planteados.

### 3.2.4.2. Tabla de hechos

La tabla de hechos o también llamada FACT almacenara todas las medidas o métricas del negocio que a su vez servirán para la posterior construcción de indicadores.

A continuación, se presenta el detalle de la tabla de hechos propuesta.

**Tabla 4.** Detalle de tabla de hechos de Edificaciones

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>COSTO TOTAL</b>	Numérico	Representa el costo total de la edificación
<b>CONSTRUCCIONES</b>	Numérico	Numero de construcciones por Provincia y/o región
<b>PERMISOS DE CONSTRUCCION</b>	Numérico	Numero de permisos de construcción por Provincia y/o región
<b>NÚMERO DE VIVIENDAS</b>		Número de viviendas por edificación
<b>PISOS</b>	Numérico	Número de pisos de la construcción
<b>METROS 2</b>	Numérico	Metros cuadrados de la edificación
<b>SUPERFICIE</b>	Numérico	Total, de la superficie de la edificación

Elaboración: Autor

### 3.2.4.3. Tablas de dimensiones

La creación de las diferentes dimensiones funciona como factores o agregaciones de diferentes aristas en función a una determinada atributo o característica. A continuación, detallaremos las diferentes dimensiones propuestas para el proyecto de investigación.

**Tabla 5.** Detalle de tabla de dimensiones de Provincias

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID PROVINCIAS</b>	Integer	Código de provincia
<b>NOMBRE PROVINCIA</b>	Varchar	Nombre de la provincia
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 6.** Detalle de tabla de dimensiones de Región

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID REGIÓN</b>	Integer	Código de la región
<b>NOMBRE REGIÓN</b>	Varchar	Nombre de la región
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 7.** Detalle de tabla de dimensiones de Construcción

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID CONSTRUCCIÓN</b>	Integer	Código de la construcción
<b>TIPO DE LA CONSTRUCCIÓN</b>	Varchar	Descripción del tipo de construcción
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 8.** Detalle de tabla de dimensiones de Recursos de construcción

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID RECURSO</b>	Integer	Código del recurso de construcción
<b>NOMBRE RECURSO</b>	Varchar	Descripción del tipo de recurso de construcción
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 9.** Detalle de tabla de dimensiones de Propósito de construcción

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID PROPOSITO</b>	Integer	Código del propósito de construcción
<b>NOMBRE DEL PROPÓSITO DE CONSTRUCCIÓN</b>	Varchar	Descripción del propósito de la construcción
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 10.** Detalle de tabla de dimensión del tipo de construcción

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID TIPO DE CONSTRUCCIÓN</b>	Integer	Código del tipo de construcción
<b>NOMBRE DEL TIPO DE CONSTRUCCIÓN</b>	Varchar	Descripción del tipo de la construcción
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 11.** Detalle de tabla de dimensión del tipo de propiedad de la edificación

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID TIPO DE PROPIEDAD</b>	Integer	Código del tipo de propiedad
<b>NOMBRE DEL TIPO DE PROPIEDAD</b>	Varchar	Descripción del tipo de la propiedad de la edificación
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

Elaboración: Autor

**Tabla 12.** Detalle de tabla de dimensión de tiempo

<b>CAMPO</b>	<b>TIPO</b>	<b>DESCRIPCIÓN</b>
<b>ID TIPO DE PROPIEDAD</b>	Integer	Código del tipo de propiedad
<b>NOMBRE DEL TIPO DE PROPIEDAD</b>	Varchar	Descripción del tipo de la propiedad de la edificación
<b>FECHA DE CARGA</b>	Datetime	Registro de la fecha de carga de información

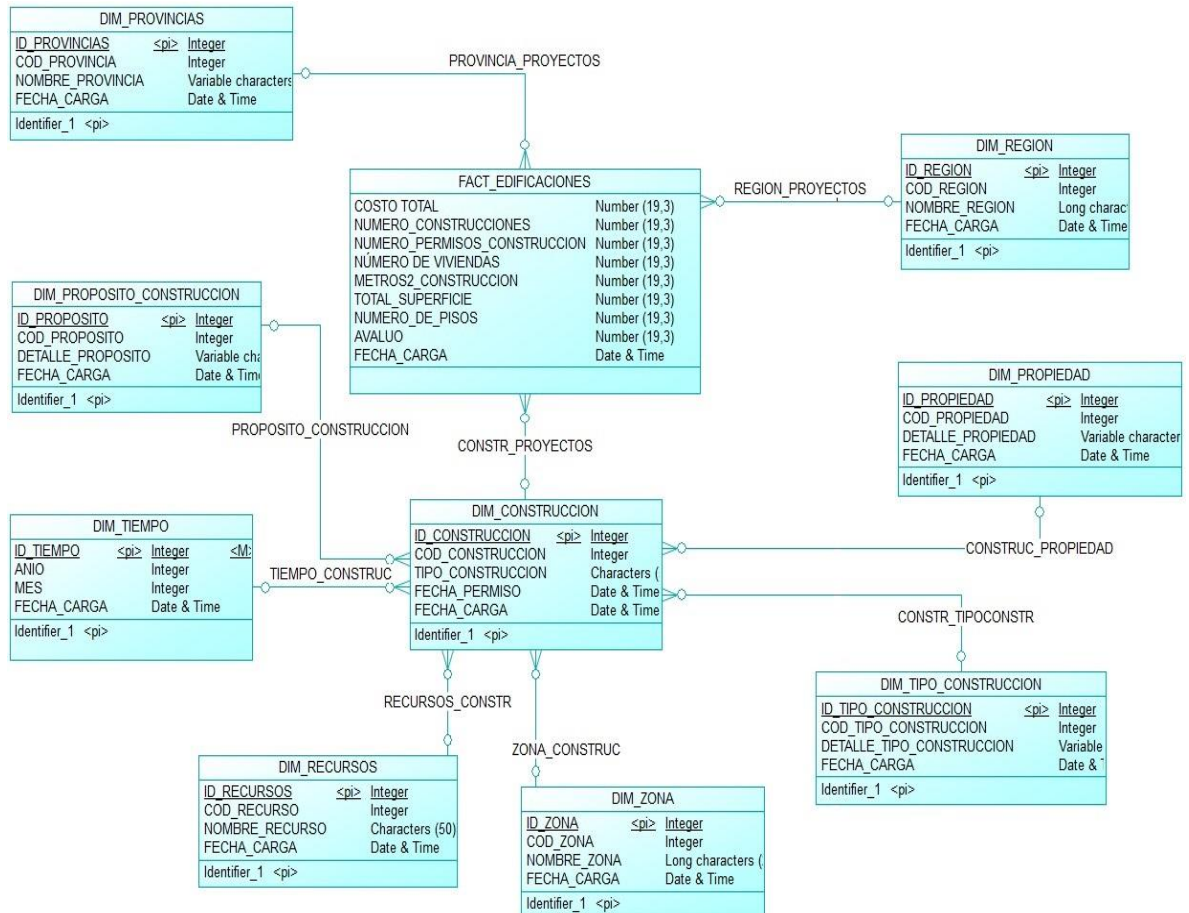
Elaboración: Autor

#### **3.2.4.4. Contexto y universo**

Para la realización del modelo de datos se utilizó el software Power Designer en donde se generan las relaciones respecto a la tabla de hechos y las respectivas dimensiones de estudio. Teniendo en cuenta la lógica de construcción basado en Kimball para la generación de estructura de Data Warehouse, en función a los datos disponibles dentro del universo de la encuesta ESED.

### 3.2.4.5. Diseño Conceptual

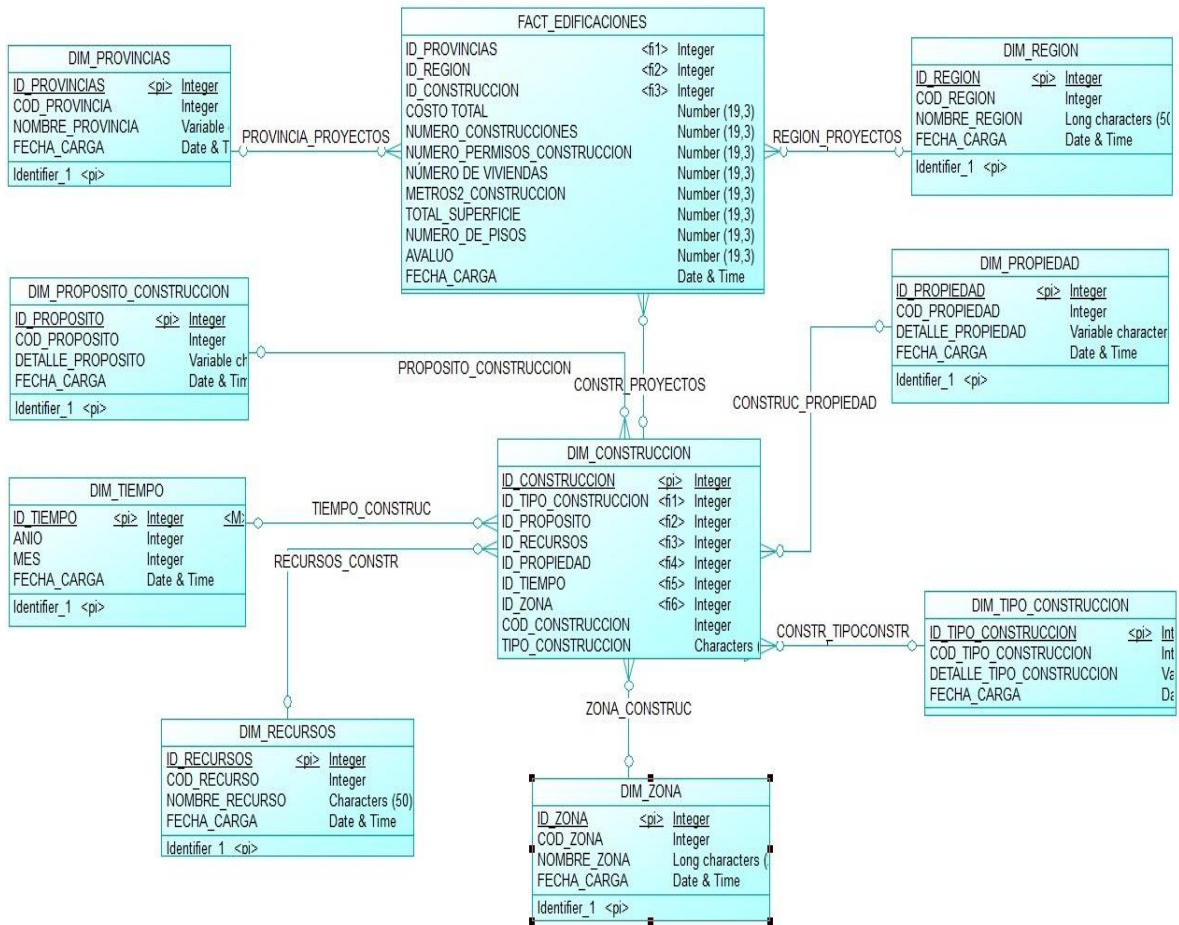
Diseño conceptual



**Figura 19.** Diseño conceptual  
Elaboración: Autor

### 3.2.4.6. Diseño Lógico

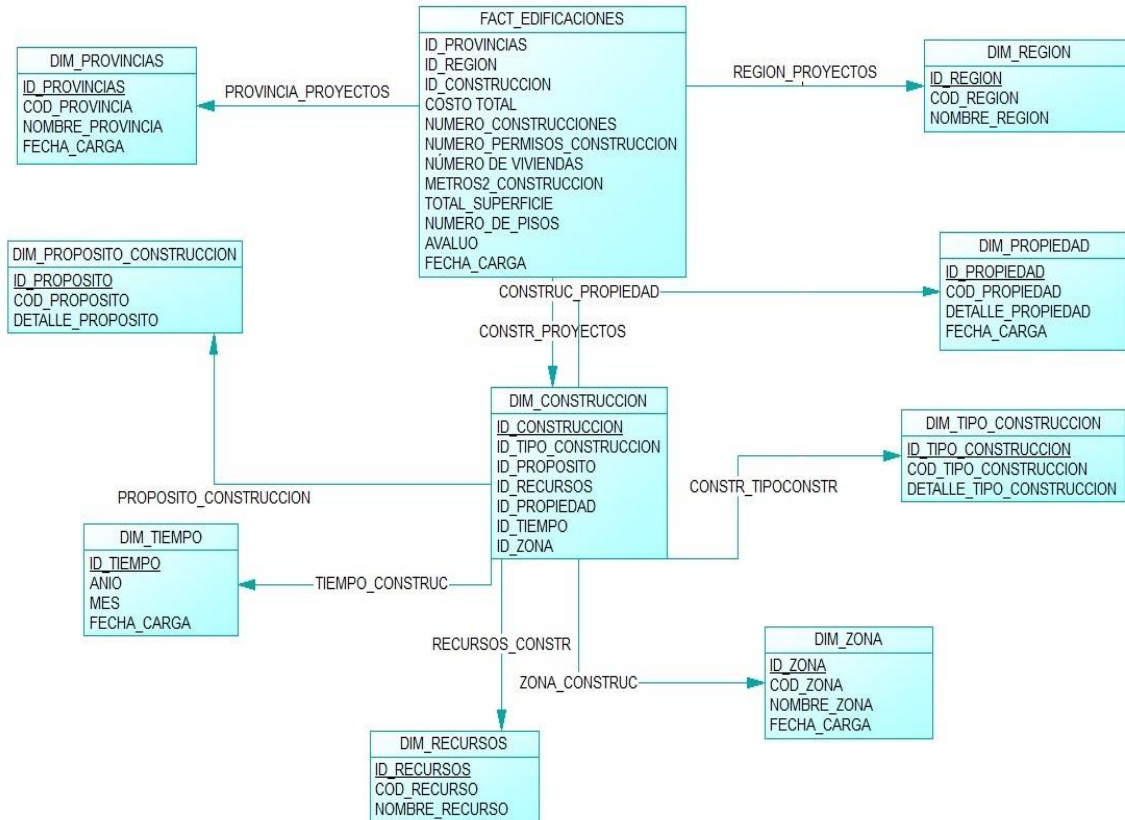
#### Diseño lógico



**Figura 20.** Diseño lógico  
Elaboración: Autor

### 3.2.4.7. Diseño Físico

#### Diseño físico



**Figura 21.** Diseño físico  
Elaboración: Autor

Como se puede observar en la figura 13. El diseño físico nos ayuda a tener una perspectiva global de todos los elementos incluidos dentro del Data Warehouse y realizar una evaluación de los mismos.

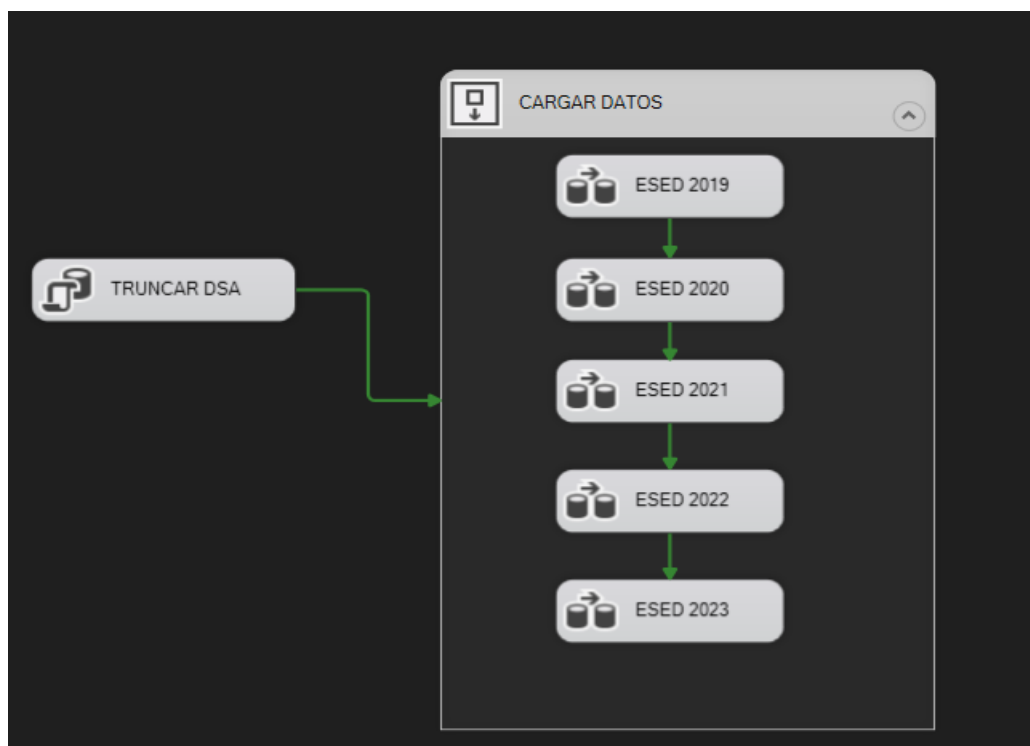
### 3.3. Diseño e Implementación del ETL

#### 3.3.1. Diseño ETL

En esta parte de la investigación se detallará el diseño, carga y ejecución del ETL, en primera instancia se alimentará de información al entorno DSA que es una carga previa de la data obtenida de la encuesta ESED.

En función a los objetivos planteados se seleccionó las variables adecuadas en torno al giro de negocio de la construcción. El proceso inicia realizando un vaciado del entorno DSA para posteriormente cargar la data proveniente de ESED incluyendo una nueva variable que describa la fecha de carga y actualización de los datos.

Carga de datos DSA



**Figura 22.** Carga de datos entorno DSA  
Elaboración: Autor

La figura 14 indica un primer flujo de ejecución en donde se procede a ejecutar una instrucción hacia la base de datos en SQL que trunque los datos anteriores para posteriormente cargar la información en orden cronológico, de manera que se pueda mantener un estándar tanto en la ejecución como en el control del proceso etl.

### 3.4. Metodología de minería de datos

Como se detalló en la metodología inicial se utilizará CRISP-DM puesto que ofrece una gran ventaja en la creación de un modelo de integración de datos en función al entendimiento y aplicación sobre el negocio.

**Tabla 13.** Principales aspectos de la metodología CRISP-DM

<b>METODOLOGÍA CRISP-DM</b>	<b>DETALLE</b>
<b>COMPRESIÓN DEL NEGOCIO</b>	Si
<b>SELECCIÓN Y PREPARACIÓN DE DATOS</b>	Si
<b>MODELADO</b>	Si
<b>EVALUACIÓN</b>	Si
<b>IMPLEMENTACIÓN</b>	Si
<b>NÚMERO DE FASES</b>	6
<b>FASES RELACIONADAS</b>	Si
<b>AMPLITUD</b>	Si
<b>ESTRUCTURA</b>	Si
<b>ESTABILIDAD</b>	Si
<b>FASES DETALLADAS</b>	Si
<b>ELECCIÓN LIBRE DE HERRAMIENTAS</b>	Si

**Fuente:** (Gironés Roig, 2013)

Elaboración: Autor

En base a la información presentada en la tabla 12 podemos concluir que la metodología que se apega más al proyecto de investigación es Crisp-DM dado se facilidad en su aplicación y comprensión.

### **3.5. Aplicación de la metodología Crisp-DM**

A continuación, se detallarán cada una de las fases de la metodología Crisp-DM y la aplicación de las mismas.

#### **3.5.1. Comprensión del Negocio**

En primera instancia se realizará un detalle general y específico de la situación contractual del sector de la construcción.

A continuación, se llevará a cabo la aplicación de cada una de las fases dentro de un proyecto de minería de datos, empezando por enunciar los objetivos, así como los requisitos del proyecto partiendo desde una perspectiva de negocio enmarcado en un análisis técnico que permita establecer un plan específico de proyecto.

La encuesta ESED menciona la siguiente misión:

Proporcionar información actualizada sobre el ritmo de crecimiento de los potenciales proyectos inmobiliarios a construirse, sean: viviendas, locales comerciales e industriales, edificios administrativos, clínicas, entre otros. (INEC, 2021)

También se menciona la siguiente visión:

Caracterizar y medir el comportamiento del sector de la construcción y del mercado inmobiliario y la relación de los distintos factores. (INEC, 2021)

#### **3.5.2. Determinar los objetivos del negocio**

Indudablemente que el principal objetivo de un proyecto basado en minería de datos es lograr generar información fiable que permita establecer posibles escenarios futuros bajo diferentes condiciones. El objetivo es generar una solución de inteligencia de negocio que permita realizar análisis más profundo sobre la situación y evolución del sector de la construcción apoyado en herramientas BI.

### **3.5.3. Contexto**

En contexto general el sector de la construcción representa un eje de desarrollo muy importante para el crecimiento de una economía del país. En referencia a la situación actual y al tema propuesto de investigación en primera instancia se cuenta con una fuente de datos primaria proporcionada por el Instituto Nacional de Estadísticas y Censos que detalla varias aristas de comparación.

Sin embargo, no existe un tablero de control que centralice toda esta información que pueda utilizarse para realizar un análisis situacional más profundo.

### **3.5.4. Objetivo del Negocio**

Como ya se ha mencionado lograr tener al alcance de la mano indicadores y métricas que permitan determinar distintos escenarios podrían mejorar y potenciar la posterior toma de decisiones enmarcado bajo estas premisas, el presente proyecto de investigación ha definido los siguientes objetivos:

- Generar indicadores de comportamiento fiables sobre la situación actual.
- Determinar los principales factores que inciden en el decrecimiento y crecimiento del sector de la construcción.

La generación de estos tableros de control puede proporcionar información fiable y de fácil acceso sobre el comportamiento de distintos segmentos del sector de la construcción.

### **3.5.5. Objetivos de la aplicación de inteligencia de negocios**

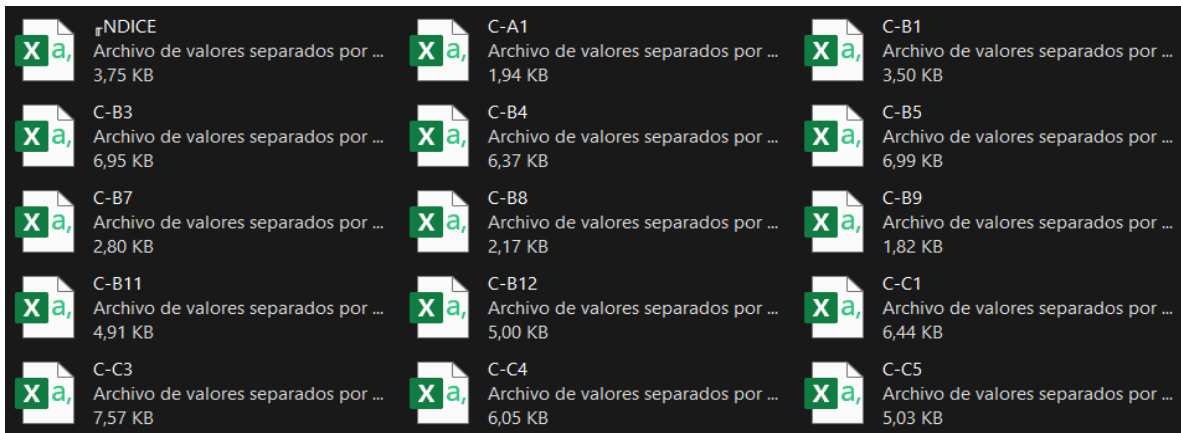
Una vez teniendo claro el panorama completo y el proceso y consolidación de información necesario.

### 3.5.6. Recolección de datos iniciales

En esta etapa de la metodología CRISP se especificará la recolección de datos y la adecuación de los mismos para su posterior procesamiento.

Como fuente principal se utilizará los datos provenientes de las estadísticas de edificaciones ESED publicada en los repositorios del INEC, cabe recalcar que esta información se encuentra almacenada en archivos tipo csv en función al año de publicación, cada uno con una denominación y nombre que se detallará a continuación.

#### Fuentes de datos



**Figura 23.** Archivos de fuente de datos  
Elaboración: Autor

#### 3.5.6.1. Comprensión de los datos

Dentro de la segunda fase de la metodología CRISP-DM es de vital importancia el entendimiento de las distintas fuentes de información, así como los recursos disponibles.

En función a lo antes detallado se logró construir un modelo de datos que se puede observar en la figura 13 que nos permitirá tener una visión mucho más clara del comportamiento y evolución de los datos a través del tiempo.

### 3.5.6.2. Descripción de los datos

Dentro de los lineamientos de investigación y los objetivos planteados para este proyecto a partir de aquí comenzaremos con la respectiva exploración y descripción de la data disponible dentro de la construcción de edificaciones en Ecuador.

La encuesta ESED nos proporciona una gran de variedad de información y variables que contribuirán a resolver la problemática de estudio a continuación, se detallarán los más importantes.

**Tabla 14.** Principales variables de la encuesta EDED

CAMPO	TIPO	DESCRIPCIÓN
<b>CODPROVF</b>	Integer	Código de provincia
<b>CODCANTF</b>	Integer	Código de cantón
<b>CODREG</b>	Integer	Código de región
<b>CAREAUR</b>	Integer	Código de área
<b>CPERM</b>	Numérico	Número de edificaciones a construir
<b>PROPIE</b>	Categorica	Propiedad de la edificación Publica / Privada
<b>CVAE</b>	Numérico	Valor total de la edificación
<b>CARCO</b>	Numérico	Área total a construir en m2
<b>CDISPUSO</b>	Categorica	Uso de la edificación
<b>CDISPRFI</b>	Categorica	Origen de financiamiento para la construcción de la edificación
<b>COSM2</b>	Numérico	Costo estimado del m2 de construcción
<b>TIPFINAN</b>	Categorica	Tipo de financiamiento para la construcción de la edificación

**Fuente:** Estadística de Edificaciones (2019-2022)

Elaboración: Autor

### 3.5.6.3. Exploración de los datos

Una vez terminada la fase de recolección y comprensión de datos, el siguiente paso es explorar toda la información con que contamos con el objetivo de generar un panorama mucho más claro lo cual nos brindará una perspectiva de la consistencia de los datos, aspecto que nos servirá de gran ayuda en las posteriores etapas de CRISP-DM.

**Tabla 15.** Número de permisos de construcción por provincia

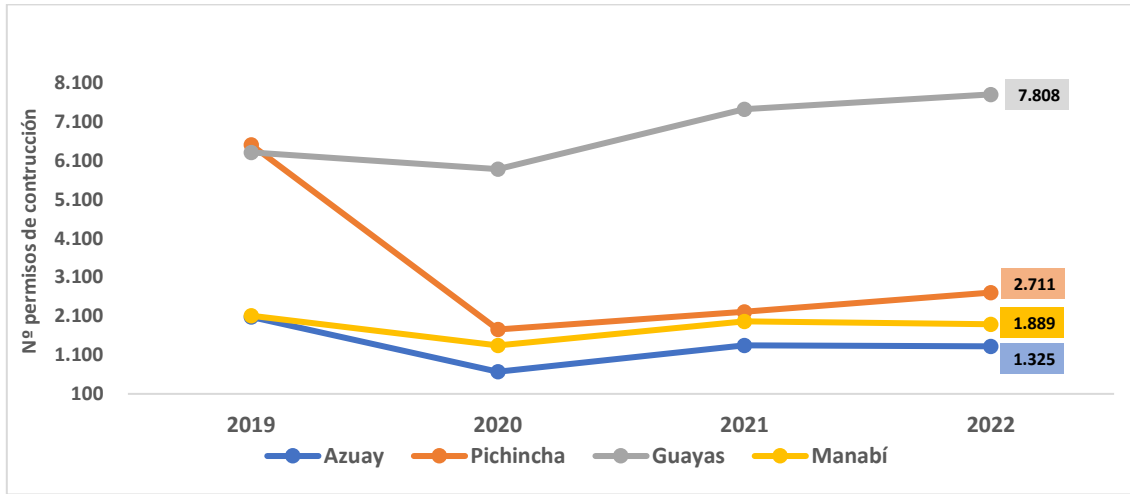
Provincia	Año				Total
	2019	2020	2021	2022	
<b>Azuay</b>	2.079	670	1.345	1.325	<b>5.419</b>
<b>Bolívar</b>	275	286	315	300	<b>1.176</b>
<b>Cañar</b>	486	442	754	584	<b>2.266</b>
<b>Carchi</b>	213	260	433	358	<b>1.264</b>
<b>Chimborazo</b>	1.072	554	897	884	<b>3.407</b>
<b>Cotopaxi</b>	695	377	488	391	<b>1.951</b>
<b>El Oro</b>	1.107	866	1.225	1.283	<b>4.481</b>
<b>Esmeraldas</b>	297	139	215	147	<b>798</b>
<b>Galápagos</b>	147	103	158	169	<b>577</b>
<b>Guayas</b>	6.314	5.884	7.426	7.808	<b>27.432</b>
<b>Imbabura</b>	1.699	1.136	1.396	1.129	<b>5.360</b>
<b>Loja</b>	1.109	1.223	1.676	1.737	<b>5.745</b>
<b>Los Ríos</b>	611	583	710	586	<b>2.490</b>
<b>Manabí</b>	2.112	1.347	1.964	1.889	<b>7.312</b>
<b>Morona Santiago</b>	344	255	449	418	<b>1.466</b>
<b>Napo</b>	246	159	197	146	<b>748</b>
<b>Orellana</b>	187	105	103	129	<b>524</b>
<b>Pastaza</b>	305	189	252	231	<b>977</b>
<b>Pichincha</b>	6.510	1.757	2.217	2.711	<b>13.195</b>
<b>Santa Elena</b>	499	310	720	545	<b>2.074</b>
<b>Santo Domingo de los Tsáchilas</b>	1.086	528	1.014	1.123	<b>3.751</b>
<b>Sucumbíos</b>	249	311	356	331	<b>1.247</b>
<b>Tungurahua</b>	1.926	1.408	1.722	1.552	<b>6.608</b>
<b>Zamora Chinchipe</b>	293	311	454	491	<b>1.549</b>
<b>Total general</b>	<b>29.861</b>	<b>19.203</b>	<b>26.486</b>	<b>26.267</b>	<b>101.817</b>

**Fuente:** Estadística de Edificaciones (2019-2022)

Elaboración: Autor

En función a la información presentada en la tabla 15. Se puede observar que en el periodo de análisis (2019 - 2022) en primera instancia se evidencia un decrecimiento importante, esto explicado debido a la pandemia ocurrida en el año 2020 posterior a ese lapso se puede visualizar crecimientos paulatinos.

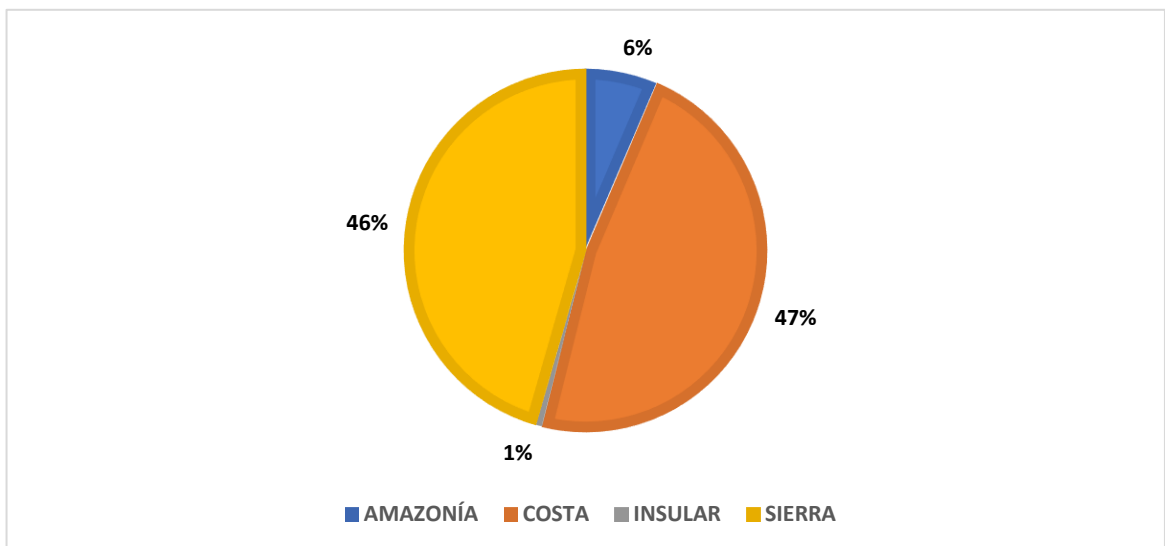
N° permisos de construcción provincias con mayor economía



**Figura 24.** Permisos de construcción en las provincias económicamente más importantes  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

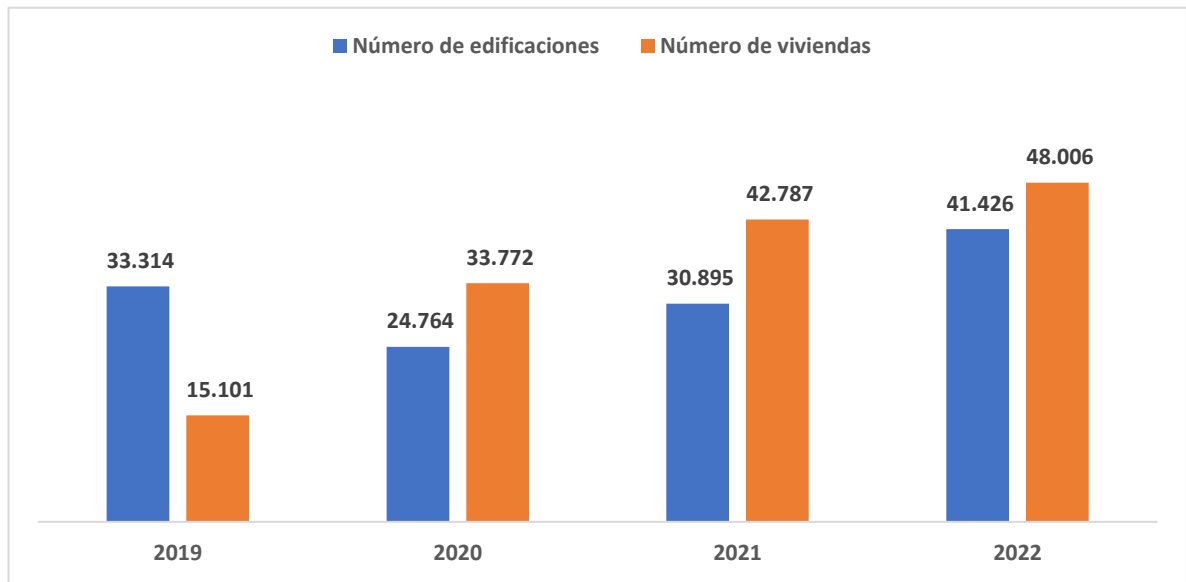
En las principales provincias del país se refleja un crecimiento importante anual luego del bajón provocado por la pandemia, algo que puede explicarse dado que son polos de desarrollo y en las periferias se este fenómeno va incrementando.

Proporción de permisos de construcción por región



**Figura 25.** Proporción de permisos de construcción por región  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

### N ° de tipo de construcción



**Figura 26.** Número de tipo de construcciones  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

#### 3.5.6.4. Verificar la calidad de los datos

Terminada la fase de exploración de los datos, es necesario verificar y comprobar la consistencia de los mismos y que los resultados posteriores se alineen a los objetivos planteados en la investigación propuesta.

En este caso partimos de la premisa que la construcción de la base de datos fue en base a los archivos csv que se mencionaron anteriormente, al encontrarse la información dispersa el ETL consolida la data en una sola fuente de información en la que se toma en consideración el tipo de dato estandarizado de manera que no ocasione problemas posteriores.

Para la base de datos se realizaron pruebas de consistencia, así como de unicidad de los datos que garantizan calidad en la información cargada dentro del Data Warehouse.

### 3.5.7. Preparación de los Datos

Para la preparación de la data se han tomado las siguientes consideraciones:

- Dado que no se cuenta con una estructura que contenga toda la información, en primera instancia se va a proceder a generar un entorno de base de datos mediante la herramienta SQL management studio.

### 3.5.8. Script SQL modelo de datos

#### 3.5.8.1. Script tabla de hechos

```
/*=====*/
/* Table: FACT_EDIFICACIONES */
/*=====*/
create table FACT_EDIFICACIONES
(
  ID_PROVINCIAS      integer          null,
  ID_REGION          integer          null,
  ID_CONSTRUCCION   integer          null,
  COSTO_TOTAL        numeric(19,3)   null,
  NUMERO_CONSTRUCCIONES numeric(19,3) null,
  NUMERO_PERMISOS_CONSTRUCCION numeric(19,3) null,
  NUMERO_DE_VIVIENDAS numeric(19,3)   null,
  METROS2_CONSTRUCCION numeric(19,3)   null,
  TOTAL_SUPERFICIE  numeric(19,3)   null,
  NUMERO_DE_PISOS   numeric(19,3)   null,
  AVALUO             numeric(19,3)   null,
  FECHA_CARGA        timestamp       null
);

/*=====*/
/* Index: RELATIONSHIP_1_FK */
/*=====*/
create index RELATIONSHIP_1_FK on FACT_EDIFICACIONES (
  ID_PROVINCIAS ASC
);

/*=====*/
/* Index: REGION_PROYECTOS_FK */
/*=====*/
create index REGION_PROYECTOS_FK on FACT_EDIFICACIONES (
  ID_REGION ASC
);

/*=====*/
/* Index: CONSTR_PROYECTOS_FK */
/*=====*/
create index CONSTR_PROYECTOS_FK on FACT_EDIFICACIONES (
  ID_CONSTRUCCION ASC
);
```

**Figura 27.** Script sql tabla de hechos edificaciones  
Elaboración: Autor

Dentro de los anexos 4 al 12 podemos ver el código de construcción de las restantes dimensiones que se incluyeron dentro del Data Warehouse.

### 3.5.8.2. Limpieza de los datos

Uno de los aspectos importantes para realizar una correcta y adecuada limpieza de datos se hizo uso de las herramientas ofrecidas por el motor de base de datos SQL SERVER, creando tablas temporales de manera que no se vea afectado el modelo de datos original.

En primera instancia se procedió a realizar una estandarización de los datos en relación a datos de texto y datos categóricos.

Limpieza de datos

id	codprovif	codcantif	codreg	careaur	fper	mes	anio	propie	cimi	piso	estru	pared	cubi
148380	El Oro	Huaquillas	Costa	Urbana	NULL	Octubre	2020	Privada	Hormigón armado	Hormigón	Hormigón armado	Ladrillo	Hormigón armado
148379	Imbabura	San Miguel De Urcuquí	Sierra	Rural	NULL	Febrero	2020	Privada	Hormigón armado	Hormigón	Hormigón armado	Bloque	Láminas metálicas (zinc)
148378	Imbabura	Pimampiro	Sierra	Urbana	2020-07-08 00:00:00.0000	Agosto	2020	Pública	Sobre pilotes de hormigón	Cerámica	Metálica	Bloque	Láminas metálicas (zinc)
148377	Imbabura	San Miguel De Urcuquí	Sierra	Rural	NULL	Febrero	2020	Privada	Hormigón armado	Hormigón	Metálica	Bloque	Hormigón armado
148376	El Oro	Huaquillas	Costa	Urbana	NULL	Octubre	2020	Privada	Hormigón armado	Madera	Hormigón armado	Ladrillo	Hormigón armado
148375	Imbabura	Pimampiro	Sierra	Urbana	2020-06-08 00:00:00.0000	Agosto	2020	Privada	Hormigón armado	Cerámica	Hormigón armado	Ladrillo	Hormigón armado
148374	Imbabura	Pimampiro	Sierra	Urbana	NULL	Agosto	2020	Privada	Hormigón ciclópeo	Hormigón	Hormigón armado	Bloque	Hormigón armado
148373	El Oro	Huaquillas	Costa	Urbana	NULL	Octubre	2020	Privada	Hormigón armado	Hormigón	Hormigón armado	Bloque	Hormigón armado
153389	11	1101	1	1	NULL	1	2021	1	0	152	121	122	123
153403	11	1101	1	1	NULL	1	2021	1	120	154	121	122	123
148372	El Oro	Huaquillas	Costa	Urbana	2020-01-09 00:00:00.0000	Septiembre	2020	Privada	Hormigón armado	Porcelanato	Hormigón armado	Ladrillo	Hormigón armado
148371	El Oro	Arenillas	Costa	Urbana	2020-06-10 00:00:00.0000	Octubre	2020	Privada	Hormigón armado	Hormigón	Hormigón armado	Bloque	Láminas metálicas (zinc)
148370	El Oro	Arenillas	Costa	Urbana	2020-07-10 00:00:00.0000	Octubre	2020	Privada	Hormigón armado	Cerámica	Metálica	Bloque	Láminas metálicas (zinc)
148369	El Oro	Arenillas	Costa	Urbana	NULL	Septiembre	2020	Privada	Hormigón armado	Cerámica	Hormigón armado	Ladrillo	Láminas metálicas (zinc)
148368	El Oro	Arenillas	Costa	Urbana	NULL	Septiembre	2020	Privada	Hormigón armado	Cerámica	Hormigón armado	Ladrillo	Hormigón armado

**Figura 28.** Errores en data inicial  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

Como se puede observar en la figura 28 varios campos importantes y que intervienen en los análisis posteriores para corregir estas observaciones se decidió manejar la información por codificación de categorías con el fin que se facilite el modelado.

### Limpeza de datos

	codreg	cod_prov	propie	CDISPUSO	CDISPRFI	CTIPOBR
1	2	7	1	1	1	1
2	1	10	1	4	1	1
3	1	10	2	1	4	1
4	1	10	1	4	13	1
5	2	7	1	1	1	1
6	1	10	1	1	1	1
7	1	10	1	4	1	1
8	2	7	1	1	1	1
9	1	11	1	2	1	2
10	1	11	1	1	9	1
11	2	7	1	1	1	1
12	2	7	1	1	1	1
13	2	7	1	1	1	1
14	2	7	1	1	1	1
15	2	7	1	1	1	1
16	2	7	1	3	1	1

**Figura 29.** Data corregida

Fuente: Estadística de Edificaciones (2019-2022)

Elaboración: Autor

Como se puede observar en la figura 29 se procedió a estandarizar la data, de manera que al usar códigos sobre las diferentes categorías se eliminan posibles errores por concepto de tipo o de longitud de datos.

Dentro del mismo proceso se decidió eliminar registros que mantenían registros nulos dado que puede influir negativamente en el resultado final de la investigación.

#### **3.5.8.3. Integración de los datos**

De acuerdo al tema planteado se realizó la integración de datos en función a los registros históricos provenientes de la encuesta ESED diseñada y publicada por el INEC.

Que posteriormente el ETL construido integra dentro de las dimensiones del Data Warehouse de manera que la información se actualice y se inserte de forma incremental conforme se vaya actualizando.

A continuación, se detalle las fuentes primarias de información utilizadas para la integración mediante el ETL.

**Tabla 16.** Número de permisos de construcción por provincia

<b>ENCUESTA</b>	<b>AÑO</b>	<b>PERIODICIDAD</b>
ESTADISTICA DE EDIFICACIONES (ESED)	2019	Anual y trimestral
ESTADISTICA DE EDIFICACIONES (ESED)	2020	Anual y trimestral
ESTADISTICA DE EDIFICACIONES (ESED)	2021	Anual y trimestral
ESTADISTICA DE EDIFICACIONES (ESED)	2022	Anual y trimestral
ESTADISTICA DE EDIFICACIONES (ESED)	2023	Anual y trimestral

Elaboración: Autor

Es importante señalar que toda esta información es consolidada dentro del Data Warehouse construido y sus respectivas dimensiones previo a un procesamiento de limpieza y depuración.

#### **3.5.8.4. Formato de los datos**

Como fase final de la integración de los datos y previo a la generación de la solución de inteligencia de negocios se procedió a validar cada uno de los atributos tengan el formato adecuado para el posterior procesamiento de los datos disponibles mismo que se detalla a continuación.

El detalle del tipo de cada uno de los datos fue explicado en la parte correspondiente a la construcción de cada dimensión y la tabla de hechos.

#### **3.5.9. Modelado**

Una vez ejecutado el proceso ETL con información limpia y depurada el siguiente proceso se centrará en la ejecución de la técnica de modelado

### **3.5.9.1. Selección de técnica de modelado**

En base a la estructura de los datos y en línea a los objetivos planteados en la presente investigación y como parte del desarrollo del modelo de bussines intelligence, se decidió utilizar dos técnicas como son:

- La generación de grupos mediante la aplicación de K medias y DBSCAN para su posterior comparativa.
- La proyección de construcción de nuevas edificaciones.

### **3.5.9.2. Validación de técnicas de modelado**

En primer lugar, se hará uso de los siguientes recursos para el análisis y validación de resultados como son Power BI, Integration Services, Analysis Services y lenguaje de programación R ya son que herramientas que agregan mucho valor a nivel de visualizaciones para la construcción de los diferentes KPIs, así como métricas o indicadores que contribuirán a validar los resultados obtenidos ya sean en la construcción de los grupos como en la evaluación de la proyección.

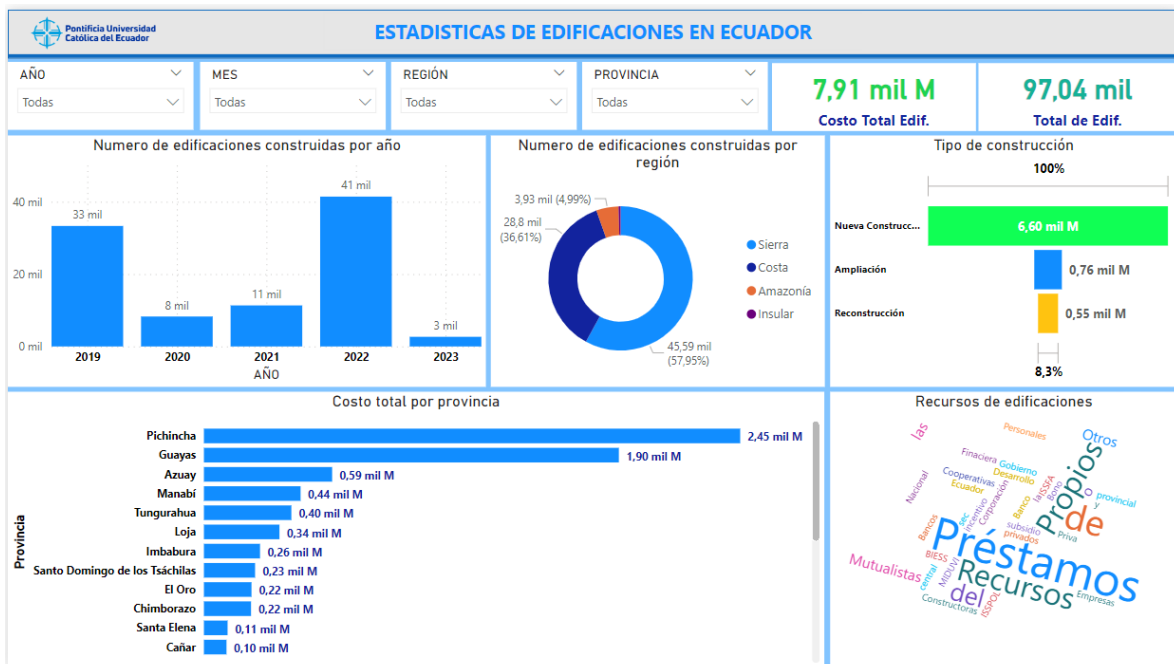
# CAPÍTULO IV

## 4. RESULTADOS

Este capítulo recogerá los principales resultados obtenidos una vez realizada la aplicación de las metodologías seleccionadas tanto para la creación del Data Warehouse y de minería de datos a través de CRISP-DM.

En primer lugar, una vez establecido e implementado la solución de inteligencia de negocios se procedió a construir un tablero de control en donde se pueda realizar un monitoreo de la evolución en función a un determinado año, mes, región y provincia de manera consolidada.

Dashboard



**Figura 30.** Dashboard de resumen general  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

La figura 30 muestra una primera aproximación de los datos ya modelados y depurados en donde podemos visualizar una desagregación general de los principales aspectos que intervienen en el análisis de las estadísticas de las edificaciones. Generando una forma mucho más dinámica de interpretar los datos.

### Dashboard



**Figura 31.** Dashboard de distribución de numero de edificaciones por provincia  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

La figura 31 nos indica una perspectiva general de las principales provincias en donde se concentran el mayor número de edificaciones construidas dentro del periodo de tiempo 2019 hasta 2023 reflejándose una mayor concentración como era de esperarse en las provincias de Pichincha y Guayas esto puede explicarse debido a que son polos de desarrollo.

De igual manera a partir de este dashboard podemos desprender mucha más información procesada a partir de minería de datos.

#### **4.1. Principales factores que inciden en el incremento de nuevas edificaciones**

La principal razón por la que se gestiona un permiso en las provincias de Pichincha y Guayas es por la construcción de una nueva edificación alcanzando el 58.42% en el periodo 2022, para el primer trimestre del año 2023 el porcentaje de nuevas edificaciones 54.92%, mayormente de carácter privado. Lo que indica que uno de los principales factores es el poder adquisitivo con que se cuenta en estas dos provincias que son consideradas como polos de desarrollo ya que el 49.68% es financiado con recursos de empresas y constructoras privadas, en tanto que el resto es financiado por recursos personales y préstamos de bancos privados.

Otro factor importante a mencionar es el propósito de la construcción en el resto de las provincias fuera de Pichincha y Guayas el principal motivo de una nueva edificación es para un propósito comercial.

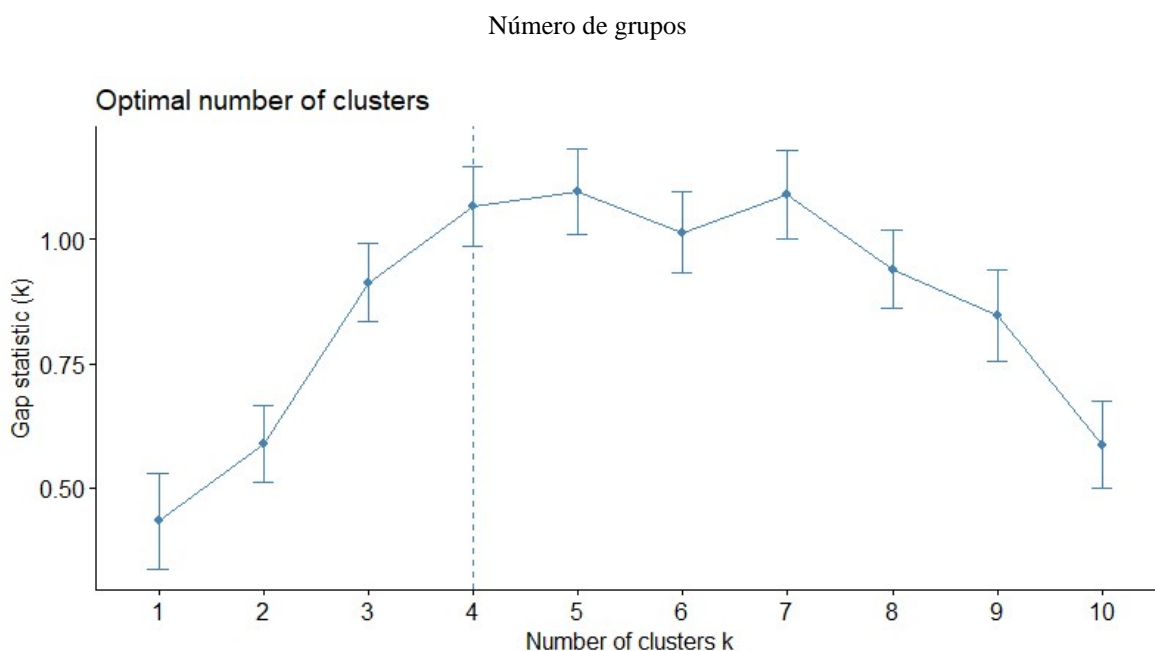
También es importante señalar como aspecto importante en el crecimiento de nuevas edificaciones la construcción de casas de salud que no necesariamente se originan desde recursos del sector público sino más bien desde capital privado, lo cual indica que un factor a considerar es la cobertura a nivel poblacional que a su vez identifica posibles y potenciales proyectos de inversión.

La superficie del terreno también es un factor determinante en el incremento de nuevas edificaciones dado que los grandes proyectos inmobiliarios requieren de grandes espacios que cubran todas las necesidades y planificaciones de construcción.

## 4.2. Resultados de los grupos utilizando K medias

Del resultado de la aplicación del algoritmo k medias se obtuvieron los siguientes resultados que se detallaran a continuación.

Como primer paso mediante el grafico del codo nos permitirá visualizar previamente el número de grupos optimo dentro de la solución de inteligencia de negocios propuesta en función a los datos que contiene el Data Warehouse.

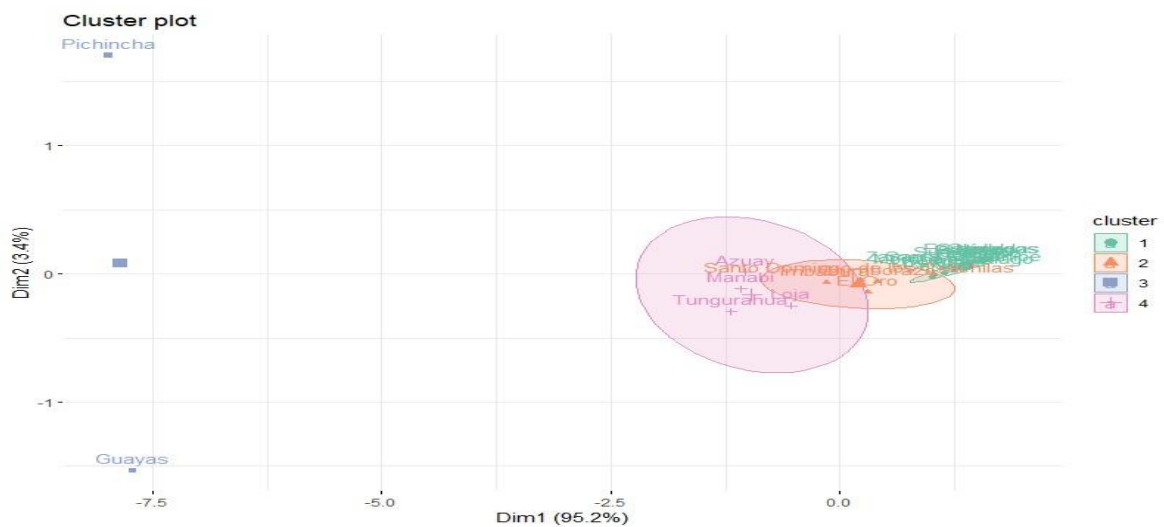


**Figura 32.** Número de grupos recomendados  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

Como muestra la figura 32 muestra el número de grupos recomendados es de 4, resultado del cálculo previo de varias funciones adecuadas a la estructura y tipo de los datos tomando en cuenta que se considera las distancias elevadas al cuadrado de los centroides de cada grupo construido de manera que se evalúa la homogeneidad y heterogeneidad entre sí. Dado los

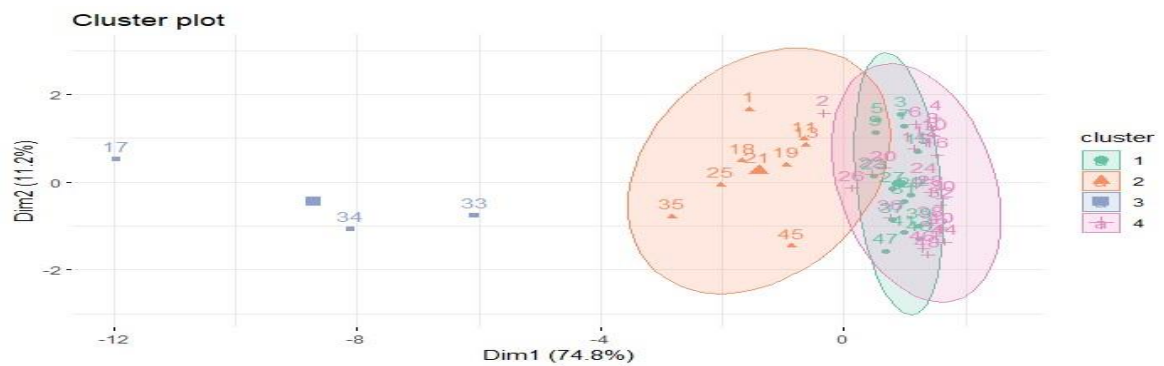
objetivos de análisis de la investigación propuesta es posible realizar una comparación modificando parámetros en la construcción de los grupos.

Gráfico de grupos



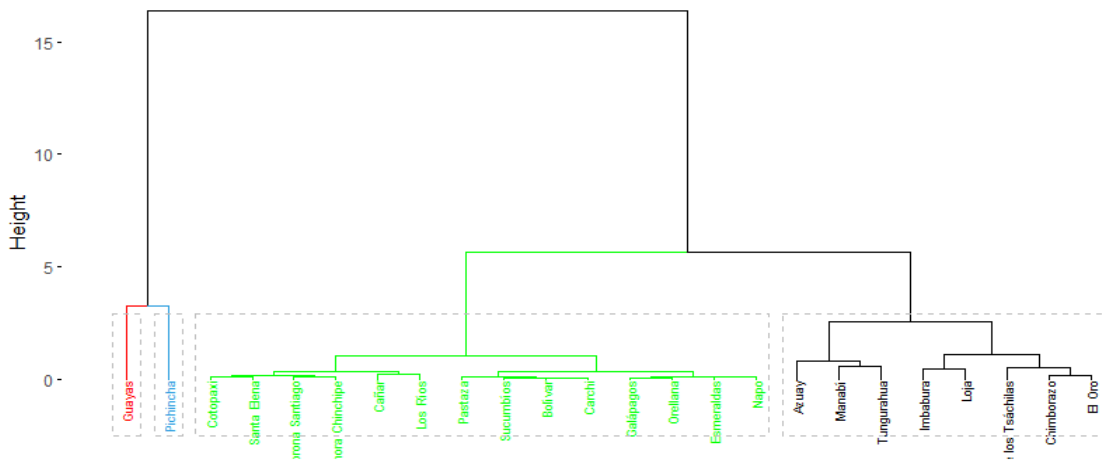
**Figura 33.** Gráfico de grupos de edificaciones  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

Gráfico de grupos



**Figura 34.** Gráfico de grupos de edificaciones  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

Dendograma

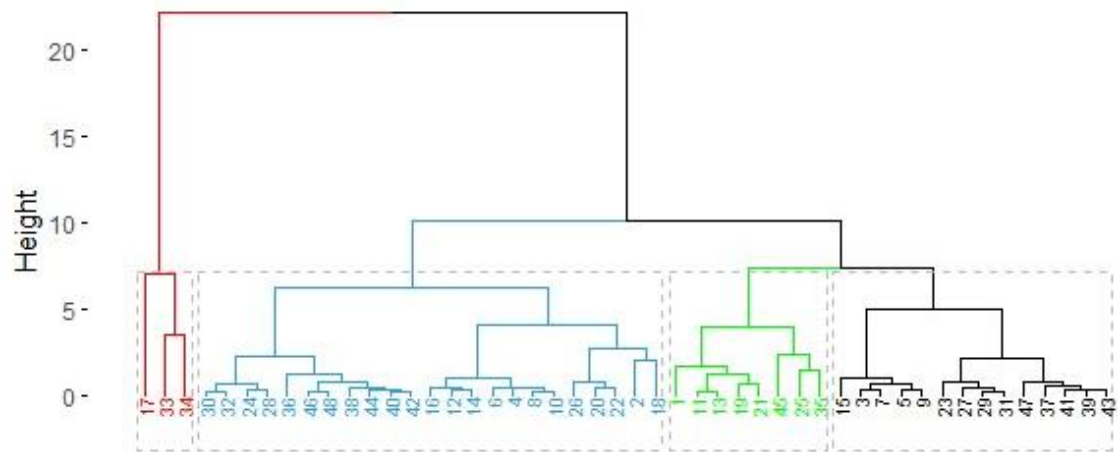


**Figura 35.** Dendograma

Fuente:

Elaboración: Autor

Dendograma



**Figura 36.** Dendograma

Fuente: Estadística de Edificaciones (2019-2022)

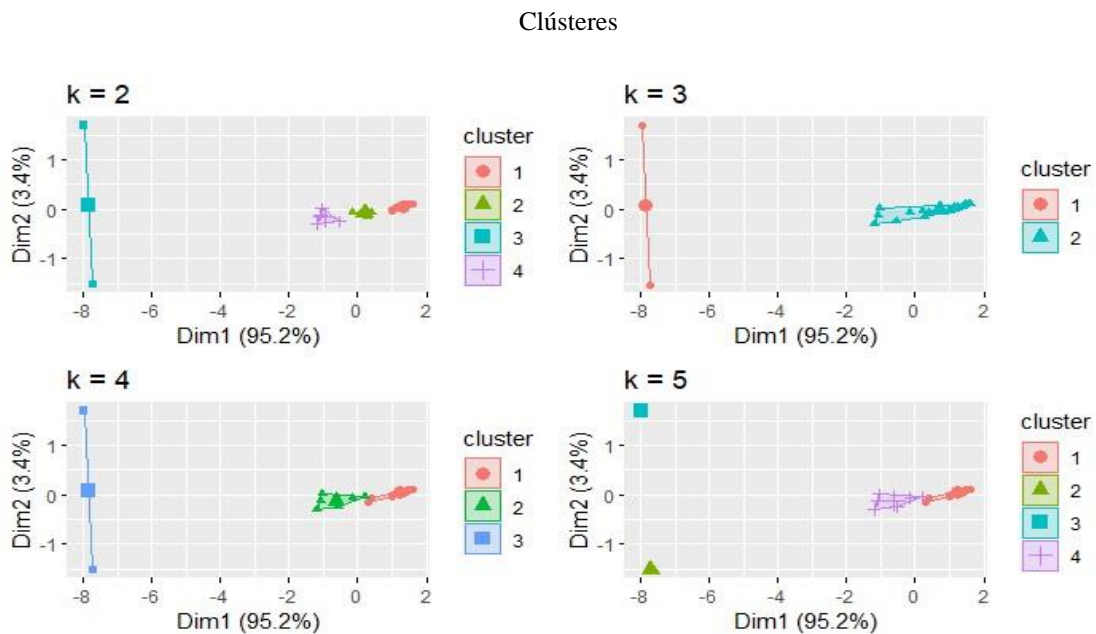
Elaboración: Autor

Como se puede visualizar el dendrograma ejemplifica de mejor manera la conformación de los grupos destacando claramente la provincia de Guayas y Pichincha como grupos excluyentes esto se puede explicar debido a que su composición es diferente al resto de provincias.

Como podemos visualizar en la siguiente tabla las provincias asignadas a cada uno de los grupos muestra una tendencia en lo referente al promedio del costo total y el promedio de permisos de edificaciones.

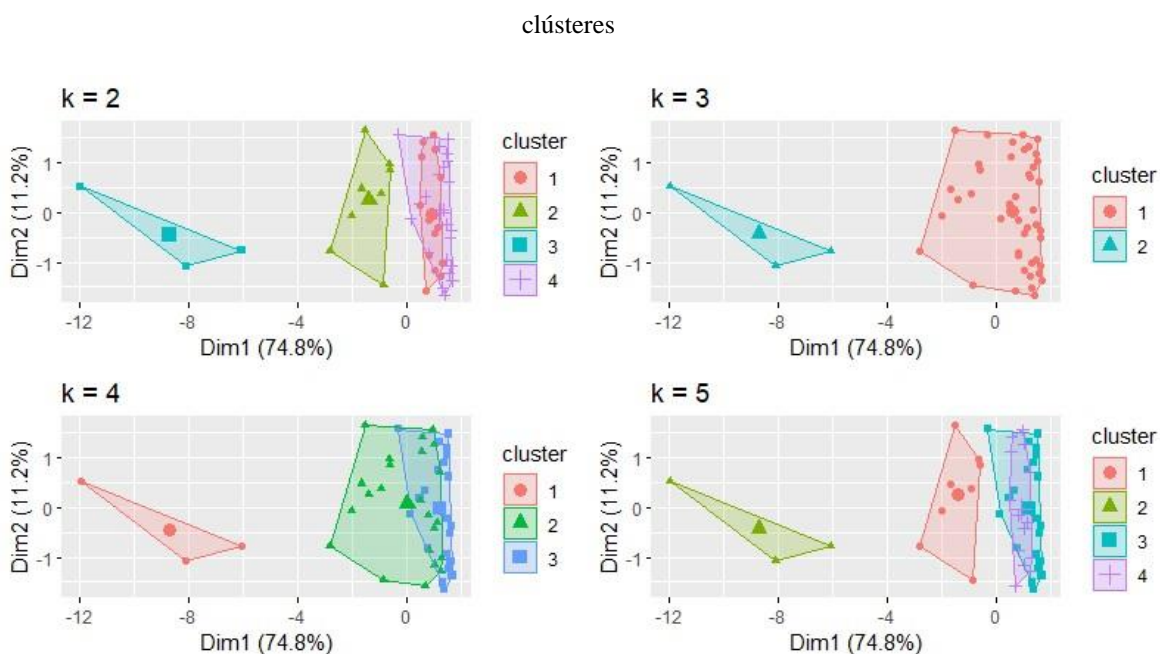
#### 4.2.1. Análisis y validación de resultados de K-medias

Una vez realizado el modelo de agrupamiento utilizando K medias se procede a validar sus respectivos resultados.



**Figura 37.** Calidad de clústeres  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

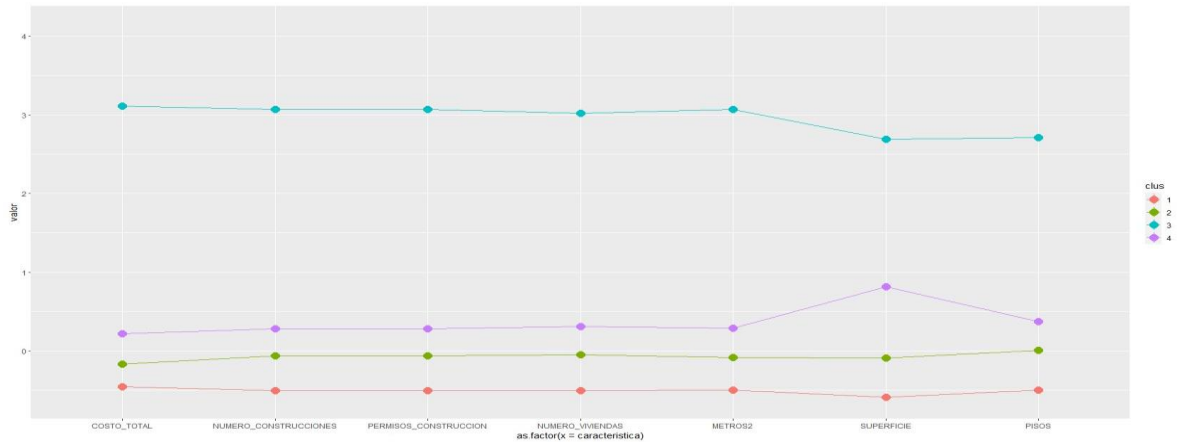
Como podemos visualizar en la figura numero 37 tenemos una comparación en función a la generación de distintos grupos en primera instancia podemos evidenciar que la data muestra una concentración marcada de acuerdo a las variables seleccionadas, de manera que no se hace un poco complicado optimizar la caracterización de los resultados.



**Figura 38.** Optimización de clústeres  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

La figura 39 muestra los grupos optimizados de manera que es mucho más sencillo visualmente identificarlos que a su vez permita realizar un análisis y comparativa en función a un diferente número de clústeres

Variabilidad de clústeres



**Figura 39.** Variabilidad entre grupos  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

Una vez realizado los cálculos correspondientes con cuatro grupos tenemos un nivel de aproximadamente del 96% de explicación entre grupos.

En primera instancia podemos determinar la media de cada uno de los grupos construidos en función a cada una de las variables incluidas. Dentro de este contexto se evidencia que la media de la variable superficie es mayor en el clúster 4 respecto a la comparativa al clúster 1 y 2.

De esta forma el algoritmo k medias nos proporciona una gran herramienta para poder planificar y determinar diferentes estrategias en función a una evaluación situacional que contribuyan en el crecimiento y desarrollo del sector de la construcción y a su vez pueda anticiparse a escenarios desfavorables.

La tabla 17 muestra en resumen la composición de los 4 grupos propuestos en función al cálculo del promedio de las principales variables de estudio de tal manera que podemos ver que el grupo 1 concentra un número mayor de provincias en donde el monto promedio del costo total supera los 100 mil dólares en provincias como Cañar, Cotopaxi, Los Ríos y Santa Elena, mientras que provincias como Pichinchas y Guayas tienen una marcada diferencia en

el monto promedio de costo total al ser cantidades elevadas correspondiente al número de permisos de construcción superando fácilmente los 30 mil permisos por año.

**Tabla 17.** Evaluación del modelo K-MEANS

MÉTRICA	VALOR
AVERAGE SILHOUETTE	0.6642652

**Fuente:**  
Elaboración: Autor

**Tabla 18.** Detalle de grupos por provincia

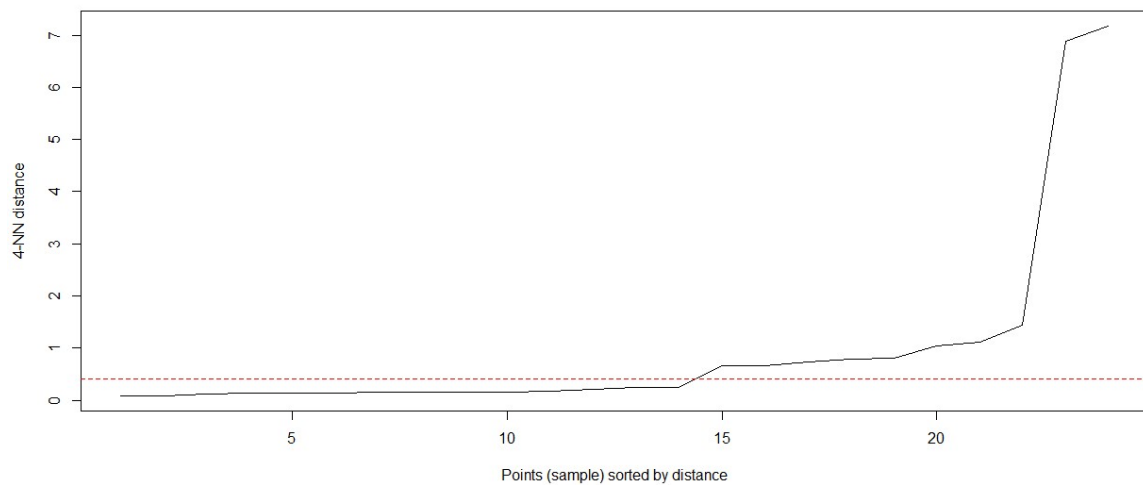
GRUPO	PROVINCIA	PROMEDIO COSTO TOTAL	PROMEDIO PERMISOS CONSTRUCCION	PROMEDIO DE METROS2
1	BOLIVAR	75,818,752	1,294	246,334
	CAÑAR	153,536,945	2,369	525,814
	CARCHI	62,442,717	1,348	235,250
	COTOPAXI	120,250,276	2,048	451,850
	ESMERALDAS	59,885,792	851	183,314
	GALAPAGOS	57,657,449	612	104,876
	LOS RIOS	136,052,922	2,670	554,784
	MORONA SANTIAGO	90,084,325	1,597	341,063
	NAPO	48,542,683	851	180,637
	ORELLANA	41,650,725	575	144,365
	PASTAZA	63,157,067	1,204	240,788
	SANTA ELENA	146,792,373	2,254	429,074
	SUCUMBIOS	67,097,998	1,334	256,308
	ZAMORA CHINCHIPE	98,328,397	1,670	390,380
<b>TOTAL GRUPO 1</b>		<b>87,235,602</b>	<b>1,477</b>	<b>306,060</b>
2	CHIMBORAZO	313,287,851	4,117	1,162,129
	EL ORO	300,592,515	4,698	1,123,821
	IMBABURA	366,089,299	7,034	1,237,939
	SANTO DOMINGO DE LOS TSACHILAS	335,247,972	4,867	1,200,293
	<b>TOTAL GRUPO 2</b>		<b>328,804,409</b>	<b>5,179</b>

3	GUAYAS	2,729,773,860	30,068	7,270,816
	PICHINCHA	3,384,318,633	33,077	8,605,996
		<b>3,057,046,247</b>	<b>31,573</b>	<b>7,938,406</b>
4	AZUAY	824,351,582	7,404	2,393,463
	LOJA	502,822,782	7,339	1,643,752
	MANABI	615,448,395	9,749	1,726,605
	TUNGURAHUA	654,821,842	7,829	2,180,107
<b>TOTAL GRUPO 4</b>	<b>649,361,150</b>	<b>8,080</b>	<b>1,985,982</b>	
<b>TOTAL GENERAL</b>	<b>468,668,881</b>	<b>5,702</b>	<b>1,367,907</b>	

**Fuente:** Estadística de Edificaciones (2019-2022)

Elaboración: Autor

### 4.3. Resultados de los grupos utilizando DBSCAN

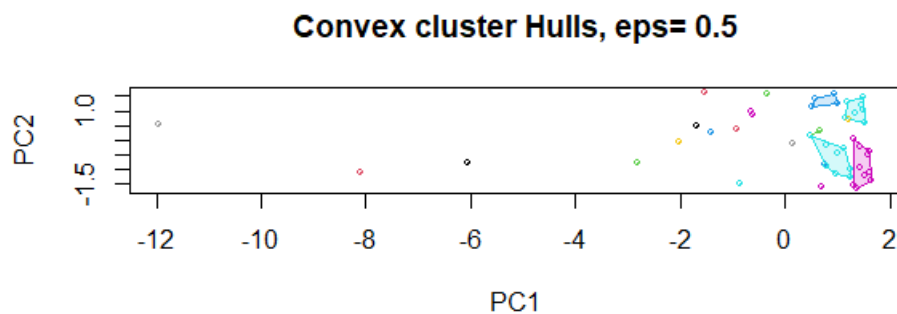


**Figura 40.** Gráfico valor óptimo de epsilon

Fuente: Estadística de Edificaciones (2019-2022)

Elaboración: Autor

En función al grafico 40 podemos determinar que el valor óptimo de epsilon oscila entre 0.01 y 0.05 en función a las reglas se ha determinado la generación de  $\text{min\_samples} = 4$ , en función a las premisas mencionadas se ha podido construir los siguientes grupos.



**Figura 41.** Grupos generados mediante el algoritmo DBSCAN  
 Fuente: Estadística de Edificaciones (2019-2022)  
 Elaboración: Autor

#### 4.3.1. Análisis y validación de resultados de DBSCAN

Una vez construido el modelo se procede a validar los respectivos resultados en función a la prevalencia y caracterización se han utilizado las siguientes métricas para su respectiva evaluación

**Tabla 19.** Evaluación del modelo DBSCAN

MÉTRICA	VALOR
AVERAGE SILHOUETTE	0.2272697

Elaboración: Autor

Para la evaluación del modelo construido en base a DBSCAN utilizaremos la métrica de del coeficiente de silueta que nos proporciona gran ayuda a la hora de determinar la calidad de la agrupación de los grupos formados en función del valor obtenido de la silueta promedio de silhouette.

La tabla 19 nos indica que el valor promedio de silhouette es de 0.22 lo que significa que la formación de 4 grupos no necesariamente es significativa para lograr explicar el caso de estudio dado que según se muestra en la figura 39 algunas observaciones quedan por fuera de pertenecer a un grupo en específico.

#### 4.3.2. Comparativa de métricas de grupos K-medias – DBSCAN

Una vez aplicado ambas técnicas de construcción de grupos y evaluado la calidad y eficiencia de cada clúster bajo las puntuaciones obtenidas después de la aplicación de la métrica podemos emitir las siguientes conclusiones.

**Tabla 20.** Evaluación de K-MEDIAS y DBSCAN

MÉTRICA	K-MEANS	DBSCAN
<b>AVERAGE SILHOUETTE</b>	0.6642652	0.2272697

Elaboración: Autor

La tabla 20 indica las puntuaciones obtenidas mediante la aplicación de ambas metodologías, la aplicación de k-medias muestra un mejor comportamiento con un coeficiente de silueta de 0.66 lo cual indica que el algoritmo aplicado funciona de mejor manera con los datos disponibles mostrando un mejor rendimiento comparado a DBSCAN que solo alcanza un 0.22 en coeficiente.

El anexo 17 muestra la solución de inteligencia de negocios en donde el código de ejecución realizado en r es trasladado a Power bi con el objetivo de que su análisis y visualización se faciliten y se puede optimizar el nivel en la toma de decisiones.

#### **4.4. Modelo de Regresión lineal múltiple**

El desarrollo y estimación del modelo de regresión nos permitirá determinar y predecir el crecimiento y evolución del sector de la construcción en función a sus datos históricos.

##### **4.4.1. Selección de la técnica de la técnica de modelado**

Dado que el objetivo es lograr predecir un comportamiento a futuro, la técnica más adecuada para llevar a cabo esto es la regresión lineal múltiple, es decir que basa su funcionamiento final en función a distintas variables independientes que se incluyen dentro del modelo y la relación entre sus variables numéricas.

##### **4.4.2. Construcción del modelo**

Para la construcción de la estimación se tomaron consideraciones como la relación que tienen las variables a incluirse dentro del modelo. El planteamiento propuesto muestra a la variable  $y$  como variable respuesta en función a las variables independientes en este caso  $x$ .

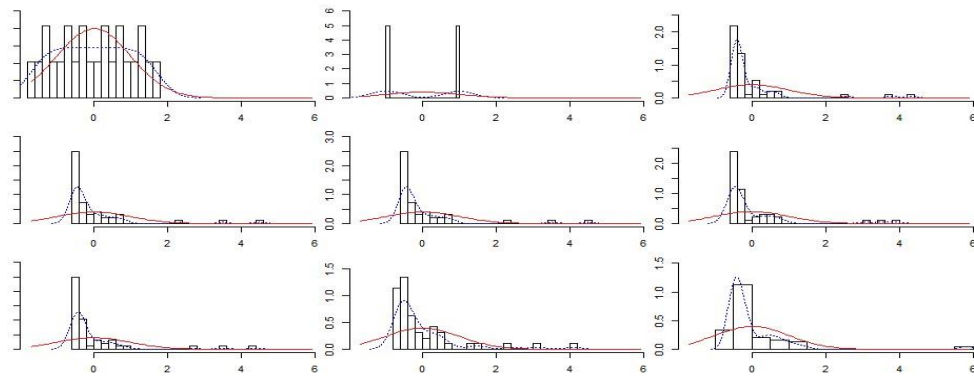
##### **4.4.3. Variables a utilizar**

Para la estimación final utilizaremos las siguientes variables.

- Número de permisos de construcción por provincia.
- Número de construcciones por provincia.
- Metros cuadrados de la edificación.

- Número de pisos.
- Costo Total.
- Tipo de propiedad (Público/Privado)
- Crecimiento poblacional anual

Correlación de variables



**Figura 42.** Relación de variables a utilizar en el modelo  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

#### 4.4.4. Análisis y validación de los resultados de la estimación lineal



**Figura 43.** Proyección de edificaciones a construir  
Fuente: Estadística de Edificaciones (2019-2022)  
Elaboración: Autor

La figura 43 muestra la proyección final de las nuevas edificaciones a construirse en los periodos 2024 y 2025 en función a las variables antes detalladas, la solución de inteligencia de negocios poder visualizar los valores de proyección de acuerdo a varios filtros como provincia y región.

En función a los resultados obtenidos las variables valor de la edificación a construir y metros cuadrados son variables importantes que contribuyen mayormente en la predicción final mismas a las que se puede cambiar parámetros para visualizar distintos escenarios.

## CAPÍTULO V

### 5. CONCLUSIONES Y RECOMENDACIONES

#### 5.1. Conclusiones

En función a las técnicas de minería de datos aplicadas dentro de la investigación sobre el desarrollo y construcción de una solución de inteligencia de negocios y la aplicación de las mismas podemos emitir las siguientes conclusiones.

- Se logro desarrollar una solución de inteligencia de negocios que integrar y consolida la información actual e histórica proveniente de la encuesta ESED de tal forma que es posible obtener indicadores y realizar análisis descriptivos a un nivel de granularidad que permite realizar comparaciones entre años, provincias o regiones etc.
- Para el sector de la construcción se vuelve indispensable contar con información histórica y actual la solución de inteligencia de negocios brinda indicadores que contribuyen a realizar un análisis más efectivo.
- Como parte del desarrollo de la solución de inteligencia de negocios, la construcción de una infraestructura que soporte las necesidades de negocio respecto al consumo de información representa una gran ventaja pues al contar con un DataWarehouse mismo que se alimenta mediante la ejecución de un proceso ETL genero un beneficio importante al lograr depurar, cargar y mejorar la calidad de datos que de forma continua de manera que se puedan evitar errores en análisis futuros.

- El desarrollo y utilización del tablero del control no solo optimizó el tiempo de consulta de información, así como lograr un análisis efectivo y mucho más amplio debido al gran alcance e importancia de los datos generando así valor agregado sobre la información procesada y analizada.
- Un punto a considerar es que las construcciones y edificaciones privadas muestran un mejor escenario en contra parte a las edificaciones públicas. Aunque tampoco son valores que se aproximan a la media del año 2019.
- Las provincias de Pichincha y Guayas son las que muestran un comportamiento hacia el alza en la construcción de nuevas edificaciones esto puede explicarse debido a que son consideradas como polos de desarrollo.
- La aplicación de las dos técnicas de agrupamiento permitió identificar a mayor profundidad una clara tendencia y diferencia que existen en las provincias de Pichincha y Guayas referente al costo y al número de permisos de construcción.
- Mediante la aplicación del modelo de predicción y el apoyo de la infraestructura de la solución de inteligencia de negocios es posible explorar los datos futuros en función a variables como la región o la provincia lo cual es un aspecto que maximiza el nivel de análisis.
- Es importante recalcar que las proyecciones realizadas mediante el uso de la data Warehouse pueden actualizarse automáticamente y si se prevén cambios en temporalidad esto no incurrirá en largos tiempos de desarrollo.

## **5.2. Recomendaciones**

- Los datos analizados cuentan con una gran diversidad que pueden ser utilizados como insumo para futuros estudios o extensiones sobre el tema de investigación.
- Tomando en consideración el volumen y diversidad de la información utilizada se vuelve indispensable contar con una perspectiva clara del alcance respecto a la metodología de minería de datos con el fin de que los resultados sean consistentes y acorde a los objetivos planteados.
- Es recomendable la aplicación de distintas herramientas y metodologías de procesamiento y modelados de datos dado que esto amplía nuestro rango de análisis y comparativa de datos.

## BIBLIOGRAFÍA

- Aspiazu Prado, A. M. (2019). Análisis de Impacto en el Desempeño de la Toma de Decisiones en un Proceso de Mesa de Ayuda, Implementando una Solución de Business Intelligence para una Empresa de Soluciones de Negocio en Ecuador. (*Tesis de grado*). Universidad Católica de Santiago de Guayaquil, Guayaquil. Obtenido de <http://repositorio.ucsg.edu.ec/bitstream/3317/12998/1/T-UCSG-PRE-ING-CIS-225.pdf>
- Bonilla Botia, L., & Briceño Díaz, F. (2006). Sistemas de información como apoyo a la toma de decisiones. *Prospectiva*, 55.
- Cano, J. (2007). *Bussines Intelligence: Competir con información*. Barcelona: Banesto Fundación Cultural.
- Cedeño Trujillo, A. (2006). MODELO MULTIDIMENSIONAL. *MODELO MULTIDIMENSIONAL*. Instituto Superior Politécnico José Antonio Echeverría, La Habana, Cuba. Obtenido de <https://www.redalyc.org/pdf/3604/360433560009.pdf>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). *Guía paso a paso de Minería de Datos*.
- Chapman, P., Clinton, J., Randy, K., Thomas, K., Thomas, R., Colin, S., & Rüdiger, W. (2000). CRISP-DM 1.0 Step-by-step data mining guide. *CRISP-DM 1.0 Step-by-step data mining guide*.
- Chavez Huapaya, S. M., & Contreras Ochoa, C. Y. (2018). Implementación de bussines intelligence, utilizando la metodología de Ralph Kimball, para el proceso de toma de decisiones del área de ventas. Empresa Yukids. (*Tesis de grado*). Universidad Autónoma del Perú, Lima. Obtenido de [https://repositorio.autonoma.edu.pe/bitstream/handle/20.500.13067/435/TESIS%202018%20CHAVEZ\\_CONTRERAS.pdf?sequence=1&isAllowed=y](https://repositorio.autonoma.edu.pe/bitstream/handle/20.500.13067/435/TESIS%202018%20CHAVEZ_CONTRERAS.pdf?sequence=1&isAllowed=y)
- Conesa Caralt, J., & Curto Díaz, J. (2010). *Introducción al Business Intelligence*. Barcelona: UOC.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Clifford, S. (2022). *Introduction to Algorithms*. Cambridge, Massachusetts, USA: The MIT Press.
- Corona Organiche, E., Jiménez Alfaro, A. J., & Cortés Barrera, G. (2020). Principales Metodologías en el Desarrollo de Proyectos de Minería de Datos. *Principales Metodologías en el Desarrollo de Proyectos de Minería de Datos*. Tecnológico de

Estudios Superiores de Ecatepec, México, México. Obtenido de <https://tecnocultura.org/index.php/Tecnocultura/article/view/9>

- Delgado Palacios, P. M. (2014). Inteligencia de Negocios para Empresas de Construcción y la Gestión de Proyectos con enfoque en las mejores prácticas. *Inteligencia de Negocios para Empresas de Construcción y la Gestión de Proyectos con enfoque en las mejores prácticas*. Universidad del Azuay, Cuenca, Azuay, Ecuador. Obtenido de <https://dspace.uazuay.edu.ec/bitstream/datos/3739/1/10407.PDF>
- Flasco. (2011). *Boletín mensual de análisis sectorial de MIPYMES Sector de la Construcción*. Quito: FLACSO - MIPRO.
- Forero Castañeda, D., & Sánchez García, J. (2021). INTRODUCCIÓN A LA INTELIGENCIA DE NEGOCIOS. *INTRODUCCIÓN A LA INTELIGENCIA DE NEGOCIOS*. Tecnología, Investigación y Academia -Red Avanzada – RITA, Bogotá, Colombia.
- Gerrero García, M., & Rodas Silva, J. (2022). Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones. *Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones*. Universidad Técnica de Manabí, Manabí, Ecuador.
- Gironés Roig, J. (2013). *Business Analytics*. Cataluña: Eureka Media SL.
- Hernández Leal, E., Duque Méndez, N., & Moreno Cadavid, J. (2017). Big Data: an exploration of research, technologies and application cases. *Big Data: an exploration of research, technologies and application cases*. TecnoLógicas, Medellín. Obtenido de <http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf>
- IBM. (2021). *Guía de CRISP-DM de IBM SPSS Modeler* (Vol. 18). Armonk: Copyright International Business Machines Corporation. Obtenido de [https://www.ibm.com/docs/es/SS3RA7\\_18.4.0/pdf/ModelerCRISPDM.pdf](https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf)
- INEC. (2021). *Estadísticas de Edificaciones*. Obtenido de <https://www.ecuadorencifras.gob.ec/encuesta-de-edificaciones-2021/>
- Jaime, P., Matto, M., & Schuster, A. (1999). Data Warehousing. (*Tesis de grado*). Universidad Nacional de la Plata, Buenos Aires, Argentina. Obtenido de <https://core.ac.uk/download/pdf/15761033.pdf>
- Jianfeng, L., Huaqin , Q., Ziyi , L., Sheng , W., Qiang , Z., & Zhengmin , H. (2022). A Density-Based Spatial Clustering of Application with Noise Algorithm and its Empirical Research. *A Density-Based Spatial Clustering of Application with Noise*

- Algorithm and its Empirical Research*. North China Electric Power University, Beijing, China.
- Kaplan, R., & Norton, D. (1997). *Cuadro de Mando Integral (The Balanced Scorecard)* (1 ed.). Barcelona: Gestión 2000.
- Laudon, K., & Laudon, J. (2012). *Sistemas de Información Gerencial*. México: Pearson.
- Mavesoy Murcia, C. (2018). MODELO BASADO EN CRISP-DM EXTENDIDO MEDIANTE PRÁCTICAS DE METODOLOGÍAS ÁGILES PARA PROYECTOS MEDIANOS DE ANALÍTICA DE DATOS. 2018. Universidad de Medellín, Medellín, Colombia. Obtenido de [https://repository.udem.edu.co/bitstream/handle/11407/6341/T\\_MIS\\_398.pdf](https://repository.udem.edu.co/bitstream/handle/11407/6341/T_MIS_398.pdf)
- Menéndez Domínguez, V. H., & Castellanos Bolaños, M. E. (2015). SPEM: Software Process Engineering Metamodel. *SPEM: Software Process Engineering Metamodel*. Universidad Autónoma de Yucatán, Mérida, México. Obtenido de [file:///C:/Users/pc/Downloads/admin,+Gestor\\_a+de+la+revista,+relais-v3-n2-92-100.pdf](file:///C:/Users/pc/Downloads/admin,+Gestor_a+de+la+revista,+relais-v3-n2-92-100.pdf)
- Microsoft. (3 de Febrero de 2016). *Power BI: analítica de negocios para obtener información de valor*. Obtenido de <https://news.microsoft.com/es-xl/power-bi-analitica-de-negocios-para-obtener-informacion-de-valor/>
- Microsoft. (2017). *Microsoft SQL Server Notas técnicas del producto*. Microsoft. Obtenido de <chrome-extension://efaidnbmnmnibpcajpcglclefindmkaj/https://info.microsoft.com/rs/157-GQE-382/images/ES-ES-CNTNT-Whitepaper-SQLServer2017ITDMWhitePaper-ES.pdf>
- Microsoft. (2023). *Power Bi*. Obtenido de Power Bi: <https://powerbi.microsoft.com/es-es/what-is-power-bi/>
- Muñoz, H., Osorio, R., & Zuñiga, L. (2016). Inteligencia de Negocios Clave del éxito en la era de la información. *Clío América*, 195.
- Navarrete Carrasco, R. C. (2002). Análisis de impacto del Bussines Intelligence: Expectativas y Realidades. (*Tesis de maestría*). Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey. Obtenido de [https://repositorio.tec.mx/bitstream/handle/11285/568332/DocsTec\\_1043.pdf?sequence=1&isAllowed=y](https://repositorio.tec.mx/bitstream/handle/11285/568332/DocsTec_1043.pdf?sequence=1&isAllowed=y)
- O'Brien, J., & Marakas, G. (2004). *Sistemas de Información Gerencial* (3 ed.). Mexico, Mexico: The McGraw-Hill Companies.

- Ortega Encina, D. A. (2013). DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE CONTROL Y PRONÓSTICO DE CONEXIONES PARA UNA COMPAÑÍA DISTRIBUIDORA DE COMBUSTIBLE. *DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE CONTROL Y PRONÓSTICO DE CONEXIONES PARA UNA COMPAÑÍA DISTRIBUIDORA DE COMBUSTIBLE*. Universidad de Chile, Santiago, Chile. Obtenido de [https://repositorio.uchile.cl/bitstream/handle/2250/114674/cf-ortega\\_de.pdf?sequence=1&isAllowed=y](https://repositorio.uchile.cl/bitstream/handle/2250/114674/cf-ortega_de.pdf?sequence=1&isAllowed=y)
- PowerData. (2017). *Procesos ETL La Base de la Inteligencia de Negocios*. Lima: PowerData.
- Prieto, A., & Martínez, M. (2004). Sistemas de información en las organizaciones: Una alternativa para mejorar la productividad gerencial en las pequeñas y medianas empresas. *Sistemas de información en las organizaciones: Una alternativa para mejorar la productividad gerencial en las pequeñas y medianas empresas*. Universidad de Zulia, Maracaibo, Venezuela. Obtenido de <https://www.redalyc.org/pdf/280/28010209.pdf>
- Ramos, S. (2016). *Data Warehouse, Data Marts y Modelos Dimensionales un pilar fundamental para la toma de decisiones*. Alicante: SolidQ Global S.A.
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 10, 12-15.
- Riscos Ramos, R. (2022). Propuesta metodológica para la utilización de las herramientas de inteligencia de negocios en la gestión de operaciones de la pequeña y mediana empresa en el sistema eléctrico peruano. *Propuesta metodológica para la utilización de las herramientas de inteligencia de negocios en la gestión de operaciones de la pequeña y mediana empresa en el sistema eléctrico peruano*. Universidad de Piura, Piura, Perú. Obtenido de [https://pirhua.udep.edu.pe/bitstream/handle/11042/5687/DOC\\_ING\\_AUT\\_2202.pdf?sequence=2&isAllowed=y](https://pirhua.udep.edu.pe/bitstream/handle/11042/5687/DOC_ING_AUT_2202.pdf?sequence=2&isAllowed=y)
- Robles Rodríguez, J., & Velásquez García, L. (2001). La estructura del mercado laboral en la industria de materiales para la construcción. *La estructura del mercado laboral en la industria de materiales para la construcción*. Universidad Autónoma Metropolitana Unidad Azcapotzalco, México. Obtenido de <https://www.redalyc.org/pdf/413/41303307.pdf>
- Rodríguez Suárez, Y., & Díaz Amador, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 74-75.

- Salazar Torres, J. (2021). Análisis y aplicación de algoritmos de minería de datos. *Análisis y aplicación de algoritmos de minería de datos*. Colegio integrado nacional oriente de caldas, Caldas, Colombia. Obtenido de <https://revistas.uniminuto.edu/index.php/Pers/article/view/2547/2450>
- Salcedo Parra, O., Galeano, R. M., & Rodriguez, L. (2010). Metodología Crisp para la implementación Data Warehouse. *Metodología Crisp para la implementación Data Warehouse*. Universidad Distrital Francisco Jose de Caldas, Bogotá, Colombia. Obtenido de <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.redalyc.org/pdf/2570/257019810005.pdf>
- Salgueiro Roldán, J. L., Cepeda Carrión, G., & Galán González, J. L. (2012). Los Sistemas de Inteligencia de Negocios como soporte a los procesos de toma de decisiones en las organizaciones. *Los Sistemas de Inteligencia de Negocios como soporte a los procesos de toma de decisiones en las organizaciones*. Universidad de Sevilla, Sevilla, España. Obtenido de [https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS\\_PEE/132art16.pdf](https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PEE/132art16.pdf)
- Sanchez, J. (2012). Aplicación de Bussines Intelligence en Unidades de Negocio (UNE´ s). *Anuario Jurídico y Económico Escurialense*, 287-300.
- Sandoval, L. (2018). ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS. *ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS*. ITCA-FEPADE, Santa Tecla, El Salvador. Obtenido de [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6\\_RT2018.pdf](chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf)
- Sathya, D., Anu, J., & Divyadharshini, M. (2017). Air Pollution Analysis using Clustering Algorithms. *Air Pollution Analysis using Clustering Algorithms*. Kumaraguru College of Technology Coimbatore, Coimbatore, India. Obtenido de <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.internationaljournalssrg.org/uploads/specialissuepdf/ICETSST-2017/2017/CSE/19%20ESEC1094.pdf>
- Silva Peñafiel, G., Zapata Yáñez, V., Morales Guamán, K., & Toaquiza Padilla, L. (2019). Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones. *Ciencia Digital*, 404.
- Torres Sánchez, M., Espinosa Cervantes, Y., Simon Cuevas, A., García Hernández, L., & Simon Cuevas, A. (2015). Modelación Tabular: una alternativa sugerente para el análisis de los datos. *Modelación Tabular: una alternativa sugerente para el análisis*

*de los datos*. Instituto de Información Científica y Tecnológica Cuba, La Habana, Cuba. Obtenido de <https://www.redalyc.org/pdf/1814/181439409001.pdf>

Umanzor, C. (2011). *¿Como enseñar a tomar decisiones acertadas? Proceso para toma de decisiones*. San Salvador: Casa Barak.

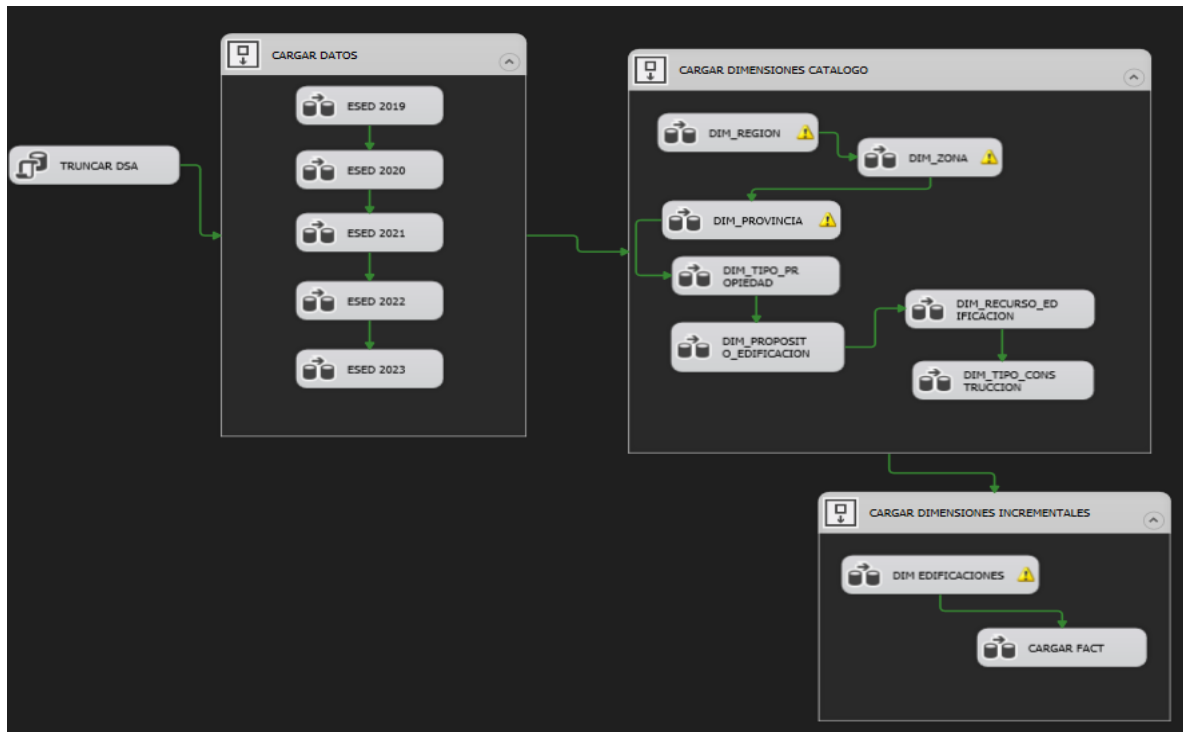
Vásquez Castrillon, J., & Sucerquia Osorio, A. (2011). *La Inteligencia de Negocios: Etapas del Proceso*. *Universidad Tecnológica de Pereira*.

## ANEXOS

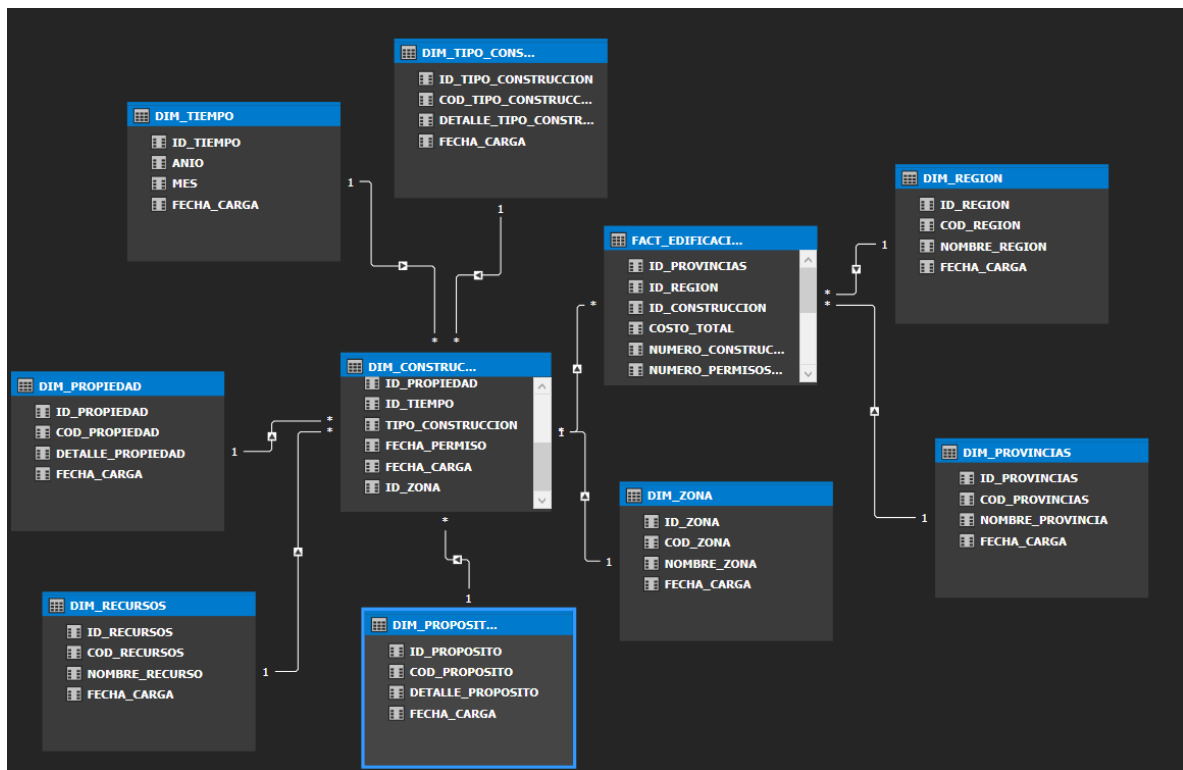
### Anexo I: Variables y descripción ESED

Nombre del campo	Descripción del campo
id	Identificador
codprovf	Provincia
codcantf	Cantón
codreg	Región
careaur	Área urbana o rural
fper	Fecha de permiso
mes	Mes
anio	Año
propie	Propiedad de la edificación: Pública/Privada
cimi	Cimientos
piso	Piso
estru	Estructura
pared	Pared
cubi	Cubierta
CPERM	Número de edificaciones a construir
CTIPOBR	Tipo de obra codificado
CORES	Código residencial de mayor área
NUVICAL	Número de viviendas por edificación
NUCUCAL	Número de cuartos por vivienda
COAMS	Código no residencial de mayor área
NUPICAL	Número de pisos por edificación calculado
NUDOCAL	Número de dormitorios por vivienda
CDISPUO	Uso de la edificación
CDISPRFI	Origen del financiamiento para edificación de mayor valor
CVAE	Valor total de la edificación calculado
CNUVICAL	Número total de viviendas calculado
CARES	Área a construir residencial en m2
CARNRES	Área a construir no residencial en m2
CARDPAR	Área a construir para parqueadero en m2
CAESV	Área para espacios verdes en m2
CSUTE	Superficie del terreno en m2
CARCO	Área total a construir en m2
CDISPSUT	Rangos de superficie total del terreno
CDISPARC	Rangos de área total a construir
SECREDIT	Segmento de crédito
NUVISEC	Número de viviendas de segmento de crédito
ESPVERD	Distancia en metros de la edificación al espacio verde público más cercano
DHAB	Documento habilitante del proyectista
CINTERN	Cobertura de internet y celular
CSERVB	Cobertura a servicios básicos
AEDIF	Tipo de acceso de la edificación
TOCUP	Tipo de ocupación de la edificación
FIDEIC	¿La edificación ha sido construida con un fideicomiso?
CASAEDIF	De las edificaciones que se proyectan construir, indique sí estas pertenecen a casas o edificios
TIPFINAN	Tipo de financiamiento con el cual se construirá la edificación
VALTERR	Valor del terreno del catastro actualizado, valor en dólares
COSM2	Costo estimado del m2 de construcción

## Anexo 2: Diseño de ETL de carga de información



## Anexo 3: Modelo Tabular Analysis Services



#### Anexo 4: Scripts sql dimensión Construcción

```
/*=====*/
/* Table: DIM_CONSTRUCCION */
/*=====*/
create table DIM_CONSTRUCCION
(
  ID_CONSTRUCCION      integer          not null,
  ID_TIEMPO            integer          null,
  ID_PROPIEDAD         integer          null,
  ID_TIPO_CONSTRUCCION integer          null,
  ID_RECURSOS          integer          null,
  ID_PROPOSITO         integer          null,
  COD_CONSTRUCCION     integer          null,
  TIPO_CONSTRUCCION    char(50)        null,
  FECHA_PERMISO        timestamp        null,
  FECHA_CARGA          timestamp        not null,
  constraint PK_DIM_CONSTRUCCION primary key (ID_CONSTRUCCION)
);

/*=====*/
/* Index: DIM_CONSTRUCCION_PK */
/*=====*/
create unique index DIM_CONSTRUCCION_PK on DIM_CONSTRUCCION (
ID_CONSTRUCCION ASC
);

/*=====*/
/* Index: RELATIONSHIP_4_FK */
/*=====*/
create index RELATIONSHIP_4_FK on DIM_CONSTRUCCION (
ID_TIPO_CONSTRUCCION ASC
);

/*=====*/
/* Index: PROPOSITO_CONSTRUCCION_FK */
/*=====*/
create index PROPOSITO_CONSTRUCCION_FK on DIM_CONSTRUCCION (
ID_PROPOSITO ASC
);

/*=====*/
/* Index: RECURSOS_CONSTR_FK */
/*=====*/
create index RECURSOS_CONSTR_FK on DIM_CONSTRUCCION (
ID_RECURSOS ASC
);

/*=====*/
/* Index: CONSTRUC_PROPIEDAD_FK */
/*=====*/
create index CONSTRUC_PROPIEDAD_FK on DIM_CONSTRUCCION (
ID_PROPIEDAD ASC
);

/*=====*/
/* Index: TIEMPO_CONSTRUC_FK */
/*=====*/
create index TIEMPO_CONSTRUC_FK on DIM_CONSTRUCCION (
ID_TIEMPO ASC
);
```

### Anexo 5: Scripts sql dimensión Propiedad

```
/*=====*/
/* Table: DIM_PROPIEDAD */
/*=====*/
create table DIM_PROPIEDAD
(
  ID_PROPIEDAD      integer          not null,
  COD_PROPIEDAD     integer          null,
  DETALLE_PROPIEDAD varchar(24)     null,
  FECHA_CARGA       timestamp        null,
  constraint PK_DIM_PROPIEDAD primary key (ID_PROPIEDAD)
);

/*=====*/
/* Index: DIM_PROPIEDAD_PK */
/*=====*/
create unique index DIM_PROPIEDAD_PK on DIM_PROPIEDAD (
  ID_PROPIEDAD ASC
);
```

### Anexo 6: Scripts sql dimensión Propósito Construcción

```
/*=====*/
/* Table: DIM_PROPOSITO_CONSTRUCCION */
/*=====*/
create table DIM_PROPOSITO_CONSTRUCCION
(
  ID_PROPOSITO      integer          not null,
  COD_PROPOSITO     integer          null,
  DETALLE_PROPOSITO varchar(50)     null,
  FECHA_CARGA       timestamp        null,
  constraint PK_DIM_PROPOSITO_CONSTRUCCION primary key (ID_PROPOSITO)
);

/*=====*/
/* Index: DIM_PROPOSITO_CONSTRUCCION_PK */
/*=====*/
create unique index DIM_PROPOSITO_CONSTRUCCION_PK on DIM_PROPOSITO_CONSTRUCCION (
  ID_PROPOSITO ASC
);
```

### Anexo 7: Scripts sql dimensión Propósito Provincias

```
/*=====*/
/* Table: DIM_PROVINCIAS */
/*=====*/
create table DIM_PROVINCIAS
(
  ID_PROVINCIAS     integer          not null,
  COD_PROVINCIA     integer          null,
  NOMBRE_PROVINCIA  varchar(50)     not null,
  FECHA_CARGA       timestamp        not null,
  constraint PK_DIM_PROVINCIAS primary key (ID_PROVINCIAS)
);

/*=====*/
/* Index: DIM_PROVINCIAS_PK */
/*=====*/
create unique index DIM_PROVINCIAS_PK on DIM_PROVINCIAS (
  ID_PROVINCIAS ASC
);
```

### Anexo 8: Scripts sql dimensión Recursos

```
/*=====*/
/* Table: DIM_RECURSOS */
/*=====*/
create table DIM_RECURSOS
(
  ID_RECURSOS      integer          not null,
  COD_RECURSO      integer          null,
  NOMBRE_RECURSO   char(50)         null,
  FECHA_CARGA      timestamp        null,
  constraint PK_DIM_RECURSOS primary key (ID_RECURSOS)
);

/*=====*/
/* Index: DIM_RECURSOS_PK */
/*=====*/
create unique index DIM_RECURSOS_PK on DIM_RECURSOS (
ID_RECURSOS ASC
);
```

### Anexo 9: Scripts sql dimensión Región

```
/*=====*/
/* Table: DIM_REGION */
/*=====*/
create table DIM_REGION
(
  ID_REGION        integer          not null,
  COD_REGION       integer          null,
  NOMBRE_REGION    varchar(50)     null,
  FECHA_CARGA      timestamp        null,
  constraint PK_DIM_REGION primary key (ID_REGION)
);

/*=====*/
/* Index: DIM_REGION_PK */
/*=====*/
create unique index DIM_REGION_PK on DIM_REGION (
ID_REGION ASC
);
```

### Anexo 10: Scripts sql dimensión Tiempo

```
/*=====*/
/* Table: DIM_TIEMPO */
/*=====*/
create table DIM_TIEMPO
(
  ID_TIEMPO        integer          not null,
  ANIO             integer          null,
  MES              integer          null,
  FECHA_CARGA      timestamp        null,
  constraint PK_DIM_TIEMPO primary key (ID_TIEMPO)
);

/*=====*/
/* Index: DIM_TIEMPO_PK */
/*=====*/
create unique index DIM_TIEMPO_PK on DIM_TIEMPO (
ID_TIEMPO ASC
);
```

## Anexo 11: Scripts sql dimensión Tipo Construcción

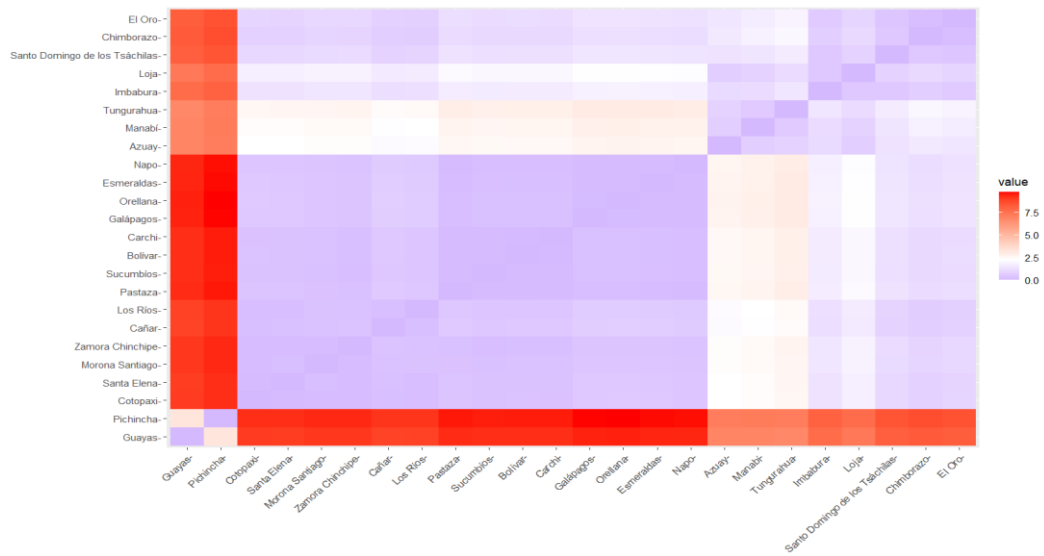
```
/*-----*/
/* Table: DIM_TIPO_CONSTRUCCION */
/*-----*/
create table DIM_TIPO_CONSTRUCCION
(
  ID_TIPO_CONSTRUCCION integer          not null,
  COD_TIPO_CONSTRUCCION integer         null,
  DETALLE_TIPO_CONSTRUCCION varchar(50) null,
  FECHA_CARGA          timestamp       null,
  constraint PK_DIM_TIPO_CONSTRUCCION primary key (ID_TIPO_CONSTRUCCION)
);

/*-----*/
/* Index: DIM_TIPO_CONSTRUCCION_PK */
/*-----*/
create unique index DIM_TIPO_CONSTRUCCION_PK on DIM_TIPO_CONSTRUCCION (
ID_TIPO_CONSTRUCCION ASC
);
```

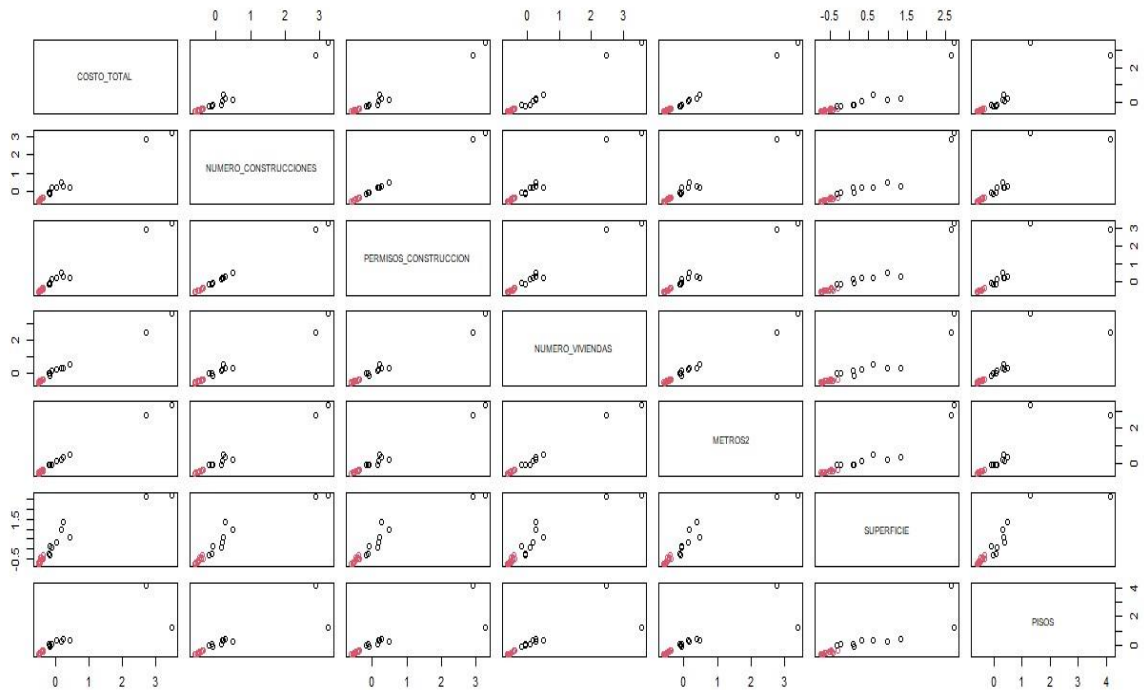
## Anexo 12: Scripts sql dimensión Zona

```
/*-----*/
/* Table: DIM_ZONA */
/*-----*/
create table DIM_ZONA
(
  ID_ZONA          integer          not null,
  COD_ZONA         integer         null,
  NOMBRE_ZONA     varchar(24)      null,
  FECHA_CARGA     timestamp       null,
  constraint PK_DIM_ZONA primary key (ID_ZONA)
);
```

### Anexo 13: Distancia euclidiana



### Anexo 14: Relaciones de distancias y principales variables



**Anexo 15:** Clústeres generados con k-medias

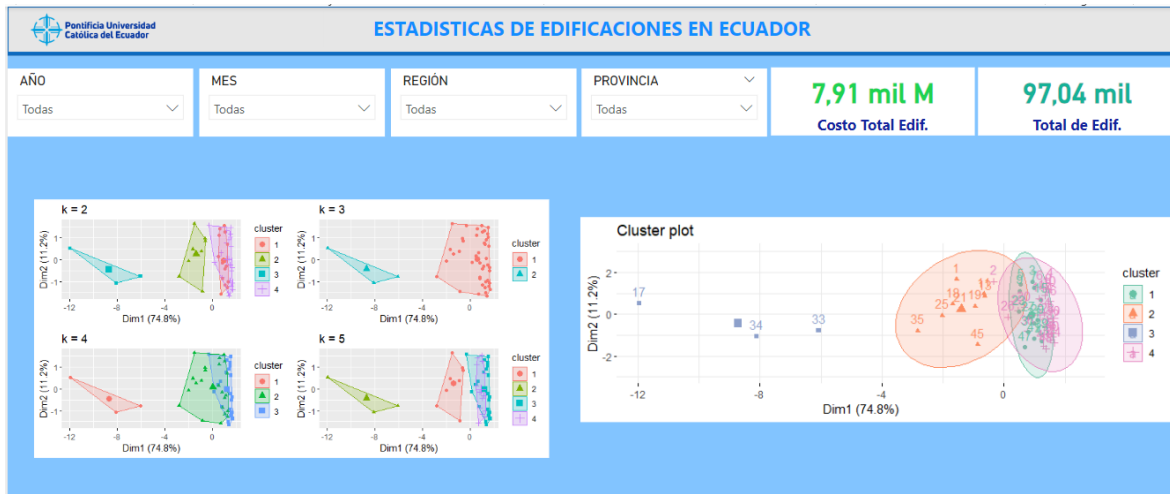


**Anexo 16:** Relación de variables a utilizar en el modelo

```
> round(cor(x = df, method = "pearson"), 3)
```

	COD_PROVINCIAS	ID_ZONA	COSTO_TOTAL	NUMERO_CONSTRUCCIONES	PERMISOS_CONSTRUCCION	NUMERO_VIVIENDAS
COD_PROVINCIAS	1.000	0.000	-0.010	-0.016	-0.016	-0.027
ID_ZONA	0.000	1.000	-0.175	-0.211	-0.211	-0.228
COSTO_TOTAL	-0.010	-0.175	1.000	0.991	0.991	0.990
NUMERO_CONSTRUCCIONES	-0.016	-0.211	0.991	1.000	1.000	0.988
PERMISOS_CONSTRUCCION	-0.016	-0.211	0.991	1.000	1.000	0.988
NUMERO_VIVIENDAS	-0.027	-0.228	0.990	0.988	0.988	1.000
METROS2	-0.022	-0.214	0.996	0.993	0.993	0.994
SUPERFICIE	-0.008	-0.234	0.929	0.946	0.946	0.928
PISOS	-0.104	-0.297	0.850	0.883	0.883	0.846
	METROS2	SUPERFICIE	PISOS			
COD_PROVINCIAS	-0.022	-0.008	-0.104			
ID_ZONA	-0.214	-0.234	-0.297			
COSTO_TOTAL	0.996	0.929	0.850			
NUMERO_CONSTRUCCIONES	0.993	0.946	0.883			
PERMISOS_CONSTRUCCION	0.993	0.946	0.883			
NUMERO_VIVIENDAS	0.994	0.928	0.846			
METROS2	1.000	0.943	0.874			
SUPERFICIE	0.943	1.000	0.859			
PISOS	0.874	0.859	1.000			

**Anexo 17:** Grupos generados y migrados a Power BI



**Anexo 18:** Resultados del modelo de regresión

OLS Regression Results

```

=====
Dep. Variable:          y          R-squared:              0.937
Model:                 OLS        Adj. R-squared:         0.935
Method:                Least Squares  F-statistic:           561.8
Date:                  Sun, 10 Dec 2023  Prob (F-statistic):    1.54e-89
Time:                  18:40:44     Log-Likelihood:        -1092.1
No. Observations:     156          AIC:                   2194.
Df Residuals:         151          BIC:                   2209.
Df Model:              4
Covariance Type:      nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-3.308e+04	6.19e+04	-0.534	0.594	-1.55e+05	8.93e+04
ANIO	16.3988	30.584	0.536	0.593	-44.029	76.826
VALOR_EDIF	0.0003	0.000	1.294	0.198	-0.000	0.001
AREA_METROS2	0.0036	8.62e-05	42.060	0.000	0.003	0.004
CRECIMIENTO_POBLACIONAL	-32.8624	131.184	-0.251	0.803	-292.056	226.331

```

=====
Omnibus:              88.232     Durbin-Watson:         2.059
Prob(Omnibus):        0.000     Jarque-Bera (JB):     2407.815
Skew:                 1.387     Prob(JB):              0.00
Kurtosis:             22.046     Cond. No.              9.05e+08
=====

```

Anexo 19: Resultados de diagnóstico de residuos del modelo

Diagnóstico residuos

