



Westfield
Business
School



Maestría en

ADMINISTRACIÓN DE EMPRESAS

**Tesis previa a la obtención de título de
Magister en Administración de Empresas**

AUTORES:

Daniel Valentín Ferez Martínez

Pedro Miguel Perugachi Isizan

Jorge Andres Pillco Duchi

Marcelo Bladimir Jácome Castañeda

TUTOR: Ing. Víctor David Arrellano Peñafiel

Consumer Lens

**PROYECTO DE TITULACIÓN – FIN DE MÁSTER
MAESTRÍA EN ADMINISTRACIÓN DE EMPRESA ENL**

Título del trabajo de titulación

Por

Nombres completos del autor(a)

Daniel Valentín Ferez Martínez

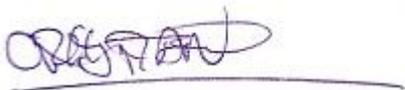
Pedro Miguel Perugachi Isizan

Jorge Andres Pillco Duchi

Marcelo Bladimir Jácome Castañeda

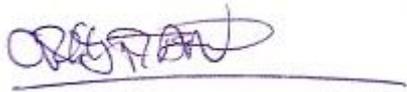
Febrero 2025

Aprobado



Cristian Melo
Presidente(a) del Tribunal
Universidad Internacional del Ecuador

Yo, Cristian Javier Melo González e Ignacio Maroto, declaramos que, personalmente conocemos que los graduandos: Daniel Valentín Ferez Martínez, Pedro Miguel Perugachi Isizan, Jorge Andres Pillco Duchi, Marcelo Bladimir Jácome Castañeda, son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Cristian Melo
Coordinador MBA UIDE



Ignacio Maroto
Provost WBS

Autoría del Trabajo de Titulación

Nosotros, Daniel Valentín Ferez Martínez, Pedro Miguel Perugachi Isizan, Jorge Andres Pillco Duchi, Marcelo Bladimir Jácome Castañeda, declaramos bajo juramento que el trabajo de titulación titulado **Consumer Lens** es de nuestra autoría y exclusiva responsabilidad legal y académica; que no ha sido presentado anteriormente para ningún grado o calificación profesional, habiéndose citado las fuentes correspondientes y respetando las disposiciones legales que protegen los derechos de autor vigentes.

Daniel Valentín Ferez Martínez

Correo electrónico: daniel.ferez@yahoo.com

Pedro Miguel Perugachi Isizan

Correo electrónico: pmperugachi@gmail.com

Jorge Andres Pillco Duchi

Correo electrónico: georgepillco@outlook.es

Marcelo Bladimir Jácome Castañeda

Correo electrónico:

Autorización de Derechos de Propiedad Intelectual

Nosotros, Daniel Valentín Ferez Martínez, Pedro Miguel Perugachi Isizan, Jorge Andres Pillco Duchi, Marcelo Bladimir Jácome Castañeda, en calidad de autores del trabajo de investigación titulado Título del trabajo de investigación Consumer Lens, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que me pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autor me corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, Febrero 2025



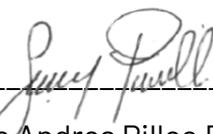
Daniel Valentín Ferez Martínez

Correo electrónico: daniel.ferez@yahoo.com



Pedro Miguel Perugachi Isizan

Correo electrónico: pmperugachi@gmail.com



Jorge Andres Pillco Duchi

Correo electrónico: georgepillco@outlook.es



Marcelo Bladimir Jácome Castañeda

Correo electrónico:

Dedicatorias y Agradecimientos

"A mi Mama que es el ejemplo de fortaleza y crecimiento personal siendo un referente para todos, a mi pareja que me apoya incondicionalmente y es mi ejemplo de tenacidad, a mi hija Malena que hace que todo tenga sentido y es la esperanza de salir adelante y buscar un mejor futuro constantemente.

Daniel Valentín Férrez Martínez

"A mis padres, por su amor incondicional, su apoyo constante y sus palabras de aliento en los momentos más difíciles. A mis hermanos, por ser mi inspiración y recordarme siempre que el esfuerzo vale la pena. Esta tesis es un reflejo del esfuerzo conjunto y del amor que me han brindado."

Pedro Miguel Perugachi Isizan

INDICE DEL DOCUMENTO

Contenido

| | |
|---|----|
| PARTE I | 12 |
| CAPITULO 1: CONCEPTOS BÁSICOS DE BUSINESS INTELLIGENCE | 12 |
| 1.1 Definición del proyecto..... | 12 |
| 1.2 Definición de la empresa o negocio..... | 12 |
| 1.3 Procesos, organización y cifras del negocio..... | 13 |
| 1.4 Vinculación con la empresa, fuentes de información y nivel de conocimiento..... | 14 |
| 1.5 Indicadores KPIs | 14 |
| CAPITULO 2: ALMACENAMIENTO DE LOS DATOS Y EL ECOSISTEMA DE DW | 16 |
| 2.1 Origen de la información..... | 16 |
| 2.2 Sistemas y destino de Carga..... | 16 |
| 2.3 Almacenamientos de datos warehouse y datamarts..... | 17 |
| CAPITULO 3: PROCESOS ETL (Extract, Transform, Load) | 20 |
| 3.1 Proceso de Carga de Datos | 20 |
| 3.2 Data Lake..... | 21 |
| 3.3 Información y calidad de la información..... | 21 |
| 3.4 Horarios de Carga y accesibilidad..... | 22 |
| 3.5 Variables disponibles y acceso a la data warehouse:..... | 23 |
| 3.6 Sistemas implicados en los procesos..... | 23 |
| CAPITULO 4: TECNOLOGÍAS PARA EL ANÁLISIS Y DISEÑO DE VISUALIZACIONES DE BI | 24 |
| 4.1 Análisis de herramientas Big Data | 24 |
| 4.2 Google Big Query alimentación de datos:..... | 25 |
| 4.3 Seleccionar los usuarios y perfiles de cada dashboard..... | 26 |
| 4.3.1 Creación de Roles y Perfiles de Usuario | 26 |
| 4.3.2. Asignación de Usuarios a Roles | 26 |
| 4.3.3 Acceso a los Dashboards en Power BI Service | 26 |
| 4.4 Tipos de usuarios: | 27 |
| 4.5 Perfiles de cada dashboard..... | 28 |
| 4.5.1 Perfil de Dashboard para Gerente | 28 |
| 4.5.2 Perfil de Dashboard para Supervisor..... | 29 |

| | | |
|--|---|----|
| 4.5.3 | Perfil de Dashboard para Sistemas (Departamento de TI) | 30 |
| 4.5.4 | Perfil de Dashboard para Cliente | 30 |
| 4.6 | Implementación en Power BI | 31 |
| 4.7 | Necesidades de información y accesos por departamento y perfil. | 32 |
| 4.8 | Alineación con los Objetivos Iniciales | 33 |
| CAPITULO 5: CREACIÓN DEL ÁREA DE BI | | 36 |
| 5.1 | Situación actual del departamento de BI. | 36 |
| 5.2 | Definición en las carencias en formación y recursos. | 37 |
| 5.3 | Coordinación de los flujos entre otras áreas. | 38 |
| 5.4 | Organigrama | 39 |
| 5.5 | Dependencia de Otras Áreas | 40 |
| 5.6 | Solución de Problemas | 41 |
| 5.7 | Recursos externos Destinados: Subcontratación | 42 |
| 5.7.1 | Software y herramientas de BI | 42 |
| 5.7.2 | Consultoría especializada | 43 |
| 5.7.3 | Capacitación y certificaciones | 43 |
| 5.7.4 | Infraestructura tecnológica | 44 |
| 5.7.5 | Reclutamiento de talento | 44 |
| 5.7.6 | Libros y recursos adicionales | 44 |
| 5.8 | Business Case de costes y beneficios | 45 |
| 5.8.1 | Objetivos del Proyecto | 45 |
| 5.8.2 | Costos Estimados | 45 |
| 5.8.3 | Beneficios Esperados | 46 |
| 5.8.4 | Análisis de Retorno de Inversión (ROI) | 46 |
| 5.8.5 | Plan de Implementación | 47 |
| PARTE II | | 48 |
| CAPITULO 1: BIG DATA | | 48 |
| 1.1 | Fuentes de información obtenidas (Estructuradas internas): | 48 |
| 1.2 | Fuentes de Información (Semiestructuradas internas) | 50 |
| 1.3 | Fuentes de información (Semiestructurada externa): | 50 |
| 1.4 | Objetivos planteados | 51 |
| 1.5 | Casos para incorporar modelo de BI: | 52 |
| 1.6 | Impacto de Big Data | 53 |
| CAPITULO 2: ARQUITECTURA DE DATOS EN ENTORNOS CLOUD: SAAS | | 54 |

| | | |
|--|--|-----------|
| 2.1 | Alternativas de cloud computing | 54 |
| 2.2 | IaaS Infraestructure as a Service | 54 |
| 2.3 | PaaS Platform as a service | 55 |
| 2.4 | SaaS Software as a Service | 55 |
| 2.5 | Selección del cloud computing | 55 |
| CAPITULO 3: PRINCIPALES TECNOLOGÍAS OPEN SOURCE DEL UNIVERSO BIG DATA | | 56 |
| 3.1 | Tipos de Frameworks Open Source | 57 |
| 3.2 | Framework ideal para Consumer Lens | 58 |
| CAPITULO 4: MACHINE LEARNING E INTELIGENICA ARTIFICIAL | | 60 |
| 4.1 | Modelo de aplicación de data Science y machine learning | 60 |
| 4.2 | Fuentes de Datos: | 61 |
| 4.3 | Área de aplicación y departamentos de negocios implicados: | 63 |
| 4.4 | Identificación de la mejora esperada: | 64 |
| 4.5 | Cuantificación de la mejora en el área de aplicación: | 64 |
| CAPITULO 5: ANÁLISIS EN TIEMPO REAL. FORECAST Y DATA STREAM ... | | 66 |
| 5.1 | Aplicación del modelo de aplicación de Real time | 66 |
| 5.2 | Fuentes de datos utilizados | 66 |
| 5.3 | Área de aplicación y departamentos de negocio implicados | 67 |
| 5.4 | Identificación de la mejora esperada | 68 |
| 5.5 | Cuantificación de la mejora en el área de aplicación a través del business case | 69 |
| 5.5.1 | Resumen Ejecutivo | 69 |
| 5.5.2 | Situación Actual y Oportunidad de Mejora | 70 |
| 5.5.3 | Beneficios Esperados (Cuantificación de la Mejora) | 70 |
| 5.5.4 | Costos del Proyecto | 70 |
| 5.5.5 | Retorno de Inversión (ROI) | 71 |
| 5.5.6 | Conclusión y Recomendación | 71 |

INDICE DE FIGURAS Y TABLAS

Espacio para el índice de figuras y tablas

| | |
|--|----|
| Ilustración 1. Logotipo de Consumer lens | 12 |
| Ilustración 2 Captación de clientes | 13 |
| Ilustración 3 KPIS Financieros | 14 |
| Ilustración 4 KPIS Comercial y Mercadeo | 15 |
| Ilustración 5 KPIS Operaciones | 15 |
| Ilustración 6 Extracción, Carga y Transformación..... | 17 |
| Ilustración 7 Almacenamiento de datos Warehouse..... | 17 |
| Ilustración 8 Tabla ETL vs Data Lake | 21 |
| Ilustración 9 Google Big Query | 25 |
| Ilustración 10 Red de accesos a Dashboard | 27 |
| Ilustración 11 Tabla de resumen de perfiles | 32 |
| Ilustración 12 Tabla de perfiles utilizados..... | 33 |
| Ilustración 13 Datawarehouse y Procesamiento de Datos | 34 |
| Ilustración 14 Esquema de objetivos del proyecto..... | 35 |
| Ilustración 15 Situación actual del departamento de BI | 37 |
| Ilustración 16 Diagrama de carencia y formación de recursos | 38 |
| Ilustración 17 Esquema del departamento de implementación | 40 |
| Ilustración 18 Esquema del proceso basado en datos..... | 41 |
| Ilustración 19 Procedimiento de gestión de datos..... | 42 |
| Ilustración 20: Ejemplo de cómo se obtiene la información en base a filtros, y agrupaciones propias de la plataforma | 48 |
| Ilustración 21 Módulo de contactos, detalla información del cliente y su histórico de compras | 49 |
| Ilustración 22 Ejemplo de cómo obtenemos la información detallada de que producto compro en pedidos y así como el histórico de valor de compra | 49 |
| Ilustración 23 Ejemplo de cómo podemos descargar la factura y ver que compro el cliente | 50 |
| Ilustración 25 Fuentes de datos internas..... | 61 |
| Ilustración 26 Ejemplo de factura para análisis | 62 |
| Ilustración 27 Obtención de información de fuentes internas | 62 |
| Ilustración 28 Información de obtención de fuentes de datos internas..... | 63 |

INTRODUCCIÓN Y OBJETO DEL PROYECTO

Implementar un sistema de Business Intelligence (BI) y Big Data que posicione a los clientes de la empresa primero con la implementación de un Datawarehouse para luego tener un análisis y conocimiento en detalle del cliente sus gustos y preferencias para aumentar la venta por medio de campañas publicitarias, la idea se complementa al aumentar el movimiento de las ventas por canales electrónicos facilitando el proceso de compra para el cliente aumentando la venta en línea para la empresa.

Consumer Lens se enfoca en crear un Data Warehouse específico para cada cliente, el objetivo de este Data Warehouse va a ser alimentado por la información de los consumidores propios del cliente siendo clasificados por sus características de compra en base a la preferencia de cada uno de sus productos.

Parte fundamental de cada proceso es la correcta ejecución y desarrollo de los procesos ELT (Extract, Load, Transform) o ECT (Extraer, Cargar, Transformar), para ello es muy importante la correcta comunicación entre el cliente y Consumer lens, el cliente de la misma forma tiene que ser muy específico en hasta donde Consumer Lens puede obtener información y que tipo de información consideran confidencial por lo que no se debe usar o manipular por parte de Consumer Lens.

PARTE I

CAPITULO 1: CONCEPTOS BÁSICOS DE BUSINESS INTELLIGENCE

1.1 Definición del proyecto

- **Nombre del proyecto**

Consumer Lens

- **Objetivo del proyecto**

Implementar un sistema de Business Intelligence (BI) y Big Data que posicione a los clientes de la empresa primero la implementación de un Datawarehouse para luego tener un análisis y conocimiento en detalle del cliente sus gustos y preferencias para aumentar la venta por medio de campañas publicitarias, la idea se complementa al aumentar el movimiento de las ventas por canales electrónicos facilitando el proceso de compra para el cliente aumentando la venta en línea para la empresa.

1.2 Definición de la empresa o negocio

- **Nombre de la empresa**



ILUSTRACIÓN 1. LOGOTIPO DE CONSUMER LENS

- **Misión y visión de la empresa**

Misión: Ser líderes en soluciones tecnológicas brindando herramientas que permiten al cliente personalizar su servicio, conociendo más al cliente y mejorando considerablemente su valor como marca.

Visión: Consolidarnos como referentes globales en tecnologías de análisis de datos y almacenamiento de información para brindar las mejores soluciones al momento de toma de decisiones, así como ayudar a empresas a personalizar y generar un mayor engagement entre nuestros clientes y sus clientes.

1.3 Procesos, organización y cifras del negocio

Procesos

- Captación de información de los clientes
- Almacenamiento de la información y correcta clasificación en base a los requerimientos del cliente
- Elaboración de campañas enfocadas en diferentes tipos de clientes con mismas características de consumo, en base a promociones planteadas por el cliente.



ILUSTRACIÓN 2 CAPTACIÓN DE CLIENTES

Organización

Estructura con una junta directiva que define la estrategia corporativa. El equipo de dirección ejecuta y monitorea el desarrollo de nuevas líneas de negocio, apoyado por equipos de proyectos y especialistas en tecnología. El proyecto se integrará a la organización a través del departamento de Implementación.

Cifras del negocio

- Facturación anual superior a \$2 millones.
- Capacidad de adaptar a otros tipos de negocio.
- Gran proyección de clientes.

1.4 Vinculación con la empresa, fuentes de información y nivel de conocimiento

- **Vinculación con la empresa**

Profesional con conocimiento directo de la empresa como parte del equipo de planificación estratégica y desarrollo de nuevos proyectos.

- **Fuentes de información y nivel de conocimiento**

Acceso a bases de datos de ventas de la empresa, estudios de mercado y análisis de rendimiento de soluciones. Nivel de conocimiento: 8/10, acceso a información clave, pero limitado a ciertos datos confidenciales.

1.5 Indicadores KPIs

Para identificar los KPI relevantes en este proyecto de expansión de Consumer Lens hacia el Sector de ventas masivas, es crucial entender que estos indicadores deben alinearse con los objetivos estratégicos de la empresa, medir el impacto de las acciones implementadas y proporcionar una base sólida para la toma de decisiones. Los KPI seleccionados abarcan aspectos financieros, comerciales, operativos, y permiten evaluar el rendimiento de la estrategia de BI y Big Data. La correcta definición y seguimiento de estos indicadores ayudará a identificar oportunidades de mejora, optimizar recursos y asegurar que el proyecto cumpla con las expectativas de crecimiento y sostenibilidad de la empresa.

KPI Financieros:

- Margen de beneficio neto (%)
- Crecimiento de ingresos trimestral
- Retorno sobre la inversión (ROI) en nuevos proyectos o campañas publicitarias
- Costos operativos totales en relación con el ingreso



ILUSTRACIÓN 3 KPIS FINANCIEROS

KPI Comercial y Mercadeo:

- Tasa de conversión de leads en clientes ventas en línea.
- Retención de clientes.
- Alcance de campañas de marketing digital de las campañas focalizadas
- Número de nuevas oportunidades de mercado identificadas trimestralmente.



ILUSTRACIÓN 4 KPIS COMERCIAL Y MERCADEO

KPI Operaciones:

- Tiempo de implementación de nuevas soluciones de BI
- Porcentaje de procesos automatizados en la recolección y análisis de datos
- Eficiencia en la gestión de campañas.
- Reducción del tiempo de respuesta en la toma de decisiones gracias al uso de BI y Big Data

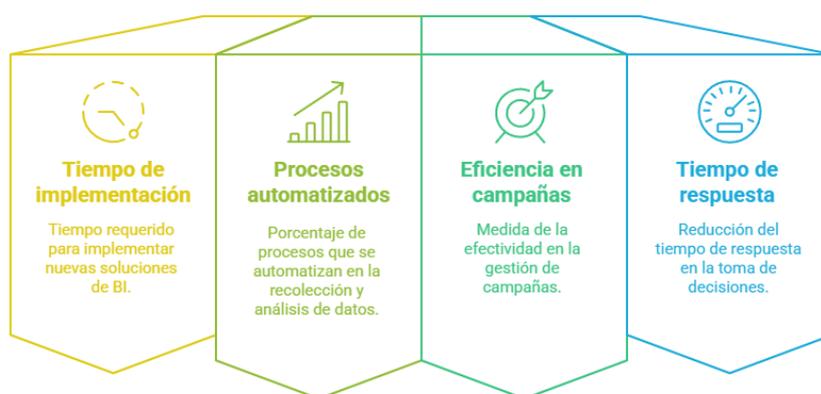


ILUSTRACIÓN 5 KPIS OPERACIONES

CAPITULO 2: ALMACENAMIENTO DE LOS DATOS Y EL ECOSISTEMA DE DW

2.1 Origen de la información

Consumer Lens se enfoca en crear un Data Warehouse específico para cada cliente, el objetivo de este Data Warehouse va a ser alimentado por la información de los consumidores propios del cliente siendo clasificados por sus características de compra en base a la preferencia de cada uno de sus productos.

Parte fundamental de cada proceso es la correcta ejecución y desarrollo de los procesos ELT (Extract, Load, Transform) o ECT (Extraer, Cargar, Transformar), para ello es muy importante la correcta comunicación entre el cliente y Consumer lens, el cliente de la misma forma tiene que ser muy específico en hasta donde Consumer Lens puede obtener información y que tipo de información consideran confidencial por lo que no se debe usar o manipular por parte de Consumer Lens.

2.2 Sistemas y destino de Carga

Extracción de la información: Consumer Lens obtendrá la información de dos formas, para comenzar y agilizar el proceso de análisis de obtendrá la información de históricos de compras por parte de los consumidores, tomando en cuenta la base de datos de compras que disponga la empresa y facturas cargadas al sistema del SRI (Servicio de Rentas Internas) del Ecuador.

La otra forma de obtener información y de forma más actual es con las compras que se realizan en la actualidad esa información es actualizada y muestra de forma más exacta las características de compras más recientes de cada cliente.

Carga de la información: Esta Etapa del proceso se clasifica la información por cliente y productos predilectos todo esto en un Data Warehouse específico del cliente, toda la información que se obtiene tanto del histórico de compras como del SRI se completa con las compras del día a día con la última información de forma constante y permanente.

Este proceso de carga es automatizado por lo que su implementación debe contar con los requisitos de carga específicos para no tener problemas de interpretación de la información a futuro.

Trasformación de la información: Esta etapa del proyecto es muy importante porque usando la información almacenada comenzamos a crear perfiles del cliente, clasificando el producto que compra más seguido con 5 estrellas, el que le sigue con 4 estrellas y así hasta 1 estrella.

Esta clasificación nos permite ligar a los clientes con los productos y cuando el cliente tenga promociones especiales por parte del proveedor de cierto producto sabemos que consumidores son los que más les interesa esta promoción enviando información de forma directa y con un link de compra en la página web para que la compra sea instantánea.

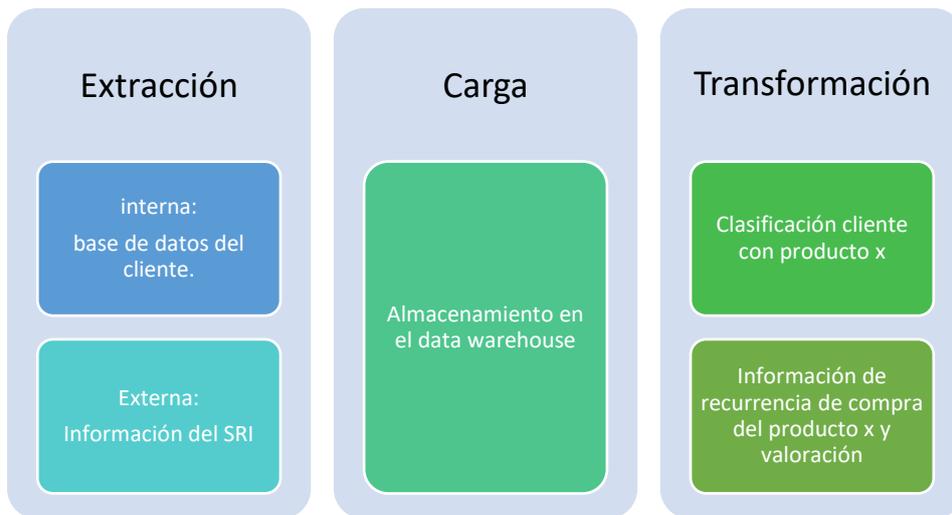


ILUSTRACIÓN 6 EXTRACCIÓN, CARGA Y TRANSFORMACIÓN

2.3 Almacenamientos de datos warehouse y datamarts.

Existen múltiples formas de almacenar la información, tanto de manera local en servidores corporativos o la opción más ampliamente utilizada en el mercado, como es el caso del “almacenamiento en la nube”. Cada una de estas opciones tiene tanto ventajas como desventajas, al contar con un servidor local es necesario contar con un departamento de TI que administre la parte informática, sin embargo, al contar con almacenamiento en la nube esta carga se reduce y únicamente con un equipo que analice los datos se puede sacar mucha partida a los datos y de manera mucho más eficiente debido a la capacidad de procesamiento que poseen empresas como Google, Amazon que brindan este servicio.



ILUSTRACIÓN 7 ALMACENAMIENTO DE DATOS WAREHOUSE

Al utilizar servicios como AWS de Amazon, los planes cuentan con herramientas para el procesamiento de datos, tal es el caso del “**Amazon Kinesis Data Analytics**” que se encarga de procesar y analizar flujos de datos en tiempo real, la gran ventaja de contar con un procesamiento en tiempo real es la velocidad de respuesta, de esta manera se podrán tomar decisiones de manera rápida y más acertadas.

Se pueden seguir enumerando incontables aplicaciones de diferentes empresas tecnológicas, lo cierto que la opción más ampliamente utilizada por su ventaja tanto a nivel de servicio como de coste, son los servicios en la nube.

2.4 Carencias y problemas en la carga de datos

La carga de datos a un data warehouse (DW) es un proceso crucial para asegurar la integración, disponibilidad y calidad de los datos para análisis y toma de decisiones. Sin embargo, este proceso puede enfrentar diversas carencias y problemas que impactan su efectividad.

Carencias de calidad en los datos

- Datos incompletos o inconsistentes: Faltan valores o estos no están alineados con los estándares esperados.
- Errores en el origen: Los datos provienen de sistemas con errores o desactualizados.
- Falta de validación: Los datos no pasan por procesos adecuados de limpieza y transformación antes de cargarse.

Problemas de integración de datos

- Heterogeneidad de las fuentes: Los datos provienen de múltiples sistemas con formatos incompatibles (bases de datos relacionales, archivos planos, API, etc.).
- Falta de estándares: No hay uniformidad en nombres de campos, tipos de datos, o definiciones.
- Problemas de duplicidad: Se generan inconsistencias debido a datos redundantes o duplicados provenientes de diferentes fuentes.

Desafíos de rendimiento

- Volúmenes de datos masivos: La carga de datos puede ser lenta o consumir demasiados recursos si el DW no está diseñado para escalar.
- Procesos ETL mal optimizados: Las transformaciones son lentas o generan cuellos de botella.
- Latencia en tiempo real: Si el DW requiere datos en tiempo real, los procesos de carga pueden no ser lo suficientemente rápidos.

Gobernanza y seguridad

- Falta de control de acceso: Usuarios no autorizados pueden alterar o visualizar datos.
- Problemas de auditoría: No hay registro de qué datos fueron cargados, cuándo o por quién.
- Cumplimiento normativo: Los datos no se manejan de acuerdo con leyes como GDPR, HIPAA u otras regulaciones locales.

Carencia de herramientas y recursos adecuados

- Tecnología obsoleta: Uso de sistemas o herramientas que no soportan las necesidades actuales del negocio.
- Falta de automatización: Muchas tareas de carga se realizan manualmente, lo que incrementa el riesgo de errores.
- Falta de personal capacitado: Equipos con conocimientos limitados en procesos ETL, manejo de grandes datos o administración de DW.

Problemas específicos en procesos ETL

- Extracción ineficiente:
 - Conexiones a sistemas fuente inestables.
 - Frecuencia de extracción incorrecta (exceso o insuficiencia).
- Transformación deficiente:
 - Lógica incorrecta que deriva en datos erróneos.
 - Transformaciones que no están alineadas con los requisitos del negocio.
- Carga incorrecta:
 - Esquemas mal diseñados en el DW que dificultan el almacenamiento.
 - Registros rechazados por errores de formato o reglas de validación.

Problemas en la arquitectura del data warehouse

- Diseño poco flexible: Estructuras rígidas que no soportan cambios o expansiones.
- Indexación ineficiente: Mala configuración de índices, lo que afecta las consultas y la carga de datos.
- Sobrecarga de consultas: La ejecución de consultas simultáneas puede interferir con los procesos de carga.

Gestión de metadatos deficiente

- Metadatos incompletos o incorrectos: Los procesos no registran información detallada sobre el origen, transformación y destino de los datos.
- Falta de documentación: Difícil identificar la procedencia o justificación de ciertos datos dentro del DW.

CAPITULO 3: PROCESOS ETL (Extract, Transform, Load)

3.1 Proceso de Carga de Datos

Los procesos de carga de datos, especialmente en entornos de ETL (Extract, Transform, Load) y data lakes, son fundamentales para gestionar, integrar y analizar grandes volúmenes de datos provenientes de diversas fuentes. A continuación, se detalla cómo funcionan ambos enfoques y sus características principales:

El proceso ETL es una metodología tradicional utilizada para mover datos desde sistemas fuente a sistemas de almacenamiento (como un data warehouse) de manera estructurada y organizada.

Extract (Extracción):

- Implica capturar datos de múltiples fuentes como bases de datos, archivos planos, APIs, o sistemas operacionales.
- Fuentes comunes: sistemas ERP, CRM, sensores IoT, logs, etc.
- Ejemplo: Recuperar datos de una base MySQL y un archivo CSV.

Transform (Transformación):

- En esta etapa, los datos se limpian, transforman y estructuran para satisfacer las necesidades del análisis.
- Actividades típicas:
 - Limpieza de datos (eliminación de duplicados, manejo de valores faltantes).
 - Aplicación de reglas de negocio.
 - Agregaciones, uniones o cálculos.
- Ejemplo: Convertir fechas a un formato estándar o calcular métricas derivadas.

Load (Carga):

- Los datos transformados se cargan en el destino final, como un data warehouse una base analítica.
- Dos enfoques de carga:
 - Carga completa: Se reescriben todos los datos.
 - Carga incremental: Solo se agregan los cambios nuevos o actualizados.
- Ejemplo: Subir los datos estructurados a un warehouse como Snowflake o Redshift.

3.2 Data Lake

Un data lake es un repositorio centralizado que permite almacenar grandes cantidades de datos en su forma cruda o sin procesar, independientemente de su estructura o formato.

a) Características principales:

- Almacenamiento escalable: Compatible con datos estructurados (bases de datos), semi-estructurados (JSON, XML) y no estructurados (imágenes, videos).
- Formato flexible: Datos pueden almacenarse en formatos como CSV, Parquet, Avro, o directamente como blobs de texto, audio o video.
- Ecosistemas populares: AWS S3, Azure Data Lake, Google Cloud Storage.

b) Proceso de carga en un data lake:

- Ingestión de datos: Datos se recopilan y almacenan directamente desde las fuentes. Ejemplo: Subir logs de servidores en tiempo real a un bucket de S3.
- Metadatos y catalogación: Los datos son etiquetados y catalogados mediante herramientas como Apache Hive o AWS Glue para facilitar búsquedas.
- Persistencia de datos crudos: Los datos permanecen sin transformar, listos para ser procesados bajo demanda.

c) Procesos posteriores (ETL en un data lake):

- En un data lake, el procesamiento puede hacerse bajo demanda: ELT (Extract, Load, Transform): Los datos se transforman después de la carga.
- Procesamiento mediante frameworks como Apache Spark o Hadoop.

| Característica | ETL | Data Lake |
|----------------------|----------------------------------|--|
| Formato de los datos | Estructurado | Estructurado, semi-estructurado, no estructurado |
| Procesamiento | Transformación antes de la carga | Transformación bajo demanda |
| Destino | Data warehouse | Repositorio masivo y flexible |
| Flexibilidad | Baja | Alta |
| Uso principal | Informes, análisis estructurados | Big data, machine learning |

ILUSTRACIÓN 8 TABLA ETL VS DATA LAKE

3.3 Información y calidad de la información

Consumer Lens obtendrá y manipulará la información directamente desde el cliente, en el Ecuador la ley tributaria exige que todo consumidor que compre en un establecimiento más de \$20,00 dólares deben recibir una factura con sus datos personales, entre los que debe incluir:

- Nombres y Apellidos,
- Número de Cédula
- Correo Electrónico
- Número de Teléfono.

Esta información personal se brinda directamente por parte del cliente por lo que su veracidad es comprobada, lo que significa que la calidad de la información es excelente.

Es importante mencionar que la compra de ciertos productos como alimenticios o de educación generan beneficios tributarios al cliente por lo que el interés de entregar correctamente la información es vital para que las y los clientes apliquen a estos beneficios económicos.

Finalmente, esta información debe subirse al SRI (Servicio de Rentas Internas) de inmediato para realizar correctamente la declaración de impuestos a final de mes, en Ecuador es obligatorio que las empresas de tamaño mediano a grande realicen la facturación electrónica lo que hace que muchas empresas incluyan en sus ERP módulos contables que manejan la información de forma efectiva y eficiente, de fácil acceso y manipulación.

3.4 Horarios de Carga y accesibilidad

La mayoría de las empresas realizan la carga de facturas de venta al SRI en horas de la noche cuando se cierran los puntos de venta, otras lo hacen de manera instantánea cuando se finaliza la venta, por lo que la carga de información se puede hacer de forma inmediata teniendo como tiempo máximo de carga de información 24 horas, el proceso de carga se hace de forma masiva por lo que se optimiza el proceso haciendo más rápida la operación.

Por otro lado, la accesibilidad a la información que puede realizarse para diferentes análisis de la información puede hacerse de inmediato después de la carga de la información es importante mencionar que el momento de la carga de la información al hacerlo correctamente siguiendo los formatos establecidos de carga, automáticamente se categoriza al producto con el cliente lo que permite tener la información de la tendencia de compra de un o una consumidora en específico.

La información obtenida en esta base de datos es accesible para todos los departamentos de la empresa que requiera los servicios de Consumer Lens para que pueda usarse para el control o planteamiento de nuevas estrategias como el departamento de la empresa considere

apropiado.

3.5 Variables disponibles y acceso a la data warehouse:

Existen variables disponibles por medio de una base de datos proporcionada por software ERP el cual proporciona la suficiente información necesario, datos tales como nombres completos, numero de cedula, dirección del domicilio, correo electrónico y numero celular en algunos casos. En ciertas ocasiones se utiliza el sistema del SRI, el cual mediante el número de cedula para obtener los nombres y apellidos completos o incluso si la misma cuenta deudas tributarias con el estado ecuatoriano.

Por lo tanto, una base de datos de una empresa de retail podría ser un buen inicio para contar con variedad de clientes, esto debido a que son negocios que cuenta con un flujo de clientes constante convirtiéndola en una opción predilecta.

3.6 Sistemas implicados en los procesos.

Efectivamente es necesario contar con sistemas en los procesos, estos sistemas podrían ser tanto para el procesamiento como para la visualización de los datos y reporting provenientes de la warehouse. Todos estos sistemas son necesarios e imprescindibles para contar con una excelente visualización de datos.

Estos sistemas ofrecen ventajas significativas tales como un mejor rendimiento, en procesamiento de grandes cantidades de volumen de datos permitiendo ahorrar grandes cantidades de tiempo, tal es el caso de Power BI, ampliamente utilizado en prácticas de Big Data pero con ciertas limitaciones en las versiones gratuitas.

CAPITULO 4: TECNOLOGÍAS PARA EL ANÁLISIS Y DISEÑO DE VISUALIZACIONES DE BI

4.1 Análisis de herramientas Big Data

Existen varias aplicaciones que nos ayudan a almacenar información en base a las operaciones que se realizan, para el modelo de negocio de Consumer Lens es muy importante optar por herramientas que permitan de forma objetiva, eficaz y eficiente procesar y analizar grandes volúmenes de datos estructurados en su mayoría, pero también los no estructurados.

En la actualidad existen varias herramientas de almacenamiento y procesamiento de datos, analicemos cada una a detalle sus bondades, ventajas, desventajas para luego de forma objetiva escoger la mejor para Consumer Lens.

| Herramienta | Descripción |
|------------------|---|
| Google Big Query | Solución de análisis en la nube altamente escalable de Google, permite ser alimentada por diferentes fuentes y sistemas que son adaptables a la aplicación. |
| Amazon Redshift | Servicio de almacenamiento de datos en la nube diseñado para análisis a gran escala |
| Apache Spark | Plataforma de análisis de datos en tiempo real que ofrece un procesamiento más rápido que la competencia |
| Snowflake | Plataforma de datos basada en la nube que permite almacenamiento de datos así como consulta y análisis de datos. |
| Hadoop | Marco de trabajo de código abierto para el procesamiento distribuido de grandes conjuntos de datos clústeres. |
| MongoDB | Base de datos NoSQL orientada a documentos. |
| Cassandra | Base de datos distribuida diseñada para manejar grandes volúmenes de datos en múltiples servidores. |

Como podemos ver las herramientas que podemos ver en el mercado hoy en día son varias y muy funcionales, siendo difícil poder escoger la mejor opción ya que en su gran mayoría todas sus propuestas son muy buenas.

Tomando en cuenta que Consumer Lens necesita alimentar información a la Base de datos por parte del cliente que pueden tener diferentes ERP'S es muy importante trabajar con una herramienta versátil que se adapte y funciones con cada ERP así como con cada módulo del mismo para que la información obtenida pueda ser manipulada y analizada por diferentes áreas de la empresa del cliente.

Desde nuestro punto de vista la herramienta que mejor se adapta a nuestro proyecto es Google Big Query, una aplicación versátil, confiable y eficiente que

se adapta de forma muy práctica a cualquier ERP alimentando su base de datos de forma ágil y eficaz.

4.2 Google Big Query alimentación de datos:

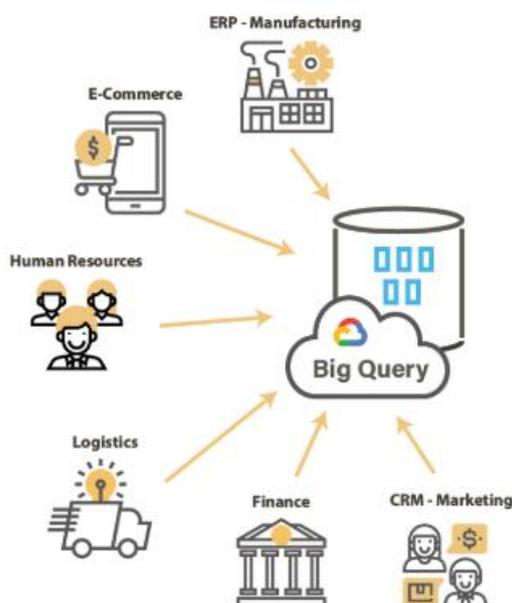


ILUSTRACIÓN 9 GOOGLE BIG QUERY

Ventajas de Google Big Query:

- Arquitectura sin servidores: No es necesario administrar o configurar servidores, Google se encarga de la infraestructura y el escalado automáticamente.
- Análisis de Datos Masivos: Es una herramienta que optimiza los procesos para el análisis y procesamiento de grandes volúmenes de datos lo cual lo hace ideal para el segmento de clientes de Consumer Lens.
- Análisis en tiempo real: Es una herramienta que permite la ingesta y análisis de datos en tiempo real de forma actualizada y confiable.
- Machine Learning integrado: BigQuery ML, permite crear y entrenar modelos de aprendizaje automático utilizando SQL.
- Seguridad y control: Ofrece encriptación de datos, control de accesos basado en roles y cumplimiento con los estándares de seguridad.
- Rapidez: Consulta y análisis en cuestión de segundos.
- Escalabilidad: Maneja datos de cualquier tamaño.
- Facilidad de uso: Ideal tanto para principiantes como para expertos en datos.

Integración de Google Big Query y Microsoft Power BI:

Para poder maximizar los beneficios de dos herramientas tan útiles como lo son Google Big Query para el almacenamiento y análisis de información y Microsoft Power BI para el manejo correcto de la información y generación de cuadros para optimizar la toma de decisiones en beneficio del cliente.

Esta integración genera beneficios de forma circular e integral, vitales para las operaciones de Consumer Lens, Siendo esta integración una solución poderosa para realizar análisis y visualización de datos, Permitiendo utilizar el rendimiento y la escalabilidad de Big Query como almacén de datos en la nube, mientras que Power BI como herramienta de visualización de dashboards e informes interactivos.

Ventajas de Integración de Big Query y Power BI:

- **Análisis en tiempo real:** Big Query puede manejar datos en streaming y Power BI permite visualizarlos en tiempo real.
- **Escalabilidad:** Se puede aalizar grandes volúmenes de datos almacenados en Big Query y visualizarlos sin necesidad de moverlos.
- **Integración sin complicaciones:** Power BI cuenta con un conector nativo para Big Query lo que simplifica la conexión y la transferencia de datos.
- **Flexibilidad de visualización:** Se puede crear Dashboards personalizados y dinámicos a partir de consultas ejecutadas en Big Query.

4.3 Seleccionar los usuarios y perfiles de cada dashboard

Seleccionar usuarios y perfiles de dashboard en Power BI implica configurar permisos de acceso y roles para garantizar que los datos y visualizaciones sean visibles solo para las personas adecuadas.

4.3.1 Creación de Roles y Perfiles de Usuario

Roles de Seguridad: Power BI permite definir roles y perfiles dentro de un informe o dashboard mediante la funcionalidad de "seguridad a nivel de fila" (Row-Level Security, RLS). Esto permite que diferentes usuarios vean solo los datos que les corresponden según su rol.

4.3.2. Asignación de Usuarios a Roles

Después de definir los roles en Power BI Desktop, debes asignar los usuarios correspondientes en el servicio de Power BI (Power BI Service).

Tipos de Usuarios a Asignar: Usuarios con Roles Específicos: Por ejemplo, un rol para "Gerentes" que vea solo ciertas regiones de ventas.

4.3.3 Acceso a los Dashboards en Power BI Service

Publicar y Compartir Dashboards: Una vez configurados los roles, puedes compartir los dashboards con los usuarios. Existen varias maneras de hacerlo:

Publicación a la web: Puedes compartir dashboards mediante un enlace, pero ten en cuenta que esta opción no es segura.

Compartir directamente: Desde Power BI Service, puedes compartir el informe o dashboard con otros usuarios específicos a través de su dirección de correo electrónico, asegurándote de que solo los usuarios autorizados tengan acceso.

Accesos de lectura vs. escritura: Ten en cuenta si solo necesitas permitir la visualización de los dashboards o si también deseas permitir que los usuarios realicen modificaciones. Para ello, asegúrate de seleccionar el tipo de acceso adecuado.

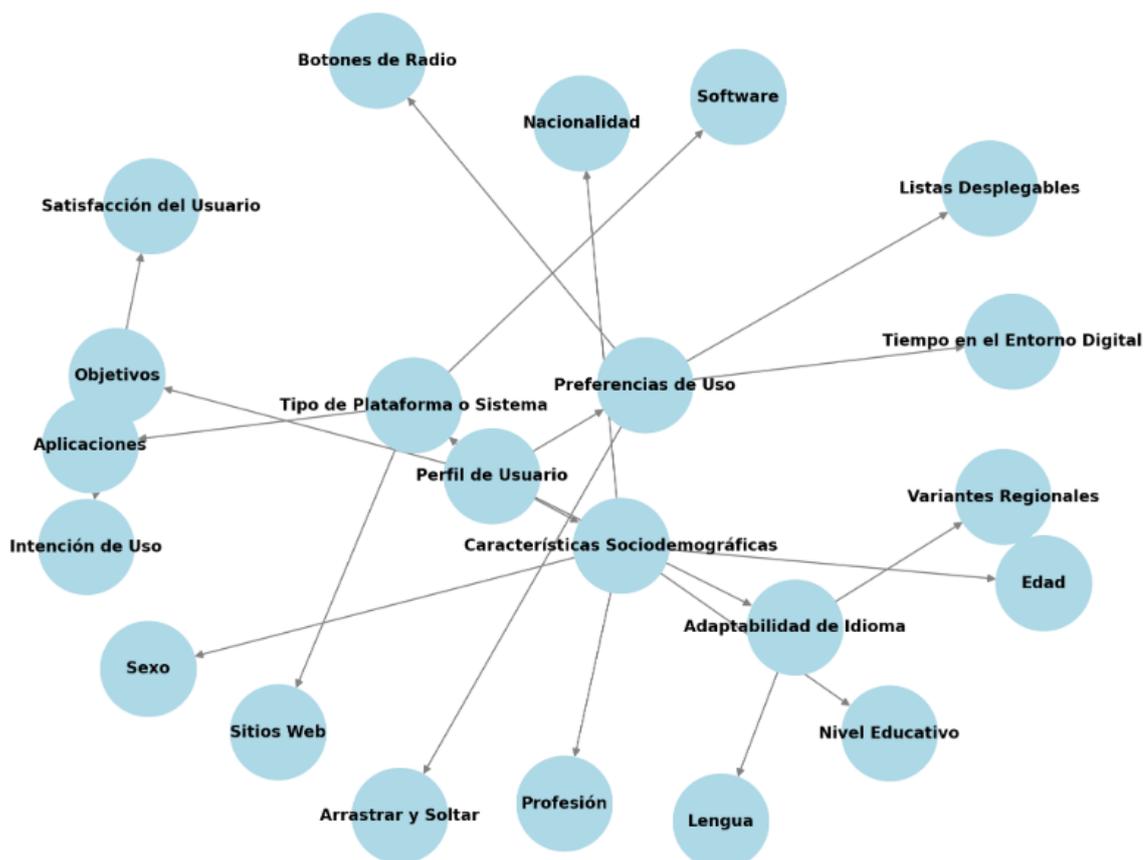


ILUSTRACIÓN 10 RED DE ACCESOS A DASHBOARD

4.4 Tipos de usuarios:

- **Usuario Temporal**

Este tipo de perfil permite acceder a la plataforma, sitio web o aplicación sin necesidad de iniciar sesión. Por ello, la interacción se realiza de forma genérica, sin personalización. Esto impide que se carguen preferencias, historiales de navegación o configuraciones predeterminadas, limitando la experiencia del usuario.

- **Usuario Móvil**

Los usuarios móviles acceden a los entornos digitales mediante un nombre de usuario y contraseña desde cualquier dispositivo o computadora. Esto les permite operar con sus configuraciones habituales y registrar sus acciones según el perfil configurado.

- **Usuario Local**

Este perfil guarda los datos de navegación directamente en el disco duro del dispositivo o computadora que se utiliza para acceder. Esta característica facilita a las empresas restringir el acceso a sus plataformas, permitiendo que los usuarios solo puedan ingresar desde un equipo autorizado o específico.

- **Usuario por Funcionalidad**

Este perfil organiza las actividades del usuario según el rol asignado. Los tipos más comunes incluyen:

- **Administrador:** Tiene control total sobre la plataforma, permitiendo cambiar configuraciones o asignar responsabilidades en cualquier momento.
- **Técnico:** Se encarga del monitoreo y de realizar mejoras técnicas, generalmente a cargo de los desarrolladores.
- **Supervisor:** Supervisa y revisa las acciones realizadas por otros usuarios.
- **Ejecutor:** Realiza las actividades principales, como modificar, ingresar o eliminar datos esenciales.
- **Lector:** También llamado "modo solo lectura". No tiene permiso para modificar la información principal, aunque puede agregar comentarios relacionados con su función.

4.5 Perfiles de cada dashboard

4.5.1 Perfil de Dashboard para Gerente

Objetivo: Proveer una vista general del desempeño de la empresa o unidad de negocio, con indicadores clave de rendimiento (KPIs).

Accesos:

- KPIs de alto nivel: Ventas, rentabilidad, eficiencia operativa, satisfacción del cliente.
- Análisis de tendencias: Gráficas que muestren el rendimiento a lo largo del tiempo (por ejemplo, gráfico de barras o líneas con datos de meses o trimestres).
- Datos comparativos: Comparación entre diferentes regiones, productos o periodos.
- Alertas y notificaciones: Indicadores de alerta si algún KPI está fuera de los valores esperados.

Permisos:

- Acceso a todos los datos: El gerente debe tener acceso a todos los datos de la organización, ya que necesita tomar decisiones estratégicas basadas en una visión global.
- Visión de alto nivel: Acceso a dashboards agregados y resúmenes.

Componentes del Dashboard para Gerente:

- Indicadores clave (KPIs) como ingresos, crecimiento, retorno sobre la inversión (ROI).
- Gráficos comparativos por áreas o unidades de negocio.
- Proyecciones de tendencias a futuro.
- Alertas de desempeño si alguna métrica cae por debajo del umbral deseado.

4.5.2 Perfil de Dashboard para Supervisor

- Objetivo: Monitorear el rendimiento de equipos o procesos específicos y tomar decisiones tácticas.
- Accesos:
 - Desempeño de equipo o departamento: Visualización de datos a nivel de equipo o área de trabajo.
 - Productividad y eficiencia: Datos sobre el rendimiento de los colaboradores o equipos bajo su supervisión.
 - Alertas operativas: Si los resultados o tareas no se están cumpliendo según lo esperado.
- Permisos:
 - Acceso a datos específicos de su área: Un supervisor no debe ver datos de toda la organización, sino solo los de su equipo o departamento.
 - Filtros dinámicos: Para que el supervisor pueda personalizar su vista de acuerdo a diferentes criterios como fechas, empleados, proyectos o tareas.

Componentes del Dashboard para Supervisor:

- Gráficos de rendimiento por empleado o equipo.
- Indicadores de productividad y eficiencia.
- Alertas para monitorear cualquier incumplimiento o desviación en tiempo real.
- Visibilidad de los proyectos o tareas en progreso.

4.5.3 Perfil de Dashboard para Sistemas (Departamento de TI)

- **Objetivo:** Monitorear la infraestructura tecnológica, el estado de los sistemas y detectar problemas técnicos.
- **Accesos:**
 - Estado del sistema: Datos sobre el estado de los servidores, bases de datos, redes y otros componentes tecnológicos.
 - Alertas y eventos: Notificaciones sobre incidencias, fallos del sistema o vulnerabilidades.
 - Rendimiento de la infraestructura: Métricas de tiempo de actividad, uso de recursos y consumo de ancho de banda.
- **Permisos:**
 - Acceso técnico y de rendimiento: El departamento de TI necesita ver datos de funcionamiento técnico y de recursos utilizados.
 - Visualización avanzada: Acceso a configuraciones detalladas, logs de eventos y métricas técnicas.

Componentes del Dashboard para Sistemas:

- Gráficos de disponibilidad de servidores y sistemas críticos.
- Alertas de incidentes en tiempo real.
- Métricas de consumo de recursos, como CPU, RAM, y almacenamiento.
- Historial de incidencias y eventos críticos.

4.5.4 Perfil de Dashboard para Cliente

- **Objetivo:** Proveer una vista sencilla y clara del servicio o producto adquirido, con datos relevantes y personalizados.
- **Accesos:**
 - Información sobre el producto/servicio: El cliente debe poder ver información relevante sobre su compra, como estado del pedido, fecha de entrega, historial de compras, etc.
 - Soporte y consultas: Visualización de tickets de soporte, preguntas frecuentes y estado de resolución.
 - Satisfacción del cliente: Si es posible, acceso a encuestas de satisfacción y feedback del servicio.
- **Permisos:**

- Acceso solo a datos relacionados con su cuenta: El cliente no debe tener acceso a datos ajenos a su propia interacción con la empresa.
- Visualización simplificada: Un dashboard con un diseño claro y fácil de usar, sin demasiados detalles técnicos.

Componentes del Dashboard para Cliente:

- Estado de pedidos o servicios contratados.
- Historial de compras o interacciones.
- Soporte y atención al cliente: Estado de solicitudes o tickets.
- Encuestas de satisfacción con las opciones de calificación o comentarios.

4.6 Implementación en Power BI

Para implementar estos perfiles en Power BI, debemos considerar las siguientes acciones:

1. Seguridad a nivel de fila (RLS):
 - Implementa reglas de RLS para garantizar que los usuarios solo vean los datos que les corresponden. Por ejemplo, para un Gerente, puedes habilitar acceso a todos los datos, mientras que para un Supervisor, solo los datos de su equipo.
2. Diseño del Dashboard:
 - Utiliza filtros interactivos y segmentaciones para permitir que los usuarios ajusten la información visible en el dashboard. Por ejemplo, un Supervisor podría filtrar por equipo o fecha.
 - Para los clientes, asegúrate de que el diseño sea sencillo, con solo los datos más relevantes, como el estado de su pedido o servicio.
3. Publicación y acceso:
 - Publica los informes de Power BI en el servicio de Power BI y asigna los roles adecuados a los usuarios según sus perfiles. Asegúrate de configurar permisos de acceso correctamente.
4. Personalización:
 - Personaliza el dashboard según las necesidades de cada rol para proporcionar la información más relevante y útil para la toma de decisiones.

Resumen de los perfiles:

| Perfil | Objetivo | Accesos y Permisos |
|------------|--|---|
| Gerente | Visión general del desempeño organizacional. | Acceso total a los KPIs, métricas de rentabilidad y tendencias. |
| Supervisor | Monitorear el desempeño de su equipo o departamento. | Acceso a datos de su área, métricas de productividad y alertas. |
| Sistemas | Monitorear la infraestructura tecnológica. | Acceso técnico a logs, servidores, y métricas de recursos. |
| Cliente | Ver el estado de pedidos y servicios contratados. | Solo acceso a su cuenta, historial de compras y soporte. |

ILUSTRACIÓN 11 TABLA DE RESUMEN DE PERFILES

Este enfoque te permite crear dashboards eficaces y ajustados a las necesidades de cada usuario dentro de la organización o fuera de ella.

4.7 Necesidades de información y accesos por departamento y perfil.

- **Necesidades de información:**

Es imprescindible la información para la idea de negocio que hemos planteado sea segura y confiable, por lo tanto, consideramos de suma relevancia que la información debe provenir de fuentes relacionadas con nuestro trabajo, con el fin de garantizar la validación de dicha información esperar los resultados correctos.

Se espera que con la información correcta podamos, contar con la gran mayoría de las siguientes características:

- Conocer mejor a nuestros clientes.
- Identificar tendencias y oportunidades
- Resolver problemas de manera más eficiente y en el menor tiempo posible.
- Tomar decisiones más acertadas con menor cantidad de recursos.
- Anticiparse a las necesidades de sus clientes buscando siempre la venta y el beneficio en el cliente.
- Eliminar áreas problemáticas antes de que afecten la reputación.
- Prever la demanda de los clientes en temporadas no esperadas.

- **Accesos por departamento y perfil:**

Al establecer el acceso por perfil es fundamental establecer una estructura organizacional en el cual existas diferentes departamentos, todos ellos manejan

diferentes funcionalidades y cualidades que necesita la empresa, todo esto permite llevar a cabo sus actividades.

Por lo tanto, es importante separar la información para el personal tanto “administrativo”, este departamento maneja información crucial e importante para la empresa, además que la información debe estar procesada, esta con el fin de que se puedan llevar rápidamente a la toma de decisiones por parte de los directivos.

El personal “operativo” está más ampliamente relacionado con la parte operativa, los cuales desarrollan y envían la información cruda y se encuentra en grandes cantidades, esta información es necesaria para su posterior procesamiento.

A continuación, se describe la tabla con sus respectivos perfiles:

| Departamento | Perfil | Descripción |
|---------------------|---------------|--|
| Directivos | Administrador | Cuenta con la gran mayoría de herramientas y datos necesarios. |
| Administrativos | Medio | <ul style="list-style-type: none"> • Encargados de depurar la información y permitir que esta sea fiable. |
| Operativo | Básico | <ul style="list-style-type: none"> • El perfil básico permite realizar consultas restringidas. |



ILUSTRACIÓN 12 TABLA DE PERFILES UTILIZADOS

4.8 Alineación con los Objetivos Iniciales

Objetivo principal: El proyecto busca implementar un sistema BI y Big Data que permita un análisis detallado del cliente, aumentando las ventas mediante campañas publicitarias efectivas y potenciando el comercio electrónico.

- **Progreso logrado:**
 - Se ha definido claramente el enfoque técnico con herramientas como Google Big Query y Power BI, lo que cumple con el requisito de integrar sistemas para el análisis avanzado.
 - El diseño de dashboards y la personalización según perfiles (gerente, supervisor, cliente, etc.) alinean las herramientas con la experiencia del cliente y la toma de decisiones estratégicas.

Datawarehouse y procesamiento de datos:

- **Progreso logrado:**
 - Se detallan los procesos ETL y la capacidad de integración con data lakes, lo que fortalece la escalabilidad y la adaptabilidad del sistema a las necesidades del cliente
 - La elección de Google Big Query, destacada por su capacidad de análisis masivo en tiempo real, está bien justificada para el modelo de negocio

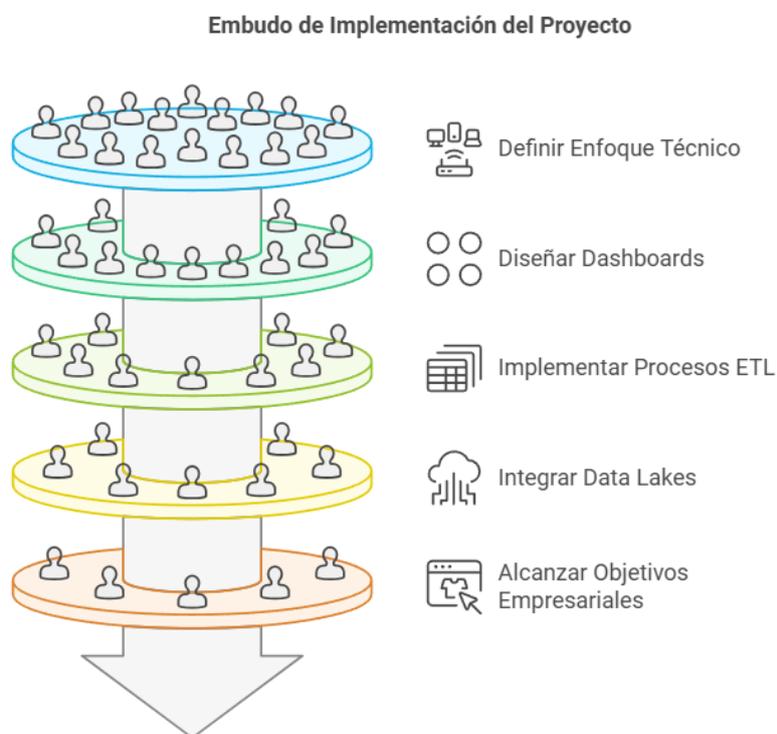


ILUSTRACIÓN 13 DATAWAREHOUSE Y PROCESAMIENTO DE DATOS

Cumplimiento de los Objetivos

- Cumplimiento técnico:**
 - La integración entre Google Big Query y Power BI permite análisis en tiempo real y generación de insights prácticos para la empresa, cubriendo los requerimientos técnicos esenciales.
- Calidad de datos:**
 - La calidad y veracidad de los datos obtenidos mediante facturación electrónica asegura la confiabilidad del sistema. Esto cumple con el estándar requerido para la toma de decisiones estratégicas.
- Escalabilidad y Flexibilidad:**
 - El proyecto ha contemplado herramientas modernas y flexibles, como Big Query y Power BI, que son escalables y permiten adaptarse a diferentes entornos de datos, tanto estructurados como no estructurados.
- Gestión de datos y accesibilidad:**
 - Se detalla el proceso de carga y acceso eficiente a los datos, lo que mejora la experiencia para todas las partes interesadas y optimiza la operación general.

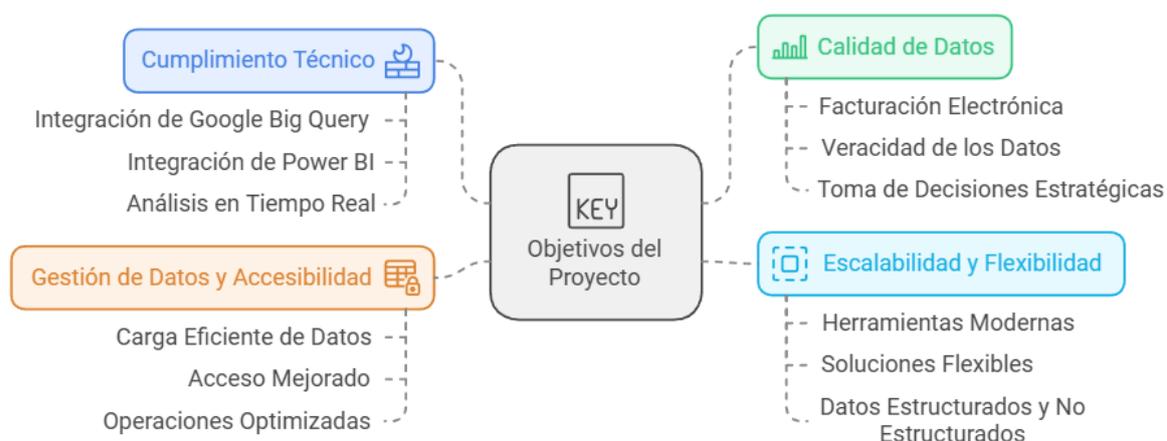


ILUSTRACIÓN 14 ESQUEMA DE OBJETIVOS DEL PROYECTO

El proyecto "Consumer Lens" está bien alineado con los objetivos iniciales, aprovechando herramientas modernas y metodologías sólidas para cumplir las metas.

CAPITULO 5: CREACIÓN DEL ÁREA DE BI

5.1 Situación actual del departamento de BI.

Consumer Lens comenzará sus operaciones con la implementación del departamento de BI en Librería Studium, una empresa reconocida en el Ecuador que busca aumentar sus ventas en línea, así como generar beneficios personalizados a cada tipo de cliente aumentando la propuesta de valor y mejorando en su servicio frente a la imagen recibida por el cliente.

En la actualidad la empresa no cuenta con un departamento de BI, quieren optar por este departamento por que hace un año implementaron un ERP nuevo para la empresa que es ODOO como beneficios de este ERP es la facilidad que brinda para la obtención de información, así como la cantidad de información útil que brinda.

Esta información debe ser bien administrada en un Datawarehouse almacenando historiales que categoricen a clientes y su consumo para generar ofertas específicas y beneficios en descuentos de sus libros preferidos, así mismo al ser una librería que vende libros escolares anticipar el cobro de la venta de libros para el próximo año escolar a cambio de descuentos especiales, esto facilitara el proceso de pago a las editoriales que brindan beneficios económicos a la librería así como tener una ventaja competitiva frente a la competencia.

La implementación de un departamento de BI permitirá que gerencia optimice sus decisiones en base a análisis de datos más precisos, optimizando la toma de decisiones basándose en datos más exactos y detallados, que permitan de forma objetiva y real la mejora de procesos operativos., lo que representaría en un futuro cercano una reducción de costos de forma positiva para la empresa.

No podemos dejar de lado que el poder implementar correctamente un departamento de BI en la librería permitirá que la misma pueda tener una correcta segmentación de sus clientes y brindar beneficios personalizados a los clientes de sus ofertas o nuevas adquisiciones generando un vínculo con el cliente más fuerte, fidelizándolo y aumentando su expectativa de compra lo que representa mayores ingresos para la empresa.

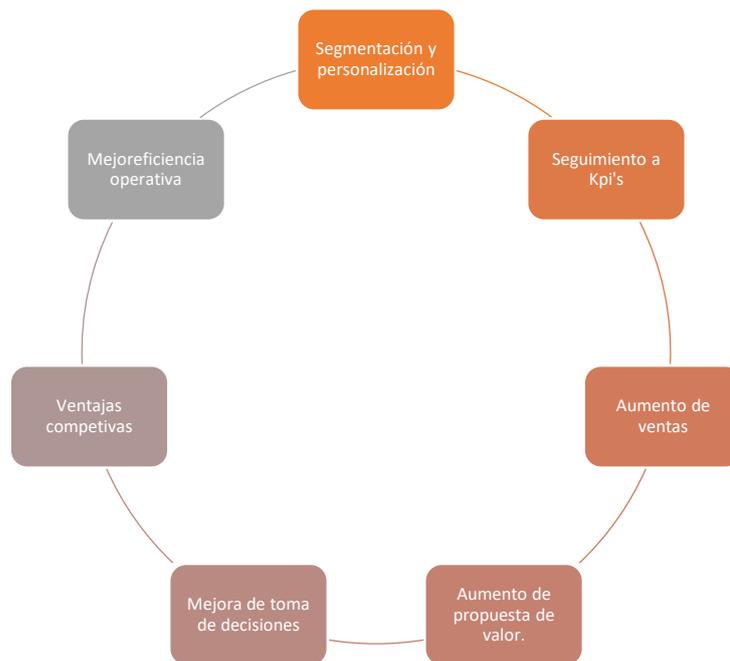


ILUSTRACIÓN 15 SITUACIÓN ACTUAL DEL DEPARTAMENTO DE BI

5.2 Definición en las carencias en formación y recursos.

Tomando en cuenta que es una implementación de un departamento prácticamente nuevo como es el de BI y que la operación se va a realizar en base a un ERP relativamente nuevo para la empresa en donde varios desafíos aparecerán por el camino es muy importante para el éxito del proyecto identificar previamente las carencias y desafíos que el proyecto representa frente al cliente que pueden representar amenazas para el éxito del proyecto, pero si nos anticipamos a las adversidades puede ser un factor fundamental para el éxito del proyecto y se vea reflejado en la satisfacción del cliente.

Al ser un departamento nuevo en la empresa la principal carencia será la de conocimiento, se debe preparar al personal desde cero lo que representa varias horas en capacitaciones de la misma forma la implementación tendrá mayores dificultades por la falta de conocimiento y manipulación de las herramientas tecnológicas brindadas a la empresa. Por lo que la obtención de la información al principio por parte del personal de la empresa puede que no sea correcta o exacta.

Por otro lado, en los directivos o supervisores, la falta de conocimiento puede traer dificultades al momento de análisis de la información brindada, así como traer dificultades al inicio ya que no saben requerir de forma precisa y concisa la información que quieren obtener por lo que se puede ofrecer mucha información que no va a ser necesaria.

Al ser un proyecto nuevo se debe implementar la estructura desde cero por lo que el proyecto puede requerir más recursos de los planteados.

Ahora las soluciones a estos problemas que se pueden presentar consideramos los siguientes:



ILUSTRACIÓN 16 DIAGRAMA DE CARENCIA Y FORMACIÓN DE RECURSOS

5.3 Coordinación de los flujos entre otras áreas.

La coordinar es imprescindible en una empresa y más aún cuando existen grupos de trabajo entre los cuales depende en gran medida la colaboración entre muchos de ellos con el fin de sacar en el mercado un producto final o servicio. Por lo tanto, hemos decidido utilizar los siguientes criterios para establecer la coordinación de los flujos entre áreas:

- **Establecer canales de comunicación claros**

Es importante tener canales de comunicación internos bien definidos para garantizar un flujo de información y comunicación entre las diferentes áreas de la empresa, por lo tanto ese proceso se volver asistemático por lo tanto debe ser confiable y seguro, ser susceptible a las fallas.

- **Mantener una comunicación simple**

Es importante ser preciso con la información que se comparte y evitar el exceso de información, por lo tanto la simplicidad permite de manera ágil y sin

desperdiciar muchos recursos la comunicación entre los miembros del equipo logrando coordinar de manera rápida y eficiente.

- **Utilizar herramientas colaborativas**

Las herramientas colaborativas permiten automatizar procesos y facilitar la comunicación entre equipos. Estas herramientas deben ser lo suficientemente rápidas, seguras que permita un rápido actuar por parte de los colaboradores o personal encargado de ejecutar dichas acciones dentro de la empresa.

- **Dar seguimiento al progreso**

El gerente de proyectos debe dar seguimiento al progreso del trabajo y al presupuesto del proyecto, efectivamente la supervisión y control de cada una de las áreas es importante. Al lograr supervisar y cubrir gran parte de las áreas, estas podrán bien ser más eficientes o ciertamente bajar su rendimiento por lo tanto es importante la constante monitorización.

- **Eliminar obstáculos**

Es importante eliminar obstáculos que no permitan desarrollarse con normalidad las actividades de la empresa, fundamental que todas las partes actúen con responsabilidad y cabalidad del caso, con una misión y visión en común de la empresa.

Pero para que todo esto funcione, necesariamente debe existes la responsabilidad y la colaboración por parte de los colaboradores con el fin de alcanzar las metas plateadas.

5.4 Organigrama

El proyecto de BI & Big Data de Consumer Lens incluye las siguientes áreas clave dentro de su organigrama:

1. **Junta Directiva:** Define la estrategia corporativa y supervisa el progreso del proyecto.
2. **Equipo de Dirección:** Ejecuta la estrategia y monitorea nuevas líneas de negocio, apoyado por equipos especializados en tecnología.
3. **Departamento de Implementación:** Responsable de integrar el proyecto a la estructura organizacional, asegurando que las soluciones se alineen con los objetivos generales.



ILUSTRACIÓN 17 ESQUEMA DEL DEPARTAMENTO DE IMPLEMENTACIÓN

5.5 Dependencia de Otras Áreas

- **Ventas y Marketing:** Proporcionan datos clave para analizar tendencias de clientes y diseñar campañas publicitarias personalizadas.
- **TI (Tecnologías de Información):** Garantiza la correcta implementación de herramientas como Google Big Query y Power BI para la integración, análisis y visualización de datos.
- **Operaciones:** Contribuyen al flujo de datos operativos necesarios para la automatización y optimización de procesos.

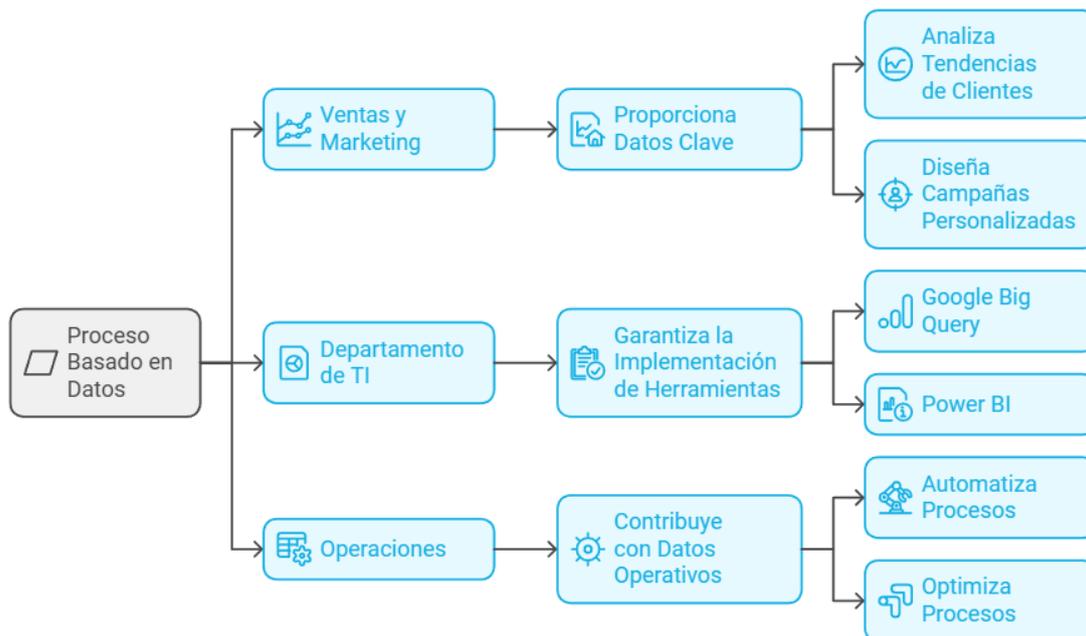


ILUSTRACIÓN 18 ESQUEMA DEL PROCESO BASADO EN DATOS

5.6 Solución de Problemas

El proyecto plantea las siguientes estrategias para abordar desafíos:

1. **Gestión de Calidad de Datos:**
 - Uso de procesos ETL para asegurar la limpieza y transformación adecuada de datos.
 - Integración de herramientas de análisis masivo como Google Big Query para garantizar precisión y confiabilidad en tiempo real.
2. **Seguridad y Control de Acceso:**
 - Implementación de reglas de seguridad a nivel de fila (RLS) en Power BI para proteger datos sensibles y garantizar accesos apropiados.
3. **Escalabilidad y Flexibilidad:**
 - Elección de plataformas escalables que se adapten a diferentes volúmenes de datos y necesidades analíticas futuras.
4. **Personalización y Eficiencia:**
 - Diseño de dashboards específicos para cada perfil de usuario (gerente, supervisor, cliente) para maximizar la utilidad de los datos y apoyar la toma de decisiones estratégicas.

Logrando la Excelencia en la Gestión de Datos

Personalización y Eficiencia

Diseñar dashboards específicos para usuarios para mejorar la toma de decisiones.



Escalabilidad y Flexibilidad

Elegir plataformas escalables para acomodar diversas necesidades de datos.



Seguridad y Control de Acceso

Aplicar seguridad a nivel de fila en Power BI para proteger datos sensibles.



Gestión de Calidad de Datos

Implementar procesos ETL e integrar herramientas de análisis para la precisión de los datos.



ILUSTRACIÓN 19 PROCEDIMIENTO DE GESTIÓN DE DATOS

5.7 Recursos externos Destinados: Subcontratación

Para implementar un departamento de Business Intelligence (BI) en una empresa como Consumer Lents, se pueden utilizar una variedad de recursos externos que abarcan tecnología, personal especializado, y servicios de consultoría.

5.7.1 Software y herramientas de BI

Estas herramientas ayudan a recopilar, analizar y visualizar datos:

- Herramientas de análisis y visualización:
 - Tableau: Visualización interactiva y análisis de datos.
 - Power BI (Microsoft): Visualización y análisis de datos con integración de Microsoft 365.
 - Qlik Sense: Herramienta avanzada de análisis asociativo.
 - Google Data Studio: Solución gratuita de Google para visualización.

- Looker: Plataforma moderna para modelado de datos y visualización.
- Bases de datos y almacenamiento de datos:
 - Amazon Redshift: Almacenamiento de datos en la nube.
 - Snowflake: Almacenamiento de datos basado en la nube.
 - Microsoft SQL Server: Bases de datos relacionales con capacidades analíticas.
 - Google BigQuery: Almacenamiento y análisis de grandes volúmenes de datos.
- Integración de datos:
 - Talend: Integración de datos y ETL.
 - Informatica: Plataforma robusta para integración y gobernanza de datos.
 - Fivetran: Sincronización automática de datos de múltiples fuentes.
 - Apache Kafka: Sistema de mensajería para procesamiento de datos en tiempo real.

5.7.2 Consultoría especializada

El soporte de consultores puede agilizar la implementación y garantizar el éxito del departamento:

- Empresas de consultoría global:
 - Accenture: Especialistas en BI y transformación digital.
 - Deloitte: Amplia experiencia en estrategias de BI y big data.
 - PwC: Consultoría en análisis de datos y automatización.
 - KPMG: Servicios de análisis de datos y BI.
- Consultoras especializadas en BI:
 - Bismart: Consultora enfocada en soluciones de BI y big data.
 - Hitachi Vantara: Consultoría para soluciones de almacenamiento y análisis.
 - Keyrus: Servicios de BI, big data y analítica avanzada.

5.7.3 Capacitación y certificaciones

Es importante capacitar a los empleados y formar un equipo especializado:

- Cursos y certificaciones:
 - Microsoft Power BI Certification: Certificaciones para dominar Power BI.
 - Tableau Certification: Certificación para especializarse en Tableau.
 - DataCamp: Cursos en línea sobre BI, ciencia de datos y análisis.
 - Coursera: Programas de BI de universidades y empresas como Google y IBM.
 - edX: Cursos sobre BI, big data y análisis de datos.
- Capacitación personalizada:

- Contratar a empresas especializadas que ofrezcan talleres personalizados.
- Participar en bootcamps intensivos sobre BI.

5.7.4 Infraestructura tecnológica

El soporte de hardware y software adecuado es esencial para implementar BI:

- Proveedores de nube:
 - AWS: Servicios para análisis y almacenamiento de datos.
 - Microsoft Azure: Infraestructura en la nube y servicios específicos para BI.
 - Google Cloud Platform: Herramientas de almacenamiento y análisis en la nube.
- Plataformas de big data:
 - Hadoop: Ecosistema para procesar grandes volúmenes de datos.
 - Apache Spark: Procesamiento de datos en tiempo real.

5.7.5 Reclutamiento de talento

Un equipo adecuado es fundamental para el éxito del departamento de BI:

- Perfiles clave:
 - Analista de BI: Responsable del análisis y visualización de datos.
 - Ingeniero de datos: Encargado de construir pipelines de datos.
 - Científico de datos: Analiza datos complejos para encontrar patrones y tendencias.
 - Gerente de BI: Coordina el departamento y alinea la estrategia con los objetivos empresariales.
- Plataformas de reclutamiento:
 - LinkedIn: Publicación de vacantes y búsqueda de perfiles especializados.
 - Glassdoor: Evaluación de candidatos con experiencia en BI.
 - Hays IT Services: Especialistas en contratación de tecnología.

5.7.6 Libros y recursos adicionales

Para líderes y empleados interesados en aprender más sobre BI:

- Libros recomendados:
 - *The Data Warehouse Toolkit* de Ralph Kimball.
 - *Storytelling with Data* de Cole Nussbaumer Knaflic.
 - *Data Strategy* de Bernard Marr.
- Comunidades y foros:
 - Kaggle: Comunidad de ciencia de datos y análisis.
 - Tableau Community: Recursos y foros para usuarios de Tableau.

- Reddit: Subreddits como r/BusinessIntelligence y r/DataEngineering.

5.8 Business Case de costes y beneficios.

Consumer Lents busca implementar un departamento de Business Intelligence (BI) para mejorar la toma de decisiones basada en datos, optimizar procesos y aumentar la competitividad en el mercado ecuatoriano. Este documento detalla los costos y beneficios asociados, así como el retorno de la inversión (ROI).

5.8.1 Objetivos del Proyecto

- Centralizar y estructurar los datos provenientes de diversas áreas (ventas, inventarios, marketing, etc.).
- Mejorar la toma de decisiones mediante reportes y dashboards en tiempo real.
- Identificar oportunidades de ahorro de costos y aumento de ingresos.
- Reducir el tiempo dedicado a análisis manual y mejorar la precisión de las proyecciones.

5.8.2 Costos Estimados

a) Costos iniciales

1. **Software de BI:**
 - Licencias de herramientas como Power BI, Tableau o Looker.
 - Power BI (20 licencias): \$3,000 anuales.
 - Alternativas como Tableau podrían costar \$5,000 anuales.
2. **Infraestructura tecnológica:**
 - Servidor en la nube (Amazon Redshift, Google BigQuery): \$6,000 anuales.
 - Almacenamiento de datos y seguridad: \$2,000 anuales.
3. **Implementación:**
 - Contratación de consultores para setup inicial (60 horas): \$15,000.
 - Diseño de ETL (Extract, Transform, Load): \$8,000.
4. **Capacitación:**
 - Cursos para empleados clave (10 personas): \$5,000.
5. **Reclutamiento:**
 - Selección y contratación de talento especializado (analistas e ingenieros de datos): \$10,000.

b) Costos recurrentes (anuales)

1. Salarios del equipo de BI:
 - Analista de BI: \$24,000.
 - Ingeniero de datos: \$30,000.
 - Gerente de BI: \$40,000.

- Total: \$94,000.
- 2. Mantenimiento de infraestructura y licencias: \$15,000.

Resumen de costos:

- Costos iniciales: \$49,000.
- Costos anuales recurrentes: \$109,000.

5.8.3 Beneficios Esperados

a) Beneficios económicos directos

1. **Optimización de inventarios:**
 - Reducción de 15% en costos de inventario debido a pronósticos precisos.
 - Ahorro estimado: \$50,000 anuales.
2. **Aumento de ventas:**
 - Identificación de productos más rentables y estrategias de cross-selling.
 - Incremento del 10% en ventas (ingreso adicional estimado): \$100,000 anuales.
3. **Reducción de costos operativos:**
 - Disminución de análisis manual y errores (20% de ahorro en tiempo de trabajo).
 - Ahorro estimado en salarios por eficiencia: \$20,000 anuales.

b) Beneficios cualitativos

1. **Toma de decisiones basada en datos:**
 - Dashboards en tiempo real para ejecutivos.
 - Mejora en la precisión de proyecciones y estrategias.
2. **Satisfacción del cliente:**
 - Personalización de ofertas y mejora en tiempos de entrega.
 - Aumento en la lealtad y retención de clientes.
3. **Ventaja competitiva:**
 - Identificación temprana de tendencias del mercado.
 - Capacidad de reaccionar rápidamente frente a cambios en la demanda.

5.8.4 Análisis de Retorno de Inversión (ROI)

Ingresos adicionales y ahorros anuales:

- Optimización de inventarios: \$50,000.
- Incremento de ventas: \$100,000.
- Reducción de costos operativos: \$20,000.

- **Total de beneficios anuales:** \$170,000.

Inversión total en el primer año:

- Costos iniciales: \$49,000.
- Costos recurrentes: \$109,000.
- **Total del primer año:** \$158,000.

ROI (a partir del segundo año):

- Beneficio neto anual: \$170,000 - \$109,000 = \$61,000.
- ROI estimado: **57% anual** a partir del segundo año.

5.8.5 Plan de Implementación

- **Fase 1: Análisis inicial (Mes 1-2):**
 - Contratación de consultores y diseño de la arquitectura BI.
 - Selección de software y herramientas.
- **Fase 2: Implementación técnica (Mes 3-5):**
 - Integración de fuentes de datos.
 - Desarrollo de pipelines de datos y dashboards iniciales.
- **Fase 3: Capacitación y pruebas (Mes 6):**
 - Entrenamiento del personal clave.
 - Pruebas y ajustes antes de la implementación total.
- **Fase 4: Lanzamiento y optimización continua (Mes 7 en adelante):**
 - Monitoreo del rendimiento del sistema.
 - Incorporación de mejoras y expansión de funcionalidades.

PARTE II

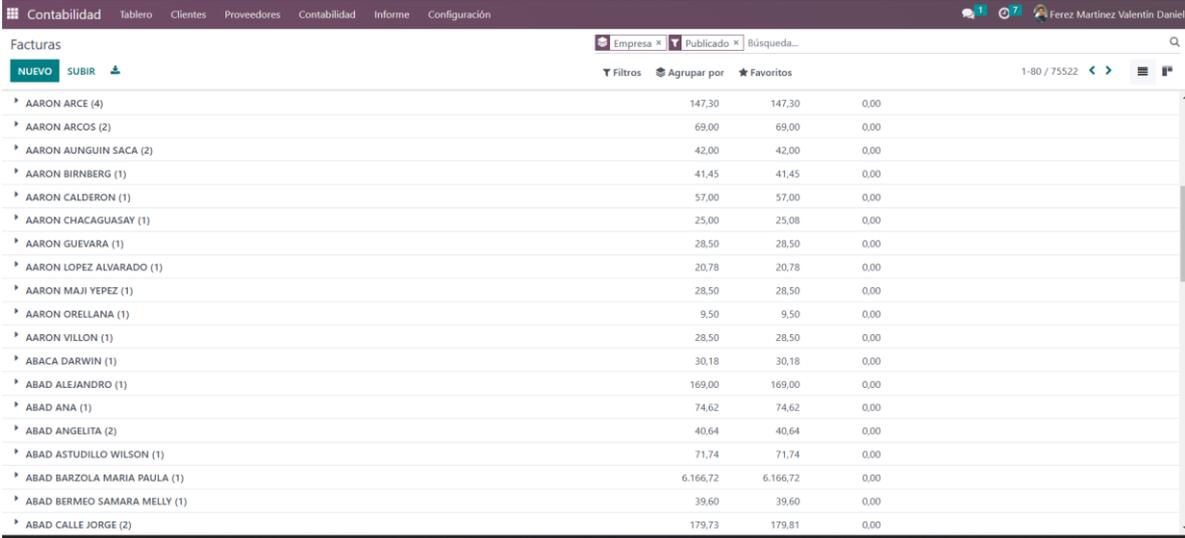
CAPITULO 1: BIG DATA

Para el Proyecto que tiene Consumer Lens con su primer cliente que es librería Studium, la información obtenida en su gran mayoría será obtenida por parte de fuentes de datos estructuradas, principalmente del ERP de la empresa.

1.1 Fuentes de información obtenidas (Estructuradas internas):

Librería Studium su ERP es Odoo un ERP bastante completo que de forma eficaz clasifica muy bien la información, la misma que es obtenida en el mismo sistema con la ayuda de filtros y otros comandos dentro del programa. El programa tiene diferentes módulos segmentando la información para que diferentes áreas de la empresa la utilicen de forma muy acertada y precisa. Módulos como contabilidad, CRM, ventas, etc.

Es importante mencionar que la información es descargable en diferentes tipos de archivos lo que permite subirlo a las otras herramientas tecnológicas que serán de ayuda para el análisis de datos y la elaboración de cuadros dinámicos y prácticos para la elaboración o supervisión de nuevas estrategias ya sean comerciales u operativas.



| Facturas | Empresa | Publicado | Búsqueda... |
|------------------------------|----------|-------------|-------------|
| NUEVO SUBIR | Filtros | Agrupar por | Favoritos |
| AARON ARCE (4) | 147,30 | 147,30 | 0,00 |
| AARON ARCOS (2) | 69,00 | 69,00 | 0,00 |
| AARON AUNGUIN SACA (2) | 42,00 | 42,00 | 0,00 |
| AARON BIRNBERG (1) | 41,45 | 41,45 | 0,00 |
| AARON CALDERON (1) | 57,00 | 57,00 | 0,00 |
| AARON CHACAGUASAY (1) | 25,00 | 25,08 | 0,00 |
| AARON GUEVARA (1) | 28,50 | 28,50 | 0,00 |
| AARON LOPEZ ALVARADO (1) | 20,78 | 20,78 | 0,00 |
| AARON MAJI YEPEZ (1) | 28,50 | 28,50 | 0,00 |
| AARON ORELLANA (1) | 9,50 | 9,50 | 0,00 |
| AARON VILLON (1) | 28,50 | 28,50 | 0,00 |
| ABACA DARWIN (1) | 30,18 | 30,18 | 0,00 |
| ABAD ALEJANDRO (1) | 169,00 | 169,00 | 0,00 |
| ABAD ANA (1) | 74,62 | 74,62 | 0,00 |
| ABAD ANGELITA (2) | 40,64 | 40,64 | 0,00 |
| ABAD ASTUDILLO WILSON (1) | 71,74 | 71,74 | 0,00 |
| ABAD BARZOLA MARIA PAULA (1) | 6.166,72 | 6.166,72 | 0,00 |
| ABAD BERMEO SAMARA MELLY (1) | 39,60 | 39,60 | 0,00 |
| ABAD CALLE JORGE (2) | 179,73 | 179,81 | 0,00 |

ILUSTRACIÓN 20: EJEMPLO DE CÓMO SE OBTIENE LA INFORMACIÓN EN BASE A FILTROS, Y AGRUPACIONES PROPIAS DE LA PLATAFORMA

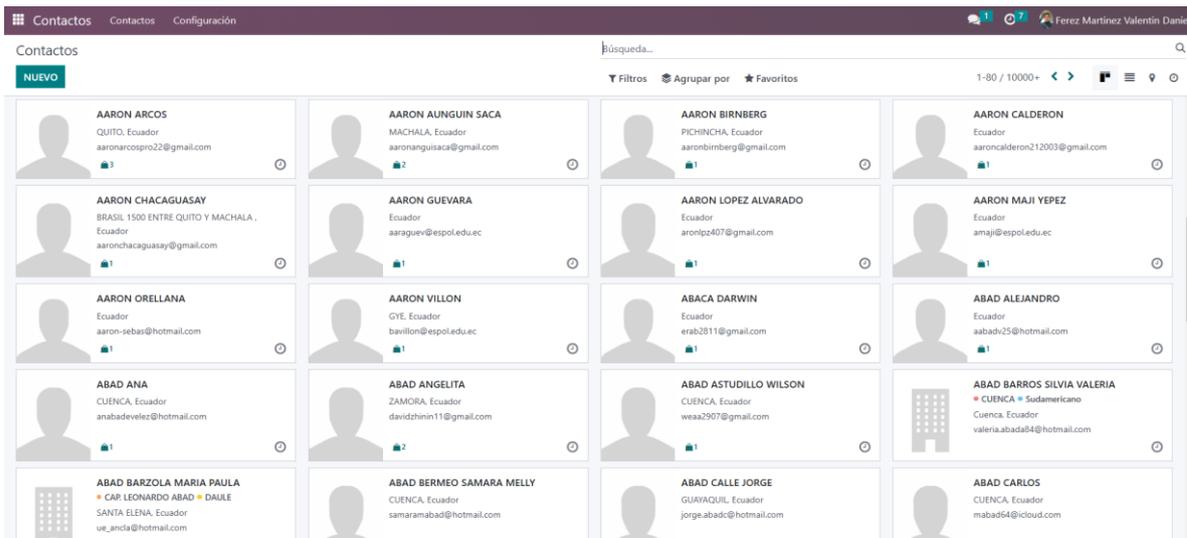


ILUSTRACIÓN 21 MÓDULO DE CONTACTOS, DETALLA INFORMACIÓN DEL CLIENTE Y SU HISTÓRICO DE COMPRAS

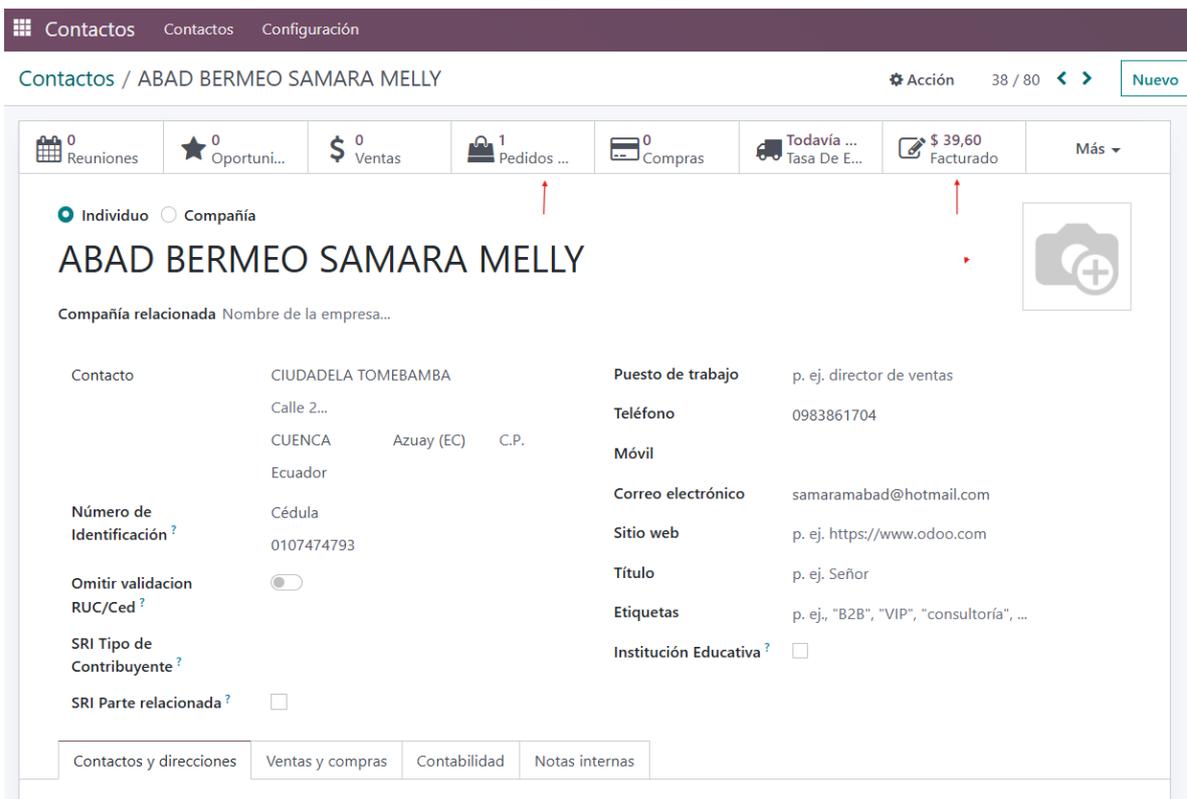


ILUSTRACIÓN 22 EJEMPLO DE CÓMO OBTENEMOS LA INFORMACIÓN DETALLADA DE QUE PRODUCTO COMPRO EN PEDIDOS Y ASÍ COMO EL HISTÓRICO DE VALOR DE COMPRA



LIBRERÍA STUDIUM CIA. LTDA.

RUC: 1791110331001

Dirección Matriz: AV. DE LA REPUBLICA
OE3-431 ENTRE AMERICA Y MAÑOSCA
QUITO

Dirección Sucursal: AV. 10 DE AGOSTO 4-
25 Y FRANCISCO MOSCOSO CUENCA
Número Contribuyente Especial: 155
Obligado a llevar contabilidad: Sí

Factura 007-057-000027455

Ambiente: Producción Emisión: Normal

Número de autorización: 2204202401179111033100120070570000274553121521418



Clave de acceso: 2204202401179111033100120070570000274553121521418

| | | | |
|---|---------------------------------|---|---------------------------------------|
| Razón Social/Nombres y Apellidos: ABAD BERMEO SAMARA MELLY Cédula: 0107474793 | Fecha de factura: 22/04/2024 | Fecha de vencimiento-Origen: 22/04/2024 Cuenca 007- 057/0926 | Reference: Cuenca 007- 057/0926 |
|---|---------------------------------|---|---------------------------------------|

| Código Principal | Descripción | Cantidad | Precio Unitario | Impuestos | Importe |
|------------------|--|----------|-----------------|-----------|----------|
| 9781108862271 | [9781108862271] AMERICAN EMPOWER STARTER A1 STUDENTS BOOK A WITH DIGITAL PACK | 1.00 | 39.600000 | IVA 0% | \$ 39.60 |

ILUSTRACIÓN 23 EJEMPLO DE CÓMO PODEMOS DESCARGAR LA FACTURA Y VER QUE COMPRO EL CLIENTE

Como podemos ver en las imágenes, esta es la forma que obtenemos información del programa que usa la empresa, esto si se maneja de forma objetiva en campañas de marketing, tendrán mejores resultados ya que las campañas van direccionadas de forma directa incentivando la compra.

1.2 Fuentes de Información (Semiestructuradas internas)

Librería Studium cambio hace más de un año de ERP por lo que disponen de bastante información en su antiguo ERP este se puede obtener la información descargándola en EXCEL, pero no está categorizada adecuadamente por lo que esta fuente de información va a requerir más trabajo.

La información en esta base de datos es muy importante porque tiene un histórico de varios años lo cual nos permite personalizar más la información de los clientes identificándolos de mejor manera.

1.3 Fuentes de información (Semiestructurada externa):

Por otro lado, tenemos las fuentes de información gubernamentales como el SRI (Servicio de Rentas Internas) Empresas como Librería Studium y otras están obligadas a facturar de forma electrónica y subir esta información de inmediato al portal del SRI para recibir la autorización de inmediato por parte de esta entidad.

Esta información es en base a lo que durante la facturación electrónica se reporta al SRI obligatoriamente, esta plataforma entrega un histórico de la misma forma, pero no está categorizado por lo que se debe añadir filtros y comandos específicos para poder interpretar y manipular la información.

1.4 Objetivos planteados

Para encontrar y definir la relación de los datos, es importante que recordemos los objetivos de este proyecto que tiene la empresa:

Objetivo 1: Enfocar la promoción de libros adelantando la venta en temporada escolar

Este objetivo es alcanzable si manejamos correctamente el historial de las compras de clientes segmentando al cliente por colegio de sus hijos y cursos, por ende, podemos antes de que se acabe el año escolar ofrecer la compra de libros del próximo año.

Por ejemplo: El señor Juan tiene un hijo en el colegio de Liga que actualmente cursa el segundo de básica, en mayo, junio se envía publicidad para que compre los libros de tercero de básica del colegio de Liga.

Beneficio para la empresa:

- Anticipa una venta lo cual le permite no utilizar el crédito de proveedores y aplicar a un descuento mayor y tener mejores condiciones de negociación.
- Reservar libros generando una mejor experiencia para el cliente.
- Mejorar el manejo de Stocks en las bodegas, no tener sobre almacenaje.
- Reducir colas y tiempo de espera del cliente que compra en los almacenes.

Objetivo 2: Aumentar la venta web de libros y test con publicidad enfocada en segmentos específicos

Después de la pandemia, la empresa empezó a notar que las ventas en línea eran una fuente de ingreso importante, de tal forma que se generó un punto de venta independiente dando como resultado uno de los puntos con mayor venta en la empresa en la actualidad, pero su mayor venta es de libros escolares.

Aprovechando la información la empresa puede realizar campañas publicitarias promocionando productos nuevos en base a los segmentos de los clientes, es decir si llegan libros de autoayuda enviar correos informando estas novedades y con un link directo de compra.

Beneficio para la empresa:

- Aumento de ventas y mayor rotación de productos.
- Negociación con proveedores para generar campañas específicas promocionando ciertas series o colecciones actuales.
- Mayor interacción con el cliente, facilitando la compra.
- Mantener al cliente informado.

- Fomentar la lectura en diferentes edades enfocándose en gustos específicos.

Objetivo 3: Crear un sistema de fidelización que premie al cliente con puntos y beneficios para futuras compras.

Para Librería Studium es muy importante fidelizar al cliente hoy por hoy la competencia es fuerte y muchas librerías son parte de corporaciones gigantes que les brindan una fortaleza económica considerable, lo cual les permite tener locales en centros comerciales, supermercados etc.

Si se crea un programa de fidelización en base a generación de puntos que les permiten acceder a descuentos especiales, así como charlas gratuitas y cursos diversos, esto fideliza al cliente al obtener un beneficio adicional por la compra recordemos que muchas compras son valores considerables en textos escolares, los mismos le dan puntos para que puedan aplicar a descuentos en libros de la preferencia del padre de familia aumentando de la misma forma las ventas.

Beneficio para la empresa:

- Fidelización de clientes
- Canal óptimo para interactuar con el cliente
- Categorización de clientes identificando cuales son los claves o tienen mayores ventas
- Generar beneficios al cliente que mejoren la imagen de la empresa.

1.5 Casos para incorporar modelo de BI:

Como hemos venido revisando, podemos ver que varios son los objetivos de la empresa los mismos que se cumplen con la incorporación de nuestro departamento de BI en la empresa. pero de forma general podemos analizar que la incorporación de este departamento será útil y óptimo para varias áreas de la empresa, permitiendo optimizar procesos y toma de decisiones de forma óptima.

Ahora con la implementación del departamento de BI la empresa puede optimizar diferentes operaciones en diferentes áreas, pero revisemos en detalle su utilidad.

- Optimización de procesos internos.
- Toma de decisiones en base a mayor información, así como más detallada.
- Prevención y visión a futuro más estructurada y con mayor probabilidad de acierto.

- Mejor resultados en la supervisión de resultados, así como revisión de estrategias, control eficaz de KPI's

1.6 Impacto de Big Data

Es importante entender que toda la información está disponible y depende de cada jefe de área y supervisor su correcto uso y manejo de la misma es por eso que este proyecto puede ser útil para todas las áreas y de diferentes formas como, por ejemplo:

- **Bodega:** optimización de rotación de Stocks control de pedidos e importaciones.
- Importaciones y compras: optimización de compras en base a históricos calculando tendencias actuales y futuras.
- **Contabilidad:** mejora del flujo de caja, optimización de procesos de cobranzas.
- **Comercial:** cumplimiento de KPI's que aumentan las ventas, información personalizada del cliente sus gustos y preferencias factor diferencial para las ventas.
- **Almacén:** aumento de ventas con ayuda de promociones y publicidad enfocada y segmentada en clientes específicos de forma personalizada, aumento de ventas web.
- **Financiero:** Mejores condiciones de negociación con el proveedor, así como los clientes de crédito, mejorando el flujo de efectivo y las proyecciones financieras.
- **Gerencia:** Optimización de toma de decisiones metas y objetivos exigentes y reales.

Finalmente, la incorporación de un departamento de Big Data y BI es factor diferencial frente a la competencia, así como una herramienta muy útil para estar a la vanguardia y poder realizar análisis profundos en base a la información que está a disposición de la empresa, recordemos que el entorno es cambiante y no adaptarse a él causa que muchas empresas cierren.

CAPITULO 2: ARQUITECTURA DE DATOS EN ENTORNOS CLOUD: SAAS

2.1 Alternativas de cloud computing

El cloud computing o computación en la nube, es una herramienta muy útil y vanguardista que permite es un modelo que permite el acceso a recursos informáticos (servidores, almacenamiento, bases de datos, redes, software) a través de Internet. Hay varias alternativas y proveedores en el ámbito del Cloud computing que se destacan en la industria por sus características y servicios ofrecidos.

Identifiquemos los tres modelos que existen de Cloud computing y sus características:

| | |
|------------------------------------|---|
| IaaS (Infrastructure as a Service) | Proporciona la infraestructura básica para que los usuarios puedan implementar y gestionar sistemas operativos, aplicaciones, y almacenamiento sin necesidad de gestionar el hardware físico. |
| PaaS (Platform as a Service) | Ofrece una plataforma completa para que los desarrolladores construyan, gestionen y desplieguen aplicaciones sin necesidad de preocuparse por la infraestructura subyacente. |
| SaaS (Software as a Service) | Ofrece aplicaciones de software completas accesibles a través de la web. El proveedor es responsable de todo, desde el hardware hasta el software. |

2.2 IaaS Infrastructure as a Service

Es un modelo muy práctico, como ventajas tiene que es fácil de personalizar y modificar, por lo que ampliar o agregar información o módulos para diferentes áreas de la empresa es muy útil. Sus costos son convenientes por que la infraestructura que requiere no es tan costosa, es un modelo que requiere estar conectados a internet siempre.

Proveedores:

| | |
|-----------------------------|--|
| Microsoft Azure | Ofrece servicios similares a AWS, con una fuerte integración con productos de Microsoft, como Windows Server, Active Directory, y SQL Server. |
| Google Cloud Platform (GCP) | Se destaca en soluciones de escalabilidad, Big Data y machine learning. |
| Amazon Web Services (AWS) | Ofrece una amplia gama de servicios de infraestructura, incluyendo Amazon EC2 (instancias de servidores virtuales) y Amazon S3 (almacenamiento). |
| IBM Cloud: | Ofrece soluciones robustas de infraestructura y herramientas para inteligencia artificial, blockchain y contenedores. |

2.3 PaaS Platform as a service

Es un modelo muy completo se enfoca en ofrecer una plataforma para que se desarrolle, gestione, construya y despliegue diferentes aplicaciones sin tener que preocuparse de la infraestructura ya establecida, Simplifica el desarrollo de aplicaciones, aunque es un modelo con poca flexibilidad y su costo no es tan conveniente.

| | |
|------------------------------|--|
| Google App Engine | Permite crear aplicaciones web y móviles utilizando varios lenguajes de programación. |
| Microsoft Azure App Services | Herramientas para desarrollar y administrar aplicaciones web. |
| Heroku: | Popular entre los desarrolladores debido a su simplicidad para implementar aplicaciones. |

2.4 SaaS Software as a Service

Es un modelo práctico que ofrece aplicaciones de software completas y accesibles a través de la web, trabaja de la mano con el hardware y software establecidos de la empresa, es de fácil acceso, así como de fácil uso y mantenimiento.

| | |
|----------------------------------|---|
| Salesforce | CRM basado en la nube, muy utilizado en ventas y marketing. |
| Google Workspace (antes G Suite) | Herramientas de productividad como Gmail, Google Docs, Sheets, Drive, etc. |
| Microsoft 365 | Incluye aplicaciones como Word, Excel, PowerPoint, Teams y OneDrive en la nube. |
| Slack | Plataforma de comunicación colaborativa en equipo. |

2.5 Selección del cloud computing

Después de analizar todas las opciones sus características ventajas y desventajas, consideramos que la mejor opción para incorporar al cliente que es Librería Studium considerando el tamaño de la empresa y que diferentes áreas dentro de la empresa en un futuro cercano van a usar y trabajar de la mano con el departamento de BI de Consumer Lens, la mejor opción es optar por una IaaS es decir infraestructura como servicio, esto porque es una herramienta que tiene una mayor escalabilidad lo cual se puede adaptar poco a poco para que se incorpore en cada área de la empresa, así mismo los proveedores que maneja trabajan bien con la arquitectura del data warehouse y del departamento de BI que ya planteamos en otros entregables, consideramos que tanto Microsoft Azure, Google Cloud Platform y Amazon Web Services son proveedores confiables que tienen excelente infraestructura y un producto idóneo para cumplir con los requisitos planteados en este proyecto.

CAPITULO 3: PRINCIPALES TECNOLOGÍAS OPEN SOURCE DEL UNIVERSO BIG DATA

Consumer Lens con su Proyecto de implementación de un departamento de BI para sus clientes, comenzando con Librería Studium consideramos muy importante la incorporación dentro de la estrategia Big Data de herramientas Open source, las mismas que son muy útiles para el manejo de información, así como de personalización de frameworks y ecosistemas de colaboración de forma simple y sin las complicaciones de desarrollo que representa el ser una herramienta Open source.

Las herramientas tienen un sin número de aplicaciones, así como de utilidades, primero la eficiencia que brinda una herramienta que no necesita niveles altos de programación, así como procesos muy estructurados, luego genera aplicaciones fáciles de usar y personalizables a los requerimientos específicos de cada área, el análisis de la información se complementa muy bien con las herramientas ya mencionadas anteriormente, en capítulos previos a este.

Al ser modificable y adaptable a los requerimientos específicos de cada área, se convierte en una herramienta que mejora continuamente adaptándose de forma rápida a los nuevos requerimientos y necesidades de la empresa, no podemos dejar de lado que estas herramientas no son costosas y no requiere de contratos demandantes que hacen que la empresa pierda muchos de los controles afectando sus finanzas, complementando el servicio brindando al ser una herramienta eficaz y eficiente que optimiza recursos.

Tomando en cuenta la información anterior, podemos hablar ahora del framework personalizable que podemos utilizar para este proyecto, un framework es un conjunto de herramientas, bibliotecas y reglas predefinidas que facilitan el desarrollo de software, el framework funciona como una estructura base sobre la cual los desarrolladores pueden construir aplicaciones sin tener que escribir todo desde cero.

El Framework brinda un sin número de ventajas como, por ejemplo:

- Código reutilizable.
- Estructura predefinida
- Estandarización del desarrollo
- Ahorro de tiempo.
- Costos bajos o moderados.

Existen diferentes tipos de frameworks, pero el que más utilizaremos es el de Backend, muy útil para manejar bases de datos así como su lógica e interpretación.

3.1 Tipos de Frameworks Open Source

Existen algunas opciones para el servicio que buscamos brindar como Consumer Lens a Librería Studium, pero analicemos a detalle cada una y cuál sería la mejor herramienta para este proyecto.

Apache Hadoop:

Es un Framework de código abierto diseñado para almacenar y procesar grandes volúmenes de datos, funciona en diferentes clusters de servidores de manera distribuida.

Fue desarrollado por Apache software foundation y se basa en el modelo de procesamiento paralelo de map reduce de Google.

- Divide los datos en fragmentos y los procesa simultáneamente en diferentes nodos.
- Puede crecer añadiendo más servidores sin afectar el rendimiento.
- Replica los datos en varios nodos para evitar pérdidas.
- Tiene capacidad de almacenamiento masivo es compatible con petabytes de datos estructurados y no estructurados.
- Al ser open source libre para modificar y se adapta a diferentes necesidades.
- Tiene actividades compartidas ideal para que diferentes módulos y áreas de la empresa funcionen unos con otros correctamente de forma enlazada.

| Ventajas | Desventajas |
|--|-------------------------------------|
| Maneja enormes volúmenes de datos | No es eficiente para datos pequeños |
| Escalable y distribuido | Requiere hardware potente |
| Resistente a fallos | Curva de aprendizaje elevada |
| Compatible con diversas fuentes de datos | Map reduce puede ser lento. |

Mongo DB:

Es una base de datos noSQL orientada a documentos, diseñada para manejar grandes volúmenes de datos de forma flexible y escalable, una herramienta que almacena los datos en JSON o BSON, Lo que le convierte es una herramienta muy ágil en la manipulación de datos.

Es una base de datos ágil, escalable y flexible, ideal para aplicaciones modernas como microservicios, big data e IoT.

- Es un modelo de datos flexible.
- Soporta grandes volúmenes de datos y consultas rápidas.
- Puede distribuir los datos en múltiples servidores fácilmente.

- Basa su almacenamiento en documentos noSQL.
- Permite replicar datos en varios servidores para evitar pérdidas.

| Ventajas | Desventajas |
|---|---|
| Modelo flexible sin esquema fijo | No es óptimo para transacciones complejas |
| Escalable horizontalmente | Usa más almacenamiento que SQL |
| Consulta y agregación potente | No soporta fácilmente relaciones complejas como SQL |
| Maneja datos semiestructurados y no estructurados | Requiere más memoria RAM para optimizar el rendimiento. |

Apache Spark:

Es un framework open source diseñado para el procesamiento rápido de datos a gran escala, es una alternativa más rápida y eficiente a Apache Hadoop, ya que permite el procesamiento en memoria Ram en lugar de depender solo del almacenamiento en disco. Es una herramienta potente para big data, machine learning y análisis de datos en tiempo real.

- Es mucho más rápido de hadoop
- Compatible con múltiples lenguajes soporta Python, java, Scala y R.
- Soporta procesamientos en tiempo real.
- Escalabilidad y tolerancia a fallos
- Integración con big Data puede trabajar con otros frameworks
- Soporta machine learning y análisis de datos complejos.

| Ventajas | Desventajas |
|---|---|
| Procesamiento en memoria rápido | Alto consumo de memoria RAM |
| Compatible con otros Frameworks como hadoop, cassandra, hbase, etc. | Configuración compleja para principiantes |
| Manejo de datos en tiempo real | Puede ser costoso en infraestructura |
| Soporte para machine learning y SQL | No incluye un Sistema de almacenamiento propio. |

3.2 Framework ideal para Consumer Lens

Después del análisis a detalle presentado, consideramos que la mejor herramienta por la que podemos optar de todas las expuestas es la de Apache Spark, consideramos que es la que mejor se adapta al proyecto por las ventajas presentadas, factores muy importantes para esta decisión es que es una herramienta flexible que no depende de un tipo de lenguaje en específico, por otro lado es una herramienta amigable que puede trabajar en conjunto con otras similares sin que se use el mismo idioma.

Aunque entendemos que una de las desventajas es que necesita bastante memoria adicional, es algo solucionable debido a los servidores que posee la empresa Librería Studium por el modelo de negocio que tienen, lo cual disminuye considerablemente esta desventaja.

Así mismo entendemos que el lenguaje, así como el manejo puede ser complicado para las y los colaboradores al principio, con un plan de capacitaciones específico y directo es algo que se puede solucionar y crear muchos beneficios a futuro.

Apache Spark con que otros elementos del entorno Big data se puede integrar para mejores resultados del proyecto.

Apache Spark es extremadamente flexible y se integra con diversas tecnologías dentro del ecosistema de Big Data, lo que lo convierte en una opción ideal para procesar, almacenar, analizar y visualizar grandes volúmenes de datos en tiempo real o por lotes. A través de estas integraciones, Spark puede realizar tareas de procesamiento por lotes (batch), streaming en tiempo real, análisis de datos estructurados y no estructurados, Machine Learning, Deep Learning y mucho más. La capacidad de interactuar con otras plataformas como Hadoop, Kafka, Hive, Delta Lake, y Cassandra permite que Spark sea aún más poderoso y versátil en entornos de Big Data.

Algunos de los frameworks con los que se puede complementar pueden ser:

- **Cassandra** es una base de datos NoSQL altamente escalable, ideal para manejar datos de alto volumen y baja latencia. Spark puede integrarse con Cassandra para realizar procesamiento distribuido de grandes volúmenes de datos y consultas en tiempo real. Esta integración se usa en aplicaciones de análisis en tiempo real y Big Data que requieren alta disponibilidad y escalabilidad.
- **Presto** es un motor de consultas distribuido para realizar consultas SQL sobre grandes volúmenes de datos. Se puede integrar con Apache Spark para permitir consultas interactivas de datos almacenados en HDFS, Cassandra, o incluso en bases de datos relacionales. La integración de Presto con Spark permite obtener tiempos de respuesta más rápidos para consultas ad hoc.
- **Delta Lake** es un almacenamiento de datos basado en Apache Parquet que permite transacciones ACID en sistemas de Big Data, proporcionando consistencia y fiabilidad en entornos de datos masivos. Se integra muy bien con Apache Spark para asegurar la integridad de los datos, permitiendo actualizaciones, eliminaciones y combinaciones de datos con garantías transaccionales (ACID).
- **Hive** proporciona un sistema de data warehouse basado en SQL para Big Data, y puede ser usado con Spark a través de Spark SQL. Esta integración permite ejecutar consultas SQL sobre grandes volúmenes de datos almacenados en HDFS o en otros sistemas de almacenamiento.

Además, Spark puede aprovechar las optimizaciones de ejecución de Hive para mejorar el rendimiento de las consultas.

CAPITULO 4: MACHINE LEARNING E INTELIGENICA ARTIFICIAL

4.1 Modelo de aplicación de data Science y machine learning

El objetivo principal de este proyecto es el de mejorar la toma de decisiones para la empresa mediante el análisis de datos optimizando las operaciones comerciales y personalizar la experiencia al cliente, aumentando el valor agregado al mejorar el servicio y de la misma forma aumentar los ingresos al crear campañas que incentiven al consumidor a comprar ofreciendo productos y ofertas que cumplan con las necesidades del cliente.

Pero al crear un departamento de BI en la empresa debemos optimizar esta inversión al generar mayores beneficios por lo que los objetivos de la empresa se vuelven varios, pero detallémoslos a continuación:

- Mejorar los pronósticos de compras anticipando las ventas y realizando pagos anticipados a los proveedores para tener mayores descuentos al no requerir de crédito, optimizar el inventario reduciendo costos por exceso o falta de stock, fomentando la venta de los productos por ofertas y campañas publicitarias.
- Análisis de clientes y segmentación, usando algoritmos de aprendizaje no supervisado segmentamos a los clientes según sus hábitos de compra, preferencias y ubicación clasificándolos en diferentes clústeres mejorando su experiencia e imagen de la empresa.
- Optimización de precios y promociones, analizando el impacto de diferentes estrategias de precios mediante modelos de regresión y series temporales, diseñando promociones efectivas según patrones de compra.
- Análisis de ventas y rendimiento del negocio. Utilizando herramientas de BI para visualizar tendencias de ventas comparando con períodos y evaluar el impacto de las campañas de marketing, identificando oportunidades de crecimiento y optimización operativa.

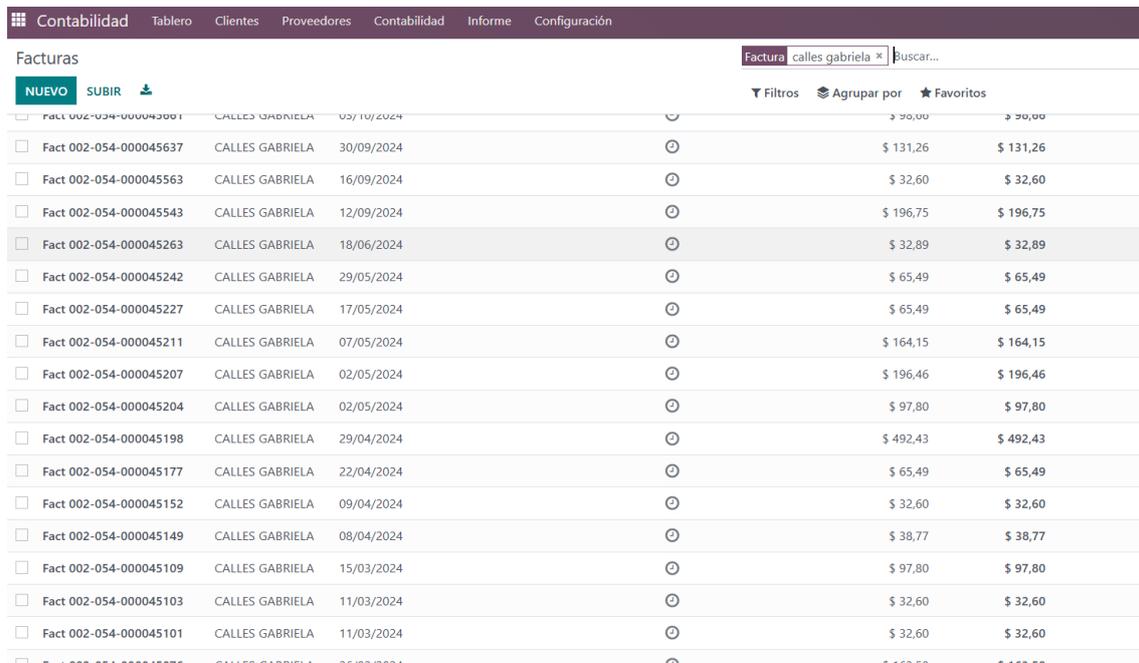
Finalmente debemos tener en claro que el enfoque basado en Data Science y Machine Learning permitirá que Librería Studium pueda tomar decisiones informadas, aumentando la eficiencia y eficacia operativa en varias áreas mejorando considerablemente la satisfacción del cliente y por ende los ingresos por ventas y optimización de costos.

4.2 Fuentes de Datos:

Para implementar el departamento de Bi y Data analytics en Librería Studium es importante usar varias fuentes de datos para poder generar perfiles más específicos y claros lo que nos lleva a clusters claros y bien detallados para enfocar los diferentes recursos y personalizar la atención de cada segmento o cluster.

Fuentes de Datos internas:

Son la principal fuente de datos de la empresa, tomemos en cuenta que es una empresa con más de 32 años en el mercado y que tiene un sistema ODOO que es bastante útil para la implementación de este proyecto ya que nos ayuda con información clara y detallada que puede ser segmentada por filtros y agrupada por clientes para especificar los requerimientos que necesitamos para generar los segmentos o clústeres que necesitamos para alguna campaña en específico.



The screenshot shows the ODOO accounting interface. At the top, there is a navigation bar with tabs for 'Contabilidad', 'Tablero', 'Clientes', 'Proveedores', 'Contabilidad', 'Informe', and 'Configuración'. Below this, the 'Facturas' (Invoices) section is active. A search bar contains 'Factura calles gabriela' and a 'Buscar...' button. There are also buttons for 'NUEVO', 'SUBIR', and a download icon. The main area displays a table of invoices with columns for invoice ID, client name, date, and amounts. The table is filtered to show only invoices for 'CALLE GABRIELA'. The table includes a 'Filtros' button, an 'Agrupar por' button, and a 'Favoritos' button. The table data is as follows:

| Factura | Cliente | Fecha | Monto | Monto |
|------------------------|----------------|------------|-----------|-----------|
| Fact 002-054-000045637 | CALLE GABRIELA | 30/09/2024 | \$ 131,26 | \$ 131,26 |
| Fact 002-054-000045563 | CALLE GABRIELA | 16/09/2024 | \$ 32,60 | \$ 32,60 |
| Fact 002-054-000045543 | CALLE GABRIELA | 12/09/2024 | \$ 196,75 | \$ 196,75 |
| Fact 002-054-000045263 | CALLE GABRIELA | 18/06/2024 | \$ 32,89 | \$ 32,89 |
| Fact 002-054-000045242 | CALLE GABRIELA | 29/05/2024 | \$ 65,49 | \$ 65,49 |
| Fact 002-054-000045227 | CALLE GABRIELA | 17/05/2024 | \$ 65,49 | \$ 65,49 |
| Fact 002-054-000045211 | CALLE GABRIELA | 07/05/2024 | \$ 164,15 | \$ 164,15 |
| Fact 002-054-000045207 | CALLE GABRIELA | 02/05/2024 | \$ 196,46 | \$ 196,46 |
| Fact 002-054-000045204 | CALLE GABRIELA | 02/05/2024 | \$ 97,80 | \$ 97,80 |
| Fact 002-054-000045198 | CALLE GABRIELA | 29/04/2024 | \$ 492,43 | \$ 492,43 |
| Fact 002-054-000045177 | CALLE GABRIELA | 22/04/2024 | \$ 65,49 | \$ 65,49 |
| Fact 002-054-000045152 | CALLE GABRIELA | 09/04/2024 | \$ 32,60 | \$ 32,60 |
| Fact 002-054-000045149 | CALLE GABRIELA | 08/04/2024 | \$ 38,77 | \$ 38,77 |
| Fact 002-054-000045109 | CALLE GABRIELA | 15/03/2024 | \$ 97,80 | \$ 97,80 |
| Fact 002-054-000045103 | CALLE GABRIELA | 11/03/2024 | \$ 32,60 | \$ 32,60 |
| Fact 002-054-000045101 | CALLE GABRIELA | 11/03/2024 | \$ 32,60 | \$ 32,60 |

ILUSTRACIÓN 24 FUENTES DE DATOS INTERNAS

Como podemos ver en la imagen el sistema nos provee de forma general el histórico de compras del cliente en este caso es Gabriela Calles, las fechas de compra, la factura y el monto de compra, información muy útil para entender cuando realiza compras.

Como podemos ver en la imagen tenemos un filtro por editorial en este caso es Cambridge, eso nos despliega diferentes productos, su precio ISBN y stock a mano.

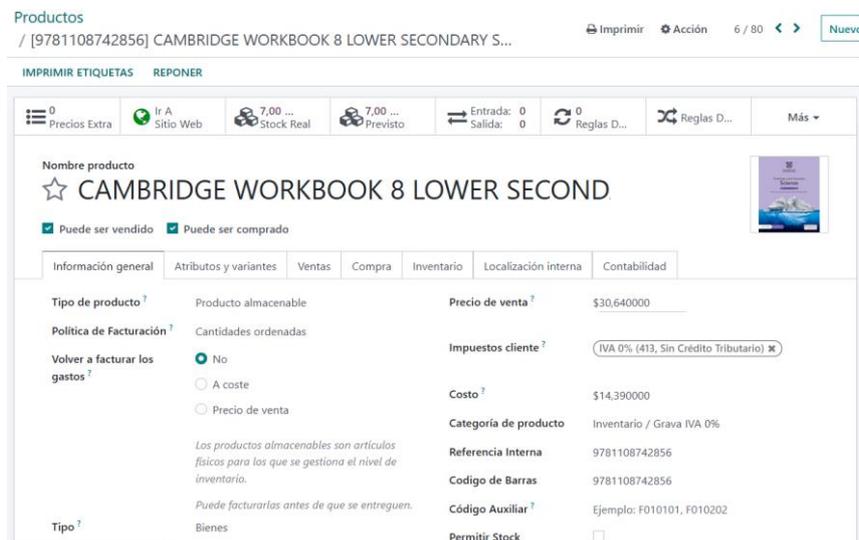


ILUSTRACIÓN 27 INFORMACIÓN DE OBTENCIÓN DE FUENTES DE DATOS INTERNAS

Sí damos click en el producto obtenemos información más específica del producto, así como cantidad en stock, costo, si tiene impuestos y si va a llegar alguna carga con este producto.

Fuentes de datos Externas:

En este caso tenemos la información de las editoriales que disponen sobre estudios de mercados, así como de análisis de ventas de sus distribuidores que son entregados obligatoriamente cada año.

4.3 Área de aplicación y departamentos de negocios implicados:

Como hemos detallado anteriormente, es muy importante por la inversión tanto en dinero como en recursos que este proyecto representa que la incorporación de la mayoría de departamentos de la empresa sea parte y utilice de forma eficaz la información y herramientas brindadas por el departamento de BI, incluso para disminuir su carga laboral y enfocarse en cometer menos errores, pero detallemos la utilidad por área.

- **Ventas:** aumento de ventas, así como optimización de ventas de la página web en la actualidad es uno de los puntos de venta más rentables y que más ventas genera, pero se debe optimizar su uso vendiendo más, de la misma forma la personalización de oferta al cliente para que tenga una mejor experiencia generando un mejor servicio y un aumento del valor agregado

- Comercial: El generar facilidades de compra, así como la anticipación de la venta ayuda mucho a que los colegios prefieran trabajar con Studium por el servicio y la facilidad de obtener los libros, haciendo que el equipo comercial obtenga más pedidos de las instituciones educativas.
- Bodega: les permite mejorar su rotación de productos, así como el control de inventarios mejorando la rapidez de entrega y eficiencia del departamento.
- Importaciones: Al tener información detallada de la rotación, así como la venta de productos por segmentación y organización de clústeres mejora el control de stock lo que hace más fácil el proceso de compra.
- Contabilidad y finanzas: Control más eficiente de cobros, así como de pagos, mejorando el flujo de caja optimización de inventarios disminuyendo los costos.
- Gerencia: Obtención de datos claros y específicos mejora la toma de decisiones, creación de campañas efectivas, segmentación y depuración de estrategias, análisis de resultados, correcciones etc.
- Marketing: mejora de campañas, campañas más específicas, medición de resultados y desempeño.

4.4 Identificación de la mejora esperada:

La implementación de Machine Learning y data science permitirá optimizar los procesos mejorando la toma de decisiones en diversas áreas del negocio, algunas de las mejoras esperadas son:

- Optimización de la gestión de inventario.
- Incremento de ventas y fidelización de clientes.
- Mejora en la planificación de estrategias de precios y promociones
- Mayor precisión en la toma de decisiones
- Mejora en la experiencia del cliente
- Reducción de riesgos y detección de anomalías.

En general la implementación de un departamento de BI y Data Análisis permitirá una mayor eficiencia operativa reducción de costos y una experiencia de compra mejorada generando un impacto positivo en la rentabilidad y crecimiento del negocio.

4.5 Cuantificación de la mejora en el área de aplicación:

- **Propuesta de valor:**

- Optimización de inventario: reducción del 20% en costos por sobre stock o falta de libros.
- Aumento de ventas: Incremento de un 20% inicialmente de ingresos mediante recomendaciones personalizadas. Mejora de la experiencia del cliente reducción de los tiempos de espera obteniendo los libros antes de lo esperado.
- **Segmentos de clientes:**
 - Análisis de clientes frecuentes aumentando la fidelización.
 - Segmentación de clientes aumentando la venta por promoción o publicidad.
- **Canales de venta y distribución**
 - Aumentar un 25% las ventas de la página web mejorando la experiencia de compra al no hacer filas y obteniendo los libros en su casa.
 - Estrategia de precios dinámicos aumentando el margen de beneficio un 10%
- **Relación con los clientes:**
 - Fidelización de clientes mejorando la experiencia de compra, automatización de recomendaciones para ofrecer libros que el cliente no conocía o no sabía que estaban disponibles en su ciudad.

CAPITULO 5: ANÁLISIS EN TIEMPO REAL. FORECAST Y DATA STREAM

5.1 Aplicación del modelo de aplicación de Real time

Comenzando con el proyecto de la incorporación de un departamento de BI para Librería Studium consideramos importante que se utilicen las herramientas de real time como una herramienta para fidelizar y crear mejores interacciones con el cliente frente a la marca de esta forma podemos aumentar las ventas por promociones enfocadas correctamente y por qué conocemos mejor al cliente sus gustos y preferencias y otra información que termina siendo trascendental al momento de la compra.

Objetivos de aplicación de Real Time:

- Fidelización de clientes por medio de un sistema de rewards por compras lo que permite identificar los gustos y preferencias de nuestros clientes dándonos un mapa para segmentar tanto nuestros clientes con los productos que se acoplan a sus requerimientos.
- Creación de clústeres y segmentos agrupando clientes por sus gustos y preferencias viendo con herramientas de tiempo real en streaming como es su comportamiento dentro de la página web así como su historial de compras y su actividad con las redes sociales de la empresa, esto permitirá entender en detalle al cliente personalizando la atención las ofertas volviendo que la marca sea el top of mind del cliente tomando en cuenta para todas sus compras de libros.
- Aumento del Engagement con el cliente, al tener herramientas de respuesta en tiempo real se optimiza la comunicación con el cliente generando una satisfacción instantánea. Facilitando los procesos de reservas de libros y posteriormente la obtención del libro para que le llegue al cliente en la comodidad de su casa sin que tenga que visitar el cliente.
- Optimización de las cobranzas de la empresa, por medio del envío de correos de forma diaria para optimizar el proceso de pago por parte del cliente.

5.2 Fuentes de datos utilizados

Para el cumplimiento de estos objetivos es muy importante que las fuentes sean confiables y accesibles en tiempo real, permitiendo que se actualicen de forma inmediata y confiable, pero detallemos que fuentes de datos son las más útiles para este proyecto.

Historial de compras: Esta información obtenemos de forma inmediata del mismos ERP de la empresa que como lo hemos detallado antes es ODOO La

información se sube de forma inmediata al terminar cada transferencia de compra, recordemos que el cliente al momento de comprar un libro entrega su información detallada como es nombre, correo, cédula, dirección, etc.

Navegación en la web: En la actualidad un 20% de las ventas de la librería se realiza en la página web, lo que permite un análisis detallado de las actividades del cliente que a su vez para su compra tienen que logearse y en el caso de que sus hijos estén en una institución educativa indicar cual y que curso eso permite a la librería en un futuro promocionar los libros del año siguiente antes de que el actual termine

Redes Sociales y Opiniones: Comentarios, reseñas y menciones sobre productos o servicios.

Sistema de fidelización: Puntos acumulados, descuentos utilizados, membresía VIP.

Disponibilidad e inventario: Datos en tiempo real de stock para ofrecer alternativas si un producto está agotado.

Fuentes de cartera de la empresa: de forma inmediata desde los 3 días antes de que la deuda llegue a su fin se enviarán correos para promover el pago por parte de los clientes de crédito.

5.3 Área de aplicación y departamentos de negocio implicados

El modelo de aplicación en tiempo real para la fidelización de clientes en una librería se implementaría dentro del área de Marketing y Experiencia del Cliente, con impacto en otros departamentos clave como Ventas, TI, Contabilidad, Logística y gerencia, pero detallemos un poco cada área.

Marketing y customer experience:

- Desarrollo de estrategias de fidelización basadas en datos en tiempo real.
- Implementación de programas de recompensas y descuentos dinámicos.
- Personalización de campañas publicitarias y promociones según el comportamiento del usuario.
- Gestión de redes sociales y análisis de comentarios/reseñas.

Ventas y Atención al Cliente

- Uso de datos en tiempo real para recomendaciones personalizadas en la tienda física y online.
- Implementación de chatbots y asistentes virtuales para resolver dudas al instante.

- Seguimiento de clientes habituales y activación de notificaciones de promociones o lanzamientos.

Tecnología y Desarrollo (TI)

- Desarrollo e integración del sistema en la web, app y sistemas internos.
- Implementación de motores de recomendación y procesamiento de datos en tiempo real.
- Seguridad de datos y cumplimiento de normativas de privacidad.
- Gestión de plataformas de geolocalización y análisis de tráfico web.

Logística e Inventario

- Sincronización del stock en tiempo real para evitar sugerencias de productos agotados.
- Análisis predictivo para la reposición de libros en función de tendencias de compra.
- Integración con proveedores y sistemas de distribución para optimizar entregas.

Gerencia, Finanzas y Análisis de Datos

- Evaluación del impacto de las estrategias de fidelización en las ventas.
- Cálculo del ROI de las promociones en tiempo real.
- Análisis de comportamiento del cliente y segmentación avanzada.
- Análisis y control de resultados y estrategias, así como el cumplimiento de objetivos.

5.4 Identificación de la mejora esperada

La aplicación de un modelo en tiempo real para la fidelización de clientes en la librería generará mejoras clave en diferentes áreas:

Experiencia del Cliente Personalizada

- Recomendaciones más precisas y en tiempo real basadas en su historial de compras y preferencias.
- Notificaciones y ofertas instantáneas cuando un cliente está en la tienda o navegando en la web/app.
- Interacción rápida con chatbots o asistentes virtuales, mejorando la atención y resolución de dudas.

Aumento en la Fidelización y Retención de Clientes

- Mayor engagement a través de programas de lealtad más dinámicos y atractivos.
- Incentivos personalizados, como descuentos en sus géneros favoritos o puntos adicionales por compras recurrentes.
- Reducción de la tasa de abandono al ofrecer promociones oportunas antes de que el cliente deje la compra incompleta.

Incremento en las Ventas y Ticket Promedio

- Cross-selling y up-selling en tiempo real, recomendando productos complementarios o ediciones especiales.
- Promociones dinámicas basadas en el comportamiento del usuario (ejemplo: descuento en libros de un autor favorito si el usuario ha comprado varios títulos similares).
- Menos productos sin vender, gracias a estrategias de ofertas inteligentes que promuevan stock no rotativo.

Optimización de la Gestión de Inventarios

- Reposición más eficiente, ya que el sistema detectará tendencias de compra en tiempo real.
- Reducción del desabastecimiento, enviando alertas de stock bajo antes de que un cliente quiera comprar un producto agotado.

Mejor Toma de Decisiones Basada en Datos

- Análisis en tiempo real de las tendencias de compra para mejorar estrategias de ventas.
- Segmentación avanzada de clientes para diseñar campañas más efectivas.
- Monitoreo de la efectividad de promociones y ajustes instantáneos según el desempeño.

Mejor gestión de cobranza:

- Optimización de los tiempos de cobranza cumpliendo con los tiempos de crédito sin generar demoras.
- Mejor flujo de caja y manejo de efectivo.

5.5 Cuantificación de la mejora en el área de aplicación a través del business case

5.5.1 Resumen Ejecutivo

Este proyecto busca mejorar la fidelización de clientes en una librería mediante un sistema en tiempo real que ofrezca recomendaciones personalizadas, promociones dinámicas y comunicación interactiva. Se espera aumentar la

retención de clientes, incrementar las ventas y optimizar la gestión del inventario mediante la aplicación de inteligencia de datos en tiempo real.

5.5.2 Situación Actual y Oportunidad de Mejora

Actualmente, la librería enfrenta los siguientes desafíos:

- Baja retención de clientes: Solo el 30% de los clientes vuelve a comprar en los primeros 6 meses.
- Promociones poco efectivas: Solo el 15% de los clientes usa cupones o descuentos.
- Ventas no optimizadas: Falta de personalización en las recomendaciones, lo que reduce el ticket promedio.
- Problemas de stock: Productos agotados sin notificación previa, lo que genera pérdida de ventas.

El modelo en tiempo real busca abordar estos problemas con una estrategia más dinámica y centrada en el cliente.

5.5.3 Beneficios Esperados (Cuantificación de la Mejora)

| Indicador | Situación Actual | Situación Esperada | Mejora (%) |
|--------------------------------|----------------------------|---------------------------|------------|
| Tasa de retención de clientes | 30% | 50% | +66% |
| Redención de promociones | 15% | 35% | +133% |
| Incremento del ticket promedio | \$25 USD | \$35 USD | +40% |
| Crecimiento en ventas anuales | \$1,000,000 USD | \$1,300,000 USD | +30% |
| Reducción de quiebres de stock | 10% de productos sin stock | 4% de productos sin stock | -60% |

- Impacto financiero estimado: Un aumento del 30% en las ventas, lo que equivale a \$300,000 USD adicionales anuales.
- Reducción de pérdidas por stock agotado: Aproximadamente \$50,000 USD menos en pérdidas por productos no disponibles.

5.5.4 Costos del Proyecto

- Para la implementación del sistema en tiempo real, se requerirá inversión en tecnología y operaciones.

| Concepto | Costo Estimado (USD) |
|--------------------------------------|----------------------|
| Desarrollo e integración del sistema | \$80,000 |
| Infraestructura tecnológica | \$30,000 |
| Capacitación del personal | \$10,000 |
| Costos operativos anuales | \$20,000 |
| Total inversión inicial | \$140,000 |

5.5.5 Retorno de Inversión (ROI)

El ROI esperado se calcula considerando el beneficio neto sobre la inversión inicial:

- Beneficio neto anual: \$300,000 USD - \$20,000 USD (costos operativos) = \$280,000 USD
- Inversión inicial: \$140,000 USD
- Retorno esperado del 200% en el primer año, con recuperación de la inversión en 6 meses.

5.5.6 Conclusión y Recomendación

La implementación del modelo en tiempo real para la fidelización de clientes tiene un alto potencial de retorno y una rápida recuperación de la inversión. Con mejoras en la experiencia del cliente, ventas y eficiencia operativa, se recomienda avanzar con el proyecto para maximizar la competitividad de la librería en el mercado.

CONCLUSIONES Y APLICACIONES

La implementación de Business Intelligence en el proyecto *Consumer Lens* ha optimizado significativamente el proceso de toma de decisiones empresariales, permitiendo analizar grandes volúmenes de datos en tiempo real y mejorar la eficiencia operativa en distintas áreas clave como logística, inventario, contabilidad y comercial. Esto ha derivado en una reducción de costos y un aumento en la productividad. Además, la personalización de la experiencia del cliente ha sido un aspecto clave, ya que la integración de BI ha permitido segmentar mejor a los consumidores, personalizar ofertas y mejorar su fidelización a través de estrategias basadas en datos.

Otro aspecto fundamental ha sido el incremento en las ventas, con una proyección de crecimiento del 30% anual, impulsado por una mejor gestión de inventarios y la personalización en la oferta de productos. Asimismo, el uso de análisis predictivo ha reducido de manera considerable las pérdidas derivadas de quiebres de stock, lo que se ha traducido en un ahorro estimado de \$50,000 USD anuales. En términos tecnológicos, la adopción de plataformas de BI en la nube ha mejorado la accesibilidad y eficiencia en el procesamiento de información, permitiendo reducir tiempos de respuesta y optimizar la infraestructura tecnológica.

Finalmente, la evaluación financiera del proyecto ha demostrado un retorno de inversión (ROI) altamente positivo. Se estima que la inversión inicial podrá ser recuperada en menos de seis meses, con un ROI del 200% en el primer año de implementación. La correcta integración de herramientas como Google Big Query y Power BI ha sido clave para lograr estos resultados, asegurando que la empresa pueda tomar decisiones estratégicas basadas en datos confiables y precisos.

Una de las aplicaciones más destacadas del proyecto es la optimización de ventas y estrategias comerciales. La implementación de *Consumer Lens* ha mejorado la rotación de inventarios y ha permitido diseñar campañas publicitarias más eficientes, aumentando la conversión de clientes y maximizando la rentabilidad. A través del análisis de datos, se han identificado patrones de compra y segmentaciones específicas que han permitido desarrollar estrategias de marketing personalizadas, incrementando la lealtad del cliente y reduciendo el desperdicio de recursos en campañas menos efectivas.

Otra aplicación clave es la automatización de reportes y monitoreo en tiempo real, lo que ha mejorado la gestión operativa de la empresa. La implementación de dashboards de BI ha permitido visualizar en tiempo real el desempeño de cada área, facilitando la supervisión de KPI's críticos y reduciendo errores en la interpretación de datos. Esto ha impactado positivamente en la eficiencia de la toma de decisiones y ha permitido una mayor alineación entre los objetivos estratégicos de la empresa y la ejecución operativa.

REFERENCIAS

- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley.
- Marr, B. (2017). *Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things*. Kogan Page.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.
- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.
- Nussbaumer Knaflic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail - but Some Don't*. Penguin Books.
- Camm, J., Cochran, J., Fry, M., Ohlmann, J., Anderson, D., Sweeney, D., & Williams, T. (2020). *Business Analytics*. Cengage Learning.
- Westerman, G. (2014). *Leading Digital: Turning Technology into Business Transformation*. Harvard Business Review Press.
- Schmarzo, B. (2013). *Big Data: Understanding How Data Powers Big Business*. Wiley.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson.
- Chen, M., Mao, S., & Liu, Y. (2014). *Big Data: A Survey*. ACM Computing Surveys.
- Morales, M., & Guerrero, C. (2017). *Gestión de datos y toma de decisiones en empresas modernas*. Alfaomega.

Turban, E., Pollard, C., & Wood, G. (2017). Information Technology for Management: On-Demand Strategies for Performance, Growth and Sustainability. Wiley.

Segura, J. (2018). La revolución del Big Data en los negocios. McGraw-Hill.

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. Wiley.

Hammer, M., & Champy, J. (2001). Reengineering the Corporation: A Manifesto for Business Revolution. HarperBusiness.