



Automatización de Reportes estadísticos del Área de Estrategia de Clientes
Persona Natural e Independientes de Bancolombia S. A: Integración de Power BI, SQL y
Python para la Optimización de Procesos de Información y Toma de Decisiones

Lukas Guerra Escobar

Ingeniero Industrial

Semestre de Industria o Práctica Empresarial

Miguel Ángel Arroyave Guerrero, Magíster (MSc) en Ingeniería Industrial

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Industrial

Medellín, Antioquia, Colombia

2025

Cita	(Guerra Escobar, 2025)
Referencia	Guerra Escobar, L. (2025). <i>Automatización de Reportes estadísticos del Área de Estrategia de Clientes Persona Natural e Independientes de Bancolombia S.A</i> [Trabajo de grado profesional]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

Este trabajo es dedicado a mi madre y mi tío que siempre estuvieron para apoyarme durante el transcurso de mi carrera.

Agradecimientos

Quiero agradecer a mi familia y amigos por aportar su grano de arena en este camino, por apoyarme y estar siempre presentes cuando los necesite.

También extiendo mi gratitud a todos los profesores que contribuyeron a mi crecimiento tanto profesional como personal, y a mi alma máter, por enseñarme a ver la vida desde diferentes perspectivas, acogerme y acompañarme durante estos años de carrera.

Finalmente, quiero agradecerme a mí mismo por no desistir, por mantenerme firme ante los desafíos y por siempre dar lo mejor de mí.

Tabla de contenido

Dedicatoria	3
Agradecimientos	3
Lista de tablas	5
Lista de figuras	6
Siglas, acrónimos y abreviaturas	7
Resumen	8
Abstract	9
1. Introducción	10
2. Objetivos	11
2.1. Objetivo General	11
2.2. Objetivos Específicos	11
2.2.1. Estudio de la documentación e información disponible:	11
2.2.2. Automatización del proceso de obtención de datos:	11
2.2.3. Relacionamiento de datos:	11
2.2.4. Diseño de informes en Power BI:	12
3. Marco Teórico	13
4. Metodología	17
4.1. Entendimiento de los datos y el negocio	19
4.2. Preparación de los datos	20
4.3. Modelado	20
4.4. Despliegue	21
5. Análisis de resultados	22
5.1. Entendimiento de los datos y el negocio	22
5.2. Preparación de los datos	24
5.3. Modelado	26
5.4. Despliegue	29
6. Conclusiones	32
Referencias	35

Lista de tablas

Tabla 1 Fases del proyecto	19
Tabla 2 Limite de filas en Excel vs filas generadas	23
Tabla 3 Resultados 2da fase	26

Lista de figuras

Figura 1 Ejemplo de esquema estrella.....	15
Figura 2 Componentes de Apache Hadoop.....	16
Figura 3 Fases de la metodología CRISP- DM.....	17
Figura 4 Diagrama de flujo para la actualización de los informes.....	22
Figura 5 Diagrama de flujo para la actualización de los informes luego de automatización.....	24
Figura 6 Cantidad de memoria RAM consumida por las consultas.....	25
Figura 7 Tablas de dimensiones.....	27
Figura 8 Filtros antiguos.....	27
Figura 9 Cantidad de filas generadas por las consultas.....	28
Figura 10 Esquema estrella aplicado en las diferentes tablas.....	28
Figura 11 Nuevo tablero de visualizaciones.....	30
Figura 12 Tablas dinámicas de informes antiguos.....	31

Siglas, acrónimos y abreviaturas

BI	Business intelligence
DAX	Data Analysis Expressions
ETL	Extract Load Transform
ODBC	Open Database Connectivity
SQL	Structured Query Language
HDFS	Hadoop Distributed File System
YARN	Yet Another Resource Negotiator
RDBMS	Relational Database Management System
CRISP-DM	Cross-Industry Standard Process for Data Mining

Resumen

En el sector financiero, la gestión de grandes volúmenes de datos representa un desafío crítico, especialmente para áreas como la de clientes, donde la información debe ser procesada y analizada de manera eficiente para facilitar la toma de decisiones estratégicas. Este proyecto tuvo como objetivo optimizar el proceso de generación, análisis y visualización de reportes estadísticos en el área de clientes de Bancolombia, utilizando herramientas avanzadas de procesamiento y visualización de datos como SQL, Power BI y Python.

El enfoque metodológico incluyó el análisis de las consultas SQL existentes, la reestructuración de estas para eliminar redundancias y mejorar la eficiencia, la automatización del proceso de obtención de datos mediante un script de Python y el diseño de un modelo de datos basado en el esquema estrella. Asimismo, se desarrolló un tablero interactivo en Power BI que centraliza la información previamente dividida en múltiples archivos de Excel, mejorando la accesibilidad y la precisión de los datos para los analistas.

Los resultados del proyecto evidencian una significativa reducción en los tiempos de procesamiento, una mayor confiabilidad en los datos reportados y una disminución de la carga de trabajo asociada a tareas manuales repetitivas. Además, se destacó el valor de la automatización y la inteligencia de negocios para potenciar la productividad organizacional y la toma de decisiones basada en datos.

Palabras clave: Inteligencia de negocio, Power BI, sector bancario, eficiencia operacional, automatización de procesos

Abstract

In the financial sector, managing large volumes of data poses a critical challenge, particularly in the customer's department, where information must be efficiently processed and analyzed to facilitate strategic decision-making. This project aimed to optimize the generation, analysis, and visualization of statistical reports in Bancolombia using advanced data processing and visualization tools such as SQL, Power BI and Python.

The methodological approach involved analyzing existing SQL queries, restructuring them to eliminate redundancies and improve efficiency, the automation of the ETL process through a Python script, and designing a data model based on the star schema. An interactive Power BI dashboard was also developed, centralizing information previously scattered across multiple Excel files, thereby enhancing data accessibility and accuracy for analysts.

The results demonstrate a significant reduction in processing times, increased reliability of reported data, and a decrease in workload associated with repetitive manual tasks. Additionally, the value of automation and business intelligence in boosting organizational productivity and enabling data-driven decision-making was highlighted.

Keywords: Business intelligence, Power BI, banking sector, Operational Efficiency, Process Automation

1. Introducción

En un contexto empresarial cada vez más competitivo y orientado al uso de datos, el sector financiero enfrenta el desafío de gestionar grandes volúmenes de información de manera eficiente (Giraldo Retuerto, Malaver Tuero, & Andrade-Arenas, 2023). Las instituciones bancarias generan millones de registros diariamente a partir de transacciones, actualizaciones de datos de los clientes y otros procesos operativos. Sin embargo, la capacidad de transformar estos datos en información útil y oportuna es lo que permite a las organizaciones tomar decisiones estratégicas basadas en evidencia, mejorar la experiencia del cliente y aumentar su competitividad.

En este contexto, la automatización de procesos y la implementación de herramientas avanzadas de inteligencia de negocios (BI) se presentan como soluciones clave para superar los retos asociados con la gestión de datos masivos. Estas tecnologías no solo facilitan un análisis eficiente y preciso, sino que también optimizan los recursos al reducir la dependencia de tareas manuales, que son propensas a errores.

Este proyecto, desarrollado en el área de clientes de una entidad bancaria, tuvo como objetivo principal optimizar el proceso de generación, análisis y visualización de reportes estadísticos, implementando un sistema más eficiente y confiable basado en Power BI, SQL y Python. El enfoque incluyó el análisis y la transformación de las consultas SQL utilizadas, el diseño de un modelo de datos acorde con los principios del esquema estrella y el desarrollo de un tablero interactivo que centralizara la información relevante para los analistas.

El nuevo enfoque no solo permitió centralizar la información previamente dividida en cinco archivos de Excel diferentes, sino que también mejoró la accesibilidad y comprensión de los datos por parte de los usuarios. Además, liberó al personal encargado de la actualización de información de tareas repetitivas, como la actualización manual de los datos mensualmente de cada uno de los archivos.

Este trabajo no solo evidencia la importancia de adoptar tecnologías modernas en la gestión de datos, sino que también subraya cómo la combinación de buenas prácticas de manejo de información y herramientas BI puede contribuir al crecimiento y a la eficiencia en el sector financiero.

2. Objetivos

A continuación, se presenta el objetivo general del proyecto, junto con los objetivos específicos formulados para alcanzarlo

2.1. Objetivo General

Automatizar el proceso de generación de reportes estadísticos mensuales del área de clientes del banco, mediante la integración de herramientas de análisis estadístico, visualización y procesamiento de datos como Power BI, SQL y Python, con el propósito de mejorar la eficiencia operativa, garantizar la precisión de los datos y facilitar el acceso a información confiable para el análisis histórico y la toma de decisiones estratégicas

2.2. Objetivos Específicos

Se formularon cuatro objetivos específicos para llevar a cabo el desarrollo del proyecto teniendo en cuenta la metodología guía escogida

2.2.1. Estudio de la documentación e información disponible:

Recolectar la información disponible sobre las cifras analizadas por el área, el almacenamiento de datos y la extracción de información, lo que permitirá una estructuración más eficiente de los nuevos métodos de obtención de datos, reduciendo los tiempos de creación y carga de la información

2.2.2. Automatización del proceso de obtención de datos:

Automatizar la generación de consultas SQL mediante el uso del lenguaje de programación Python, con el propósito de garantizar la actualización oportuna de las tablas base al mes concurrente. Esto permitirá integrar las bases de datos con Power BI de manera eficiente, asegurando la disponibilidad de datos precisos y actualizados para la creación del tablero que permita el análisis y la toma de decisiones estratégicas

2.2.3 Relacionamiento de datos:

Relacionar las tablas obtenidas con los diferentes filtros de segmentación utilizados en el área, mediante la herramienta Power BI, con el propósito de centralizar la información

proveniente de diversas tablas lo que permitirá la creación de un informe interactivo y ágil, optimizando el análisis y estudio de los datos

2.2.4 Diseño de informes en Power BI:

Diseñar las tablas y gráficos necesarios, con el fin de asegurar una visualización adecuada de los datos solicitados por los usuarios del área. De esta manera, se facilitará el acceso a la información de forma clara y efectiva, apoyando el análisis y la toma de decisiones estratégicas.

3. Marco Teórico

La creciente generación masiva de datos y la popularidad que ha alcanzado el análisis de datos, gracias a los resultados positivos obtenidos en compañías de alto desempeño, han llevado a las organizaciones a transformar sus métodos de toma de decisiones. Estas han pasado de un enfoque basado en la intuición a uno fundamentado en la evidencia obtenida a partir de los datos (Rodrigo , Hettiarachchi , & Kumarasinghe , 2024). En este contexto, la gestión eficiente de los datos, que permita a las organizaciones tomar decisiones informadas y generar valor en sus operaciones diarias, se ha convertido en un desafío estratégico (Giraldo Retuerto, Malaver Tuero, & Andrade-Arenas, 2023).

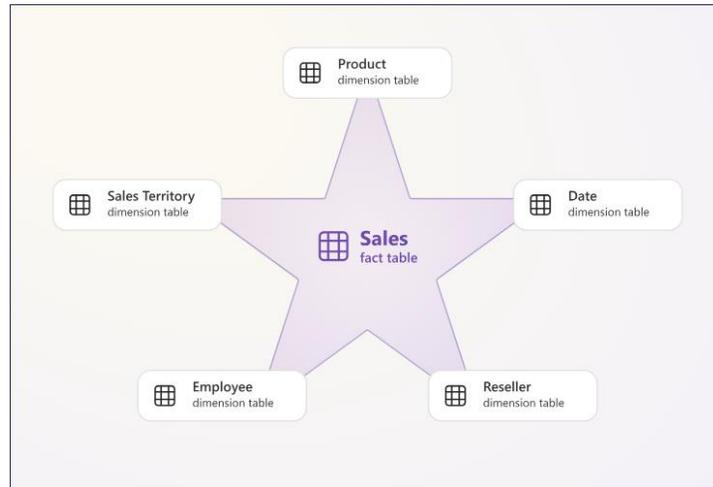
El sector financiero está profundamente vinculado al fenómeno del big data debido al acceso a múltiples fuentes de información. Transacciones diarias, actualizaciones de datos personales y muchas otras actividades generan miles de millones de registros diariamente, los cuales son procesados para extraer información valiosa basada en la actividad histórica de los clientes, con el propósito de respaldar una toma de decisiones más informada. Sin embargo, la diversidad y heterogeneidad de estas fuentes representan un desafío, ya que la pérdida de datos y la falta de confiabilidad se convierten en factores críticos para las empresas (Morshadul, Popp, & Oláh, 2020).

Considerando lo anterior, resulta evidente que la información del cliente desempeña un papel crucial en el sector financiero. Además, la adopción de técnicas avanzadas para su análisis, como la inteligencia de negocios (BI, por sus siglas en inglés), se torna indispensable para impulsar el crecimiento corporativo (Tavera Romero, Ortiz, Khalaf, & Ríos Prado, 2021). Tal como lo demuestran los autores Oluwatoyin et al. (2022), el uso de herramientas tecnológicas de inteligencia de negocios, como Power BI, es altamente eficiente para el análisis de grandes volúmenes de datos provenientes de clientes del sector bancario. Por otra parte, la integración de estas herramientas con la automatización de tareas repetitivas permite un incremento en la productividad organizacional, como se evidencia en los hallazgos de Bras, Pereira, Fonseca, Ribeiro, y Bianchi (2024).

Power BI es una suite de servicios de software, aplicaciones y conectores que trabajan en conjunto para transformar orígenes de datos no relacionados en información coherente. Los datos pueden cargarse desde fuentes locales, como hojas de Excel, o desde datos basados en la nube (Microsoft, 2024). Las conexiones a las bases de datos se establecen mediante herramientas como ODBC (Open Database Connectivity), una interfaz que permite a las aplicaciones de escritorio basadas en Windows conectarse a diferentes entornos de computación (Microsoft, 2024). Además, Power BI incorpora herramientas como las Expresiones de Análisis de Datos (DAX), que consisten en funciones, operadores y valores diseñados para realizar cálculos avanzados y consultas en tablas y columnas relacionadas dentro de modelos tabulares, optimizando así la creación de informes (Microsoft, 2024). Los modelos de datos en Power BI pueden estructurarse de diversas formas (Asteria, 2024), destacándose entre ellas el esquema estrella y el esquema de copo de nieve. De estos, el esquema estrella es el más recomendado para desarrollar modelos semánticos optimizados en Power BI, debido a su simplicidad y eficacia. El esquema estrella es un enfoque de modelado ampliamente utilizado en almacenes de datos relacionales, y su estructura se adapta perfectamente a los principios de Power BI. Para construir un esquema estrella, las tablas del modelo deben clasificarse como dimensiones o hechos. Las tablas de dimensiones contienen una columna clave que actúa como identificador único, junto con otras columnas que describen atributos específicos de cada clave (como nombre, latitud, valor, costo, etc.). Por otro lado, las tablas de hechos registran observaciones o eventos (transacciones, cantidades, temperaturas, entre otros), con columnas de

medidas numéricas y claves de dimensiones que se relacionan con las tablas de dimensiones (Microsoft, 2024). La *Figura 1* ilustra un ejemplo del esquema estrella.

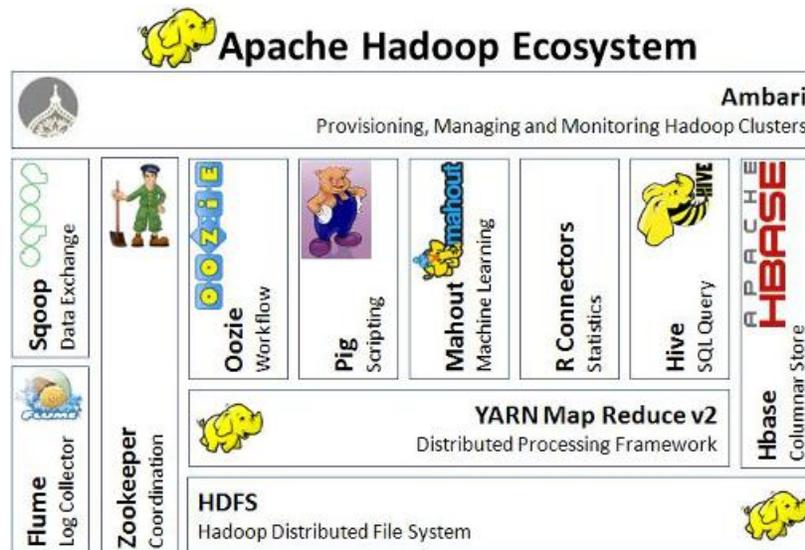
Figura 1 Ejemplo de esquema estrella



Fuente: Microsoft. (2024). Power BI. Recuperado el 12 de 2024, de Descripción de un esquema de estrella e importancia para Power BI: <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

Por otra parte, las organizaciones adoptan diversas formas de almacenamiento y organización de datos, que van desde bases de datos relacionales (RDBMS) hasta marcos de código abierto como Apache Hadoop. Este último es un marco de código abierto diseñado para almacenar y procesar grandes volúmenes de datos a través de diferentes módulos, como HDFS (Hadoop Distributed File System), YARN (Yet Another Resource Negotiator), MapReduce y Hadoop Common. Apache Hadoop también facilita la creación de clústeres de múltiples computadoras, lo que permite analizar grandes conjuntos de datos de manera paralela y eficiente, mejorando significativamente la velocidad de procesamiento (Amazon Web Services, 2024). La *Figura 2* muestra los componentes de un ecosistema Apache Hadoop.

Figura 2 Componentes de Apache Hadoop



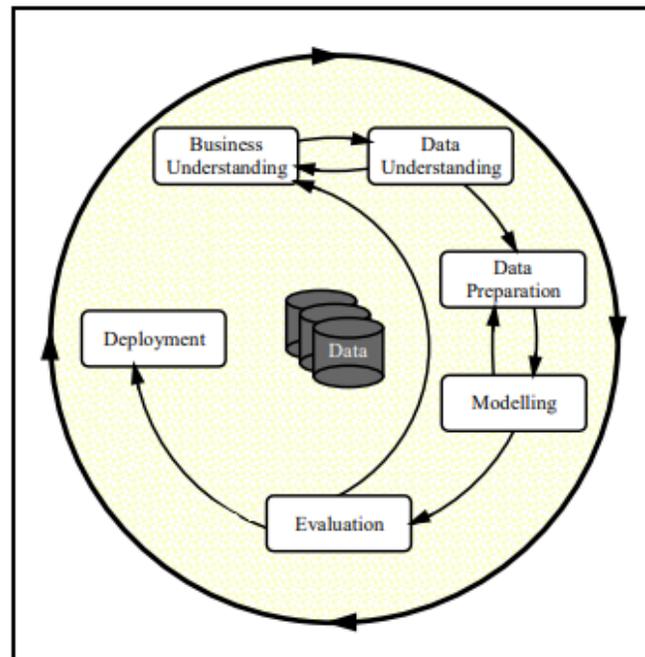
Fuente: Representación de ecosistemas Apache Hadoop. (s.f). Recuperado de Google. Recuperado el 12 de 2024, de <https://www.mikelnino.com/2015/04/ecosistema-hadoop-pila-tecnologica-apache-big-data.html>

Los avances tecnológicos proporcionan a las organizaciones herramientas para aumentar su productividad y eficiencia operativa. Sin embargo, la correcta implementación de estas herramientas para el desarrollo de productos que realmente generen valor representa un reto. Superar este desafío puede traducirse en un aumento de su competitividad y flexibilidad en un entorno cada vez más cambiante y competitivo.

4. Metodología

Para el presente proyecto se utilizó la metodología CRISP-DM (*Cross-Industry Standard Process for DataMining*) como guía principal. Esta metodología que se ha convertido en un estándar de facto en los procesos de minería de datos (Schröer, Kruse, & Gómez, 2021) consta de varias fases, representadas en la *Figura 3*. Las secuencias mostradas en la figura no son estrictas; más bien, representan las dependencias más importantes y frecuentes entre las fases. Sin embargo, para un proyecto específico, la decisión de cuál será la siguiente fase por ejecutar dependerá de los resultados obtenidos en cada etapa (Wirth & Hipp, 2000)

Figura 3 Fases de la metodología CRISP- DM



Fuente: Wirth, R., & Hipp, J. (2000). Towards a Standard Process Model for Data Mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 1, págs. 29-39.

A continuación, se presenta una breve descripción de cada fase:

- Entendimiento del negocio (Business Understanding): En esta fase se establecen los objetivos del proyecto y los requerimientos desde la perspectiva del negocio, para luego convertirlos en un problema de minería de datos y desarrollar un plan preliminar del proyecto.
- Entendimiento de los datos (Data Understanding): Esta fase comienza con una recolección inicial de los datos, para identificar posibles problemas de calidad de los datos y familiarizarse con ellos.
- Preparación de los datos (Data Preparation): En esta fase se realizan todas las actividades necesarias para construir el conjunto de datos final. Es probable que esta fase se repita varias veces durante el desarrollo del proyecto, ya que no siempre se realiza en un solo paso. Aquí se incluyen actividades como la limpieza de datos, la selección de atributos y las pruebas de validación de datos, las cuales tienen una alta probabilidad de necesitar ser ejecutadas más de una vez.
- Modelado (Modeling): En esta fase se selecciona la técnica de modelado adecuada. Todas las técnicas de minería de datos pueden ser utilizadas (técnicas de clasificación, agrupamiento, regresión, etc.).
- Evaluación (Evaluation): Para esta fase debe haber más de un modelo aplicado con alta calidad desde una perspectiva analítica. Antes de proceder al despliegue, es importante evaluar completamente el modelo para asegurarse de que cumpla con los objetivos del negocio.
- Despliegue (Deployment): Dependiendo de la complejidad del problema, esta fase puede implicar algo sencillo, como generar un reporte, o algo más complicado, como la implementación de un proceso de minería de datos repetitivo. En esta fase es crucial que tanto los usuarios como los desarrolladores tengan claro qué pasos seguir para realizar un adecuado despliegue de los modelos implementados. (Wirth & Hipp, 2000) & (Schröer, Kruse, & Gómez, 2021)

En este trabajo, las fases del proyecto se estructuraron y adaptaron de acuerdo con las necesidades específicas del mismo. La *Tabla 1* presenta una descripción de las actividades realizadas en cada fase, así como su relación con los objetivos específicos planteados para la ejecución del proyecto.

Tabla 1 Fases del proyecto

<i>Fase</i>	<i>Actividades</i>	<i>Objetivo específico</i>
Entendimiento de los datos y el negocio	Reunión con la asesora de Bancolombia para comprender las principales cifras.	Estudio de la documentación e información disponible
	Revisión de la documentación disponible del banco sobre el almacenamiento y la extracción de datos.	
	Estudio de los métodos actuales de obtención de información para identificar oportunidades de mejora.	
Preparación de los datos	Estructuración de nuevas consultas optimizadas para la alimentación del tablero de visualizaciones.	Automatización del proceso de obtención de datos
	Parametrización de las consultas SQL para la posterior automatización del proceso de obtención de datos mediante el lenguaje de programación Python.	
	Desarrollo del código para la obtención de datos históricos	
Modelado	Establecimiento de la conexión entre la fuente de datos y el tablero, asegurando una carga precisa y confiable.	Relacionamiento de datos
	Creación de tablas de dimensiones para facilitar la relación entre las tablas principales mediante columnas comunes.	
	Elaboración del esquema estrella de relaciones entre las tablas para garantizar una óptima integración de los datos.	
Despliegue	Entrevistas con los miembros del área para identificar las visualizaciones más importantes y los indicadores clave requeridos.	Diseño de informes en Power BI
	Creación del esquema del tablero, definiendo las visualizaciones y elementos interactivos	

4.1. Entendimiento de los datos y el negocio

En esta etapa, se llevaron a cabo reuniones con el asesor de prácticas asignado por Bancolombia para obtener una comprensión integral del proceso. Durante estas reuniones, se abordaron temas clave como el procedimiento para la actualización mensual de los informes, las herramientas utilizadas para la gestión y creación de estos, así como la información contenida en

cada informe y las principales cifras y siglas empleadas por el banco para identificar los rubros más relevantes.

Posteriormente, se revisó la documentación disponible sobre el ecosistema de almacenamiento de datos del banco y las mejores prácticas relacionadas con la obtención y manejo de datos. Esto permitió realizar un análisis detallado de los métodos actuales de obtención de información, identificando posibles problemas de calidad en las consultas SQL y las oportunidades de mejora correspondientes.

4.2. Preparación de los datos

Tras el análisis realizado sobre las consultas SQL utilizadas para la obtención de información, se implementaron mejoras basadas en buenas prácticas y en las oportunidades de optimización identificadas. Como parte de este proceso, se parametrizaron las variables mensuales en las consultas.

Una vez estructuradas las nuevas consultas optimizadas y parametrizadas, se procedió a desarrollar el script de Python necesario para integrarlas y automatizar el proceso de obtención de datos. Dado que el banco utiliza un ecosistema Hadoop para el almacenamiento y procesamiento de la información, se empleó el controlador ODBC (Open Database Connectivity) proporcionado por el banco para establecer la conexión entre Python y la base de datos.

4.3. Modelado

Se estableció la conexión entre el tablero de Power BI y la base de datos del banco mediante el controlador ODBC. Este paso permitió cargar al tablero las tablas generadas por el script de Python en la base de datos.

Con las seis tablas de hechos necesarias para la creación de los informes ya cargadas, se procedió a elaborar las tablas de dimensiones en formato Excel. Estas tablas facilitaron la implementación del esquema estrella en Power BI, lo que permitió optimizar las relaciones entre las tablas y garantizar una integración eficiente y precisa de los datos en los informes requeridos.

4.4. Despliegue

Se llevaron a cabo entrevistas con algunos miembros del área de clientes para recopilar sus opiniones y determinar qué gráficos y visualizaciones consideraban más relevantes para incluir en el tablero. Con base en esta retroalimentación, se procedió a utilizar herramientas avanzadas de Power BI, como fórmulas DAX (Data Analysis Expressions), marcadores y botones, para diseñar y personalizar los gráficos y tablas requeridos. Estas visualizaciones se organizaron en diferentes hojas del informe, asegurando una experiencia interactiva y funcional.

5. Análisis de resultados

En esta sección se presentan los resultados obtenidos en cada una de las fases del desarrollo del tablero. Este análisis permitirá comprender el impacto del nuevo enfoque implementado en la generación de informes y la visualización de la información, destacando las mejoras en eficiencia, precisión y accesibilidad que aporta a la organización.

5.1. Entendimiento de los datos y el negocio

Durante las reuniones realizadas, se identificó cómo se llevaba a cabo la ingestión de datos hacia los cinco archivos de Excel que contenían las tablas dinámicas con la información de los reportes estadísticos mensuales. Este proceso utilizaba seis consultas en Impala SQL, donde el procedimiento para actualizar manualmente las fechas de ingestión de las tablas fuente al mes en que era solicitado el reporte resultaba ser lento y propenso a errores. Además, estas consultas generaban nueve tablas en la base de datos, que debían descargarse y anexarse manualmente a los archivos de Excel cada mes. La *Figura 4* representa el proceso que debía seguirse con cada consulta para la actualización mensual de los informes.

Figura 4 Diagrama de flujo para la actualización de los informes



Fuente: Elaboración propia

Cuando la información anexada excedía el límite de filas permitido por Excel, era necesario eliminar datos históricos más antiguos o, en algunos casos, segmentaciones consideradas menos relevantes para los analistas del banco. Esto con el fin de ajustar el volumen de datos en ocasiones donde no era posible eliminar datos históricos. La *Tabla 2* muestra la cantidad de filas arrojadas por una de las consultas donde la información histórica es imprescindible, para este caso algunas segmentaciones fueron removidas. Además, debido al gran volumen de datos manejado en los archivos de Excel, los filtros aplicados a las tablas dinámicas se volvían extremadamente lentos, afectando la eficiencia en el análisis de la información.

El análisis de la documentación sobre la extracción y manejo de datos en el ecosistema Hadoop del banco reveló que las consultas SQL no seguían las buenas prácticas recomendadas. Estas consultas omitían pasos clave para optimizar el procesamiento de los datos o incluían información innecesaria en los cruces de tablas fuente, lo que aumentaba la complejidad y el tiempo de ejecución. Sin embargo, estas deficiencias en ocasiones eran justificadas por las limitaciones de Excel en cuanto a la cantidad máxima de filas permitidas por hoja.

Se identificaron redundancias, como consultas SQL que contenían información duplicada y eran utilizadas para alimentar diferentes tableros de Excel, así como segmentaciones obsoletas que ya no eran relevantes para el análisis. A partir de esta información y la documentación proporcionada por el banco, se lograron identificar mejoras significativas en la generación de datos. Se encontró la posibilidad de reducir el número de consultas de seis a cinco y el número de tablas generadas en la base de datos de nueve a seis. Estas tablas optimizadas alimentan directamente el tablero de Power BI sin omitir ningún tipo de información previamente analizada.

Filas admitidas en Excel por hoja	Filas arrojadas por la consulta de rentabilidad en Noviembre
1,048,576	1,306,036

Tabla 2 Limite de filas en Excel vs filas generadas

5.2. Preparación de los datos

Con base en los hallazgos identificados en la fase anterior, se procedió a la reestructuración de las consultas SQL. Este proceso incluyó la eliminación de filtros irrelevantes, la incorporación de filtros esenciales para los analistas que anteriormente se habían omitido debido a las limitaciones de filas en Excel, la implementación de buenas prácticas en la escritura de consultas, y en algunos casos, la reorganización de cómo se agrupaban los datos. Estos ajustes permitieron que las tablas generadas por las consultas se adaptaran de manera más eficiente al nuevo entorno de visualización en Power BI.

Adicionalmente, la parametrización implementada para las fechas de actualización mensual no solo eliminó la posibilidad de errores humanos, sino que también redujo drásticamente el tiempo necesario para ejecutar una consulta. El tiempo de ejecución pasó de un rango de 5 a 10 minutos a tan solo 1 segundo, lo que facilitó la conexión automática entre el código Python y la base de datos. La *Figura 5* ilustra el nuevo procedimiento para la actualización de los informes, mostrando cómo los pasos se redujeron a la mitad en comparación con el proceso anterior.

Figura 5 Diagrama de flujo para la actualización de los informes luego de automatización



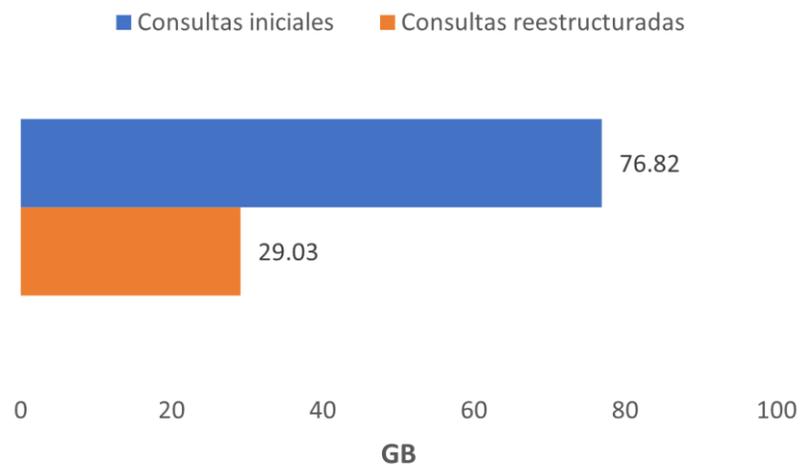
Fuente: Elaboración propia

Dado que algunas tablas fuente actualizan su información histórica mensualmente, el código fue diseñado para extraer datos correspondientes a todos los meses del año anterior al momento de ejecutar el script, así como desde enero del año actual hasta el mes solicitado. Aunque este enfoque incrementó el tiempo necesario para generar las tablas que alimentan los informes, ya que estas tablas se crean de forma temporal y se actualizan completamente cada mes, también mejoró significativamente la confiabilidad de los datos presentados. Este diseño asegura que, si en algún momento las tablas fuente son modificadas o reemplazadas, la información histórica se

actualice automáticamente para reflejar dichos cambios, siempre que las consultas SQL sean ajustadas adecuadamente.

Por último, la *Figura 6* compara la cantidad de memoria total consumida durante una ejecución mensual (para el mes de noviembre) entre las consultas anteriores y las reestructuradas. Aunque la memoria total consumida aumenta al ejecutar el script para múltiples meses, la nueva metodología distribuye mejor el consumo de memoria en los servidores del banco gracias a la obtención de información mes a mes, optimizando el rendimiento general del sistema.

Figura 6 Cantidad de memoria RAM consumida por las consultas



Fuente: Elaboración propia

La *Tabla 3* presenta un resumen de los resultados obtenidos en la fase de preparación de los datos

Tabla 3 Resultados 2da fase

Resultados
Eliminación de intervención humana en la actualización de las consultas SQL
Automatización de la generación de información y carga de datos
Aumento en la confiabilidad de la información
Reducción de los recursos consumidos por las consultas

5.3. Modelado

Una vez creadas las tablas necesarias para alimentar el archivo de Power BI, se estableció la conexión a la base de datos utilizando el controlador ODBC, permitiendo así la carga de los datos. El tiempo requerido para este proceso es muy variable, dependiendo del estado de los servidores del banco. En condiciones óptimas, la actualización de los datos podía completarse en aproximadamente 1 hora, mientras que, en momentos de alta concurrencia en los servidores, este tiempo podía extenderse a 4 a 5 horas. A pesar de esta variabilidad, el nuevo método representa una mejora significativa en comparación con el proceso antiguo, en el cual la actualización mensual de los informes para un solo mes podía tardar entre 5 y 8 horas debido a las limitaciones de Excel para manejar grandes volúmenes de datos, además de que requería la ocupación de una persona en todo momento para realizar el proceso. El nuevo proceso permite actualizar toda la información histórica en todos los informes sin ningún tipo de intervención humana durante el proceso, mejorando tanto la eficiencia como la integridad de los datos.

Con base en la reestructuración de las consultas SQL, donde todas las categorías fueron reemplazadas por números o claves para reducir el peso de las consultas, se procedió a la creación de tablas de dimensiones. Estas tablas permiten una integración más eficiente de los datos entre las tablas de hechos y facilitan la normalización de los nombres de las categorías. La *Figura 7* Tablas de dimensiones y la *Figura 8* ilustran esta transformación, mostrando una comparación entre las

nuevas tablas de dimensiones con valores normalizados y un ejemplo de cómo se presentaban los filtros en los informes anteriores. Este enfoque no solo mejora el entendimiento de los datos, sino que también aporta claridad y eficiencia a la visualización de la información y las segmentaciones.

Figura 7 Tablas de dimensiones

Ing Rec	
cod	Ing rec
1	Nomina
2	Pensionado
-1	No Aplica

vin des	
cod	tipo
0	Desertores
1	Vinculados
-1	No Aplica

Genero	
cod	genero
1	Hombre
0	Mujer
-1	No registra

Fiel	
cod	fiel
1	Fiel
0	No Fiel

Fuente: Elaboración propia

Figura 8 Filtros antiguos

nomina	nomina
0	0
1	1
	(en blanco)

pensionados	Pensionados
0	PENSIONADO
1	(en blanco)

- (Todas)
- Pensionado
- Nómina
- (en blanco)

Fuente: Elaboración propia

Posteriormente, las tablas de dimensiones fueron integradas al informe de Power BI, estructurando el esquema estrella junto con las tablas de hechos generadas a partir de las consultas SQL previamente cargadas. La Figura 10 muestra el esquema estrella resultante con las diferentes tablas. Este diseño, además de simplificar y facilitar el entendimiento del modelo de datos, optimiza la velocidad de las consultas sobre grandes volúmenes de información, como los más de 100 millones de filas cargadas en el informe (Figura 9).

Figura 9 Cantidad de filas generadas por las consultas



Fuente: Elaboración propia

Figura 10 Esquema estrella aplicado en las diferentes tablas



Fuente: Elaboración propia

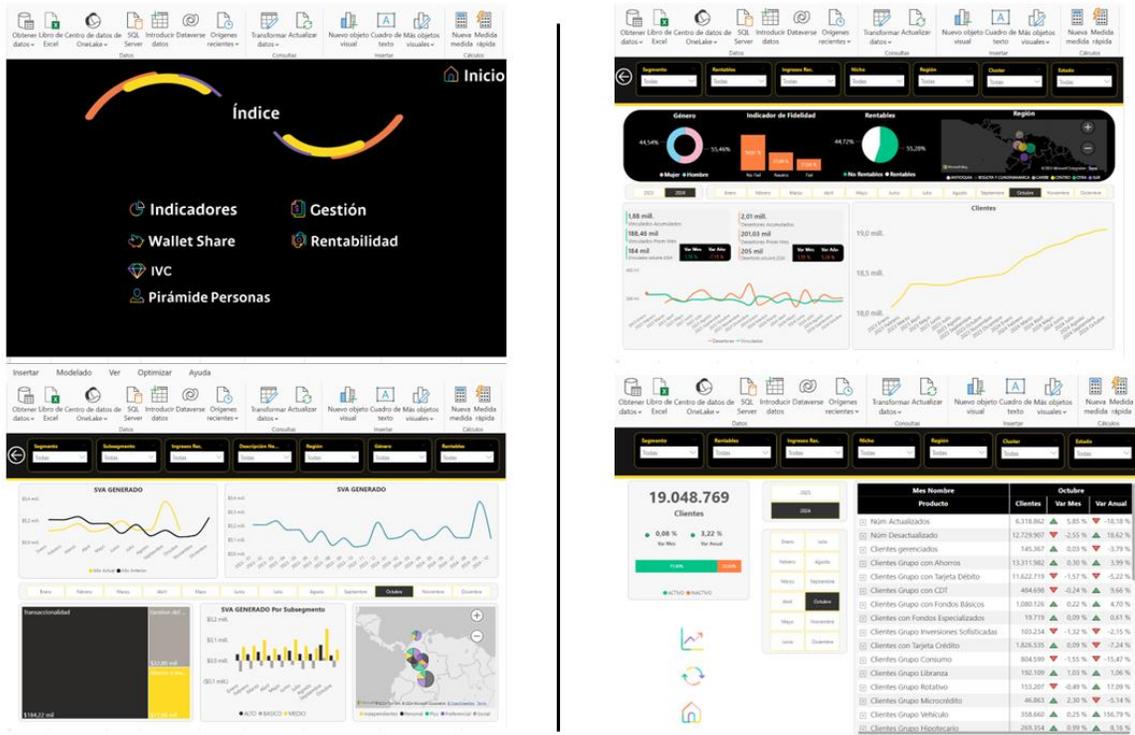
5.4. Despliegue

En la etapa de despliegue, el primer paso consistió en recolectar información sobre las segmentaciones más importantes para los usuarios finales. A través de entrevistas con diferentes miembros del área, se identificaron las principales métricas y gráficos que debían mostrarse en el tablero. Este enfoque colaborativo permitió diseñar un tablero adaptado a las necesidades específicas de los usuarios, optimizando su utilidad y relevancia.

El diseño del tablero utilizó fórmulas DAX de inteligencia de tiempo para facilitar la comparación rápida entre valores actuales y pasados. Además, se incluyeron marcadores y botones interactivos que permiten a los usuarios navegar de manera ágil entre las diferentes páginas y cifras de los informes. Este nuevo enfoque logró centralizar la información, reduciendo significativamente el tiempo dedicado a la búsqueda y filtrado de datos. Mientras que anteriormente un usuario tardaba entre 7 y 15 minutos en abrir los diferentes archivos de Excel y aplicar los filtros necesarios, ahora este tiempo se redujo a un promedio de 2.5 minutos al navegar por las diferentes páginas del tablero y aplicar los filtros correspondientes.

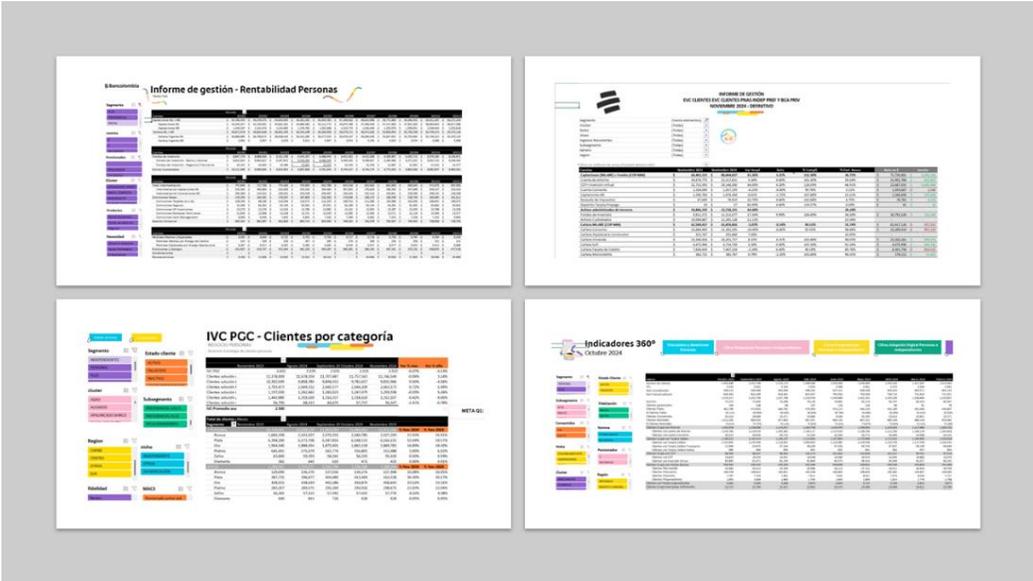
Adicionalmente, el tablero permite la visualización inmediata de cifras que anteriormente solo podían obtenerse aplicando múltiples filtros, mejorando la eficiencia y la experiencia de usuario. La *Figura 11* presenta algunas de las páginas del tablero de visualización, evidenciando una clara mejora en la accesibilidad y visualización de los datos en comparación con los informes antiguos mostrados en la *Figura 12*.

Figura 11 Nuevo tablero de visualizaciones



Fuente: Elaboración propia`

Figura 12 Tablas dinámicas de informes antiguos



Fuente: Elaboración propia`

6. Conclusiones

- Como parte del entendimiento de los datos y del proceso actual de generación de reportes, se analizaron las consultas SQL empleadas para alimentar los cinco archivos de Excel con tablas dinámicas que presentaban la información histórica y mensual. Durante esta etapa, se identificaron diversas deficiencias, como la generación de información duplicada, la falta de optimización en la estructura de las consultas y el elevado nivel de intervención humana requerido para elaborar los reportes.

Los resultados evidenciaron un proceso lento y propenso a errores, con un consumo elevado tanto de tiempo como de recursos. Sin embargo, el análisis exhaustivo de las consultas SQL, las tablas fuente asociadas y la documentación del banco sobre el manejo de datos permitió reestructurar dichas consultas. Esto incluyó la eliminación de segmentaciones irrelevantes y la reducción del número de consultas de nueve a cinco, logrando disminuir el consumo de recursos sin comprometer la integridad de la información.

En conclusión, el análisis continuo de los métodos de obtención de datos y la implementación de buenas prácticas en su gestión son esenciales para mejorar la eficiencia en la extracción y análisis de información que en un contexto como el del banco, donde se manejan volúmenes de datos del orden de millones, son mejoras que resultan fundamentales para optimizar su operación diaria.

- En el proceso de preparación de los datos, las consultas SQL fueron adaptadas para generar tablas más ligeras en términos de información contenida. Esto incluyó la sustitución de categorías por identificadores numéricos, la aplicación de buenas prácticas documentadas por el banco, y la eliminación de filtros desactualizados e irrelevantes para los analistas. Adicionalmente, se desarrolló un script en Python que permitió parametrizar y automatizar la ejecución de las consultas, facilitando la actualización de la información histórica. Esto no solo eliminó el riesgo de errores humanos, sino que también redujo en su totalidad el tiempo requerido para ajustar los parámetros mensuales de las consultas.

Es posible concluir que la integración de herramientas modernas como Python con las capacidades de manejo de datos del banco demuestra ser una estrategia pertinente para optimizar procesos repetitivos. Esto no solo incrementa la productividad de los colaboradores, sino que también mejora la precisión y la confiabilidad de la información generada

- En la etapa de modelado de la información, se elaboraron tablas de dimensiones que correspondían a los identificadores numéricos de las categorías implementados en las consultas SQL. Esto permitió posteriormente la creación del esquema estrella en Power BI, integrando las tablas de hechos o principales generadas a partir de dichas consultas. Este enfoque facilitó la simplificación y normalización de los nombres de las categorías en todas las tablas principales, además de optimizar los tiempos de consulta sobre el gran volumen de datos cargados en el tablero.

En conclusión, es evidente que, para las empresas del sector bancario, donde se manejan grandes volúmenes de datos con múltiples segmentaciones, resulta imprescindible adoptar nuevas estrategias de integración y manejo de información si se quiere garantizar un análisis profundo tanto de sus clientes como de sus finanzas.

- A través de las entrevistas realizadas para estructurar el diseño del tablero, se recopiló la información necesaria para definir los gráficos y las principales métricas a mostrar. Esto permitió desarrollar un tablero interactivo y funcional adaptado a las necesidades de los analistas del área. El tablero de Power BI, además de centralizar la información de todos los informes, empleó fórmulas DAX de inteligencia de tiempo para facilitar la rápida visualización de cifras históricas y su comparación con las actuales de manera gráfica, mejorando significativamente la experiencia de los usuarios y el acceso a los datos.

En conclusión, el aprovechamiento de herramientas de visualización como Power BI, en combinación con el acceso a grandes volúmenes de datos, representa una herramienta sumamente valiosa para aumentar la agilidad en el acceso a la información y apoyar procesos efectivos de toma de decisiones.

- El desarrollo de este proyecto permitió optimizar significativamente el proceso de generación, análisis y visualización de reportes estadísticos en el área de clientes del banco, alcanzando así el objetivo general de implementar un sistema más eficiente y confiable para gestionar grandes volúmenes de datos. No obstante, se identificaron limitaciones relacionadas con el acceso a ciertas fuentes de datos que, de haber estado disponibles, habrían posibilitado la integración de información adicional relevante para los analistas. Con más tiempo para gestionar el acceso a estas fuentes, el alcance del tablero podría haberse ampliado aún más.

El tablero fue diseñado dentro de un plazo de tiempo limitado, lo que lo convierte en un producto funcional que los analistas y programadores del área pueden seguir mejorando continuamente.

En conclusión, la modernización del proceso de generación de reportes, la centralización de la información y la adopción de herramientas tecnológicas avanzadas resultaron fundamentales para mejorar la productividad y confiabilidad en la gestión de datos en un entorno bancario, caracterizado por grandes volúmenes de información y altas demandas analíticas. Este enfoque no solo satisface las necesidades actuales, sino que también establece un modelo replicable en otros proyectos del sector, demostrando cómo las buenas prácticas en el manejo de datos y el uso de herramientas innovadoras pueden maximizar el impacto operativo y estratégico de la organización.

Referencias

- Amazon Web Services. (2024). *¿Qué es Hadoop?* Recuperado el 12 de 2024, de <https://aws.amazon.com/es/what-is/hadoop/>
- Astera. (Marzo de 2024). *Esquema de estrella vs. Esquema de copo de nieve: 4 diferencias clave.* Recuperado el 12 de 2024, de <https://www.astera.com/es/type/blog/star-schema-vs-snowflake-schema/#>
- Bras, J. C., Pereira, R. F., Fonseca, M., Ribeiro, R., & Bianchi, I. S. (2024). Advances in auditing and business continuity: A study in financial companies. *Journal of Open Innovation: Technology, Market,, 10*(2). Obtenido de https://www.sciencedirect.com/science/article/pii/S2199853124000982?ref=pdf_download&fr=RR-2&rr=8f9f7c3e9d4a0a3e
- Giraldo Retuerto, M., Malaver Tuero, B., & Andrade-Arenas, L. (2023). Business Intelligence Implementation using Power BI for Decision Making in Peruvian Banking Systems. *International Journal of Engineering Trends and Technology, 97-108.* Obtenido de https://www.researchgate.net/publication/370620812_Business_Intelligence_Implementation_using_Power_BI_for_Decision_Making_in_Peruvian_Banking_Systems
- Microsoft. (2024). *C++*. Recuperado el 12 de 2024, de Conceptos básicos de ODBC: <https://learn.microsoft.com/es-es/cpp/data/odbc/odbc-basics?view=msvc-170>
- Microsoft. (2024). *DAX*. Recuperado el 12 de 2024, de Información general sobre DAX: <https://learn.microsoft.com/es-es/dax/dax-overview>
- Microsoft. (2024). *Power BI*. Recuperado el 12 de 2024, de ¿Qué es Power BI?: <https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>
- Microsoft. (2024). *Power BI*. Recuperado el 12 de 2024, de Descripción de un esquema de estrella e importancia para Power BI: <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

- Morshadul, H. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data. *Journal of Big Data*, 7(1). Obtenido de <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00291-z>
- Oluwatoyin, A. M., Misra, S., Wejin, J., Gautam, A., Behera, R. K., & Ahuja, R. (2022). Customer Churn Prediction in Banking Industry Using Power Bi. *Proceedings of Third International Conference on Computing, Communications, and Cyber-Security*, (págs. 767-774). Obtenido de https://link.springer.com/chapter/10.1007/978-981-19-1142-2_60
- Representación de ecosistemas Apache Hadoop. (s.f.). *Recuperado de Google*. Recuperado el 12 de 2024, de <https://www.mikelnino.com/2015/04/ecosistema-hadoop-pila-tecnologica-apache-big-data.html>
- Rodrigo , A., Hettiarachchi , H., & Kumarasinghe , D. (January de 2024). An Exploration on Utilization of Business Analytics in Sri Lankan Banks. *Journal of Business and Technology*, 89-99. Obtenido de <https://jbt.sljol.info/articles/10.4038/jbt.v8i1.114>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526-534. Obtenido de <https://www.sciencedirect.com/science/article/pii/S1877050921002416>
- Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Ríos Prado, A. (2021). Web Application Commercial Design for Financial Entities. *Tech Science Press*, 67(3). Obtenido de <https://www.sciencedirect.com/org/science/article/pii/S1526149220002040>
- Wirth, R., & Hipp, J. (2000). Towards a Standard Process Model for Data Mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 1, págs. 29-39. Obtenido de <https://www.semanticscholar.org/paper/CRISP-DM%3A-Towards-a-Standard-Process-Model-for-Data-Wirth-Hipp/48b9293cfd4297f855867ca278f7069abc6a9c24>