



UNIVERSIDAD NACIONAL
DE
MAR DEL PLATA



FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INFORMÁTICA

**“Implementación de una herramienta de Data
Analytics para la toma de decisiones y aplicación
de Business Intelligence para la predicción en una
PyME local”**

Autores

Franco Martucci (franco.martucci.21@gmail.com)

Jessica Elizabeth Miguel (miguel.jessica.e@gmail.com)

Director

MBA Lic. Fernando Genin

Proyecto Final para optar al grado de Ingeniero en Informática

Mar del Plata, marzo del 2024



RINFI es desarrollado por la Biblioteca de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata.

Tiene como objetivo recopilar, organizar, gestionar, difundir y preservar documentos digitales en Ingeniería, Ciencia y Tecnología de Materiales y Ciencias Afines.

A través del Acceso Abierto, se pretende aumentar la visibilidad y el impacto de los resultados de la investigación, asumiendo las políticas y cumpliendo con los protocolos y estándares internacionales para la interoperabilidad entre repositorios



Esta obra está bajo una [Licencia Creative Commons Atribución- NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).



UNIVERSIDAD NACIONAL
DE
MAR DEL PLATA



FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INFORMÁTICA

**“Implementación de una herramienta de Data
Analytics para la toma de decisiones y aplicación
de Business Intelligence para la predicción en una
PyME local”**

Autores

Franco Martucci (franco.martucci.21@gmail.com)

Jessica Elizabeth Miguel (miguel.jessica.e@gmail.com)

Director

MBA Lic. Fernando Genin

Proyecto Final para optar al grado de Ingeniero en Informática

Mar del Plata, marzo del 2024

Índice

1. Agradecimientos.....	1
2. Resumen del proyecto.....	2
2.1. Problemas detectados.....	2
2.2. Soluciones implementadas.....	2
2.3. Beneficios en la organización.....	2
3. Introducción.....	3
4. Gestión del proyecto.....	5
4.1. Objetivos.....	6
4.1.1. Objetivo general.....	6
4.2. Análisis FODA.....	8
4.3. Metodología de Trabajo.....	10
4.3.1. Proceso y gestión de tareas.....	11
4.3.2. Gestión de versiones.....	13
4.3.3. Integración continua.....	14
4.3.4. Ambiente de desarrollo.....	14
4.4. Planificación inicial y estimaciones previas.....	16
4.5. Análisis de los tiempos finales vs estimados.....	21
5. Marco de referencia.....	22
5.1. Los datos como activo estratégico.....	22
5.1.1. Calidad de los datos.....	22
5.2. Ingeniería de datos.....	23
5.2.1. Ciclo de vida.....	23
5.2.2. Sistemas origen de los datos.....	24
5.2.3. Almacenamiento.....	25
5.2.4. Ingesta.....	25
5.2.5. Transformación.....	26
5.2.6. Disponibilización o despliegue de los datos.....	27
5.3. Indicadores claves de desempeño y de resultado (KPI y KRI).....	28
5.4. Inteligencia de negocio y análisis de datos.....	28
5.4.1. Series de tiempo y su aplicación.....	30
5.4.1.1. Algunos algoritmos de predicción.....	31

5.5. Transformación digital en organizaciones y empresas.....	33
5.5.1. Datos como impulsores de la transformación digital.....	35
5.5.2. Desafíos a nivel organizacional al aplicar una estrategia de datos.....	36
6. Sistema vs software comercial de visualización.....	37
7. Análisis de madurez.....	40
7.1. Frameworks de madurez de datos, etapas y aspectos principales.....	40
7.2. Análisis de madurez de datos para Embalajes Marplatenses S.A.....	42
7.3. Conclusiones a partir del análisis.....	43
8. El sistema.....	44
8.1. La solución propuesta a alto nivel.....	45
8.2. Backend.....	45
8.2.1. Respuesta a los interrogantes planteados en el marco teórico.....	45
8.2.2. Estructura del backend.....	47
8.2.3. Transformaciones en los datos y obtención de indicadores.....	49
8.3. Frontend.....	51
8.3.1. Objetivos de la Interfaz de Usuario.....	52
8.3.2. El proyecto frontend.....	55
8.3.3. Ejemplificando el funcionamiento de una página.....	57
8.4. Seguridad.....	57
9. Algoritmos de predicción.....	58
9.1. Análisis exploratorio.....	58
9.2. Elección de los algoritmos de predicción.....	63
9.3. Resultados obtenidos.....	64
10. Retrospectiva del proyecto.....	75
10.1. Estimaciones, objetivos y planificación.....	75
10.2. Alcance de los objetivos.....	76
10.3. Trabajo en equipo, metodología y contacto con el cliente.....	77
11. Conclusiones.....	78
12. Glosario.....	82
13. Bibliografía.....	86

1. Agradecimientos

En primer lugar, deseamos expresar nuestro profundo agradecimiento a la educación pública y, en particular, a la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata. No podemos poner en palabras lo que esta institución ha significado para nosotros, tanto en términos personales como académicos.

A nuestros estimados docentes, queremos manifestar nuestra gratitud por su calidez humana y la excelencia profesional que demostraron en cada clase. Valoramos enormemente todo lo que compartieron con nosotros y la forma en que nos inspiraron.

A nuestros compañeros de estudio, les agradecemos por los momentos compartidos durante nuestro camino académico. Los mates y las risas hicieron que el camino fuera mucho más llevadero.

A nuestro director, Fernando Genin, por haber confiado en nosotros para el desarrollo de este proyecto. Su apoyo constante, orientación y contribuciones invaluable han sido determinantes en cada etapa de este viaje. Nos sentimos afortunados de contar con la sabiduría y el conocimiento que nos ha transmitido.

Al referente funcional y gerente de la empresa Embalajes Marplatenses, Mariano Roveretti, por su tiempo y acompañamiento en la realización del proyecto, aportando su experiencia y conocimiento en el ámbito financiero.

Queremos extender un agradecimiento especial a nuestras familias y amigos, quienes han sido un pilar fundamental en esta carrera que elegimos. Gracias por brindarnos ánimo en los momentos de frustración, por cada palabra de aliento, por darnos confianza cuando las cosas se volvían difíciles, por los momentos de estudio en donde nunca nos faltó un mate o un café. Cada gesto y palabra nos hizo sentir que todo era posible. Estamos agradecidos porque hayan estado presentes en cada celebración de un examen aprobado, por los abrazos y por esas palabras que siempre nos impulsaron. Esperamos que se sientan orgullosos por lo que hemos logrado juntos.

Muchas gracias a todos.

2. Resumen del proyecto

2.1. Problemas detectados

Este proyecto surge como respuesta a varios puntos de dolor identificados en la empresa Embalajes Marplatenses S.A., ubicada en Mar del Plata. Estos desafíos incluyen la generación manual de reportes y elementos visuales propensos a errores, la multiplicación de información a partir de tableros no automatizados, la toma de decisiones en situaciones de incertidumbre y la inversión de tiempo por parte del gerente en el mantenimiento de herramientas manuales. La perspectiva aportada por una formación técnica e ingenieril permitió identificar estas áreas problemáticas. Para sintetizar, las dificultades identificadas se centran en obstáculos para la toma de decisiones a nivel gerencial, indicando una falta de fundamentos en el modelo operativo para la toma de decisiones.

2.2. Soluciones implementadas

Para abordar estos problemas, se ideó una solución a través de una plataforma web, encargada de procesar los datos y generar los elementos de información necesarios para la toma de decisiones. Trabajando en colaboración con el director de este proyecto y el gerente de Embalajes Marplatenses S.A., se identificaron factores críticos como los *stakeholders* (involucrados), los recursos disponibles y los riesgos potenciales.

Se trabajó en múltiples iteraciones dentro de un marco de trabajo ágil, donde se validaron constantemente los resultados del sistema con el referente funcional y gerente de la empresa, Lic. Mariano Roveretti. La validación se efectuó tanto en términos de usabilidad como de indicadores obtenidos. Una vez alcanzados los primeros hitos del sistema, se lanzó una primera versión que Embalajes Marplatenses S.A. pudo incorporar a sus análisis financieros.

2.3. Beneficios en la organización

Este proyecto, con bases en el desarrollo de un sistema, no solo proporciona una solución a nivel de software, sino que también introduce un cambio en los procedimientos internos de la empresa. Representa una transformación en la forma en que se manejan las operaciones. Este proyecto implica una nueva perspectiva en relación a los datos, resaltando la importancia de considerarlos como un activo estratégico. La transformación de datos en información valiosa modifica la manera en que Embalajes Marplatenses S.A. opera, elevando su nivel de madurez hacia una mentalidad orientada por los datos (*Data-Driven*), donde las decisiones se basan en evidencia y análisis sólidos.

3. Introducción

Embalajes Marplatenses S.A. es una PyME ubicada en la ciudad de Mar del Plata, especializada en la producción de cajas de cartón corrugado desde 2018. La empresa^[1] abastece a numerosas compañías en el mercado local y sus alrededores. Su instalación abarca un terreno de 30.000 metros cuadrados, de los cuales 10.000 corresponden a la nave industrial. Posee un inmenso potencial debido a su ubicación estratégica. Es la única planta productora de cajas de cartón corrugado en la ciudad y sus cercanías, lo que simplifica y reduce los costos de transporte de sus productos. Está situada en el Parque Industrial Gral. Savio, en la RP 88 km 6.5, calle 6 esquina 7.



La visión de la compañía es convertirse en líder en el sector de envases de cartón, con planes de expansión en el mercado de microcorrugado. Su misión es ofrecer a sus clientes una asociación estratégica, proporcionándoles soluciones innovadoras, prácticas y funcionales. Además, Embalajes Marplatenses S.A. tiene la ambición de expandirse en el futuro, por lo que la implementación de una solución de gestión de datos se vuelve esencial para lograr estos planes prometedores.





Actualmente, Embalajes Marplatenses S.A. opera con un sistema de gestión que almacena una amplia gama de datos relacionados con facturación, ventas, órdenes de producción, cheques, cobranzas, proveedores, clientes y niveles de inventario. Este sistema incluye una función de generación de informes que permite exportar datos de diferentes áreas y períodos. Anteriormente, el gerente de la empresa dependía de estos informes para crear tableros de control que le permitieran analizar la salud financiera de la compañía y tomar decisiones informadas. Este proceso presentaba numerosos problemas, además de requerir una gran cantidad de tiempo.

La herramienta desarrollada ha transformado esta situación. En la actualidad, el tiempo que antes se dedicaba a la laboriosa tarea de construir y mantener estos **tableros ahora se invierte en actividades estratégicas, de análisis financiero y en la creación de planes de acción.** Esto significa que el gerente puede enfocarse en decisiones más importantes en lugar de lidiar con la creación y actualización constante de tableros, que rápidamente quedaban desactualizados.

Además, este sistema sienta las bases para que Embalajes Marplatenses S.A. adopte un enfoque orientado por datos (*data driven*). Los esfuerzos de este proyecto se han centrado en crear una única fuente de información precisa para la consulta y la toma de decisiones, evitando la duplicación de datos y mitigando la pérdida de información que solía ocurrir cuando los tableros se almacenaban localmente.

4. Gestión del proyecto

El proyecto dio su puntapié inicial con la puesta en común de las partes, reuniéndose varias veces el equipo y el referente funcional. Algunas de estas reuniones tuvieron lugar en la empresa, para conocer de primera mano su funcionamiento. Se pudo relevar información acerca de la cultura y los procesos de trabajo, así como también de sus principales puntos de dolor. Durante esta etapa inicial, el foco estuvo en identificar los objetivos del proyecto y lograr un consenso en cuanto al alcance de la solución. El plan de trabajo marcó el punto de partida para la creación de un diagrama de Gantt, en el que se detallaron las actividades necesarias para alcanzar los objetivos.

En un entorno de avances tecnológicos constantes y cambios en las necesidades empresariales, la adaptabilidad del equipo es crucial. Sin embargo, la inexperiencia en planificación, estimaciones de tiempos, habilidades de negociación y las limitaciones de agenda por compromisos laborales añaden un desafío adicional a la capacidad de respuesta del equipo ante estos cambios. Como consecuencia, el equipo se descoordinó en una etapa temprana del proyecto, desencadenando en una falta de cumplimiento con lo pactado con el cliente en determinadas instancias y en cancelar sesiones de revisión en donde se debían presentar avances. La sinergia de trabajo que los integrantes presentaban previo al inicio del proyecto los llevó a confiarse y subestimar la cobertura de los entregables.

Inicialmente, el proceso de relevamiento de requerimientos no fue una tarea que se desviara demasiado de los tiempos esperados, pero al momento de comenzar con las implementaciones, comenzaron a surgir los primeros problemas, siendo necesario replanificar en varias instancias.

Ejemplificando estos escenarios, al comienzo se planteó construir prototipos del sistema sin funcionalidad que provean un primer acercamiento al diseño. A partir de estos prototipos es que surgió desde el negocio la necesidad de rediseñar algunos indicadores, y en otros casos, anexas funcionalidad.

Otro ejemplo es el caso de los resúmenes de venta. Se requirió incorporar un ABM de artículos no contemplado en la estimación de tareas inicial. Los artículos juegan un papel fundamental en la confección de indicadores de venta, dado que el negocio crea un nuevo artículo según los clientes que posee en su cartera. Esta regla no fue identificada en el proceso de relevamiento, por lo que demandó horas adicionales para elicitar, diseñar e implementar una solución.

En este punto fue necesario realizar replanificaciones, priorizaciones e incluso invertir más tiempo del estimado inicialmente para que el proyecto, así como también el ritmo de trabajo, vuelvan a encarrilarse.

Considerando estos cambios, el proceso de replanificación constó de priorizar nuevos requerimientos. A su vez, volver a priorizar requerimientos ya relevados negociando con el cliente, tomando en cuenta qué objetivos consideraba esenciales. De esta forma, se intercambiaron tareas en el *backlog* tratando de preservar lo más posible los tiempos estimados.

El aprendizaje más significativo proviene de haber enfrentado desafíos de gestión en las primeras etapas del proyecto. El haberlos identificado, replanificado y reorganizado para restablecer los plazos permitió, al mismo tiempo, recuperar la confianza del cliente.

Como resultado, la planificación se volvió un proceso medianamente continuo y adaptable. Estuvo sujeta a ajustes que permitieron reorganizar y modificar el plan según las circunstancias cambiantes y la inexperiencia del equipo en temas de gestión y liderazgo. Esta capacidad fue esencial para cumplir los objetivos en un entorno tan dinámico.

Al haber presentado el protocolo con cierto grado de avance en las tareas, el diagrama de Gantt incluido en esa entrega no sufrió demasiadas modificaciones, dado que la necesidad de replanificar y reordenar el proyecto fue en etapas previas.

4.1. Objetivos

Inicialmente, el objetivo de este proyecto fue solventar las necesidades de integración de datos, disminuyendo la incertidumbre respecto de la toma de decisiones a nivel gerencial. Estas decisiones son fundamentales para definir futuros planes de acción. Por otra parte, generar herramientas de predicción que les permita a los mandos medios contemplar posibles comportamientos financieros atípicos o anómalos.

Después de iniciar el proceso de relevamiento y tras llegar a un acuerdo con el referente funcional, este objetivo se materializó en la construcción de un sistema que permita la recolección, procesamiento y visualización de datos para la toma de decisiones en la capa gerencial de la organización. Esto se complementó con la implementación de algoritmos de pronóstico para analizar las ventas futuras.

4.1.1. Objetivo general

El objetivo principal del proyecto implicó la creación de perfiles económicos y financieros de los clientes de la PyME y la generación de gráficos para representar diversas clasificaciones de datos relevantes según las necesidades corporativas. Esto proporciona una base sólida para evaluar los comportamientos financieros y económicos y respaldar la toma de decisiones.

Adicionalmente, el referente funcional tenía un interés particular en analizar las ventas según los materiales utilizados en la fabricación de las cajas. Esta perspectiva enriquece

significativamente el análisis financiero, ya que aborda aspectos clave como las ventas, los ingresos y los plazos de cobro, elementos cruciales en un entorno marcado por la inflación y la necesidad de tomar decisiones rápidas.

Por otro lado, con miras a ampliar la perspectiva futura de las ventas, surgió la necesidad de incorporar un modelo de predicción. Este modelo toma en consideración la facturación en pesos argentinos y la cantidad de metros cuadrados de material utilizado desde febrero de 2022 hasta el momento de la consulta, y proyecta las ventas para los próximos meses. La implementación de este algoritmo de pronóstico complementa el sistema. Aprovecha las tendencias actuales de *business intelligence* que se encuentran en pleno auge y además permite la evaluación de la calidad y cantidad de datos para su implementación. No obstante, llevar a cabo un análisis detallado y exhaustivo de la aplicación de estas técnicas está fuera del alcance de los objetivos establecidos para este proyecto. Un examen más profundo demandaría la asignación de recursos adicionales y podría transformarse en un proyecto independiente debido a su magnitud y complejidad.

Dadas las características mencionadas del proyecto, donde se destaca su enfoque financiero, resultó esencial que los integrantes del equipo desarrollaran habilidades de gestión y liderazgo. Esto permitió negociar efectivamente con el referente funcional para adaptar los requerimientos del proyecto según fuese necesario. **La necesidad del cliente de tomar decisiones rápidas se transformó en una capacidad que el equipo también tuvo que adquirir, resultando en una competencia indispensable. El referente requería respuestas inmediatas para sus necesidades.**

Este proceso de gestión implicó no sólo la comprensión de los aspectos técnicos del proyecto, sino también la capacidad de comunicarse eficazmente con el demandante y otros interesados, gestionar el tiempo y los recursos de manera eficiente, y resolver conflictos de manera constructiva. **Además, la habilidad para adaptarse a cambios inesperados y mantener un enfoque orientado a resultados fue fundamental para garantizar el éxito del proyecto.**

4.1.2. *Objetivos específicos*

A partir de lo mencionado anteriormente y tras un proceso continuo de recopilación y análisis de requerimientos, se han establecido los siguientes objetivos específicos:

1. Desarrollo de una aplicación web de fácil acceso para los mandos medios, permitiendo la consulta de datos y la ingesta de informes del sistema contable de la empresa, que cuenta con una función de generación de informes en formato XLS.
2. Creación de una única fuente confiable de datos, desde la cual se realicen consultas y se originen decisiones respaldadas por evidencia.

3. Construcción de indicadores clave de desempeño empresarial, relevantes para la dirección de la empresa. Estos indicadores abarcan aspectos como:
 - Tiempo promedio desde la recepción de cheques hasta su liquidación.
 - Tiempo promedio que un cliente demora en pagar sus facturas.
 - Comportamiento promedio de facturación y cobranza por cliente.
 - Clasificación de clientes según el monto de facturación y cobranza en un período específico.
 - Monto total adeudado a la fecha.
 - Cantidad de facturas pendientes.
 - Factura más antigua sin pagar.
 - Porcentaje de metros vendidos por material y por período.
 - Porcentaje del monto vendido por material y por período.
 - Total de metros y monto a nivel de planta por período.
 - Total de cajas de cartón corrugado producidas por período.
 - Histórico de cobranzas de los últimos 12 meses.
 - Histórico de facturación de los últimos 12 meses, incluyendo y excluyendo el IVA.
4. Utilización de los indicadores mencionados anteriormente para crear un perfil de cliente que permita su clasificación en función de los tiempos de liquidación, tanto a nivel de cheques como de pagos. Esta clasificación se basará en los criterios de buenos, promedio o inferiores a lo deseado, según lo determinen los mandos medios. Será visible tanto a nivel general (lista de todos los clientes) como a nivel individual por cliente.
5. Desarrollo de un módulo de análisis enfocado en el tipo de embalaje, es decir, en el análisis de las ventas de artículos específicos.
6. Implementación de un modelo predictivo basado en series temporales para el análisis de las ventas y metros cuadrados vendidos en cada período.

4.2. *Análisis FODA*

Fortalezas:

- Disponibilidad del gerente de la empresa para responder dudas sobre el negocio.
- Buena comunicación y sinergia entre los miembros del equipo.
- El software es una solución personalizada, lo que asegura que satisfará las necesidades del cliente.
- El director del proyecto cuenta con una sólida experiencia en los principios de transformación digital dentro de las organizaciones.

Oportunidades:

- Transformación cultural y operativa a nivel empresarial con posibilidad de: ahorrar tiempo a nivel gerencial dado que no se deben crear los tableros manuales y tampoco realizar el mantenimiento; inversión del tiempo ahorrado en actividades estratégicas y de análisis de los resultados del sistema; generación de una única fuente de verdad que permita crear nuevos indicadores de negocio.
- El proyecto puede abrir posibilidades de desarrollo de nuevos sistemas que automaticen tareas, dado que la empresa posee muchos procesos manuales. Por ejemplo, aplicar el mismo concepto de solución de datos para otras áreas de la empresa como el de producción.
- Desarrollar competencias del equipo al integrar y aplicar conocimiento académico en la metodología de análisis de puntos de dolor de flujos de información, así como en las propuestas de solución ante las situaciones encontradas.
- Elevación en el grado de madurez estableciendo una arquitectura de ignición que permita establecer las bases hacia un comportamiento *data driven*.

Debilidades:

- Curva de aprendizaje de los miembros del equipo en temas de data, analítica, inteligencia de negocio y gestión.
- Restricciones temporales surgidas de los compromisos laborales de los integrantes del equipo, las cuales limitan el tiempo disponible para llevar a cabo el proyecto.
- El software desarrollado debe integrarse con sistemas preexistentes que conforman la infraestructura del cliente, principalmente, para el consumo de datos. De este punto surgen limitaciones ligadas a comprender y acoplarse a estos sistemas.

Amenazas:

- Cambios en la infraestructura del cliente que impliquen la readaptación del producto o, incluso, su obsolescencia.
- Variación en la visión de la empresa respecto de los objetivos a medir dado que el software tiene bases financieras.
- Posibilidad de que los resultados del análisis predictivo no sean precisos debido a insuficiencia o baja calidad de datos.

4.3. Metodología de Trabajo

El equipo estuvo compuesto por cuatro miembros, dos de cuales llevaron a cabo tareas de planificación, gestión, elicitación y estimación de requerimientos, coordinación de sesiones de revisión, desarrollo de código, presentaciones ante el cliente y autores de este documento. Estas actividades descritas y conocidas desde el punto académico resultaron de gran aprendizaje para el equipo. La metodología de desarrollo de un proyecto varía significativamente de la teoría a una implementación en un proyecto real, dado que la gestión y liderazgo son habilidades que se adquieren con la práctica. Por otro lado, el director de este proyecto cumplió el rol de referente técnico y el gerente de la empresa se desempeñó como referente funcional.

Es de destacar que el perfil del referente funcional es económico. Esto impactó de forma directa sobre los requerimientos, dado que la naturaleza de este perfil junto con el contexto socioeconómico del país produjeron que los requerimientos no fueran estables en gran parte del proyecto. Esta falta de estabilidad se tradujo en retrasos en algunas de las tareas. El punto de inestabilidad no se relaciona con la concepción del requerimiento en sí. El requerimiento primario continuaba siendo el mismo; lo que varió fue la forma de implementarlo. Por ejemplo, inicialmente, algunos requerimientos se basaban en datos desde un punto específico hasta la actualidad (de manera histórica). Sin embargo, consideraciones económicas y la necesidad del negocio de mantener una representación precisa de los indicadores llevaron a contemplar intervalos de tiempo en estos requerimientos. Aunque el objetivo del indicador permanecía invariable, su representación se volvió considerablemente más compleja.

Se optó por implementar una metodología de trabajo ágil, en gran medida debido a esta característica cambiante de los requerimientos del proyecto y la falta de claridad en algunos conceptos comerciales. Se requería una estrecha colaboración con el referente funcional y su equipo.

Se seleccionó el marco de desarrollo ágil *scrum*, con *sprints* de 15 días como patrón de proceso. Esta elección se basó en dos razones principales: la capacidad de adaptarse a cambios en los requerimientos y la entrega incremental de funcionalidades. También se consideraron las ceremonias y reuniones características de *scrum*. Se implementaron reuniones breves y frecuentes entre los miembros del equipo, así como sesiones de revisión con el referente funcional.

Scrum^[2] es un marco de trabajo que puede ser ajustado según el contexto, dependiendo de los requerimientos del proyecto y de las capacidades del equipo. Teniendo esto en cuenta, se buscó adaptarlo de manera consciente a las circunstancias de un equipo de cuatro personas. Para esto, se mantuvieron los principios clave de Scrum, como la priorización, la iteración y la

comunicación con el cliente, a la vez que se simplificaron las ceremonias y roles para que se ajusten mejor al tamaño del equipo.

En las reuniones internas de equipo, el foco estuvo en responder a las siguientes preguntas para cada miembro (Pressman, R. S., 2010)^[3]:

1. “¿Qué lograste desde nuestra última reunión de equipo? ¿Cuál fue tu progreso?”
2. ¿Cuáles obstáculos encontraste? ¿Qué te impidió avanzar?”
3. ¿Cuáles son tus planes hasta nuestra próxima reunión de equipo?”

Es importante destacar que las reuniones internas de *scrum* no se llevaron a cabo de manera diaria, sino que las se programaron dos veces a la semana de manera intercalada. Esto permitió avanzar lo suficiente entre reuniones, teniendo en cuenta las responsabilidades laborales individuales de cada integrante. Además, el tamaño del equipo hizo que sea sencillo encontrarse en constante comunicación por canales de mensajería para estar totalmente alineados, solucionar problemas de manera rápida cuando fuese necesario y mejorar de manera continua la forma de trabajo al detectar y solucionar fallas en los métodos de desarrollo.

Inicialmente, el flujo de trabajo en las primeras iteraciones del proyecto se mantuvo estable. Sin embargo, luego se experimentaron desafíos debido a varios factores. La limitada disponibilidad del equipo por compromisos laborales, junto con bloqueos ocasionados por la necesidad de revisar y retrabajar partes importantes del proyecto implicó suspender el desarrollo de los requerimientos pendientes en el *backlog*, así como la investigación requerida para abordar ciertos requerimientos. Esto resultó en retrasos en los *sprints* y una falta de coordinación en el equipo.

Después de varias iteraciones, se logró retomar un ritmo de trabajo más organizado a medida que se completaron los requerimientos funcionales centrales del sistema. Adicionalmente, los integrantes del equipo adquirieron una mayor experiencia tanto en gestión como en aspectos técnicos, lo que les permitió remediar los retrasos surgidos y proseguir con la construcción del producto de manera eficiente.

4.3.1. *Proceso y gestión de tareas*

El proceso de desarrollo se estructuró en dos componentes fundamentales: el desarrollo del backend, que es el módulo de gestión de datos, y del frontend, módulo de visualización de datos.

La coordinación precisa entre estos dos aspectos resultó esencial, ya que la lógica que se estaba elaborando en el módulo de gestión de datos necesitaba una implementación simultánea en el de visualización para lograr entregas incrementales efectivas. Esta interdependencia garantizó la

coherencia y la funcionalidad adecuada en las entregas a lo largo del proyecto, permitiendo un desarrollo ágil y cohesivo.

Nuevamente, y en concordancia con lo mencionado en la sección 4 sobre la gestión del proyecto, coordinar el desarrollo entre ambos módulos no resultó trivial. Al comienzo de la implementación del sistema, el equipo experimentó contratiempos debido a la dificultad para paralelizar ambos procesos de desarrollo. Esto surge debido a que, previo al cálculo de los indicadores, fue necesario implementar la ingesta de los datos en el módulo de gestión. Una estrategia que se utilizó para superar esta fase fue la utilización de datos de ejemplo para probar la funcionalidad del módulo de visualización y validar requerimientos. Esto permitió que las tareas del módulo de gestión avancen hasta alinearse con el de visualización. Una vez superada la fase de reorganización y priorización de tareas, se lograron alcanzar las entregas funcionales, incrementales y efectivas que se tenían como objetivo.

Tanto para la implementación del backend como para la del frontend se llevó a cabo la gestión de un *backlog* que permitió priorizar tareas y medir el progreso (ver Figura 1 para un ejemplo^[4]).

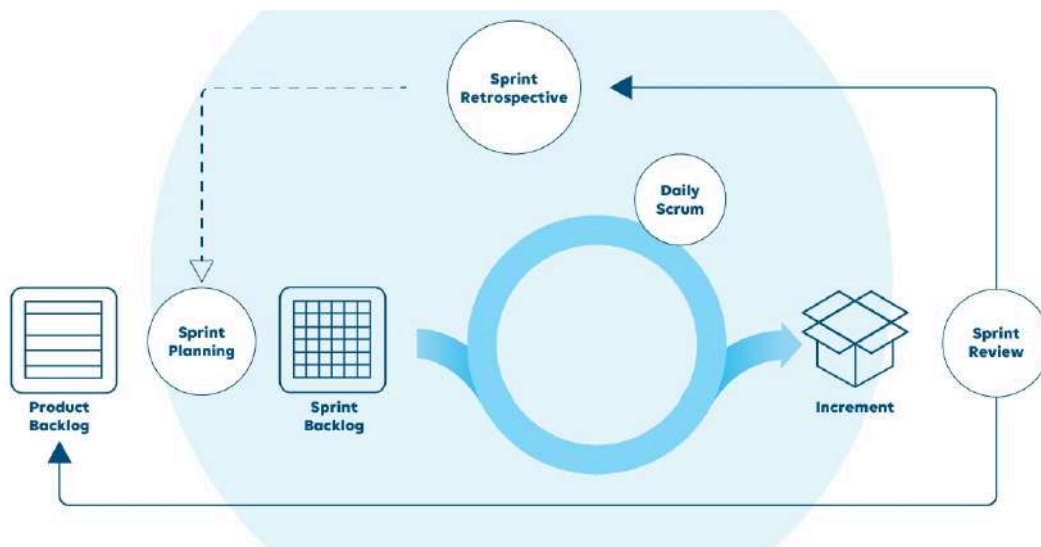


Figura 1: Scrum Framework

Conjuntamente, al final de cada fase incremental del proceso, el desarrollo se sometió a la validación por parte del referente funcional. En base a los comentarios recibidos, se llevó a cabo la revisión del *sprint* correspondiente a la iteración y se continuó con el flujo de trabajo dentro del marco ágil. Estas reuniones sirvieron, además, para evacuar dudas y discutir cambios o nuevas funcionalidades con el referente funcional, sirviendo efectivamente como una vía de comunicación bidireccional. Esto es fundamental en *scrum* para garantizar que el producto se ajuste a las necesidades del cliente.

4.3.2. Gestión de versiones

Git fue la herramienta utilizada para llevar adelante la gestión de versiones del software. Se adoptó un flujo de trabajo de ramas, con base en el popular modelo Gitflow.

Flujo de trabajo de ramas de función en Git

La idea principal del flujo de trabajo de ramas de función es que el desarrollo de una función debe llevarse a cabo en una rama especializada en lugar de en la rama *main* o principal.

Este aislamiento permite que se trabaje en una función concreta sin modificar el contenido del código base principal, lo que también implica que la rama *main* no contendrá en ningún caso código erróneo, lo que supone una gran ventaja para los entornos de integración continua.

Gitflow

Gitflow^[5] es un modelo alternativo de creación de ramas en Git en el que se utilizan ramas de función y varias ramas principales.

Este flujo de trabajo no añade ningún concepto o comando nuevo, aparte de los que se necesitan para el flujo de trabajo de ramas de función. Lo que hace es asignar funciones muy específicas a las distintas ramas y definir cómo y cuándo deben estas interactuar.

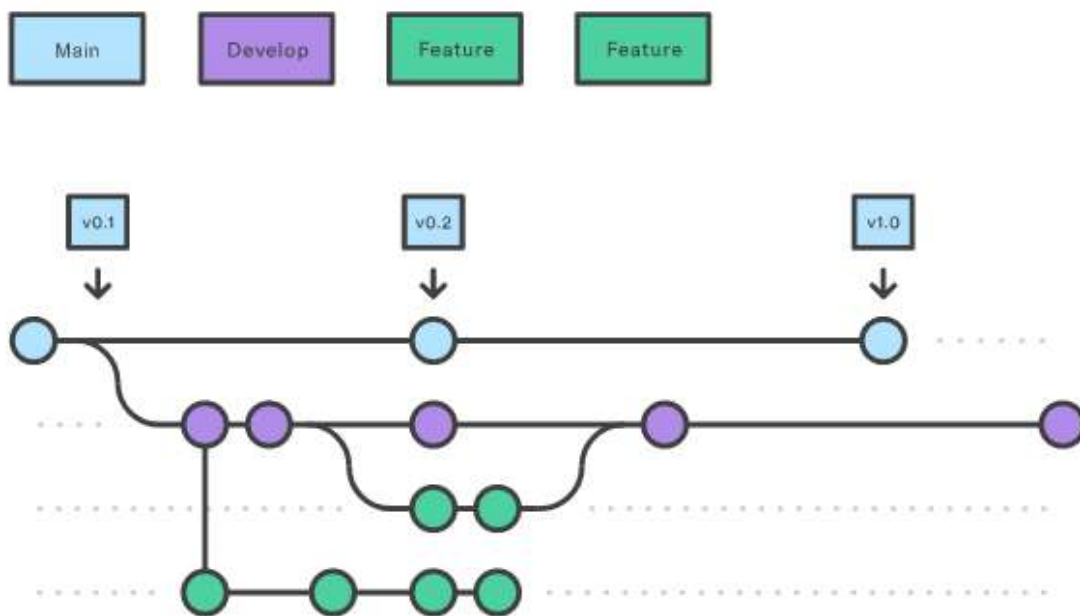


Figura 2: Gitflow^[5]

Ramas principales y de desarrollo

La rama *main* o principal^[6] almacena el historial de publicación oficial y la rama *develop* o de desarrollo sirve como rama de integración para las funciones.

Todas las funciones nuevas deben residir en su propia rama, que se puede enviar al repositorio central para copia de seguridad/colaboración. Sin embargo, en lugar de ramificarse de *main*, las ramas *feature* utilizan la rama *develop* como rama primaria. Cuando una función está terminada, se vuelve a fusionar en *develop*. Las funciones no deben interactuar nunca directamente con *main*.

Ramas de corrección

Las ramas de mantenimiento o de *hotfix* sirven para reparar rápidamente las publicaciones de producción. Son muy similares a las ramas *feature*, salvo por el hecho de que se basan en la rama *main* y no en la *develop*. Esta es la única rama que debería bifurcarse directamente a partir de *main*. Una vez aplicada la corrección, debería fusionarse en *main* y *develop*.

4.3.3. Integración continua

Se utilizó la tecnología Vercel para habilitar la integración continua en el flujo de desarrollo. Vercel opera conectándose a un repositorio Git. Al detectar cambios en el repositorio, automáticamente inicia el proceso de construcción y despliegue de la aplicación.

Este enfoque de integración continua permite mantener distintas versiones de la aplicación, siempre actualizadas con las ramas principales del repositorio y disponibles en tiempo real para pruebas (*development*) y para el usuario final (*main*). Los beneficios para el cliente son evidentes, al disponer de manera rápida de mejoras y correcciones.

Además, esta práctica ayuda con la detección temprana de problemas, facilitando la corrección oportuna. La integración continua no solo optimiza la eficiencia del proceso de desarrollo, sino que también garantiza una entrega constante de calidad para los usuarios finales.

4.3.4. Ambiente de desarrollo

Para gestionar el ambiente de desarrollo, se seleccionó la herramienta Docker Compose, que posibilita la creación de entornos de desarrollo reproducibles. De esta manera, cada miembro del equipo pudo ejecutar exactamente la misma configuración en su máquina, eliminando problemas de compatibilidad y conflictos de dependencias.

Docker y Contenedores

Docker^[7] es un estándar para contenedores de software, que son artefactos estandarizados para el empaquetado de software. Proporciona una manera unificada de obtener o crear cualquier servicio (contenedor) para luego poder ejecutarlo en un ambiente local virtual.

Docker Compose

El archivo Compose^[8] proporciona una manera de documentar y configurar todas las dependencias de servicios de una aplicación (bases de datos, APIs de servicios web, etc.).

Utilizando la sintaxis YAML, se declara una lista de servicios y para luego ejecutarlos con un solo comando: *docker-compose up*. Compose es una herramienta para definir y ejecutar aplicaciones Docker multi-contenedor, con una configuración y una infraestructura adecuadas.

La solución establece un ambiente de desarrollo definido mediante Compose, que contiene tres servicios: el servidor, una aplicación desarrollada en Node.js; la base de datos, que es MySQL; y el administrador de la base de datos, Adminer.

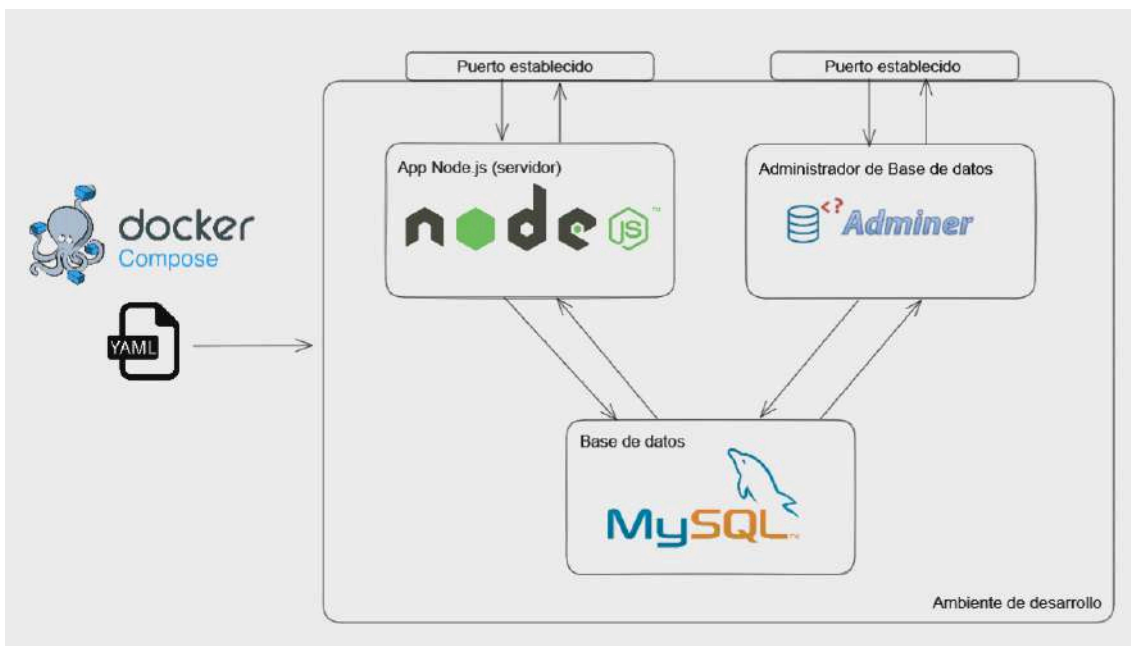


Figura 3: Ambiente de desarrollo. Fuente: Elaboración propia

Ventajas

- No es necesario instalar ni mantener el software localmente en la máquina del desarrollador.
- Se puede incluir todo el entorno de desarrollo local en control de versiones, facilitando la colaboración entre desarrolladores en un proyecto.
- Se puede iniciar todo el entorno de desarrollo local con un solo comando: *docker-compose up*.

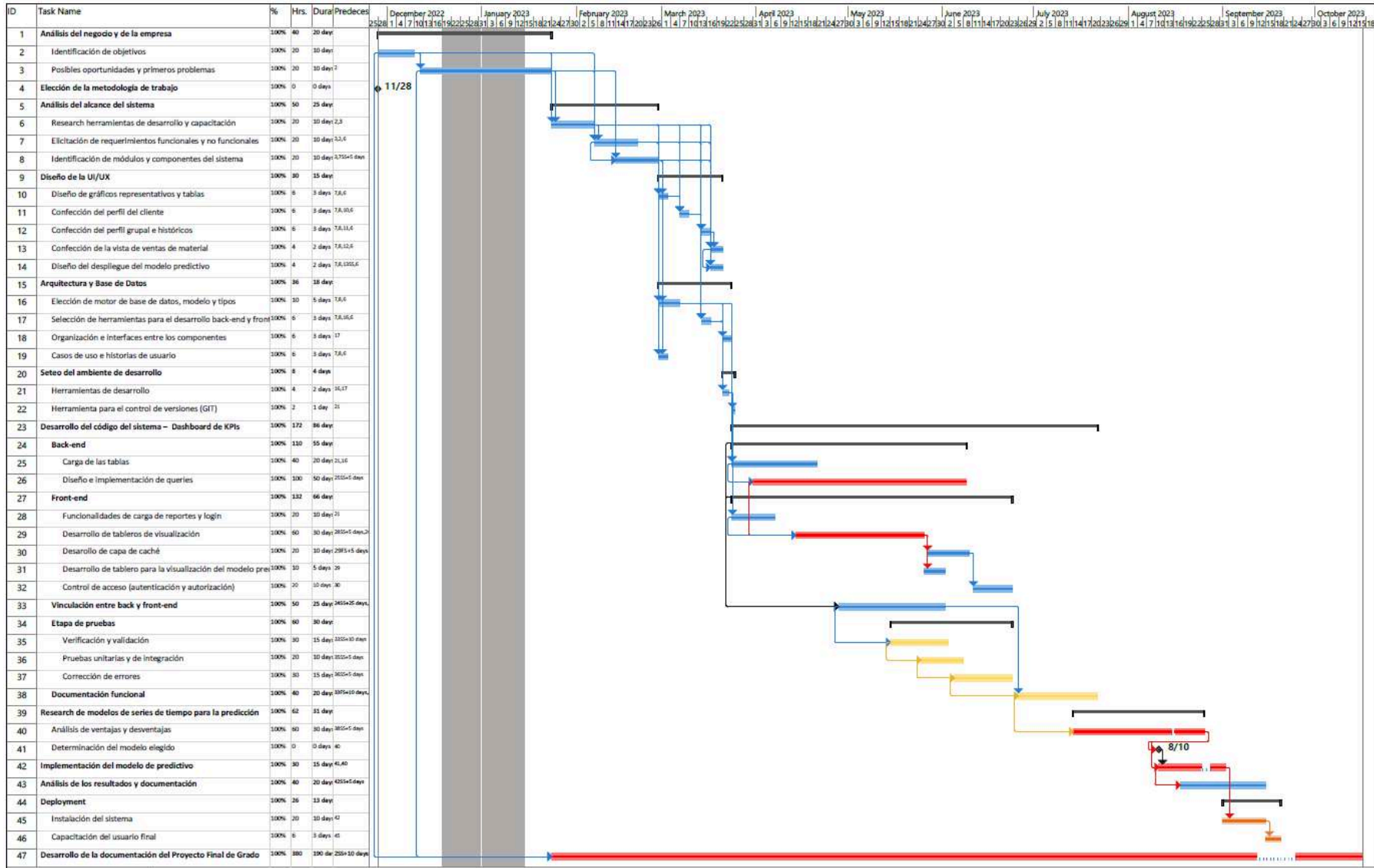
4.4. Planificación inicial y estimaciones previas

La planificación inicial de este proyecto se vio notablemente influenciada por el progreso inicial en la realización de las tareas presentadas junto con el protocolo. Esta dinámica conllevó a que la estimación de ciertos hitos se cumpla puntualmente en las etapas iniciales. Además, es fundamental destacar que a lo largo del desarrollo del proyecto, el equipo adquirió una mayor experiencia y conocimiento en el área correspondiente, desempeñando las actividades de forma más ordenada.

Este incremento en la experiencia permitió que se corrigieran muchas de las estimaciones iniciales que no habían sido realizadas con la precisión requerida. Estos ajustes se llevaron a cabo a medida que avanzaba el proyecto, en consonancia con la presentación del protocolo correspondiente.

A continuación, se puede visualizar el diagrama de Gantt con la estimación inicial considerando las actividades ya realizadas.

En el diagrama siguiente, se observa la distribución final de las actividades y su duración.



La creación e implementación de consultas requirió diez horas adicionales al tiempo estimado, debido a desafíos de rendimiento en las consultas diseñadas. Algunas de ellas resultaron ser ineficientes en términos de tiempo, lo que provocaba demoras significativas en la obtención de resultados. Se realizaron ajustes en el diseño de las consultas relacionadas con los datos del perfil del cliente para abordar este problema.

Dado lo mencionado previamente, a pesar de los esfuerzos por optimizar las consultas, persistían demoras en la obtención de datos desde el backend. Para encarar este problema, se introdujo una capa de caché en el frontend, logrando mejoras significativas en el rendimiento para accesos posteriores una vez recibidos los datos iniciales. En el diagrama de Gantt subsiguiente, esta acción se representa como la tarea número 30, agregándose a la estructura jerárquica del frontend con una dedicación adicional de veinte horas al proyecto.

La incorporación de nuevas actividades, junto con la necesidad de reingeniería en las consultas, generó un desplazamiento en el inicio de la etapa de pruebas respecto a la estimación inicial. Además esta tarea, que se estimó originalmente a cuarenta horas, se redujo a veinte, ya que muchas de las pruebas unitarias se llevaron a cabo inmediatamente después del desarrollo, integrándose con otras actividades. Las pruebas de integración fueron las únicas que se mantuvieron en esta fase. Este corrimiento también se produjo en las tareas de verificación y validación, corrección de errores y de documentaciones intermedias. El equipo reconoce que la cadena de eventos que llevó al retraso en el inicio de las actividades se debió a la limitada capacidad para manejar contratiempos y a la centralización de tareas en cada integrante. Si bien, la centralización de actividades en cada miembro continuó siendo un factor común durante el desarrollo del proyecto, surgieron planes de contingencia para remediar los contratiempos. Como ya se detalló anteriormente, el proceso de priorización y replanificación resultó clave para solventar la falta de avance en los entregables y comenzar a cumplir con los hitos de forma más precisa. Además, se asignaron horas extras no previstas en la planificación inicial para compensar la falta de respuesta al cliente.

En el ámbito del modelado de series temporales, el equipo se enfrentó a dificultades en el análisis de ventajas y desventajas. En principio, la comprensión de los modelos tomó más tiempo del previsto, ya que resultó complejo comprender cómo se parametrizan y cuáles son las características que justifican la elección de un modelo sobre otro. Además, la cantidad de datos y el ruido presente en la muestra complicaron el análisis. Lo que inicialmente se estimó en cuarenta horas se extendió a sesenta horas. En relación a la implementación del modelo, también se vio afectada su duración, extendiéndose diez horas más de lo estimado por las razones anteriormente detalladas.

En la fase de implementación, el equipo experimentó un desplazamiento en el inicio previsto de la instalación del sistema y la capacitación del usuario final. Además, estas tareas requirieron menos horas de las inicialmente estimadas, reduciendo la duración del proceso en diez horas.

La documentación del presente documento se vio interrumpida en las últimas iteraciones debido a complicaciones de tiempo por parte del equipo, lo que resultó en un aumento de veinte horas en la duración de esta actividad. Finalmente, la comparación entre lo estimado y el resultado real de las actividades se puede observar en el diagrama siguiente:

Name	% Complete: Current	% Complete: Previous	% Complete: Diff	Dur. Hrs (Number): Current	Dur. Hrs (Number): Previous	Dur. Hrs (Number): Diff	Duration: Current	Duration: Previous	Duration: Diff	Predecessor: Current	Predecessor: Previous	Predecessor: Diff
1 Análisis del negocio y de la empresa	100	100	0	40	40	0	020 days	20 days	0d			Equal
2 Identificación de objetivos	100	100	0	20	20	0	010 days	10 days	0d			Equal
3 Posibles oportunidades y primeros problemas	100	100	0	20	20	0	010 days	10 days	0d	2	2	Equal
4 Elección de la metodología de trabajo	100	100	0	0	0	0	00 days	0 days	0d			Equal
5 Análisis del alcance del sistema	100	100	0	50	50	0	025 days	25 days	0d			Equal
6 Research herramientas de desarrollo y capacitación	100	100	0	20	20	0	010 days	10 days	0d	2,3	2,3	Equal
7 Elicitación de requerimientos funcionales y no funcionales	100	100	0	20	20	0	010 days	10 days	0d	3,2,6	3,2,6	Equal
8 Identificación de módulos y componentes del sistema	100	100	0	20	20	0	010 days	10 days	0d	3,7SS+5d	3,7SS+5d	Equal
9 Diseño de la UI/UX	100	100	0	30	30	0	015 days	15 days	0d			Equal
10 Diseño de gráficos representativos y tablas	100	100	0	6	6	0	03 days	3 days	0d	7,8,6	7,8,6	Equal
11 Confección del perfil del cliente	100	100	0	6	6	0	03 days	3 days	0d	7,8,10,6	7,8,10,6	Equal
12 Confección del perfil grupal e históricos	100	100	0	6	6	0	03 days	3 days	0d	7,8,11,6	7,8,11,6	Equal
13 Confección de la vista de ventas de material	100	100	0	4	4	0	02 days	2 days	0d	7,8,12,6	7,8,12,6	Equal
14 Diseño del despliegue del modelo predictivo	100	100	0	4	4	0	02 days	2 days	0d	7,8,13SS,6	7,8,13SS,6	Equal
15 Arquitectura y Base de Datos	100	94	6	36	36	0	018 days	18 days	0d			Equal
16 Elección de motor de base de datos, modelo y tipos	100	100	0	10	10	0	05 days	5 days	0d	7,8,6	7,8,6	Equal
17 Selección de herramientas para el desarrollo back-end y fr	100	100	0	6	6	0	03 days	3 days	0d	7,8,16,6	7,8,16,6	Equal
18 Organización e interfaces entre los componentes	100	100	0	6	6	0	03 days	3 days	0d	17	17	Equal
19 Casos de uso e historias de usuario	100	70	30	6	6	0	03 days	3 days	0d	7,8,6	7,8,6	Equal
20 Seteo del ambiente de desarrollo	100	100	0	8	8	0	04 days	4 days	0d			Equal
21 Herramientas de desarrollo	100	100	0	4	4	0	02 days	2 days	0d	16,17	16,17	Equal
22 Herramienta para el control de versiones (GIT)	100	100	0	2	2	0	01 day	1 day	0d	21	21	Equal
23 Desarrollo del código del sistema - Dashboard de KPIs	100	51	49	172	160	12	1286 days	80 days	6d			Equal
24 Back-end	100	86	14	110	100	10	1055 days	50 days	5d			Equal
25 Carga de las tablas	100	100	0	40	40	0	020 days	20 days	0d	21,16	21,16	Equal
26 Diseño e implementación de queries	100	80	20	100	90	10	1050 days	45 days	5d	25SS+5d	25SS+5d	Equal
27 Front-end	100	64	36	132	100	32	3266 days	50 days	16d			Equal
28 Funcionalidades de carga de reportes y login	100	50	50	20	20	0	010 days	10 days	0d	21	21	Equal
29 Desarrollo de tableros de visualización	100	80	20	60	60	0	030 days	30 days	0d	28SS+5d,2	28SS+5d,2	Equal
30 Desarrollo de capa de caché	100	0	0	20	0	20	010 days	0 days	10d	29FS+5d		Different
31 Desarrollo de tablero para la visualización del modelo p	100	0	0	10	10	0	05 days	5 days	0d	29	29	Equal
32 Control de acceso (autenticación y autorización)	100	0	0	20	0	20	010 days	0 days	10d	30		Different
33 Vinculación entre back y front-end	100	75	25	50	50	0	025 days	25 days	0d	24SS+25d	24SS+25d	Equal
34 Etapas de pruebas	100	0	100	60	50	10	1030 days	25 days	5d			Equal
35 Verificación y validación	100	0	100	30	30	0	015 days	15 days	0d	33SS+10d	33SS+10d	Different
36 Pruebas unitarias y de integración	100	0	100	20	40	-20	-2010 days	20 days	-10d	35SS+5d	33SS+5d	Different
37 Corrección de errores	100	0	100	30	30	0	015 days	15 days	0d	36SS+5d	34SS+5d	Different
38 Documentación funcional	100	0	100	40	40	0	020 days	20 days	0d	33FS+10d	31FS+10d	Different
39 Research de modelos de series de tiempo para la predicción	100	10	90	62	40	22	2231 days	20 days	11d			Equal
40 Análisis de ventajas y desventajas	100	10	90	60	40	20	2030 days	20 days	10d	38SS+5d	36SS+5d	Different
41 Determinación del modelo elegido	100	0	100	0	0	0	00 days	0 days	0d	40	38	Different
42 Implementación del modelo de predictivo	100	0	100	30	20	10	1015 days	10 days	5d	41,40	39	Different
43 Análisis de los resultados y documentación	100	0	100	40	40	0	020 days	20 days	0d	42SS+5d	40SS+5d	Different
44 Deployment	100	0	100	26	40	-14	-1413 days	20 days	-7d			Equal
45 Instalación del sistema	100	0	100	20	30	-10	-1010 days	15 days	-5d	42	40	Different
46 Capacitación del usuario final	100	0	100	6	10	-4	-43 days	5 days	-2d	45	43	Different
47 Desarrollo de la documentación del Proyecto Final de Grado	100	30	70	380	360	20	20190 day	180 days	10d	25SS+10d,2	25SS+10d,3	Equal



Task: Previous Milestone: Current Placeholder: Duration: Previous Placeholder: Start: Current Inactive Task: Previous Inactive Milestone: Current
 Task: Current Summary: Previous Placeholder: Duration: Current Placeholder: Finish: Previous Inactive Task: Current Inactive Summary: Previous
 Milestone: Previous Summary: Current Placeholder: Start: Previous Placeholder: Finish: Current Inactive Milestone: Previous Inactive Summary: Current

4.5. Análisis de los tiempos finales vs estimados

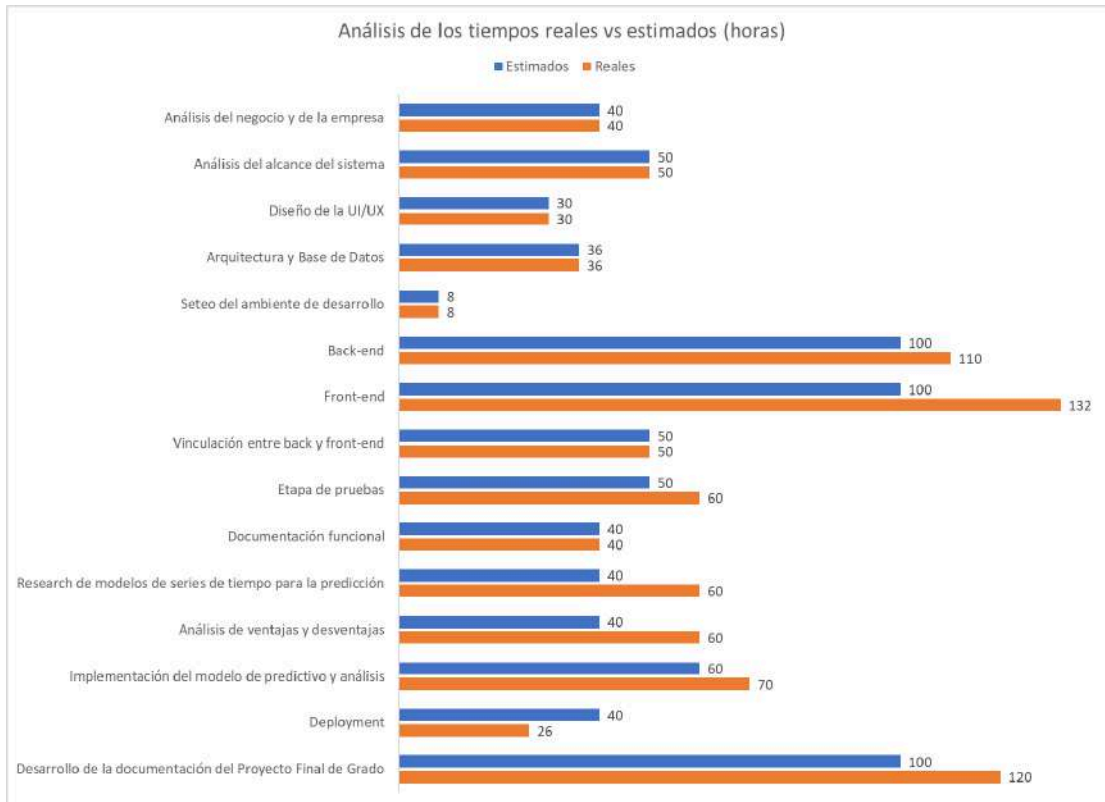


Figura 4: Análisis de los tiempos estimados vs. finales. Fuente: Elaboración propia

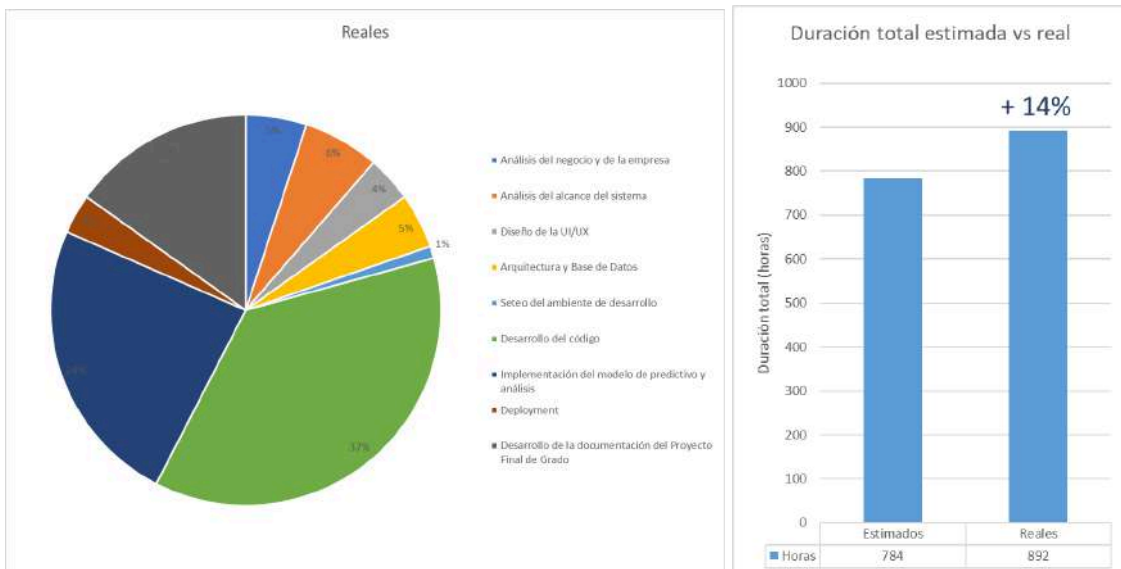


Figura 5: Porcentaje de consumo de horas en actividades principales. Fuente: Elaboración propia

5. Marco de referencia

En la actualidad, el constante avance tecnológico ha llevado a un rápido incremento en la recopilación y utilización de datos a través de dispositivos, sensores y servicios por parte de individuos, organizaciones y gobiernos^[9]. Esto ha impulsado a muchas organizaciones a reconocer la importancia estratégica de los datos para tomar decisiones más informadas, mejorar la eficiencia operativa, establecer prioridades, desarrollar productos y evaluar riesgos y procesos de manera más efectiva.

5.1. *Los datos como activo estratégico*

Los datos, más que simplemente registros neutrales, se perfilan como un activo estratégico crucial. Su correcta explotación no solo impulsa la generación de *insights* valiosos, sino que también desencadena oportunidades para alinear objetivos, fomentar cambios de interés y catalizar transformaciones significativas en diversos aspectos, desde lo social y político hasta lo económico.

5.1.1. *Calidad de los datos*

La calidad de los datos implica optimizarlos para que cumplan con el estado deseado y se centra en responder a la pregunta: (Reis, J., & Housley, M., 2022) "¿Qué se obtiene en relación a lo que se espera?"^[10] Es esencial que los datos se ajusten a las expectativas establecidas por los metadatos empresariales. ¿Concuerdan los datos con la definición acordada por la empresa?

La generación de valor a través de los datos está intrínsecamente ligada a su calidad. Por lo tanto, la calidad de los datos es el factor determinante que influye en la confiabilidad de las decisiones tomadas con base en ellos. Existen ciertas características que los datos deben poseer para que su uso como activos estratégicos aporte un alto nivel de certeza en una organización^[11]. Estas características esenciales que deben cumplir son^[12]:

- **Exactitud:** Grado en el que los datos representan correctamente entidades de la vida real. ¿Los datos recopilados son objetivamente correctos? ¿Hay valores duplicados? ¿Son exactos los valores numéricos?
- **Compleitud:** Se refiere a si todos los datos están presentes o si son suficiente. ¿Están completos los registros? ¿Todos los campos obligatorios contienen valores válidos?
- **Oportunidad:** Vigencia de los datos y frecuencia con la que cambian. ¿Están los registros disponibles de manera oportuna?

5.2. Ingeniería de datos

La ingeniería de datos, según Joe Reis & Matt Housley^[13], se dedica a construir las bases necesarias para la ciencia de datos y el análisis. Implica el diseño, desarrollo y gestión eficiente de la infraestructura y arquitectura de datos, asegurando la disponibilidad y calidad necesarias para impulsar procesos analíticos y toma de decisiones informadas. Se destaca por su contribución en la estructuración y optimización de flujos de datos para facilitar la extracción de información valiosa en entornos empresariales y científicos.

Existen diferentes definiciones de ingeniería de datos, pero con la que este equipo comparte visión es la siguiente:

Reis, J., & Housley, M., 2022: “La ingeniería de datos es el desarrollo, implementación y mantenimiento de sistemas y procesos que reciben datos en crudo y generan información de alta calidad y consistente que respalda casos de uso posteriores, como análisis y aprendizaje automático. La ingeniería de datos es la intersección de la seguridad, la gestión de datos, DataOps, la arquitectura de datos, la orquestación y la ingeniería de software. Un ingeniero de datos gestiona el ciclo de vida de la ingeniería de datos, comenzando por obtenerlos de los sistemas fuente y terminando con proporcionar datos para casos de uso, como análisis o aprendizaje automático.”

—”Fundamentals of Data Engineering. Plan and Build Robust Data Systems” por Joe Reis & Matt Housley^[14]

5.2.1. Ciclo de vida

El ciclo de vida de la ingeniería de datos (Joe Reis & Matt Housley^[15]) se refiere al conjunto de fases y procesos que abarcan desde la adquisición de datos desde diversas fuentes hasta la entrega y uso final de la información procesada. Este ciclo engloba actividades como la extracción, transformación, carga (ETL), la gestión de la calidad de los datos, el almacenamiento, la orquestación, y finalmente, la disponibilidad de los datos para su aplicación en análisis, *machine learning* u otros casos de uso.

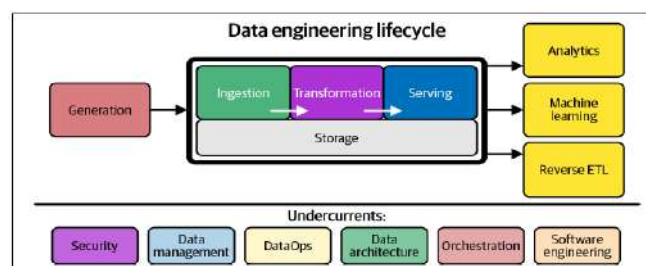


Figura 6: Data Engineering Lifecycle. Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering^[15]

En el inicio del ciclo de vida, se adquieren los datos desde uno o varios sistemas fuente y se almacenan. La siguiente fase implica la transformación de los datos, preparándolos para su disponibilidad, permitiendo así que científicos de datos y/o analistas puedan utilizarlos. Estos profesionales son quienes identificarán los puntos que aportan valor a los negocios y generarán *insights*.

Diversas disciplinas acompañan en el trasfondo del ciclo de vida de la ingeniería de datos. Sin seguridad, gestión de datos, DataOps, arquitectura de datos, orquestación e ingeniería de software ninguna etapa del ciclo de vida puede operar de manera adecuada o eficiente.

5.2.2. *Sistemas origen de los datos*

Los sistemas fuente^[16], como distintos dispositivos, sensores o bases de datos, son el punto de partida en el ciclo de vida de la ingeniería de datos. Un aspecto importante es que aquellos usuarios que consumen estos datos de las fuentes no necesariamente son los propietarios o administradores de dichas fuentes. En otras palabras, siempre que se emplee una fuente específica en un proyecto de datos, los ingenieros deben tener una comprensión sólida de cómo operan esas fuentes. Esto implica entender cómo se generan los datos, así como sus características, como la velocidad de generación, la frecuencia, los tipos, los formatos, y demás.

Un ejemplo típico de fuente de origen sería un sistema o aplicación que guarda sus datos en una base de datos, justamente como ocurre con el sistema origen de este proyecto.

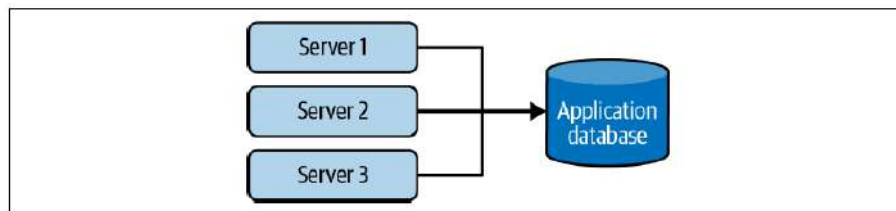


Figura 7: The data engineering lifecycle is a subset of the full data lifecycle. Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering^[17]

Según Joe Reis & Matt Housley^[18], existen algunos aspectos al evaluar los sistemas fuente. Este equipo consideró algunos de ellos en el proceso de elicitación. A continuación, se presentan algunos de ellos, partiendo de los siguientes interrogantes:

- ¿Cuáles son las características esenciales de la fuente de datos? ¿Es una aplicación? ¿Una serie de dispositivos?
- ¿Cómo se persisten los datos en el sistema fuente? ¿Los datos se almacenan a largo plazo o son temporales y se eliminan rápidamente?

- ¿Qué nivel de consistencia pueden esperar los ingenieros de los datos de salida? Si se están ejecutando verificaciones de calidad de datos, ¿con qué frecuencia ocurren inconsistencias, como nulos donde no se esperan, formato deficiente, etc.?
- Si hay cambios en el esquema (por ejemplo, se agrega una nueva columna), ¿cómo se maneja esto y se comunica a las partes interesadas en la secuencia de desarrollo?
- ¿Con qué frecuencia se debe extraer datos del sistema fuente?
- ¿Quién/cuál es el proveedor de datos que los transmitirá para el consumo?

Estos fueron los puntos clave que el equipo consideró críticos, aunque existen otros puntos mencionados por los autores de "Fundamentals of Data Engineering. Plan and Build Robust Data Systems". En la sección que aborda el sistema, se detallarán las respuestas a estos interrogantes al implementar la solución propuesta.

5.2.3. Almacenamiento

En principio, es necesario almacenar los datos extraídos desde los sistemas fuente. Si bien, el almacenamiento es una etapa del ciclo de vida de la ingeniería de datos, es común que se encuentre involucrada con otros procesos, como la ingesta, la transformación y el despliegue. La manera en la que se almacenan los datos influye en cómo serán utilizados en todas las etapas restantes.

Siguiendo las directrices propuestas por los autores Joe Reis y Matt Housley^[19], también hay aspectos clave a tener en cuenta al seleccionar la plataforma de almacenamiento. Una vez más, se destacan los que el equipo considera esenciales:

- ¿Es compatible esta solución de almacenamiento con las velocidades de escritura y lectura requeridas por la arquitectura?
- ¿El almacenamiento creará un cuello de botella para los procesos en la secuencia de desarrollo?
- ¿Manejará este sistema de almacenamiento el escalamiento?

En línea con lo propuesto anteriormente, en la sección "Sistema" se dará respuesta a estos puntos.

5.2.4. Ingesta

La ingesta^[20] es donde los datos se recopilan, importan o cargan en un sistema o aplicación. Este proceso, junto con los sistemas fuente, constituye un punto crítico, ya que si la entrada inicial no funciona correctamente, todo el producto de datos puede quedar comprometido.

La ingesta es el transporte de información de un punto a otro, moviendo los datos desde los sistemas fuente hacia el almacenamiento.

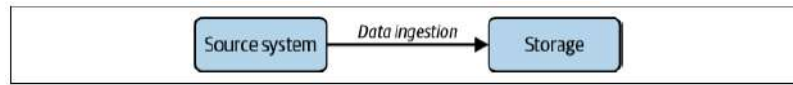


Figura 8: The data engineering lifecycle is a subset of the full data lifecycle. Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering^[21]

Los aspectos clave a responder según Joe Reis & Matt Housley:

- ¿Cuáles son los casos de uso para los datos que se están ingestandando? ¿Es posible reutilizar estos datos en lugar de crear múltiples versiones del mismo conjunto de datos?
- ¿Cuál es el destino de los datos después de la ingesta?
- ¿Con qué frecuencia será necesario acceder?
- ¿Es necesario transformarlos? ¿Cómo se encara?

5.2.5. Transformación

En esta etapa los datos comienzan a generar valor, convirtiéndose en los activos estratégicos destacados en el resumen del proyecto. En términos simples, los datos pasan de ser *raw* o crudos a *curated* o valiosos. A través de procesos de modelado y transformación, se refinan para ser consumibles.

Durante esta fase^[22], los datos se ajustan a los tipos esperados por sus consumidores, como la conversión de cadenas de caracteres a formatos numéricos o de fechas. También se eliminan datos corruptos o incorrectos. Sin embargo, estas transformaciones son solo el comienzo. En etapas posteriores, es posible realizar modelados más complejos, normalizar o ajustar los esquemas existentes.

El verdadero material de ignición de esta etapa es la lógica de negocio. En este punto, los datos empiezan a revelar la narrativa de la lógica empresarial que representan. Es como si los datos comenzaran a “hablar” sobre los detalles que encierran acerca de la lógica empresarial.

Las consideraciones clave a la hora de transformar que recomiendan Joe Reis & Matt Housley:

- ¿Es la transformación lo más simple y autoaislada posible?
- ¿Qué reglas comerciales respaldan las transformaciones?



Figura 9: Tall Street Mojo. Data Transformation^[23]

5.2.6. Disponibilización o despliegue de los datos

El valor de los datos se desata en esta etapa, siempre y cuando sean consumidos y consultados. Aquellos datos que no encuentran uso, no aportan valor. Según la bibliografía consultada^[24], los proyectos de datos se consideran riesgos para las empresas. Esta designación se basa en la inversión significativa en la recopilación de datos, con gastos considerables en almacenamiento y procesamiento, que a menudo no se consumen de manera efectiva.

No obstante, el equipo discrepa en algún punto con este material. Incluso si un proyecto de datos no tiene éxito al no consumirlos de manera efectiva (debido a errores de consistencia, falta de datos maestros, carencia de una sólida arquitectura que respalde los productos de datos o fallas en la calidad), siempre revelarán algo sobre el negocio. Los *pain points* o puntos de dolor de las empresas surgirán, incluso si los datos no pueden "contar", predecir, o explicar. En última instancia, los proyectos de datos siempre ofrecerán una perspectiva de la realidad inmersa en la lógica empresarial. Sin embargo, con lo que sí concuerda el equipo respecto de la bibliografía es que los proyectos de datos deben ser intencionales, con un propósito.

Dentro de lo que es el despliegue de los datos, se pueden mencionar algunas aplicaciones:

Business Intelligence (BI)^[25]: La inteligencia empresarial utiliza los datos a modo descriptivo. Mediante *BI*, es posible realizar un análisis del estado pasado y actual de una organización o empresa. Debe aplicar la lógica de negocio pertinente para poder desplegar la información. A medida que una empresa desarrolla su madurez en datos (un término que se explorará en detalle más adelante en este documento), evolucionará desde una fase en la que un equipo especializado desarrolle informes y tableros en función de las necesidades, a un estadio de "autoservicio de análisis", también conocido como *self-service analytics* o *self-service BI*.

Análisis operativo^[26]: En contraste, el análisis operativo emplea datos de manera prescriptiva, generando informes que delinear planes de acción. Ejemplos de este tipo de análisis incluyen la evaluación de inventarios o *stocks*, el monitoreo de aplicaciones y el seguimiento de sistemas en general.

Machine learning: Según Amazon^[27], *Machine learning* (ML) es la ciencia de desarrollo de algoritmos y modelos estadísticos que utilizan los sistemas de computación con el fin de llevar a cabo tareas sin instrucciones explícitas. Los sistemas de computación utilizan algoritmos de *machine learning* para procesar grandes cantidades de datos históricos e identificar patrones. Esto les permite generar resultados con mayor precisión a partir de un conjunto de datos de entrada.

5.3. *Indicadores claves de desempeño y de resultado (KPI y KRI)*

Según David Parmenter^[28], autor de “Key Performance Indicator: Developing, Implementing, and Using Winning KPIs”, los indicadores claves de rendimiento (KPIs, por sus siglas en inglés) ofrecen información en tiempo real sobre cómo se está desempeñando la organización en factores críticos de éxito. Tomar medidas basadas en los KPIs puede aumentar drásticamente el rendimiento. Son aquellos indicadores que se centran en los aspectos del rendimiento organizativo más críticos para el éxito actual y futuro de la organización.

Los indicadores claves de resultado^[29] (KRIs, por sus siglas en inglés) son medidas que a menudo se confunden con los KPIs. La característica común de estas medidas es que son el resultado de muchas acciones llevadas a cabo por diferentes equipos durante un período de tiempo. De ahí el uso del término "resultado", y son buenas medidas resumen, de ahí el término "clave". Estos indicadores clave de resultado deben revisarse típicamente en las reuniones de negocio bimensuales o trimestrales, brindando una comprensión de cómo avanza la organización con su estrategia. Los KRIs siempre son medidas del pasado.



Figura 10: Conocimiento del negocio^[30]

Después de analizar lo descrito anteriormente, este proyecto finalmente proporcionó a Embalajes Marplatenses S.A. indicadores clave de resultado.

5.4. *Inteligencia de negocio y análisis de datos*

La inteligencia de negocios (BI, por sus siglas en inglés) y la analítica empresarial se han convertido en elementos fundamentales en la estrategia operativa de las empresas.

El *Data Warehousing Institute (tdwi.org*^[31]) define inteligencia de negocios como “la combinación de tecnología, herramientas y procesos que permiten transformar los datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial”.

En definitiva, la integración de la inteligencia de negocios en la estrategia empresarial es esencial para optimizar el uso de recursos, supervisar el cumplimiento de los objetivos corporativos y potenciar la toma de decisiones acertadas, con el fin de alcanzar resultados superiores.

Desde una perspectiva gerencial y empresarial, la evaluación de datos históricos y actuales, así como el análisis del desempeño, ofrece a los usuarios de los sistemas de inteligencia de negocio la capacidad de obtener conocimientos (*insights*) que les permitan tomar decisiones más informadas y fundamentadas. En conclusión, la inteligencia de negocios se refiere a las operaciones de captura, acceso, comprensión y conversión de los datos crudos en información accionable con el propósito de mejorar su desempeño o rendimiento. Con esta información, se construyen los tableros que desplieguen los KPIs y KRIs comentados anteriormente.

Respecto del análisis de datos o analítica de datos, existen numerosas definiciones. La ISACA (organización profesional de impacto mundial en sistemas de información^[32]) publicó en 2011 un artículo con la definición de análisis de datos: “implica los procesos y actividades diseñados para obtener y evaluar datos y extraer información útil”. Las técnicas más utilizadas incluyen consultas e informes (*querying and reporting*), visualización, minería de datos, análisis predictivo, lógica difusa, optimización, así como el procesamiento en tiempo real de audio, video o fotografía, entre otros.

A medida que la investigación progresaba para aclarar algunas definiciones, el equipo se enfrentó a cierta ambigüedad al tratar de distinguir entre inteligencia de negocios y análisis de datos, ya que ambos casos involucran la generación de informes y visualizaciones. ¿Son definiciones equivalentes? En el siguiente párrafo se pudo hallar una respuesta.

Analítica Descriptiva (*Descriptive Analytics*^[33]): implica la preparación y análisis de datos históricos con el fin de identificar patrones y tendencias. Al generar informes, se obtiene un conocimiento profundo a partir de dichos datos. Emplea datos para describir y explicar eventos pasados. Por esta razón, el análisis descriptivo responde a la pregunta: ¿qué sucedió?.

La analítica descriptiva emplea técnicas como modelos de regresión, modelado y visualización de datos, y en este último proceso es en dónde está estrechamente vinculada a la inteligencia de negocios. El enfoque principal de la inteligencia de negocios es proporcionar información procesable para apoyar la toma de decisiones empresariales estratégicas. Involucra la presentación de datos históricos y actuales de manera accesible para facilitar el análisis y la

planificación. Por otro lado, la analítica descriptiva se centra en describir y comprender los datos, identificando patrones y tendencias en conjuntos de datos históricos. Su objetivo es explicar "qué sucedió" mediante el análisis detallado de eventos pasados.

La inteligencia de negocios involucra herramientas como *dashboards* e informes interactivos para proporcionar una visión general accesible y fácil de entender. En cambio, la analítica descriptiva utiliza técnicas como modelos de regresión y visualización de datos para describir y explicar eventos pasados en detalle.

Ambas disciplinas son complementarias y se integran para ofrecer una visión completa de los datos y apoyar la toma de decisiones empresariales. Mientras que la inteligencia de negocios se centra en la accesibilidad y la acción estratégica, la analítica descriptiva profundiza en la comprensión detallada de los eventos pasados.

5.4.1. *Series de tiempo y su aplicación*

Manu Joseph^[34], autor de "Modern Time Series Forecasting with Python", presenta una definición concisa de las series de tiempo. Según él, son un conjunto de observaciones registradas secuencialmente en el tiempo. Esta definición resalta la importancia del factor temporal en este contexto.

Iniciar la comprensión de la serie de tiempo implica embarcarse en lo que los profesionales denominan el análisis exploratorio de datos (EDA^[35]) y conduce a una predicción más sencilla de encarar. Consta de diferentes procesos, comenzando por determinar los componentes de la serie de tiempo. En este punto se busca definir la tendencia, la estacionalidad y la irregularidad o ruido de la serie.

- Tendencia: es el patrón ascendente o descendente a largo plazo.
- Estacionalidad: es un movimiento recurrente o periódico que presenta la variable de la serie.
- Ruido: es aquello remanente que permanece una vez removidas la tendencia y la estacionalidad.

Una vez identificados los componentes de la serie temporal, se inicia la etapa de visualización para revelar patrones, anomalías y establecer supuestos preliminares sobre los datos mediante una variedad de representaciones gráficas.

Concluida la fase de visualización, el siguiente paso es descomponer la serie en sus componentes principales: tendencia, estacionalidad y ruido. Finalmente^[36], el análisis exploratorio concluye abordando la detección y tratamiento de valores atípicos que puedan estar presentes en las series temporales. Estos valores se destacan por su distancia significativa con

respecto al resto de los puntos observados. La técnica más popular para la detección de valores atípicos es la desviación estándar.

Las series temporales encuentran aplicaciones en diversas áreas, como la clasificación basada en series temporales, la interpretación de causalidades y, específicamente para este proyecto, la capacidad de realizar predicciones.

5.4.1.1. Algunos algoritmos de predicción

Antes de decidir qué algoritmos de predicción emplear, es importante crear conjuntos de datos de validación y prueba^[37]. Estos conjuntos no participan en el entrenamiento de los modelos. En cambio, los datos de validación se reservan para evaluar la calidad de las predicciones. En otras palabras, permiten determinar qué tan acertadamente los modelos seleccionados están realizando predicciones. Por otro lado, los datos de prueba son esenciales para realizar la comparación entre diversas alternativas de modelado. Se utilizan para ajustar hiperparámetros y características, proporcionando una indicación clara de cómo se desempeña el modelo en datos no vistos.

Cuando se trata de predicciones es común separar los datos más recientes, especialmente al trabajar con series temporales en donde se debe considerar la variable del tiempo. A su vez, la bibliografía consultada recomienda que tanto el conjunto de validación como el de prueba sean de igual tamaño, para que todo lo relacionado con decisiones de modelado reflejen los datos de prueba.

Una vez determinados los conjuntos anteriores, es necesario revisar qué métricas medirán los resultados de las predicciones desarrolladas^[38]. Dos de las más utilizadas en los pronósticos son:

- Error absoluto medio (MAE, de sus siglas en inglés): Es el promedio de la diferencia absoluta entre el valor predicho y el valor real en cada punto *forecasteado*.

$$MAE = \frac{1}{N \times L} \sum_{j=1}^N \sum_{i=1}^L |f_{ij} - y_{ij}|$$

Donde f e y son los valores *forecasteados* y observados (reales), respectivamente.

- Error Cuadrático Medio (MSE, de sus siglas en inglés): Es el promedio del cuadrado del error entre el pronóstico (f) y el valor observado (y).

$$MSE = \frac{1}{N \times L} \sum_{j=1}^N \sum_{i=1}^L (f_{ij} - y_{ij})^2$$

En ambos casos, N es el número de series temporales, L es la longitud de la serie limitada al conjunto de prueba.

Es tiempo de presentar algunos de los modelos de pronóstico más comúnmente utilizados, específicamente dentro de la categoría de modelos clásicos. Vale la pena recordar que los modelos más avanzados basados en aprendizaje automático, que son considerablemente más poderosos, no son aplicables al conjunto de datos disponible de Embalajes Marplatenses S.A. Esto se debe a que dicho conjunto de datos es reducido y presenta algunas carencias que impiden su aplicación.

- Suavizado exponencial (ETS^[39]): Cuenta con diversas variantes, entre las que se incluyen el suavizado exponencial simple, el doble y el estacional de Holt-Winters. Este método ha estado en uso desde la década de 1950 y se ha demostrado altamente eficaz en la generación de pronósticos. Se basa en dos principios: el primero, tomando como base el método Naïve o del ingenuo, en donde el último valor observado se pronostica; y el segundo, el modelo del promedio móvil que calcula la media entre las últimas n observaciones. El suavizado exponencial realiza una combinación de ambos principios y sostiene que “toda la historia es importante, pero la historia reciente es más importante”, Joseph, M. (2022).

Tal es así, que el método se calcula como el promedio ponderado de las observaciones afectadas por diferentes pesos, que van disminuyendo su valor a medida que se avanza en la historia de las observaciones. La fórmula:

$$f_t = \alpha \cdot y_{t-1} + \alpha \cdot (1 - \alpha) \cdot y_{t-2} + \alpha \cdot (1 - \alpha)^2 \cdot y_{t-3} + \dots$$

Donde α es el parámetro de suavizado, $0 \leq \alpha \leq 1$ y permite establecer con qué velocidad deben disminuir los pesos que afectarán a los valores observados y .

Lo anterior puede crecer en complejidad e incorporar varios parámetros dependiendo de las características de la serie temporal en concreto.

- Modelos Autorregresivos de Promedio Móvil Integrado (ARIMA^[40]): A diferencia de los modelos ETS que están basados en la tendencia y la estacionalidad, ARIMA se centra en la autocorrelación, es decir, el grado de relación que existe entre el valor actual de la serie y los valores pasados.

Dentro de la familia ARIMA podemos encontrar:

1. Modelos de autorregresión AR(p):
 - Utilizan regresión lineal con p pasos temporales anteriores o p rezagos.
 - Se representan matemáticamente con una constante (c) y un error (ϵ) en el tiempo t .
2. Modelos de media móvil MA(q):

- En lugar de utilizar valores observados pasados, se utilizan los últimos q errores en la predicción.
 - Se modelan con el ruido blanco (ϵ) y una constante (c).
3. Modelos de media móvil autorregresiva ARMA(p,q):
- Combinación de modelos AR(p) y MA(q).
 - La fórmula general es $y_t = AR(p) + MA(q)$.

5.5. Transformación digital en organizaciones y empresas

En la actualidad, la mayoría de las empresas cuenta con cierta digitalización, como por ejemplo, la utilización de correo electrónico, sitio web, perfiles en redes sociales, implementación de *softwares* contables o de administración de recursos, etc. Este proceso no implica considerarlas como empresas digitales, puesto que existen una serie de normas que las compañías deben cumplir para alcanzar esa definición^[41].

La transformación digital conlleva una reorganización en las tareas habituales, y según Luis Joyanes Aguilar, una “concientización de la necesidad de digitalizarse y transformarse digitalmente. La empresa digital trae aparejada un nuevo modelo de economía, economía digital.”

La transformación digital, según la definición de Margaret Rouse^[42], implica la reinención de una organización mediante la utilización de tecnología digital para mejorar su desempeño y servicio a sus partes constituyentes. En este contexto, lo "digital" se refiere al uso de tecnologías que generan, almacenan y procesan datos. Rouse destaca que la simple implementación de tecnología no constituye en sí misma una transformación digital, pero sí tiene el poder de cambiar fundamentalmente una organización para aprovechar el potencial de las tecnologías disponibles.

La transformación digital se ha vuelto esencial en el panorama empresarial actual para garantizar la competitividad en un entorno globalizado, interconectado y digital^[43]. La presencia en un mercado omnicanal, donde los clientes tienen acceso a través de diversos canales se ha convertido en una necesidad para adaptarse a las expectativas cambiantes y aprovechar las oportunidades en evolución. Es un proceso que va más allá de la implementación de tecnologías y abarca la reconfiguración integral de las operaciones y estrategias empresariales para abrazar la era digital.

Gil Press publicó en 2016 un artículo en la revista *Forbes*^[44] planteando tres estrategias y tres tecnologías que impactan en la transformación digital:

Estrategias:

1. La transformación digital se posicionará como el principal impulso estratégico para la mayoría de los CEOs.
2. Las iniciativas de transformación digital se consolidarán, dando lugar a estrategias de negocios con una visión muy amplia durante el año 2016.
3. La transformación digital exigirá nuevas habilidades y competencias, con un cambio hacia inversiones en tecnologías de la información (*TI*).

Tecnologías:

1. *Big Data Analytics* será el soporte fundamental de la transformación digital.
2. Internet de las cosas (*IoT*) actuará como un catalizador para expandir la transformación digital a todos los sectores de la economía.
3. La Inteligencia Artificial (*IA*) impulsará nuevos flujos de ingresos en el ámbito de la transformación digital.

Daniel Newman, publicó en Forbes las siguientes tendencias para 2017^[45]:

1. La adaptabilidad es más vital para el éxito que nunca. Cita el caso del éxito de *cloud computing* a lo largo de 2016. Newman también considera que la transformación digital no se trata solo de tecnología, sino de reunir el poder de la tecnología con una cultura que abrace el cambio que puede conducir a la organización.
2. La importancia del crecimiento de la experiencia de usuario. La experiencia del cliente (incluyendo empleados) es el objetivo final de cualquier transformación digital.
3. La innovación debe suceder rápidamente. La innovación proactiva es una de las mejores maneras de mantenerse competitivo en un mercado en evolución.
4. Trabajo remoto. Los jóvenes profesionales prefieren la flexibilidad a la compensación. La tecnología portable y la conexión a internet permiten a las empresas conectarse y retener a los mejores talentos en cualquier parte del mundo.
5. El advenimiento de la realidad aumentada (*RA*) y la realidad virtual (*RV*).
6. Las APIs representan una herramienta clave para adoptar la transformación digital.
7. *Big Data and Analytics*. En la era de la transformación digital todo gira en torno a los datos que permiten medir todo, o casi todo.
8. La transformación digital está conducida por *IoT*.
9. Las máquinas inteligentes y la inteligencia artificial están despegando en gran medida. Su relación con la tecnología continúa evolucionando. Pronto las máquinas podrán aprender y adaptarse a sus entornos.
10. Destrucción de silos. El rol de los CIO ha cambiado radicalmente en los últimos años. Con el advenimiento de los nuevos roles del CDO (*Chief Digital Officer*) y el CCO (*Chief Customer Officer*) se ve la importancia de la transformación digital, no solo en la

tecnología de una compañía, sino en toda la organización completa. Un enfoque en romper los silos dará a la innovación más espacio para crecer y la colaboración se hará más fácil.



Figura 11: Transformación digital^[46]

5.5.1. Datos como impulsores de la transformación digital

En su libro "The Digital Transformation Playbook: Rethink your business for the digital age", David L. Rogers^[47] ofrece una guía completa sobre cómo las empresas pueden adaptarse en la era digital. Rogers destaca el cambio radical en el papel de los datos, que pasan de ser simples insumos a convertirse en activos estratégicos clave que impulsan la innovación y el valor en todas las industrias. Esta transformación de mentalidad está llevando a las empresas a reevaluar cómo gestionan y utilizan los datos, expandiendo su impacto más allá de los límites de una sola industria y hacia un panorama empresarial más amplio. Este cambio de enfoque hacia los datos reconoce su valor estratégico para impulsar la innovación y la creación de valor en todos los ámbitos empresariales.

Repensar los datos^[48]:

Anteriormente, los datos eran considerados como herramientas para medir y gestionar procesos. En la actualidad, se han convertido en activos estratégicos para la creación de valor en las empresas. Sin embargo, para aprovechar al máximo el potencial de los datos, las empresas deben cambiar su enfoque. Esto implica no solo la capacidad de generar y almacenar datos, sino también la habilidad de transformarlos en información valiosa. Este cambio de paradigma representa un desafío de liderazgo para las empresas, que deben adoptar una mentalidad centrada en los datos y desarrollar estrategias efectivas para gestionar e invertir en ellos como activos empresariales.

En la siguiente tabla, David L. Rogers lista los cambios desde una mentalidad “analógica” de los datos hacia una “cultura digital”. Estos cambios reflejan la transformación en que las empresas perciben, gestionan y utilizan los datos en el entorno empresarial digital actual.

AS IS: era analógica	TO BE: era digital
Los datos son caros de generar en la empresa	Los datos se generan continuamente en todas partes
El desafío de los datos es almacenarlos y gestionarlos	El desafío de los datos es convertirlos en información valiosa
Las empresas solo hacen uso de datos estructurados	Los datos no estructurados son cada vez más utilizables y valiosos
Los datos se gestionan en silos operativos	El valor de los datos está en conectarlos a través de silos
Los datos son una herramienta para optimizar procesos	Los datos son un activo intangible clave para la creación de valor

Data: Changes in Strategic Assumptions from the Analog to the digital Age^[49]

5.5.2. *Desafíos a nivel organizacional al aplicar una estrategia de datos*

El hecho de implementar una **estrategia basada en datos**^[50] va más allá de las cuestiones técnicas y aborda los cambios culturales y operacionales dentro de la empresa. La necesidad de reorganizar los procesos comerciales para aprovechar plenamente la riqueza de datos disponibles se convierte en un aspecto decisivo en este proceso.

Es posible enumerar los desafíos más importantes expuestos por David L. Rogers^[51]:

1. Incorporar *skills* de datos: esto implica encontrar a las personas con las habilidades adecuadas. Contar con analistas de datos capaces de realizar el análisis técnico, así como educar a toda la organización sobre la importancia y aplicación de los datos en su trabajo diario. Es definitorio desarrollar una cultura empresarial que adopte el pensamiento analítico y promueva la integración entre los analistas cuantitativos y los tomadores de decisiones empresariales para aprovechar plenamente el potencial de los datos como activo estratégico.
2. Conectar silos: se refiere a las barreras internas dentro de las organizaciones que dificultan el intercambio de datos entre diferentes departamentos. Esto puede manifestarse en la separación de datos en distintas bases de datos o sistemas, lo que dificulta la obtención de una visión completa y unificada del cliente o del negocio en su conjunto. Estas divisiones pueden estar reforzadas por la tendencia de cada departamento a considerar sus datos como propiedad exclusiva. Para superar este desafío, es necesario fomentar la colaboración y la integración entre los diferentes equipos y departamentos, con el fin de unificar los datos y aprovechar todo su potencial para la toma de decisiones.

3. Intercambio de datos con socios comerciales: destaca la importancia de compartir datos no solo dentro de la organización, sino también con colaboradores externos. Este intercambio se ha convertido en un elemento importante en las negociaciones y acuerdos comerciales, especialmente para empresas que no tienen acceso directo al punto de venta final. Al asociarse con socios comerciales, las empresas pueden obtener acceso a datos valiosos que les permiten mejorar su eficiencia operativa y comprender mejor las necesidades y comportamientos de los clientes.
4. Seguridad y privacidad: surge a medida que las empresas aumentan la recopilación y utilización de datos, especialmente datos de clientes. Esta creciente dependencia de los datos también conlleva riesgos adicionales en términos de seguridad y privacidad.

A partir de lo anterior, es importante recalcar una vez más el desafío estratégico que implica el cambio cultural y operativo requerido en todas las organizaciones para maximizar el potencial de los datos. En un entorno donde los datos siguen aumentando su volumen, cada empresa se ve obligada a reconsiderar su enfoque hacia estos activos. No importa la industria en la que operen, todas las empresas tienen acceso a una cantidad sin precedentes de datos. **La clave radica en desarrollar una visión clara y la capacidad necesaria para convertir estos datos en innovación y valor**^[52]. Al adoptar una mentalidad que coloca a los datos en el centro de la toma de decisiones, las organizaciones pueden no solo optimizar sus procesos y mejorar la experiencia del cliente, sino también transformar la forma en que aprenden y se innovan.

6. Sistema vs software comercial de visualización

Este segmento tiene como propósito abordar uno de los interrogantes señalados durante la revisión del protocolo. El equipo consideró que era un aspecto intrigante para analizar, por lo que se llevó a cabo una evaluación entre los beneficios y las posibles desventajas asociadas tanto al desarrollo de un sistema propio como al uso de un software comercial para la presentación de indicadores.

	<i>Beneficios</i>	<i>Peso en términos de política de la empresa</i>
<i>Sistema propio</i>	<p>Personalización total</p> <p>Se tiene la capacidad de adaptar el sistema exactamente a las necesidades específicas y requerimientos empresariales, manteniendo un control total sobre la apariencia y</p>	Alto

	funcionalidad.	
	<p>Integración personalizada Permite la integración directa con otras aplicaciones internas y bases de datos sin depender de las integraciones estándar de un software comercial.</p>	Medio
	<p>Menor costo a largo plazo Una vez desarrollado, el costo de mantenimiento puede ser menor a largo plazo en comparación con las suscripciones a software comercial, que suelen tener tarifas recurrentes.</p>	Alto
<i>Software comercial</i>	<p>Implementación rápida Las interfaces de usuario amigables de los software comerciales permiten una implementación rápida sin necesidad de un desarrollo extenso.</p>	Alto
	<p>Amplia variedad de conectores Los software comerciales suelen ofrecer una amplia variedad de conectores predefinidos para integrarse con diferentes fuentes de datos y aplicaciones.</p>	Bajo
	<p>Actualizaciones y mantenimiento automático Las actualizaciones y el mantenimiento son manejados por el proveedor, liberando al equipo interno de la carga de mantener el sistema.</p>	Medio

	<i>Contraindicaciones</i>	<i>Peso en términos de política de la empresa</i>
<i>Sistema propio</i>	Alto costo inicial	Medio a bajo (en el caso

	El desarrollo personalizado puede ser costoso, especialmente en términos de tiempo y recursos. Los costos de desarrollo, prueba e implementación inicial pueden ser significativos.	de este proyecto)
	<p>Tiempo de desarrollo extensivo</p> <p>El tiempo necesario para desarrollar un sistema personalizado puede ser extenso, y durante este período, puede que no se tenga acceso inmediato a las funciones de visualización.</p>	Alto
	<p>Menos funcionalidades listas para usar</p> <p>Puede llevar más tiempo implementar nuevas características en comparación con un software comercial ya que este último ofrece una variedad de funciones listas para usar.</p>	Medio
Software comercial	<p>Limitaciones de personalización</p> <p>Es posible encontrar limitaciones en términos de personalización, ya que se está utilizando una plataforma estándar.</p>	Alto
	<p>Costos recurrentes</p> <p>Las suscripciones a software comercial generalmente implican costos recurrentes, que pueden aumentar con el tiempo.</p>	Alto
	<p>Dependencia del proveedor</p> <p>Está sujeto a las actualizaciones y cambios del proveedor del software, lo que puede afectar al flujo de trabajo.</p>	Alto

Después de evaluar detenidamente, se optó por el desarrollo interno de un sistema propio. Esta decisión se basa en la necesidad de una personalización precisa, el control total sobre la apariencia y funcionalidad, así como el menor costo a largo plazo. Adicionalmente, la

flexibilidad del desarrollo propio se ajusta a las necesidades específicas de la empresa y permite una óptima adaptación a sus demandas particulares.

Además de resaltar las ventajas a nivel de software, se hizo hincapié en el cambio cultural potencial que este enfoque podría impulsar dentro de la organización. Adoptar un sistema propio no solo implica una mejora en las capacidades tecnológicas, sino que también representa un cambio hacia una mentalidad orientada hacia los datos, lo que transforma la forma en que la empresa aborda la toma de decisiones. Este aspecto se considera fundamental para el crecimiento continuo y la capacidad de adaptación de la empresa en un contexto donde se enfatiza cada vez más la importancia de los datos.

7. Análisis de madurez

Este segmento tiene como objetivo definir la madurez de datos en una empresa y realizar un análisis sobre el estado de Embalajes Marplatenses S.A. después de implementar la solución de datos. El equipo considera muy importante el análisis de madurez, ya que aporta tanto la perspectiva ingenieril como la de gestión de proyectos, más allá de las habilidades técnicas requeridas para el desarrollo del sistema. Por esta razón, el proyecto involucró una fase inicial de análisis exhaustivo de requerimientos, investigación y posterior análisis de madurez para evaluar el impacto que esta solución de datos tuvo en la organización.

7.1. Frameworks de madurez de datos, etapas y aspectos principales

En primer lugar, es importante destacar que existen diversos *frameworks* para evaluar la madurez de los datos de una organización. La perspectiva presentada por Joe Reis y Matt Housley^[53], por ejemplo, se centra en la ingeniería de datos, ya que estos autores buscan situar el papel del ingeniero de datos dentro de este modelo.

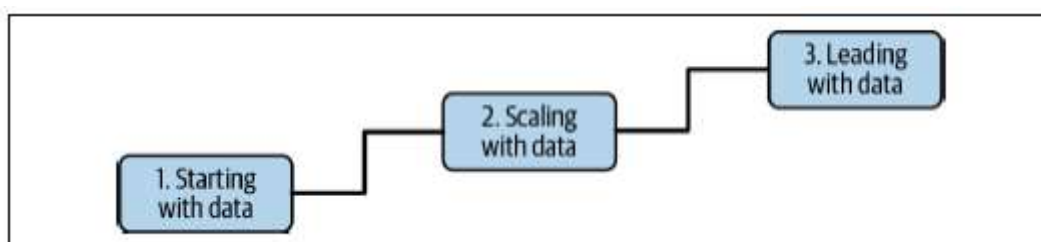


Figura 12: Our simplified data maturity model for a company. Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering^[54]

Lo que ellos plantean es un modelo de tres etapas, que a continuación se enumeran:

Etapa 1 - Comenzando con datos:

- Puede que la organización no conozca los objetivos que está persiguiendo, o si los tiene, puede que estén poco definidos.
- La infraestructura necesaria para una arquitectura de datos puede estar planificándose o en las primeras etapas de desarrollo. Tampoco se tiene un uso exhaustivo de esta arquitectura.
- Equipo de datos muy reducido, medido en unidades y generalistas.

Etapa 2: Escalando con datos:

- Las solicitudes de datos dejan de ser puntuales para pasar a ser un procesos formales.
- Se explota la arquitectura de ignición y se la comienza a escalar. Se prepara el terreno para que el futuro de la organización sea *data-driven*.
- Se cuenta con equipos con *expertise* en las etapas del ciclo de vida.

Etapa 3: Liderando con datos:

- La organización se considera *data-driven* o impulsada por datos. Se cuenta con la democratización en datos necesaria para realizar análisis complejos, como *machine learning*.
- Anexar nuevas fuentes de datos se realiza de forma fluida, aportando mucho valor a la organización.
- La gobernanza y la calidad del dato toman protagonismo.
- Los equipos de datos profundizan su *expertise*.

Para ahondar un poco más sobre la madurez de datos en las organizaciones, se investigó sobre aspectos importantes a evaluar sobre esta temática. El Departamento de Gasto Público y Reforma de Irlanda suele publicar diversos documentos. Uno de ellos titulado “Data Management Maturity Assessment^[55]” donde se define y detalla el *framework* de madurez de datos y se aportan dos aspectos principales a evaluar dentro de los servicios públicos. Este departamento define al *framework* (marco) de la madurez de datos como: “Método para que las organizaciones evalúen su capacidad actual de gestión de datos al examinar sus procesos y procedimientos de gestión de datos en diversas áreas y comparar estos procesos y procedimientos con niveles predefinidos de madurez (...) Un objetivo clave de una evaluación de madurez de datos es proporcionar una comprensión de los procesos, procedimientos y normas de datos existentes que están en juego en la organización.”

Los dos aspectos principales a evaluar:

- “En primer lugar, proporciona una vista de los procesos, prácticas, capacidades, flujos de trabajo y gobernanza de gestión de datos existentes dentro de una organización. Evalúa diferentes procesos en estas áreas e identifica un nivel asociado de madurez de datos en áreas definidas.
- En segundo lugar, ayuda en la definición de un nivel de madurez deseado en un área de datos, incluida la elaboración de un plan de mejoras necesario para alcanzar este objetivo. Las evaluaciones de madurez de datos suelen ser el punto de partida desde el cual una organización puede desarrollar una estrategia de datos.”

El equipo considera que esta definición de marco de madurez de datos junto con los aspectos principales a evaluar son aplicables a todo tipo de organización, más allá de que sean públicas o privadas.

7.2. *Análisis de madurez de datos para Embalajes Marplatenses S.A.*

En la siguiente tabla y alineado con lo anteriormente definido en función de ambas fuentes bibliográficas, se expone el análisis desarrollado para Embalajes Marplatenses S.A.

Previo a la implementación del sistema	Posterior a la implementación del sistema	Peso del logro
La empresa no contaba con una clara definición sobre los objetivos a medir. Fue necesario brindar orientación al gerente (también referente funcional del proyecto). Esto se hizo para proporcionar claridad sobre las posibilidades y limitaciones en el desarrollo de indicadores.	A través del proceso de elicitación, se pudo identificar los indicadores clave y agregar otros nuevos que proporcionan un mayor valor a la empresa.	Alto
No se contaba con ningún tipo de infraestructura que permita desplegar una solución de datos.	Se desarrolló una arquitectura reducida de ignición, es decir, una arquitectura de datos que permite comenzar a encender los motores hacia un comportamiento data-driven.	Alto

Los reportes se realizaban de manera manual y ad-hoc por parte del gerente.	Los reportes se encuentran automatizados. Sin embargo, cualquier desarrollo adicional debe ser implementado por el equipo de datos. No se cuenta con un proceso <i>self-service-BI</i> .	Alto
Ausencia de equipo de datos y personal de sistemas.	La empresa cuenta con un equipo de sistemas de datos, que no es interno. Este equipo continúa proveyendo mantenimiento evolutivo y correctivo.	Medio
Ausencia de única fuente de verdad que permita la centralización de la información.	Inicia el proceso de democratización de datos mediante la creación de una base de datos que centraliza información más importante de la empresa en el ámbito financiero.	Alto
Toma de decisiones basadas en la experiencia y la intuición de los mandos medios.	Toma de decisiones fundamentadas, con base en información confiable y precisa.	Alto
Dificultad en la identificación y auditoría de los datos debido a la presencia de múltiples reportes.	Proceso de curado de datos que los deja listos para el despliegue. Incorporación de trazabilidad del dato.	Medio

7.3. Conclusiones a partir del análisis

La tabla refleja una transformación significativa en Embalajes Marplatenses S.A. en términos de madurez de datos. En la etapa inicial, la empresa carecía de infraestructura para una solución de datos, los reportes eran manuales y no existía un equipo dedicado a la gestión de datos. Sin embargo, con la implementación de una arquitectura inicial y la automatización de reportes, se ha logrado un avance considerable.

Aunque la empresa aún se encuentra en una fase temprana del desarrollo en su capacidad de datos, se han sentado las bases para el crecimiento futuro. La creación de una base de datos única (y en este caso, centralizada), junto con el inicio del proceso de democratización de datos

son indicadores positivos. Además, la toma de decisiones se fundamenta en datos, y marca un cambio significativo en relación a la dependencia de la experiencia y la intuición.

Es importante reconocer que, en esta etapa, Embalajes Marplatenses S.A. ha superado obstáculos iniciales como la ausencia de arquitectura, falta de trazabilidad del dato, redundancia, etc. y ha creado los cimientos para el avance continuo en su madurez de datos. Aunque nuevos desafíos se presentan (como la falta de un equipo de sistemas propio), son áreas clave que pueden abordarse en las próximas etapas de desarrollo.

En la empresa se ha experimentado un cambio importante en su enfoque operativo y cultura organizacional, impulsada por un incremento notable en la madurez de su gestión de datos. Tras la implementación de una arquitectura de datos inicial y la automatización de informes, se ha forjado un nuevo paradigma empresarial. **Hoy en día, la toma de decisiones se apoya en datos, lo que representa un cambio de relevancia en la mentalidad de la organización.** Se ha establecido una nueva dinámica donde la información es el núcleo de las operaciones y la estrategia empresarial. **Este cambio refleja una evolución cultural en la empresa, donde se ha adoptado una mentalidad proactiva hacia los datos, reconociendo su papel fundamental como impulsores de la innovación y la competitividad.** Embalajes Marplatenses S.A. ha transitado hacia un nuevo modelo de toma de decisiones, impulsado por una cultura organizacional transformada que valora y aprovecha el potencial de los datos.

8. El sistema

Uno de los desafíos iniciales que enfrentó el equipo fue la integración con la fuentes de datos existente en la empresa. En la actualidad, esta fuente consiste en la exportación de informes desde un sistema de gestión que almacena datos de distintas áreas de la organización. A partir de estos reportes, se comenzó a delinear un primer modelo de datos, basado en la ingesta cruda de estos documentos. El paso siguiente fue modelarlos en busca de los primeros indicadores de negocio. Cabe resaltar que el proceso entre la carga cruda de los datos y la obtención de los indicadores fue necesario implementar un proceso de curado que garantice el cálculo de estos indicadores mediante procesos de limpieza.

Para la obtención de lo que se consideraban KPIs (luego, KRIs) fue fundamental y determinante trabajar junto con el gerente de la empresa, quien tomó el rol de referente funcional. Es de importancia, una vez logrados los procesos técnicos, validar junto con el gerente los valores obtenidos y realizar el contraste con lo que se esperaba por parte del cliente, tanto a nivel sistema como a valor esperado de los indicadores.

8.1. La solución propuesta a alto nivel

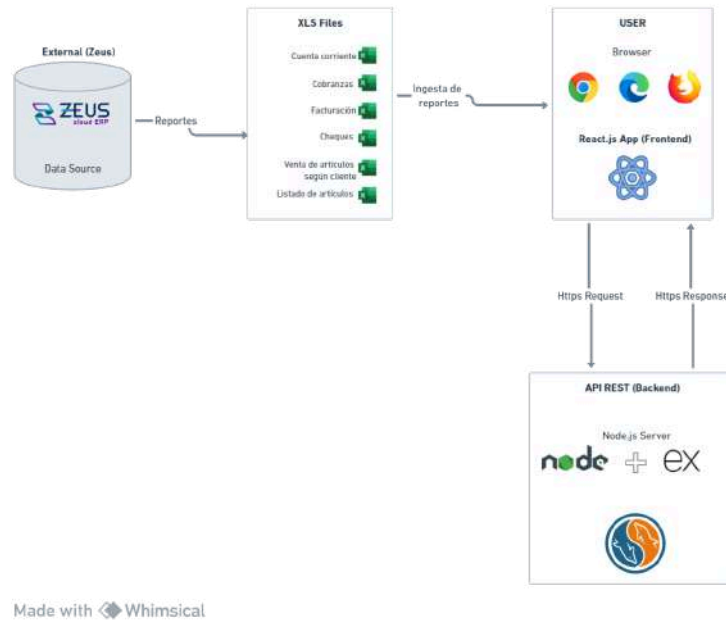


Figura 13: Solución propuesta a alto nivel. Fuente: Elaboración propia

[Solución a alto nivel](#)

8.2. Backend

El backend se compone de un servidor Node.js, donde se ha desarrollado una API REST que desempeña un papel central en la ingestión y procesamiento de datos. Esta API actúa como la interfaz entre el frontend y la base de datos MySQL que sirve como sistema de persistencia. La funcionalidad clave de la API incluye la ingestión de archivos con datos crudos provenientes del sistema contable de origen. Este proceso de ingestión garantiza que los datos se integren de manera efectiva en el sistema, preparándolos para su almacenamiento y posterior acceso. La elección de MySQL como sistema de persistencia se basa en su confiabilidad y capacidad para gestionar transacciones complejas. En conjunto, esta configuración backend establece una sólida infraestructura para el manejo eficiente de datos en la aplicación.

8.2.1. Respuesta a los interrogantes planteados en el marco teórico

Características esenciales de la fuente de datos:

- Naturaleza de la fuente: La fuente de datos es una aplicación empresarial que almacena información crítica relacionada con facturación, cobranza, cheques, cuentas corrientes, stock y ventas.
- Persistencia de datos: Los datos se almacenan de manera permanente en la aplicación.

- Acceso a datos: El acceso a los datos se realiza a través de una funcionalidad de reportería, ya que no se pudo obtener acceso a través de una API.
- Consistencia de datos: Aunque los datos presentan cierto nivel de consistencia, la información es agregada por especialistas, lo que implica que posibles errores humanos no son fácilmente detectables por el equipo.
- Comunicación en la secuencia de desarrollo: No hay una comunicación directa con el proveedor del sistema origen, lo que puede limitar la capacidad de gestionar cambios y resolver problemas de manera eficiente.
- Frecuencia de extracción: La extracción de datos se realiza mensualmente, ya que los informes deben generarse con esta periodicidad.
- Manejo de cambios en el esquema: No se proporciona información sobre cómo se manejan los cambios en el esquema de datos ni cómo se comunican a las partes interesadas.
- Verificación de calidad de datos: No se menciona la realización de verificaciones de calidad de datos, lo que podría implicar posibles inconsistencias no detectadas.

Compatibilidad del sistema de almacenamiento:

- Velocidades de escritura y lectura: La solución de almacenamiento, basada en una base de datos relacional MySQL, es compatible con las velocidades de escritura y lectura requeridas por la arquitectura.
- Cuello de botella para procesos: En este caso, el almacenamiento no crea un cuello de botella ya que la base de datos relacional es suficiente para la solución propuesta.
- Escalabilidad futura: El sistema de almacenamiento no manejará la escala futura debido a las características del sistema. No se proporcionan detalles sobre cómo se gestionarían aumentos significativos en la carga de datos.
- Frecuencia de acceso y modo de ingesta: Dado que la frecuencia de acceso es mensual y la ingesta de datos se realiza por lotes, la compatibilidad con la base de datos relacional MySQL es adecuada para esta solución.

Casos de uso para los datos e ingesta:

- Uso: Los datos son aplicados en la generación de informes para áreas específicas como facturación, cobranza, pagos, perfiles financieros y ventas.
- Reutilización de datos: Uno de los objetivos principales es establecer una única fuente de verdad, eliminando redundancias y multiplicidad en los informes utilizados por los mandos medios. Esto hace que el conjunto de datos resultante sea altamente reutilizable.

- Destino de los datos después de la ingesta: Posterior a la ingesta, los datos son almacenados en una base de datos relacional con modelado elemental, facilitando la creación de consultas para obtener los indicadores necesarios.
- Frecuencia de acceso a los datos: La frecuencia de acceso a los datos es mensual, alineándose con la estructura de almacenamiento en la base de datos relacional MySQL. Este enfoque es adecuado para satisfacer las necesidades de acceso.
- Transformación de datos: La transformación de datos es una parte esencial del proceso y se aborda mediante controladores de ingesta y transformación. La metodología de la transformación se detalla en la siguiente sección.

Transformaciones:

- Simplicidad y autoaislamiento de la transformación: Las transformaciones aplicadas son concebidas con la mayor simplicidad y autoaislamiento posible. Se implementaron puntualmente a las entidades de manera aislada durante su ingesta. Para obtener indicadores, se aplicaron estas transformaciones entre entidades a través de consultas SQL.
- Reglas comerciales respaldando las transformaciones: Las transformaciones están respaldadas por reglas comerciales establecidas por el negocio. Específicamente, aquellas relacionadas con el perfil comercial del cliente surgieron de un análisis realizado por el referente funcional. Este análisis tuvo como objetivo obtener un indicador que clarificase los comportamientos financieros. Las reglas de negocio fueron fundamentales en la construcción de todos los indicadores.

8.2.2. Estructura del backend

En el backend, se emplea Express para manejar las rutas y Sequelize como ORM (*Object-Relational Mapping*) para la interacción con la base de datos MySQL. La autenticación se gestiona mediante Passport, y se utilizan archivos Excel para la carga de datos.

En la configuración del servidor, se establecen sesiones, CORS (*Cross-Origin Resource Sharing*) y se gestionan las rutas con Express, un conocido *framework* de Node.js. También se realiza la sincronización del modelo de datos con la base de datos MySQL, y el servidor se inicia en un puerto específico.

Las rutas son manejadas con Express Router y Passport para la autenticación de Google. Se incluyen rutas para cargar archivos Excel, realizar consultas a la base de datos y otras operaciones.

Los modelos Sequelize representan las tablas en la base de datos MySQL, definiendo los campos y tipos de datos correspondientes.

Se implementan *middlewares* para gestionar la carga de archivos Excel y otros aspectos de la lógica de negocio.

Los controladores se encargan de la lógica específica de cada ruta, así como la ingesta de archivos Excel. Por ejemplo, el controlador *upload* maneja la carga de archivos Excel a la base de datos, mientras que otros controladores gestionan la creación, actualización y eliminación de registros en la base de datos.

La configuración se ha modularizado, almacenando credenciales y configuraciones en diferentes archivos para mantener un código más organizado.

Resumiendo, se dispone de un backend estructurado con Express, Sequelize, Passport y otras herramientas para gestionar la lógica de la aplicación, la autenticación y la interacción con la base de datos MySQL.

A continuación, se presenta una representación de la estructura jerárquica de las carpetas que conforman el backend. Esto ofrece una visión de cómo está organizado el sistema en términos de directorios y subdirectorios.

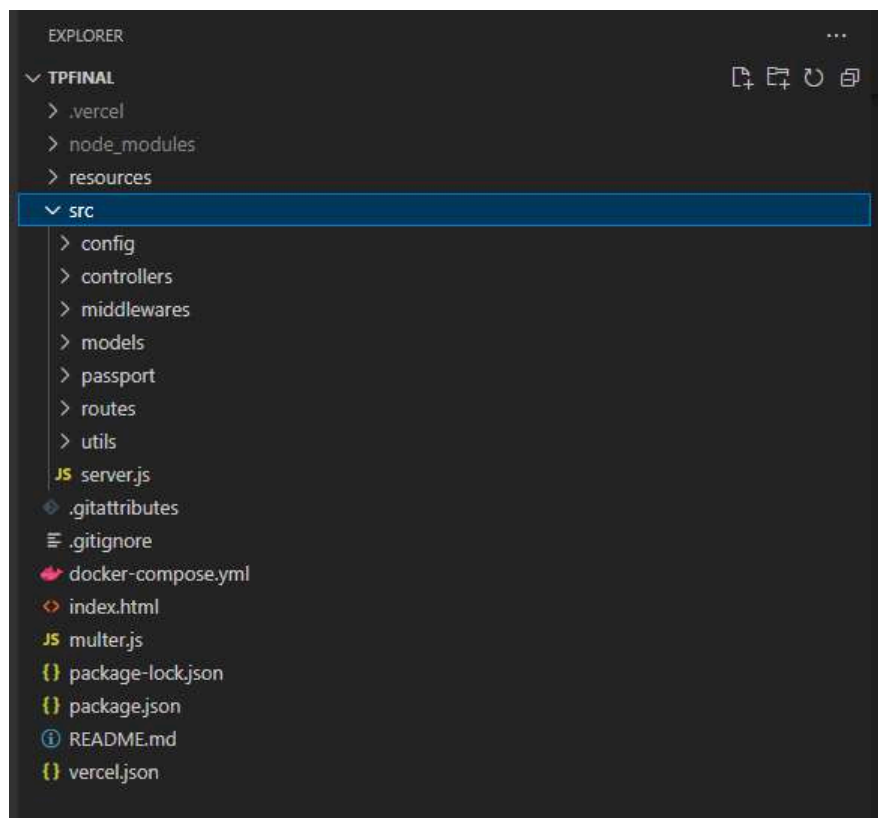


Figura 14: Estructura del backend. Fuente: Elaboración propia

8.2.3. Transformaciones en los datos y obtención de indicadores

La ingesta y transformaciones de los datos tienen lugar en los controladores anteriormente descritos. En el backend, la lógica de los controladores del sistema está basada en la ingesta de datos desde los diferentes archivos Excel y en las consultas a la base de datos para la obtención de los indicadores.

Cada controlador en el sistema desempeña un papel específico en la gestión de la ingesta y transformación de datos para las diversas tablas del modelo. Por ejemplo, hay controladores dedicados a la cuenta corriente, las cobranzas, la facturación, entre otros. Sin embargo, el controlador del *dashboard* es un caso particular. En este contexto, su responsabilidad se limita a consultar la base de datos y crear una nueva tabla destinada a almacenar indicadores que describen el perfil financiero de cada cliente. A diferencia de otros controladores, este no participa en la ingesta de datos, centrándose exclusivamente en las transformaciones, consultas y almacenamiento de indicadores.

A continuación, se detalla el proceso de ingesta y transformación inicial:

1. Requerimientos e importaciones:
 - Importa los módulos necesarios, incluyendo el modelo de base de datos, bibliotecas para el manejo de fechas, manipulación de archivos Excel (xlsx), y otras utilidades.
2. Obtención de la lista de archivos:
 - Escanea el directorio específico donde se almacenan los archivos de carga.
 - Obtiene la lista de nombres de archivos presentes en el directorio.
3. Lectura y procesamiento del archivo Excel:
 - Lee el contenido del archivo Excel seleccionado.
 - Realiza una serie de transformaciones en los datos, como saltar las primeras filas del encabezado y ajustar el formato de las fechas.
4. Iteración sobre filas del archivo Excel:
 - Itera sobre las filas del archivo Excel para procesar los datos.
 - Identifica y procesa información relacionada con la tabla que se ingestará.
5. Creación de objetos y almacenamiento en la base de datos:
 - Para cada conjunto de datos, crea objetos correspondientes para prepararlos para la inserción.
 - Almacena los objetos en la base de datos utilizando el modelo de la tabla con la que trabaja el controlador.
6. Respuesta HTTP:

- Responde con un mensaje de éxito si la operación de ingesta y transformación se realizó sin problemas.
7. Manejo de errores:
- Implementa un manejo de errores para gestionar situaciones en las que la carga de datos no se completa correctamente.

Los controladores, al ejecutarse la ingesta, procesan los archivos Excel, realizan las transformaciones necesarias en los datos y los almacenan de manera estructurada en la base de datos. Este enfoque facilita la posterior consulta y análisis de la información financiera contenida en los archivos.

Los controladores se encargan de obtener indicadores financieros, considerando los parámetros de la *request* según sea necesario y en función de la tabla del modelo a la que representan. Para el caso de la obtención de los indicadores del perfil financiero del cliente (recibiendo por parámetro el id del cliente), el proceso se lleva a cabo de la siguiente manera:

1. Configuración y validación de parámetros:
 - Importa los módulos necesarios y establece la configuración inicial.
 - Valida la presencia del parámetro "id" en la consulta. Si no se proporciona, devuelve un mensaje de error.
2. Consulta de filtros y datos del cliente:
 - La clasificación del cliente se lleva a cabo mediante filtros predefinidos por el usuario. A modo de ejemplo, un filtro consta de tres valores que determinan la clasificación A, B y C en función de la suma de los días que el cliente tardó en emitir un cheque y en liquidarlo.
 - Realiza una consulta a la base de datos para obtener los detalles de un filtro específico utilizando el parámetro "id" del filtro.
 - Recupera información relevante del cliente, como su identificación, nombre, fechas límite, etc.
3. Consulta de indicadores financieros:
 - Realiza consultas a la base de datos para obtener varios indicadores financieros específicos del cliente. Estos indicadores incluyen el promedio de días de pago, la antigüedad de la factura más antigua, el saldo actual, el número de facturas impagas, entre otros.
4. Validación y ajuste de valores nulos:
 - Realiza validaciones para manejar posibles valores nulos en los indicadores obtenidos.
5. Verificación de existencia del cliente:

- Verifica si el cliente ya existe en la tabla destino de los indicadores y si la información ha sido actualizada.
6. Categorización del cliente:
 - Según los umbrales definidos en el filtro, determina la categoría financiera del cliente (A, B, o C) en función de la suma del promedio de días de pago y el promedio de días de chequeo.
 7. Inserción o actualización en la tabla destino de los indicadores:
 - Si el cliente no existe, inserta un nuevo registro en la tabla con los indicadores calculados y la categorización.
 - Si el cliente ya existe, actualiza los valores existentes en la tabla con los indicadores y la categorización actualizados.
 8. Respuesta HTTP con indicadores actualizados:
 - Finalmente, responde con los indicadores financieros actualizados para el cliente.

Este controlador realiza un proceso exhaustivo de obtención, cálculo y categorización de indicadores financieros para un cliente específico, asegurando que la información esté actualizada en la base de datos.

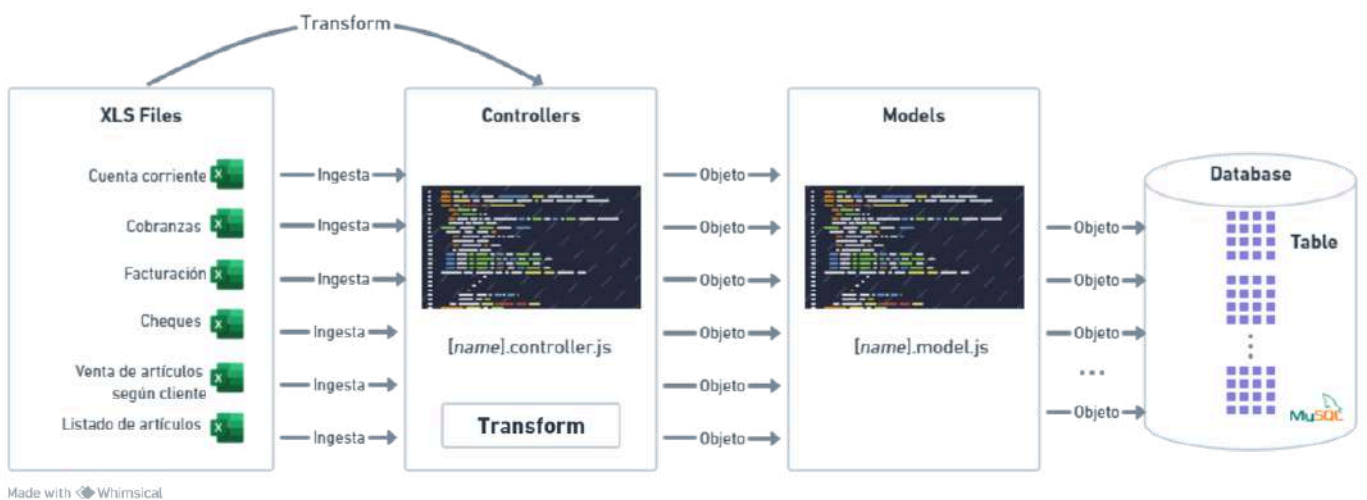


Figura 15: Diseño del backend. Fuente: Elaboración propia

[Diseño del backend](#)

8.3. Frontend

Para la interfaz de usuario, se optó por desarrollar una solución web, considerando la necesidad del *Product Owner* de acceder fácilmente a la información desde varias máquinas con conexión a internet. La elección de una interfaz web no solo cumplió con este requisito, sino que la

naturaleza intuitiva de las aplicaciones web simplifica la curva de aprendizaje del usuario, así como simplifica significativamente el acceso al sistema y a las nuevas características, gracias a la integración continua, y eliminando la necesidad de instalar y configurar el sistema en las máquinas del cliente. Esto se traduce en una experiencia mucho más eficiente.

8.3.1. *Objetivos de la Interfaz de Usuario*

Desde un principio, se tuvo en cuenta que la interfaz de usuario (*UI*) es el medio a través del cual el usuario comprende la aplicación. Por ende, se decidió desarrollarla con extremo cuidado para garantizar una experiencia de usuario fluida y brindar una *UI* que fuera intuitiva, fácil de utilizar, estéticamente agradable, con interacciones ágiles y retroalimentación adecuada. Con el objetivo de lograr una experiencia positiva del usuario, se incorporaron conceptos fundamentales de *UI* y *UX*, disciplinas actualmente cruciales en la industria.

Con el propósito de hacer la interfaz intuitiva y fácil de usar, se optó por construir la aplicación recurriendo a interacciones básicas y mediante la agregación de componentes simples muy reutilizables, para que la *UI* sea altamente consistente y comprensible de un vistazo. El contenido se organizó en secciones comunes para proporcionar cohesión, evitando la saturación de información en la pantalla y disponiendo los elementos de manera clara y precisa para facilitar la asimilación de la información. La navegación entre secciones se diseñó para ser siempre accesible, permitiendo al usuario navegar fácilmente de una sección a otra. Además, se implementó una pantalla de bienvenida que resume todas las opciones de interacción disponibles y explica qué esperar en cada una de ellas.



Figura 16: Pantalla de bienvenida del sistema. Fuente: Elaboración propia

Para garantizar esta interacción fluida, se adoptaron diversas estrategias para reducir los tiempos de carga, culminando en un sistema de almacenamiento local de caché y configuraciones del usuario.

Asimismo, se consideró esencial proporcionar retroalimentación adecuada, como es el caso de un indicador de carga, que permite al usuario comprender que el sistema está procesando la información y que debe esperar unos segundos antes de que esté disponible. También se implementaron mensajes de error claros que comunican cualquier problema, explicando la causa cuando es posible y ofreciendo alternativas al usuario.

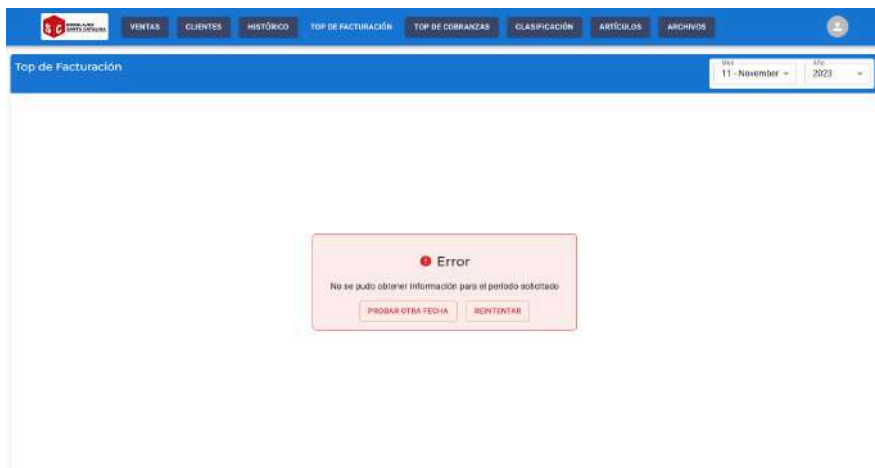


Figura 17: Mensaje de error en la pantalla del Top de Facturación. Fuente: Elaboración propia

Con el objetivo de ofrecer una experiencia visual agradable, a pesar de no contar con un diseñador, se optó por un diseño minimalista que se centra en la información y utiliza componentes básicos pero altamente reutilizables para lograr el máximo impacto visual con el menor esfuerzo posible.

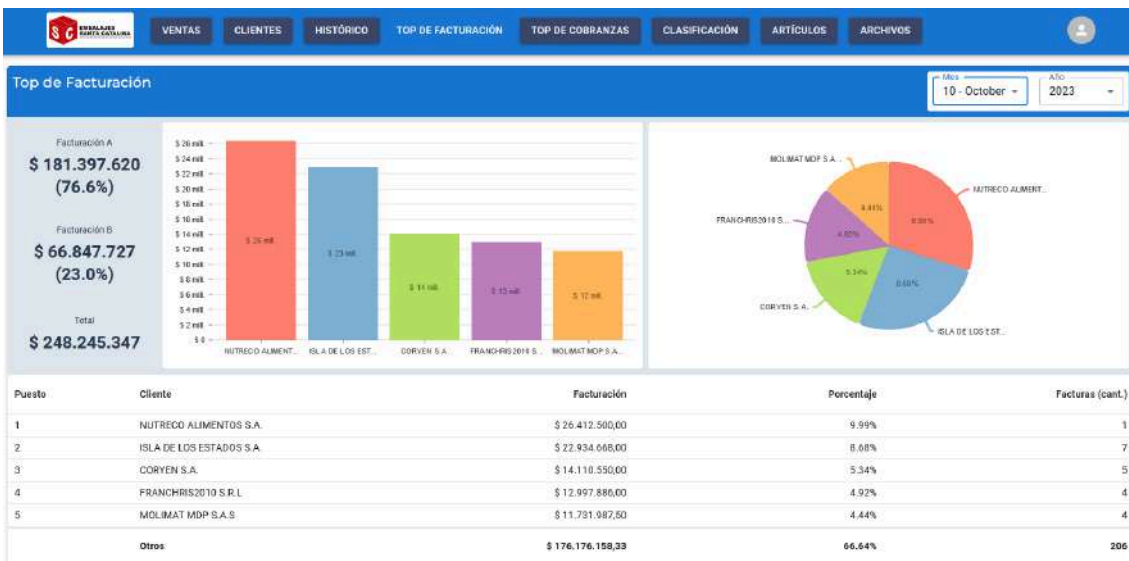


Figura 18: Top de Facturación. Fuente: Elaboración propia. *Los números visualizados son ficticios

Dado que la aplicación web está destinada exclusivamente a un ámbito empresarial, ciertos requisitos comunes en aplicaciones web no fueron necesarios en este desarrollo. Aspectos como optimización para motores de búsqueda (SEO) y el diseño *responsive* se consideraron de menor relevancia. Se optó por utilizar *React* en lugar de *Next*, ya que las opciones más avanzadas de este último, como la generación estática de páginas web, no aportan ventajas significativas en este contexto. Aunque se prescindió del diseño *responsive*, se establecieron bases para facilitar su implementación, manteniendo la flexibilidad del sistema para adaptarse a posibles cambios en el futuro. Esto permitió dedicar más tiempo de desarrollo a perfeccionar el caso de uso principal de la aplicación.

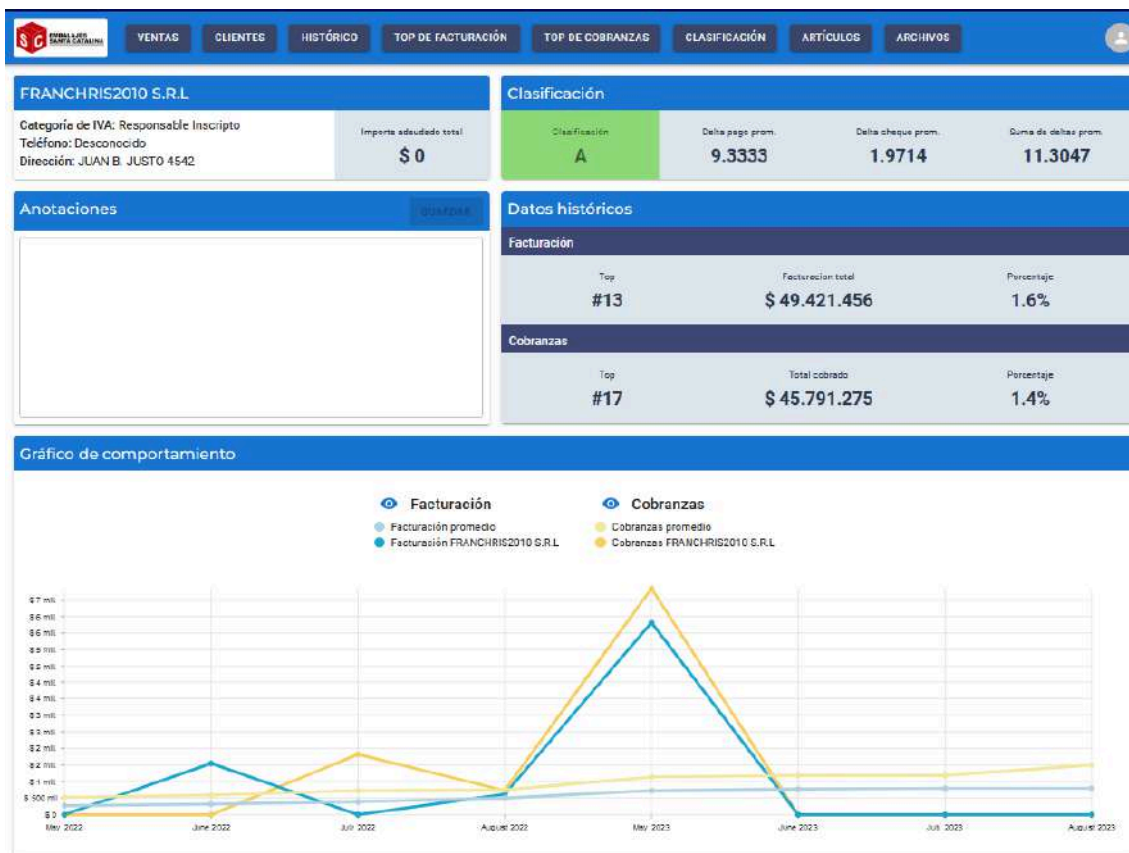


Figura 18: Perfil de cliente. Fuente: Elaboración propia. *Los números visualizados son ficticios

Principios técnicos

- Definición de una estructura robusta y flexible desde el inicio, de fácil comprensión, capaz de perdurar a lo largo del proyecto y, al mismo tiempo, fácil de migrar a otras tecnologías si fuera necesario.
- Adopción de una arquitectura limpia y funcional para reducir los tiempos de desarrollo, centrándonos en la reutilización de elementos y minimizando retrabajos.

- Implementación de componentes básicos y sencillos, fácilmente reutilizables en diversos contextos para agilizar la construcción y modificación de la interfaz de usuario.
- Creación de un circuito base para el consumo, almacenamiento y visualización de la información, asegurando su reutilización en todas las páginas de la aplicación.
- Mínimo número de dependencias posible para aumentar la flexibilidad del sistema. Siempre que los plazos lo permitieran y no se comprometiera el rendimiento del sistema, se prefirieron soluciones propias por sobre agregar dependencias al sistema.
- Aislamiento de dependencias, evitando su uso directo para facilitar su cambio en caso de ser necesario.

8.3.2. *El proyecto frontend*

Durante la configuración del proyecto frontend, se buscó alcanzar un equilibrio entre tecnologías estables y ampliamente utilizadas en la industria, y aquellas menos reconocidas o más modernas, pero que ofrecen ventajas significativas en desempeño o simplicidad de uso.

- Se comenzó con la herramienta de construcción *Vite*, que asegura una experiencia de desarrollo rápida y libre de problemas, permitiendo la generación ágil de distribuciones de producción.
- Se eligió *React*, un estándar en la industria para el desarrollo web, gracias a su paradigma de componentes y *DOM* virtual, facilitando la construcción de interfaces de usuario y el manejo de interacciones complejas a partir de elementos simples.
- La inclusión de *TypeScript*, una extensión de *JavaScript* con tipado estático, contribuyó a incrementar la velocidad y reducir errores durante el desarrollo.
- *Material UI (MUI)*, una librería de componentes gráficos reutilizables. A pesar de tener una curva de aprendizaje, resultó beneficioso al ahorrar tiempo en el desarrollo de componentes comunes utilizados en la mayoría de las aplicaciones web, como botones o campos de texto.
- *Nivo*, una librería para dibujar gráficos, como los de barra o torta. A pesar de requerir tiempo para aprender y configurar, ofreció funcionalidades invaluable en el trabajo con gráficos.

- *Zustand*, para el manejo del estado interno de la aplicación, se eligió por su *API* sencilla en comparación con estándares como *Redux*, manteniendo un nivel de desempeño y funcionalidad similar.
- Además, se integraron utilidades como enrutadores, manejadores de fecha, *JWT*, entre otros.

El proyecto se organiza en *features* o características, como por ejemplo Ventas o Clientes. Es decir, para cada característica, se agrupan sus elementos específicos, como páginas o componentes. Esto simplifica la búsqueda de elementos al trabajar en una determinada característica.

También se dispone de un directorio *common* donde, al igual que con las características, se agrupan elementos, pero esta vez de uso común en todo el proyecto o por lo menos en más de una característica.

Otros dos elementos principales en la organización del proyecto son los *services* o servicios, que agrupan la lógica para realizar peticiones o intercambiar información con el *backend*, y los *stores* o tiendas, que agrupan la lógica que se encarga de almacenar de manera eficiente la información en memoria para su acceso a lo largo del proyecto.

También se distingue el archivo *App.tsx*, que es el punto de entrada a la aplicación, desde cuyo análisis exhaustivo se puede comprender toda la lógica y el contenido del proyecto *frontend*.

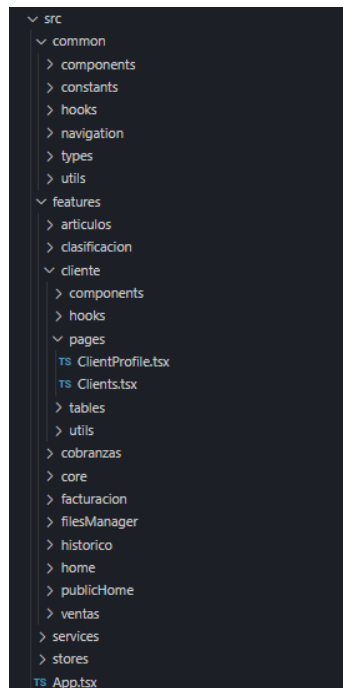


Figura 19: Estructura del frontend. Fuente: Elaboración propia

8.3.3. Ejemplificando el funcionamiento de una página

1. Dibujado inicial de la página, brindando retroalimentación lo más rápido posible aunque aún no se disponga de la información necesaria
2. Obtener la información necesaria, aquí hay dos posibilidades:
 - a. Obtener la información del caché: en caso de que el usuario recientemente haya accedido a la información, esta se encontrará disponible en el caché
 - b. En caso contrario, se muestra el estado de carga para que el usuario comprenda que se está haciendo una consulta que demorara unos instantes, y se interactúa con el servicio correspondiente para realizar la petición (solicitud al backend). Si la petición se desarrolla satisfactoriamente, se oculta el estado de carga y se muestra la información. Si no, se muestra un mensaje de error con todos los detalles disponibles y se da la posibilidad al usuario de reintentarlo.
3. Aguardar interacciones del usuario y actuar en consecuencia. Por ejemplo, en el caso de la página de ventas, donde se puede acceder a distintos resúmenes mensuales, el usuario puede interactuar con un selector de fecha. Al seleccionar una fecha, ahora es necesario mostrar una información distinta y así se reiniciará el ciclo descrito desde el paso 2.

8.4. Seguridad

La seguridad en aplicaciones es un desafío complejo. Ante esta dificultad, se optó por emplear prácticas seguras respaldadas por la industria. Al hacerlo, se aseguró un enfoque robusto y probado para garantizar la protección de los datos.

Uno de los aspectos destacados en este ámbito es la implementación del inicio de sesión mediante Google, que ofrece una seguridad adicional y comodidad para nuestros usuarios. Cuando un usuario elige iniciar sesión mediante su cuenta de *Google*, la aplicación *backend* facilita el flujo de autenticación. *Google* actúa como proveedor de identidad, permitiendo que los usuarios utilicen sus credenciales de *Google* para acceder a la aplicación.

Beneficios de la autenticación con *Google*:

- Seguridad Reforzada: Al aprovechar la robusta infraestructura de autenticación de *Google*, se reduce el riesgo de vulnerabilidades y aseguramos un proceso de inicio de sesión seguro.

- **Conveniencia del Usuario:** Al permitir a los usuarios utilizar sus credenciales de *Google*, se elimina la necesidad de recordar otra contraseña, mejorando la experiencia del usuario y fomentando la adopción de la aplicación.

Una vez que el usuario ha sido autenticado exitosamente por *Google*, el backend genera un JWT (*JSON Web Token*). El *JWT* es un estándar abierto que permite transmitir información segura entre dos partes de manera compacta y autenticada. Encapsula información segura sobre la identidad del usuario y sus permisos de autorización. En este caso, se utiliza para representar la autorización del usuario autenticado y su identidad verificada y así se convierte en la clave de acceso para el frontend al interactuar con el backend: el frontend adjunta el *JWT* a las solicitudes para demostrar la autenticidad y autorización del usuario.

9. Algoritmos de predicción

Es importante recordar que la implementación de algoritmos de predicción en este proyecto tiene como objetivo proporcionar una visión general del estado de los datos para facilitar la aplicación de *Business Intelligence*. Sin embargo, un análisis detallado y exhaustivo de la aplicación de estas técnicas se encuentra fuera del alcance de los objetivos establecidos para este proyecto.

Lo segundo a destacar es que la calidad de los datos y la presencia de meses con información faltante para ciertos materiales, como se estableció en los objetivos, presentaron desafíos significativos. Esta dificultad radica en la naturaleza dinámica de los materiales con los que opera la empresa, con adiciones y modificaciones constantes. Como resultado, se encontraron lagunas en los datos que resultaron difíciles de completar de manera precisa, incluso mediante el uso de librerías y técnicas de generación de datos.

Debido a esta complejidad y la dificultad para obtener datos completos y precisos para cada material, se tomó la decisión de enfocar el análisis y la predicción en el monto mensual facturado y en los metros cuadrados mensuales vendidos. Estos dos indicadores proporcionan una perspectiva valiosa y más generalizable, permitiendo realizar pronósticos algo significativos a pesar de las limitaciones en la integridad de los datos.

9.1. Análisis exploratorio

En primera instancia, se graficaron las dos variables a pronosticar: monto mensual y metros mensuales, obteniendo lo siguiente.

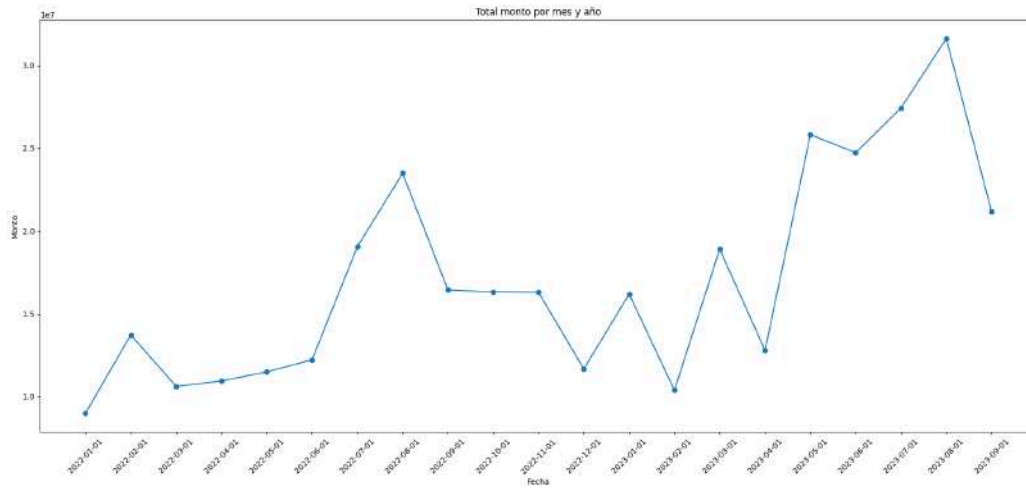


Figura 20: Gráfico del monto mensual. Fuente: Elaboración propia

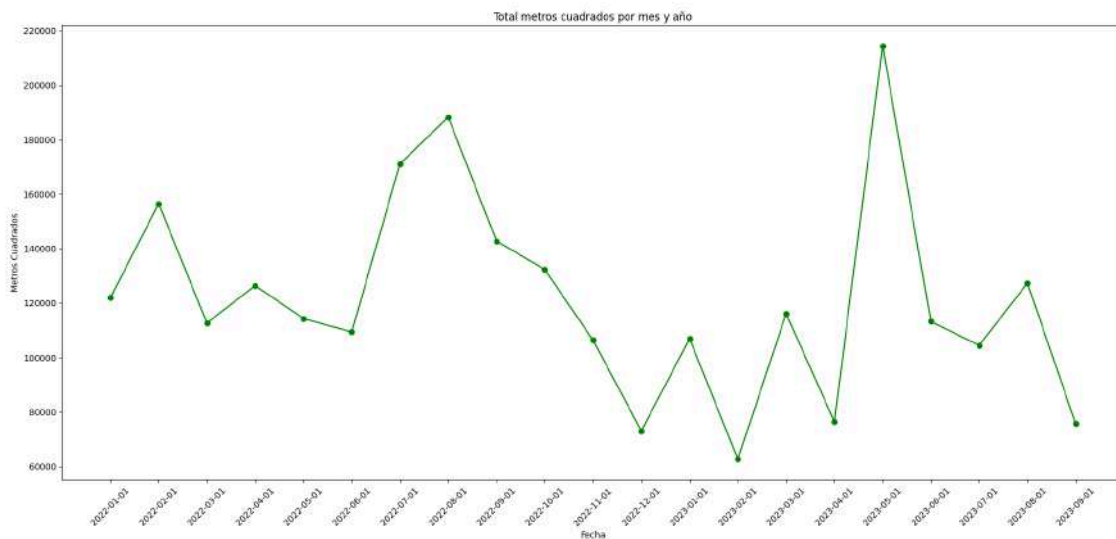


Figura 21: Gráfico de los metros mensuales. Fuente: Elaboración propia

Esta técnica brinda una visión inicial del comportamiento de las series. No obstante, como se vió en el marco de referencia, es crucial aplicar métodos adicionales para comprender a fondo el comportamiento de las series. Por lo tanto, el siguiente paso en el análisis exploratorio implica descomponer ambas series en sus componentes de tendencia y estacionalidad, proporcionando una visión panorámica de estas series.

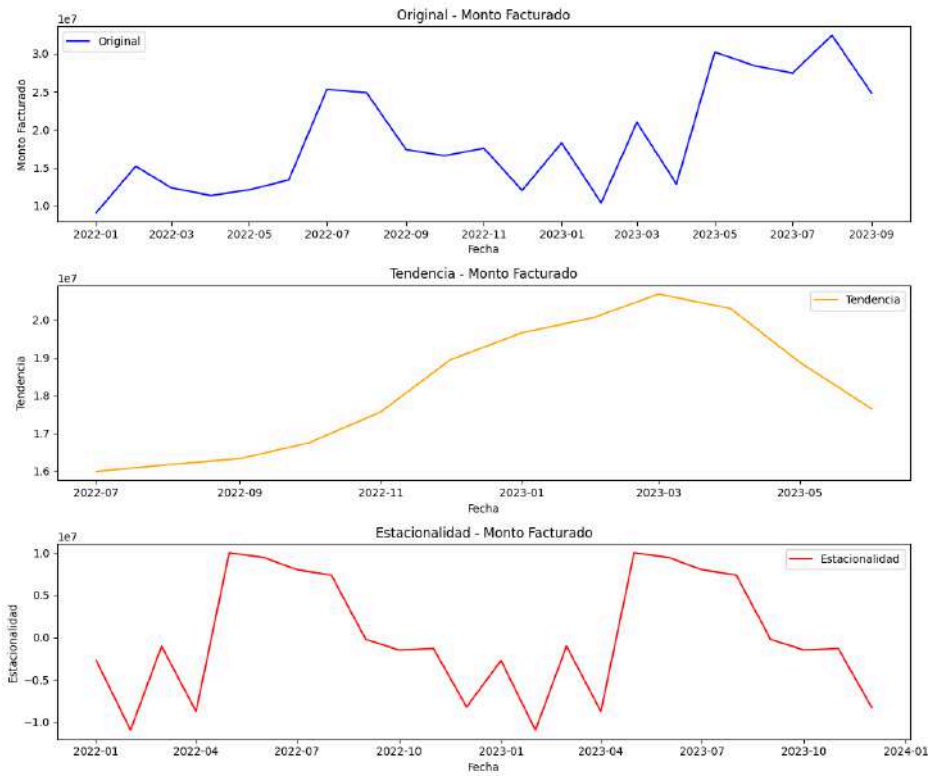


Figura 22: Descomposición del monto mensual facturado. Fuente: Elaboración propia

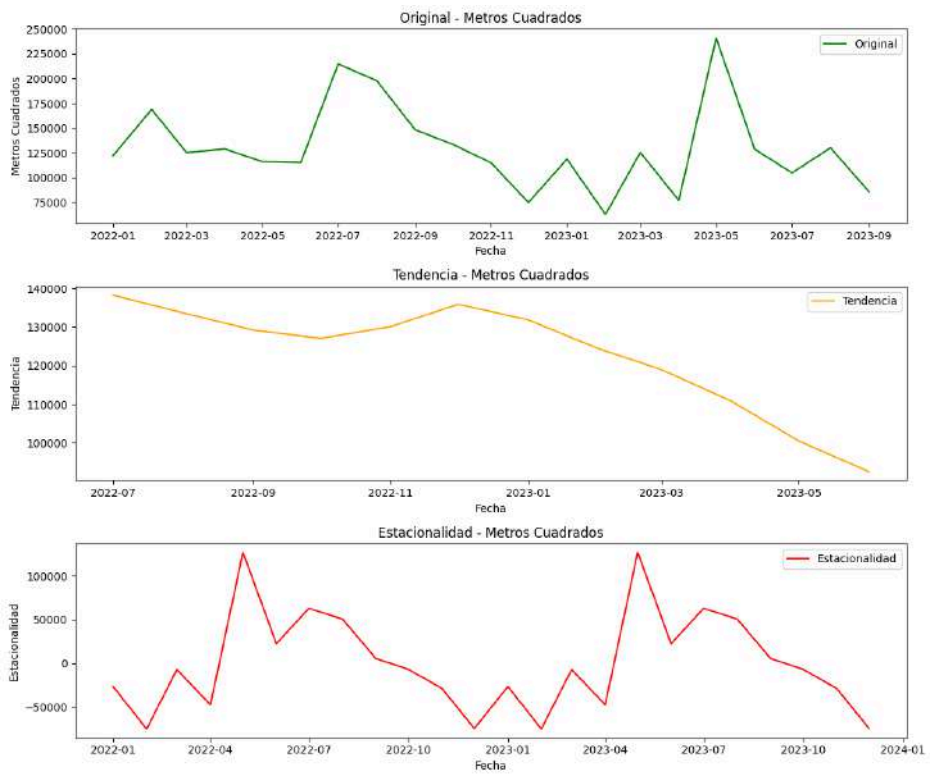


Figura 23: Descomposición de metros cuadrados vendidos. Fuente: Elaboración propia

Ahora es el momento de evaluar la autocorrelación. Recordando, la autocorrelación cuantifica la relación entre el valor actual de una serie temporal y sus valores pasados. La autocorrelación positiva significa que un valor alto en el presente probablemente resultará en un valor alto en el futuro y viceversa. La autocorrelación negativa es lo opuesto: un valor alto hoy implica un valor bajo mañana, y un valor bajo hoy implica un valor alto mañana. El eje de las x representa los *lags*, y el eje de las y muestra el nivel de autocorrelación.

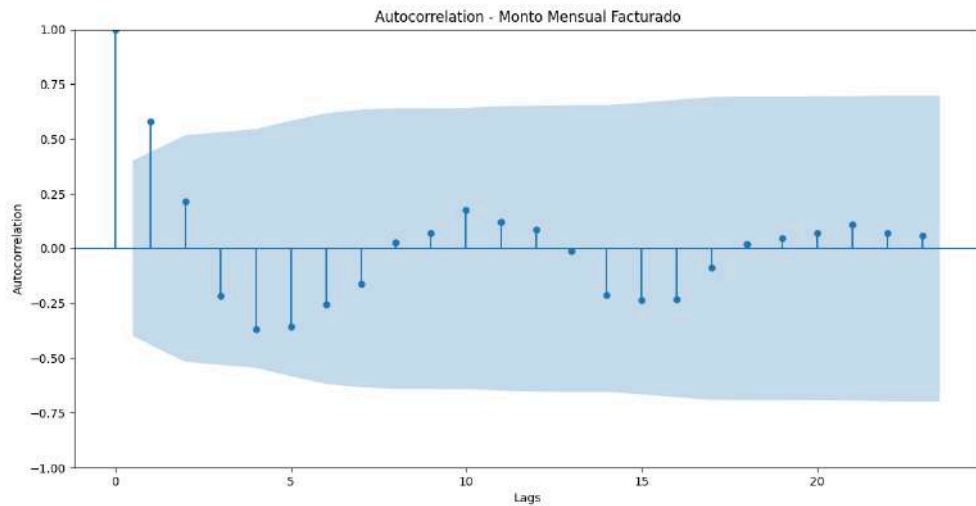


Figura 24: Autocorrelación de la serie de monto mensual facturado. Fuente: Elaboración propia

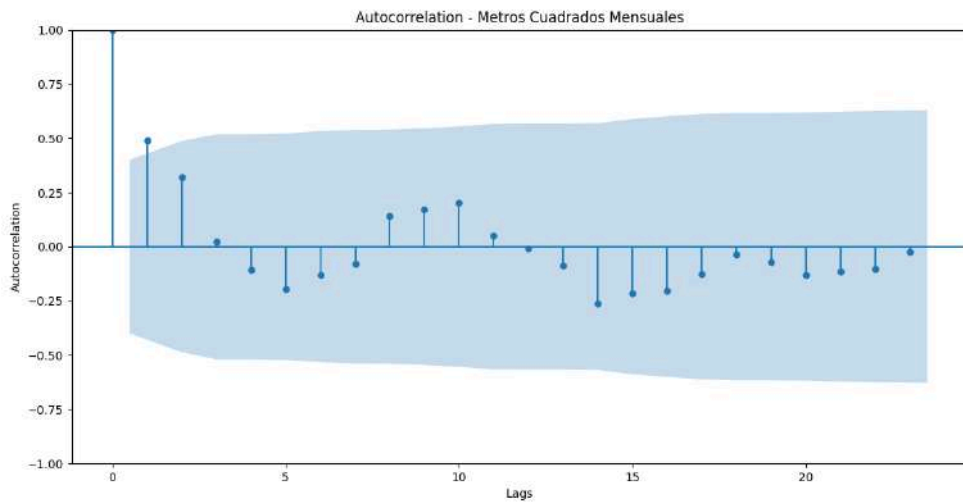


Figura 25: Autocorrelación de la serie de metros cuadrados mensuales vendidos. Fuente: Elaboración propia

Con base en las descomposiciones de las dos series temporales (monto mensual facturado y metros cuadrados) y de los gráficos de autocorrelación, es posible extraer algunas conclusiones generales:

Monto Mensual Facturado:

- Serie Original:
 - Se observa una tendencia general al alza en el monto mensual facturado, indicando un crecimiento en las ventas a lo largo del tiempo.
 - Existe cierta estacionalidad, sugiriendo patrones recurrentes en los ingresos.
- Tendencia:
 - La tendencia revela un aumento sostenido en el monto facturado, lo cual es positivo desde el punto de vista financiero.
- Estacionalidad:
 - La componente estacional muestra fluctuaciones regulares que podrían estar relacionadas con factores estacionales, como temporadas de mayor demanda.
- Autocorrelación:
 - Se observa un pico significativo en el primer *lag*, lo que podría indicar una fuerte relación entre el monto facturado en un mes y el mes anterior.
 - Patrones repetitivos en los *lags* podrían sugerir algún tipo de estacionalidad.
 - La falta de autocorrelación en los *lags* intermedios podría indicar que el monto facturado de un mes no está fuertemente influenciado por los valores exactos de varios meses anteriores.

Metros Cuadrados Mensuales (m²):

- Serie Original:
 - La serie original de metros cuadrados muestra cierta variabilidad a lo largo del tiempo, pero no presenta una tendencia clara.
- Tendencia:
 - La componente de tendencia no muestra una dirección evidente, sugiriendo que la cantidad de metros cuadrados no experimenta un cambio constante.
- Estacionalidad:
 - La estacionalidad parece fluctuar, indicando que puede haber patrones periódicos en la distribución de metros cuadrados.

- Autocorrelación:
 - Un pico en el primer *lag* podría indicar una relación fuerte entre los metros cuadrados en un mes y el mes anterior.
 - Al igual que en el caso anterior, patrones en los *lags* podrían sugerir estacionalidad en la serie temporal.

Conclusiones generales:

- La serie de monto mensual facturado muestra un crecimiento positivo a lo largo del tiempo, con evidencia de patrones estacionales.
- La serie de metros cuadrados no presenta una tendencia clara, pero muestra cierta variabilidad estacional.

9.2. Elección de los algoritmos de predicción

Considerando las características de las series temporales de monto mensual facturado y metros cuadrados mensuales vendidos, los algoritmos que se consideraron implementar son los siguientes:

Monto Mensual Facturado:

ARIMA (Autoregressive Integrated Moving Average):

- Ventajas: Efectivo para modelar series temporales con patrones de estacionalidad y tendencia. Puede proporcionar predicciones precisas.
- Desventajas: Requiere que los datos sean estacionarios. Puede ser complejo de interpretar.

Suavización Exponencial (Exponential Smoothing):

- Ventajas: Simple y adaptable a patrones estacionales y tendencias. Fácil de entender y aplicar.
- Desventajas: Puede no funcionar bien con datos altamente irregulares.

Metros Cuadrados Mensuales Vendidos:

Suavización Exponencial (Exponential Smoothing):

- Ventajas: Simple y adaptable a patrones estacionales y tendencias. Fácil de entender y aplicar.
- Desventajas: Puede no funcionar bien con datos altamente irregulares.

Procesos de Media Móvil (Moving Average Processes):

- Ventajas: Puede adaptarse a cambios en la media y proporcionar predicciones suavizadas.
- Desventajas: Puede no ser eficaz para datos con patrones no lineales o cambios abruptos.

9.3. Resultados obtenidos

Se dividieron las series temporales en los conjuntos mencionados en el marco teórico:

- Conjunto de entrenamiento
- Conjunto de validación

De manera gráfica:

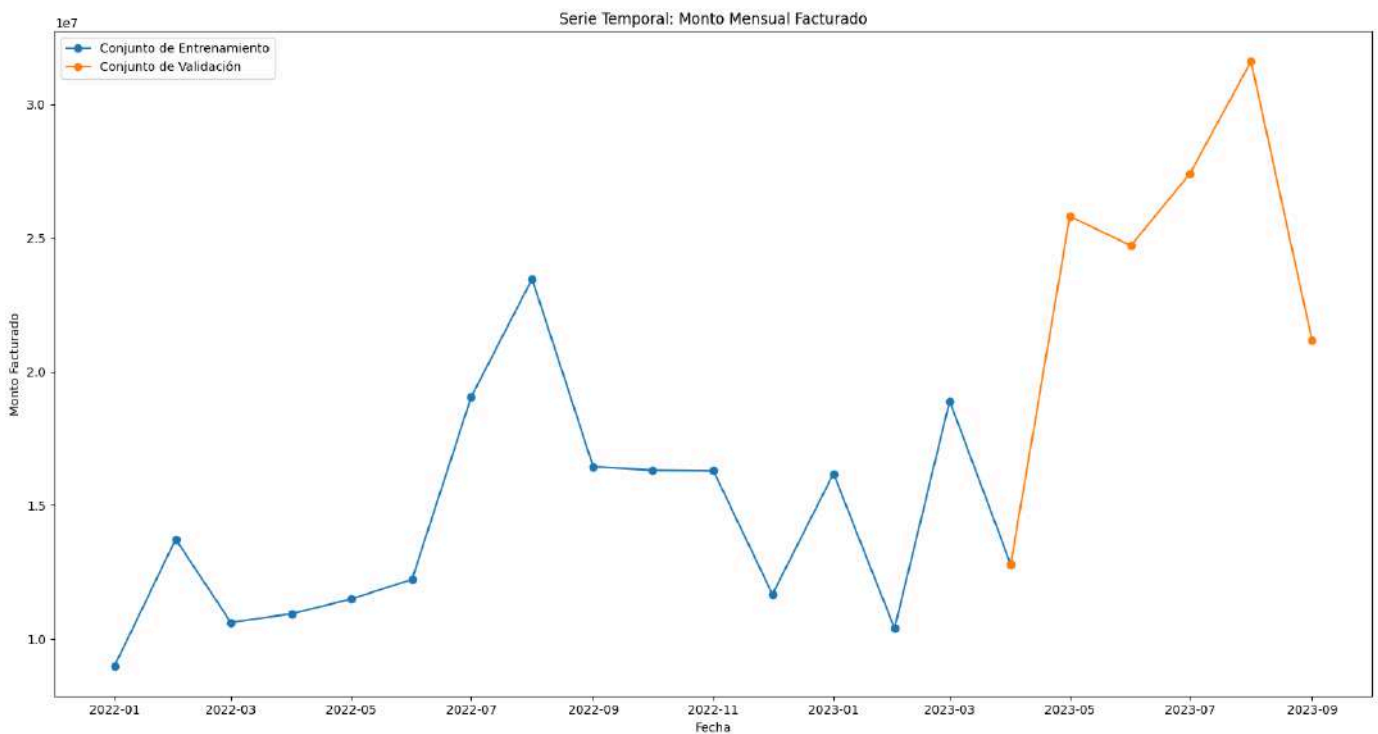


Figura 26: Serie temporal Monto Mensual Facturado. Fuente: Elaboración propia

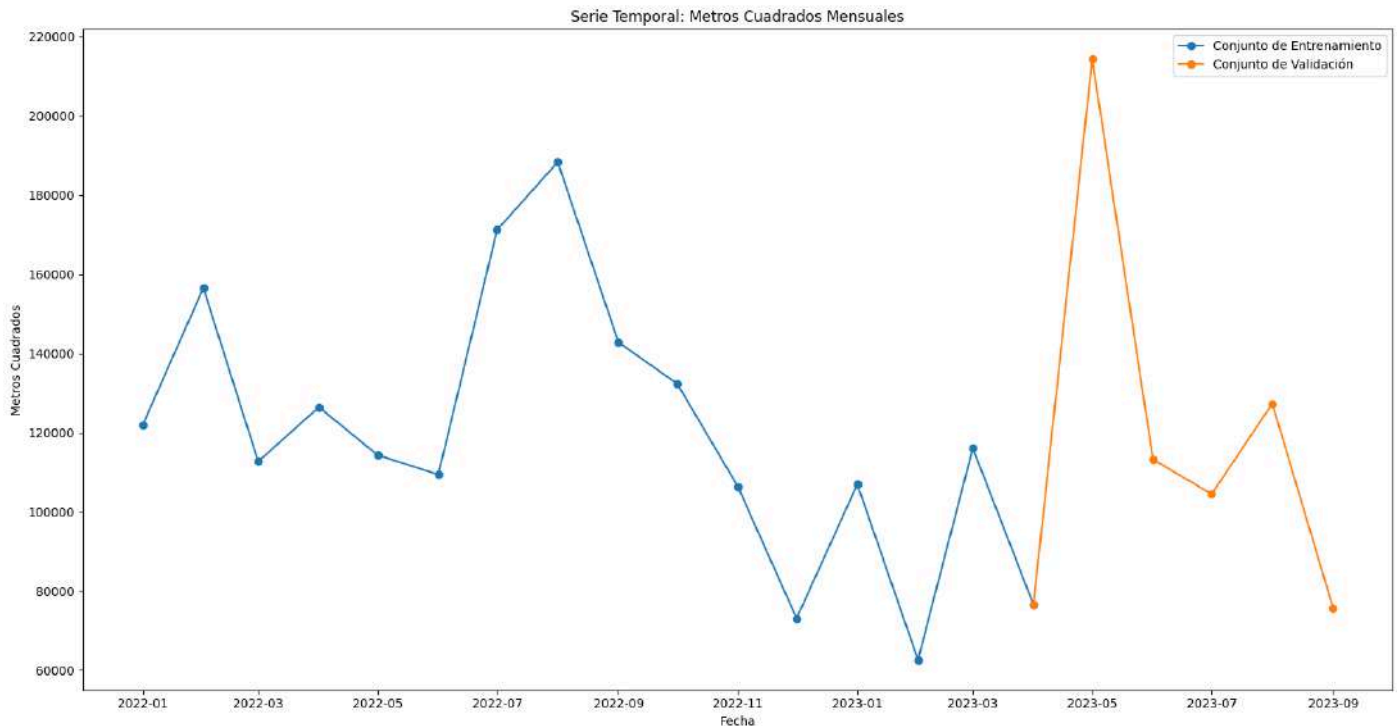


Figura 27: Serie temporal Metros Cuadrados Mensuales. Fuente: Elaboración propia

ARIMA (Autoregressive Integrated Moving Average)

Se procedió a la aplicación del método ARIMA a la serie temporal de montos de ventas mensuales. Primero, se evaluó la estacionariedad de la serie, un requisito fundamental para la aplicación de ARIMA. La prueba de Dickey-Fuller arrojó un valor de estadística de prueba de aproximadamente -2.2 y un valor p de 0.18 en el conjunto de entrenamiento. Dado que el valor p es superior al umbral común de 0.05, no se puede rechazar la hipótesis nula, indicando que la serie temporal no es estacionaria.

En respuesta a la falta de estacionariedad, se aplicó una diferenciación de primer orden en la serie temporal para lograr la estacionariedad necesaria. Posteriormente, se ajustó un modelo ARIMA con parámetros $p=1$, $d=1$ y $q=1$, donde p es el orden de los términos autorregresivos, d es el orden de diferenciación, y q es el orden de los términos de media móvil.

Las predicciones se generaron para los próximos dos meses posteriores al último registro en los datos. Cabe destacar que se optó por reservar un 30% de los datos originales para constituir un conjunto de validación, mientras que los dos meses siguientes se designaron como conjunto de prueba. Este enfoque permitió evaluar la capacidad predictiva del modelo en un escenario más realista, considerando la variabilidad presente en los datos de ventas mensuales.

Los resultados se visualizaron gráficamente, mostrando tanto los datos originales como las predicciones para los conjuntos de entrenamiento y validación.

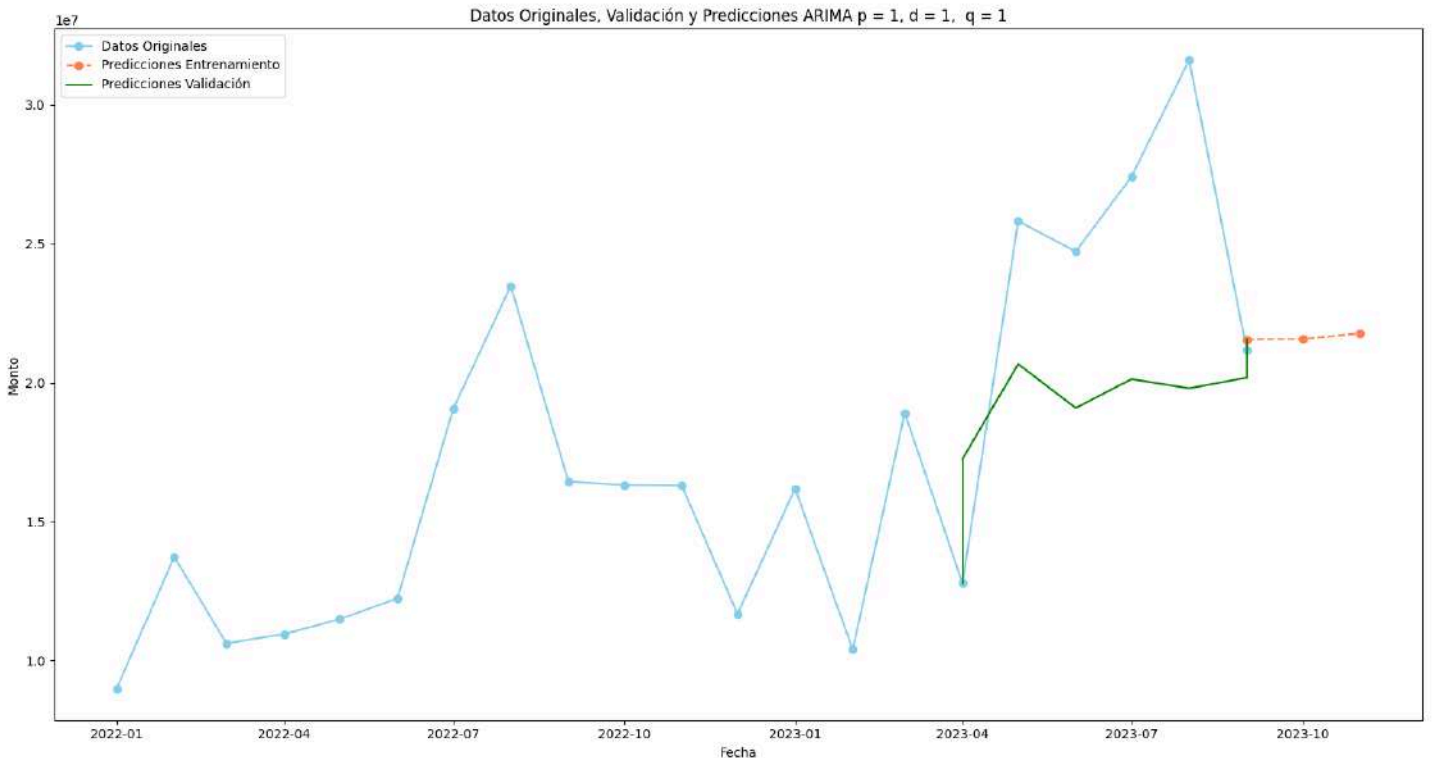


Figura 28: ARIMA con parámetros $p=1$, $d=1$ y $q=1$. Fuente: Elaboración propia

Suavización Exponencial (Exponential Smoothing) Primer Orden:

Se procedió a aplicar el método de suavizado exponencial de primer orden, que integra tanto el nivel como la tendencia de la serie, proporcionando una estimación suavizada.

El parámetro α en el suavizado exponencial controla la velocidad de adaptación al nivel de la serie temporal. Un valor más alto de α da más peso a los datos más recientes, lo que hace que el modelo sea más sensible a cambios recientes. En cambio, un valor más bajo da más peso a los datos históricos, resultando en un suavizado más lento. En este caso se utilizó un α igual a 0.3. Se generaron predicciones para los próximos dos meses (octubre y noviembre), extrapolando más allá del último registro en los datos de entrenamiento, permitiendo anticipar la evolución de la serie temporal a corto plazo.

Con el objetivo de evaluar la capacidad predictiva del modelo en un escenario realista, se reservó un 30% de los datos originales para constituir un conjunto de validación. Los dos meses siguientes se designaron como conjunto de prueba, proporcionando una oportunidad para evaluar la precisión del modelo en períodos futuros.

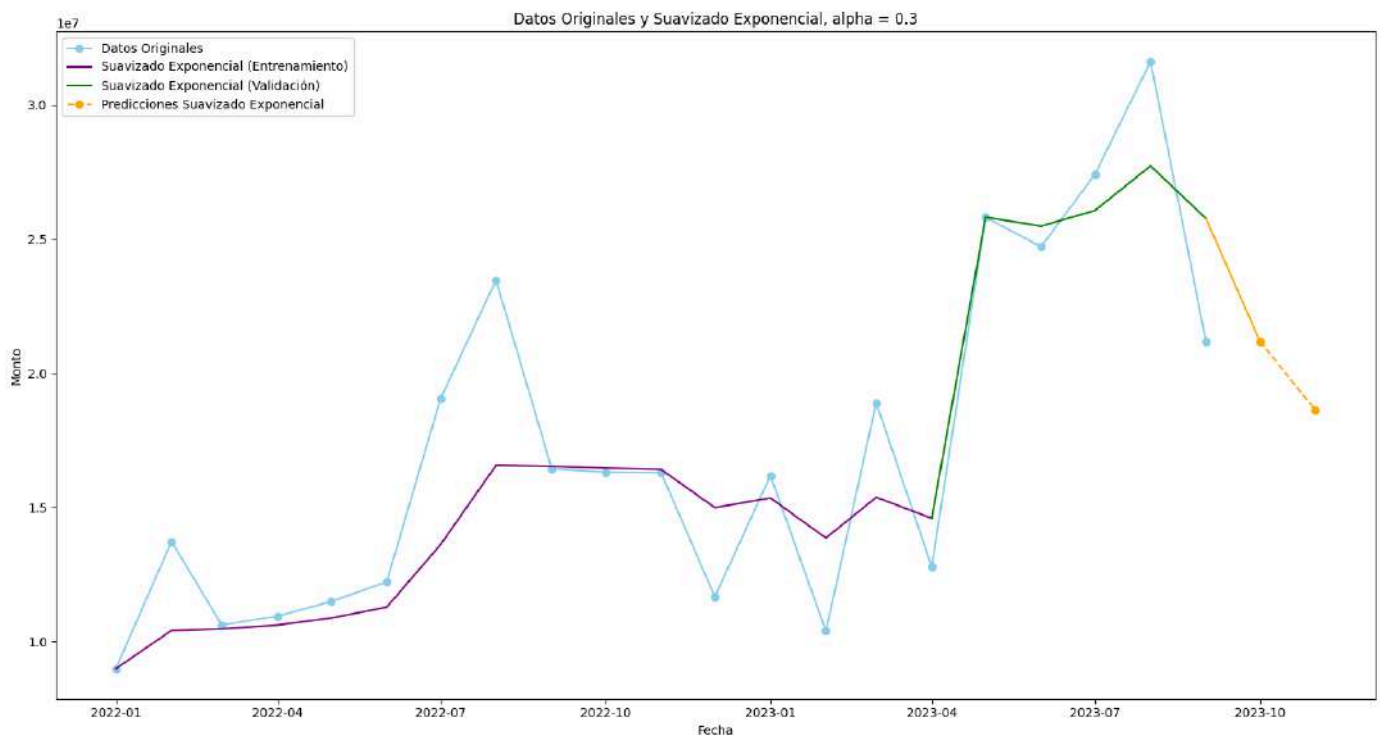


Figura 29: ETS Primer orden, alpha = 0.3. Fuente: Elaboración propia

Suavización Exponencial (Exponential Smoothing) Segundo Orden:

Se aplicó el suavizado exponencial de segundo orden, donde se ajustó el parámetro α a un valor específico y el parámetro β a otro valor particular. El parámetro α fue establecido en 0.3, mientras que el parámetro β se fijó en 0.2. Estos valores fueron seleccionados considerando la necesidad de lograr un suavizado adecuado y la captura de la tendencia en la serie temporal. La combinación de α y β permitió adaptar la metodología de suavizado exponencial de segundo orden para tener en cuenta tanto el nivel como la tendencia de los datos, proporcionando así predicciones que incorporan la información de cambio en la tendencia de la serie temporal.

Se observa que al comparar las predicciones obtenidas mediante el suavizado exponencial de primer y segundo orden, no se evidencian cambios significativos en los resultados. Con el objetivo de explorar más a fondo las posibles mejoras en la predicción, se procedió a aplicar el suavizado exponencial de tercer orden.

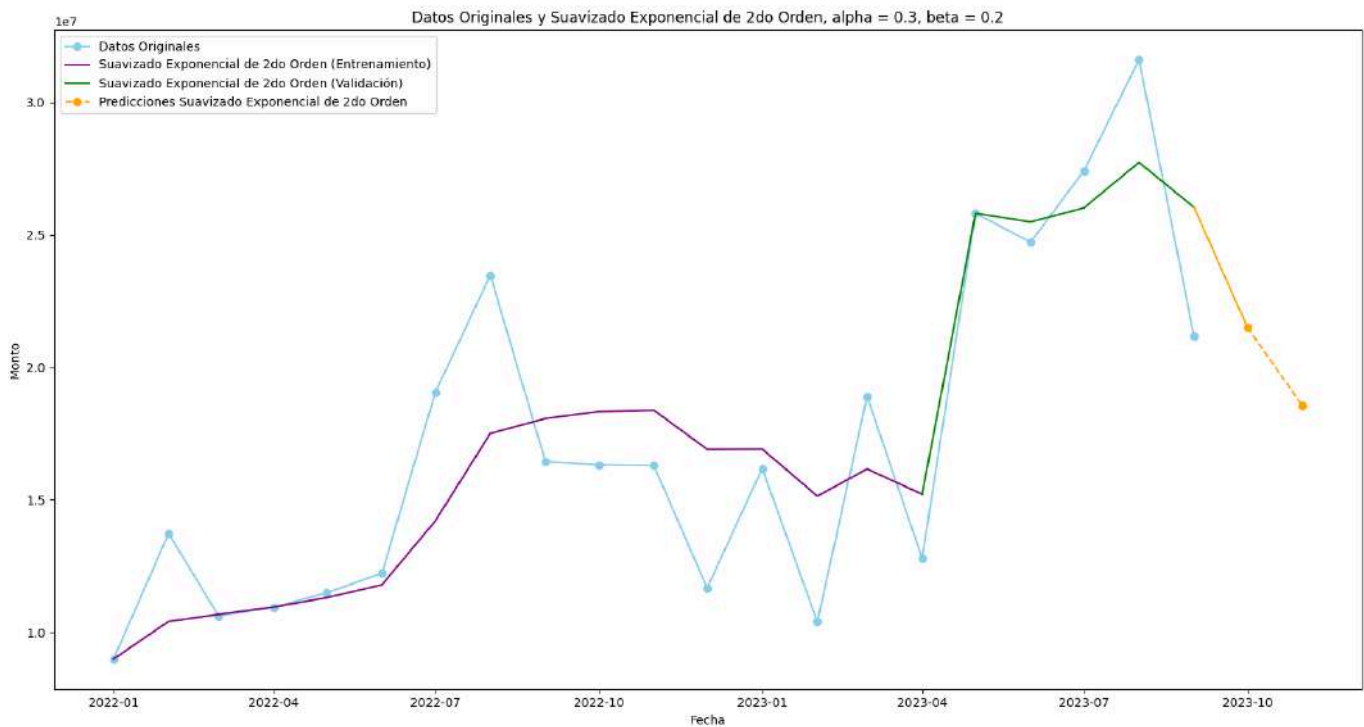


Figura 30: ETS Segundo orden, $\alpha = 0.3$, $\beta = 0.2$. Fuente: Elaboración propia

Suavización Exponencial (Exponential Smoothing) Tercer Orden o de Holt-Winters:

Después de aplicar suavizado exponencial de primer y segundo orden, se exploró la implementación del suavizado exponencial de tercer orden o de Holt-Winters con el propósito de mejorar la precisión de las predicciones. Se ajustaron los parámetros del modelo utilizando la biblioteca correspondiente, lo cual generó la advertencia de convergencia: "Optimization failed to converge. Check mle_retvals." Los valores óptimos estimados para los parámetros del modelo son: $\alpha=0.535$, $\beta=0.0001$, $\gamma=0.0001$, $\text{initial_level} = 14182796.22$, $\text{initial_trend} = 210351.71$, y $\text{initial_seasons} = [2436019.31, 1082681.30, 704174.20, -758899.06, -656642.21, -2807333.54]$. Es importante destacar que la estacionalidad se configuró para abarcar 12 meses, aunque la cantidad de puntos de datos es limitada a menos de dos años. Dada esta limitación, la división entre el conjunto de validación y entrenamiento se realizó considerando una estacionalidad de 6 meses en lugar de 12.

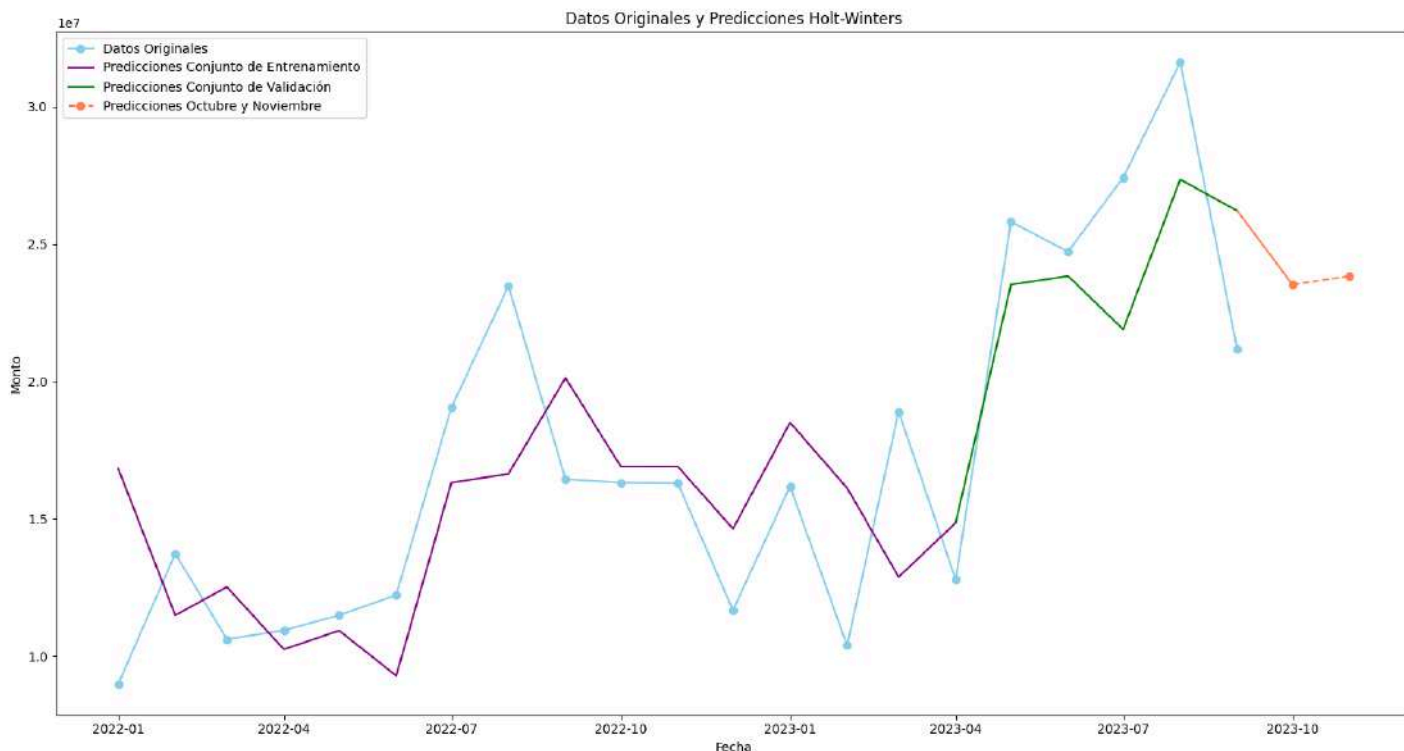


Figura 31: ETS Tercer orden, $\alpha=0.535$, $\beta=0.0001$, $\gamma=0.0001$. Fuente: Elaboración propia

Elección del modelo más adecuado:

Después de probar los cuatro modelos y ajustar sus parámetros correspondientes, se elaboró una tabla que presenta las métricas relevantes para evaluar la precisión de cada uno en términos de predicción. Esta tabla ofrece una visión comparativa que facilita la identificación del modelo más preciso.

Modelo	MAE	RMSE	R²	MAPE %
<i>ARIMA(1,1,1)</i>	2.56E+07	3.76E+07	0.367	23.3
<i>ES: Alpha = 0.3</i>	1981103.77	2826091.07	0.796	11.59
<i>ES: Alpha = 0.3 y Beta = 0.2</i>	16764727.68	22413631.17	0.815	16.33
<i>ES: Alpha = 0.535, Beta = 0.0001, Gamma = 0.0001</i>	3862794.28	4893579.06	0.38	23.7

Tras evaluar cuatro modelos diferentes, cada uno con sus respectivas configuraciones, se destaca que el modelo de Suavizado Exponencial con Alpha igual a 0.3 y Beta igual a 0.2 presenta las métricas más favorables en comparación con los demás. Este modelo exhibe el menor Error Absoluto Medio (MAE) y Raíz del Error Cuadrático Medio (RMSE), indicando una menor diferencia entre las predicciones y los valores reales. Además, el coeficiente de determinación (R^2) es significativamente alto, alcanzando un 79.6%, lo que sugiere que el modelo explica una proporción sustancial de la variabilidad en los datos. Asimismo, el Error Porcentual Absoluto Medio (MAPE) es bajo, señalando una precisión relativa elevada en las predicciones.

Es fundamental reconocer que la calidad de las predicciones se ve directamente afectada por la escasez de datos disponibles para evaluar los patrones estacionarios. El análisis exploratorio previo reveló indicios de estacionalidad y tendencias, pero la limitada cantidad de puntos de datos compromete la capacidad de los modelos para capturar con precisión estos patrones complejos. Este desafío se evidencia especialmente en el caso del modelo de Suavizado Exponencial de Holt-Winters, donde la falta de puntos para evaluar los patrones estacionarios contribuye a una mayor complejidad y una menor precisión en las predicciones. Es importante tener en cuenta esta limitación al interpretar los resultados y considerar la posibilidad de recopilar más datos en el futuro para mejorar la capacidad predictiva de los modelos.

Además, es importante destacar que en este análisis no se consideraron variables significativas como el Índice de Precios al Consumidor (IPC). La omisión de estas variables clave se debió a la necesidad de evitar una mayor complejidad en el escenario de predicción, dada la escasa cantidad de puntos disponibles. La inclusión de más variables podría haber incrementado la complejidad de los modelos y, dado el limitado conjunto de datos, no se consideró justificado. En futuros análisis, sería valioso explorar la inclusión de estas variables para evaluar su impacto en la precisión de las predicciones, siempre manteniendo un equilibrio entre la complejidad del modelo y la cantidad de datos disponibles.

Serie temporal Metros Cuadrados Mensuales:

Suavización Exponencial (Exponential Smoothing) Primer Orden:

El suavizado exponencial de primer orden con un valor de alpha igual a 0.1 demostró ser la configuración más efectiva para modelar y prever la serie temporal de metros cuadrados dentro del enfoque de este algoritmo. Se llevaron a cabo pruebas con diferentes valores de alpha, y los

resultados indicaron que esta configuración específica proporciona las métricas más favorables, incluyendo el MAE más bajo, MAPE más bajo, el R^2 más alto y el RMSE más bajo.

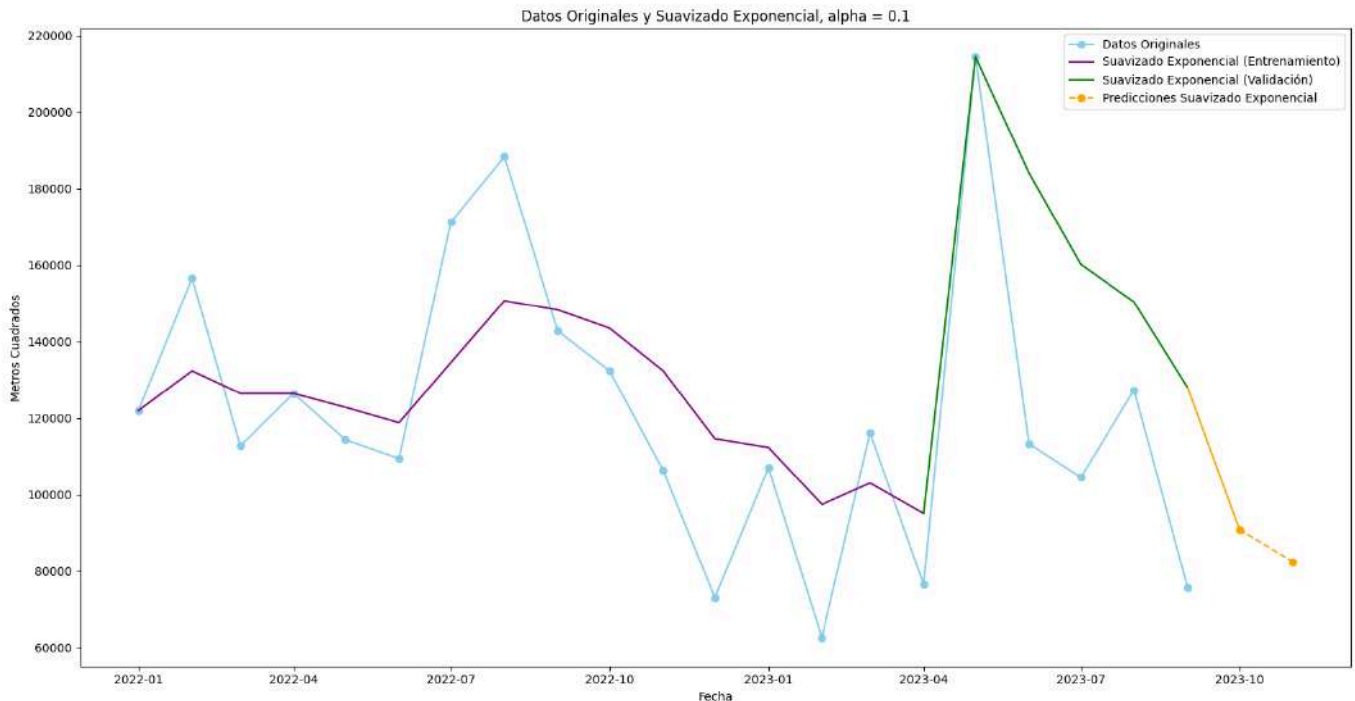


Figura 32: ETS Primer orden, alpha = 0.1. Fuente: Elaboración propia

Este algoritmo de suavizado exponencial se ajusta bastante a las características de los datos originales, capturando algunas de las tendencias y patrones subyacentes de la serie temporal. La elección de alpha en 0.1 permite un equilibrio adecuado entre la adaptabilidad del modelo a cambios en los datos y la estabilidad en la predicción, dentro de las particularidades ya mencionadas respecto de los datos.

Suavización Exponencial (Exponential Smoothing) Segundo Orden:

Durante el proceso de evaluación, se probaron diversas combinaciones de alpha y beta para encontrar la configuración óptima. Los resultados resaltan la importancia de ajustar estos parámetros según las características específicas de los datos. En este caso, la elección de parámetros más bajos parece haber dado como resultado predicciones un poco más acertadas.

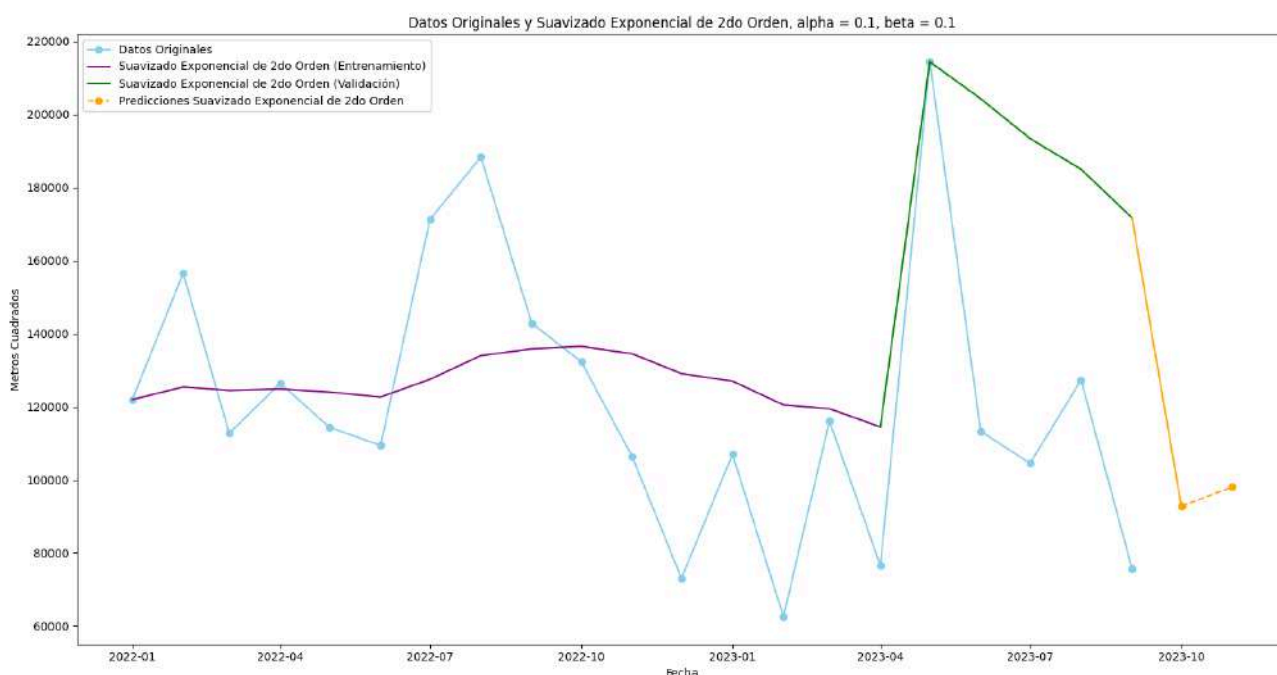


Figura 33: ETS Primer orden, $\alpha = 0.1$, $\beta = 0.1$. Fuente: Elaboración propia

Suavización Exponencial (Exponential Smoothing) Tercer Orden o de Holt-Winters:

Se procedió a aplicar el suavizado exponencial de tercer orden o método de Holt-Winters a los datos originales proporcionados. La configuración resultante para los parámetros de suavizado (α , β y γ) se fiteó utilizando las bibliotecas adecuadas.

La configuración final de los parámetros es la siguiente:

- Smoothing Level (Alpha): 0.146
- Smoothing Trend (Beta): 0.0001
- Smoothing Seasonal (Gamma): 0.0001
- Damping Trend: No aplicado
- Initial Level: 149640.75
- Initial Trend: -4172.70
- Initial Seasonal Components: [22631.62, 6957.67, 6593.83, 2722.66, -11786.16, -27119.62]

Es de destacar que el proceso de fiteo y ajuste de parámetros se realizó utilizando librerías específicas para series temporales. A pesar de los resultados, se observa una advertencia de convergencia durante la optimización. La elección de utilizar un modelo de tercer orden permite capturar tanto la tendencia como la estacionalidad en los datos, lo que es determinante para la precisión de las predicciones a largo plazo. Pero

nuevamente hay que considerar que los patrones de estacionalidad pueden utilizarse efectivamente contando con una cantidad considerable de puntos, y este no es el caso. El método de Holt-Winters se adapta relativamente bien a la complejidad de los datos originales, proporcionando predicciones que consideran tanto la tendencia como los patrones estacionales.

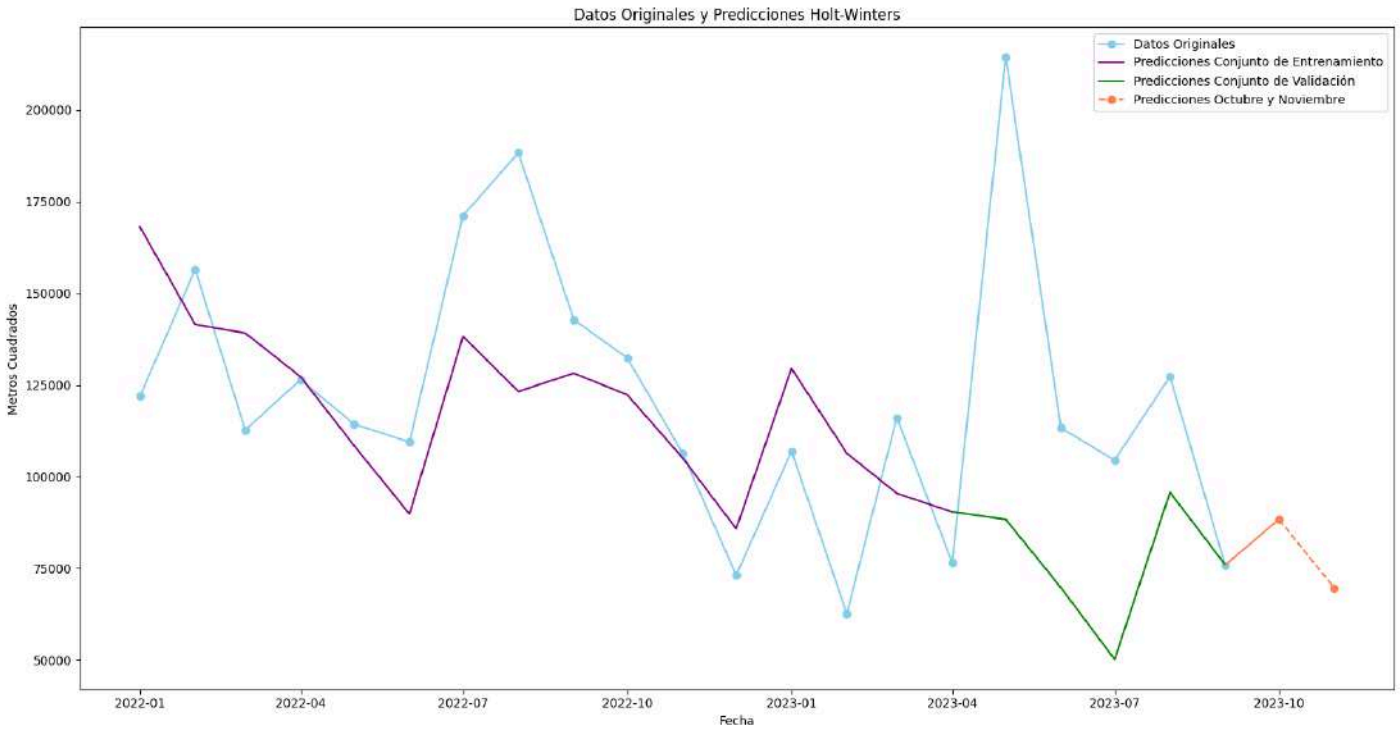


Figura 34: ETS Tercer orden, $\alpha=0.146$, $\beta=0.0001$, $\gamma=0.0001$. Fuente: Elaboración propia

Procesos de Media Móvil (Moving Average Processes):

La ventana de tamaño 3 aplicada en este modelo significa que se utilizó un enfoque de media móvil con un promedio de los últimos 3 puntos para prever el siguiente. Esta técnica es efectiva para suavizar las fluctuaciones en los datos y resaltar las tendencias a largo plazo.

Sin embargo, visualmente, esta ventana puede no capturar de manera precisa los cambios rápidos o las fluctuaciones abruptas en los datos originales, ya que la media móvil tiende a suavizar estas variaciones.

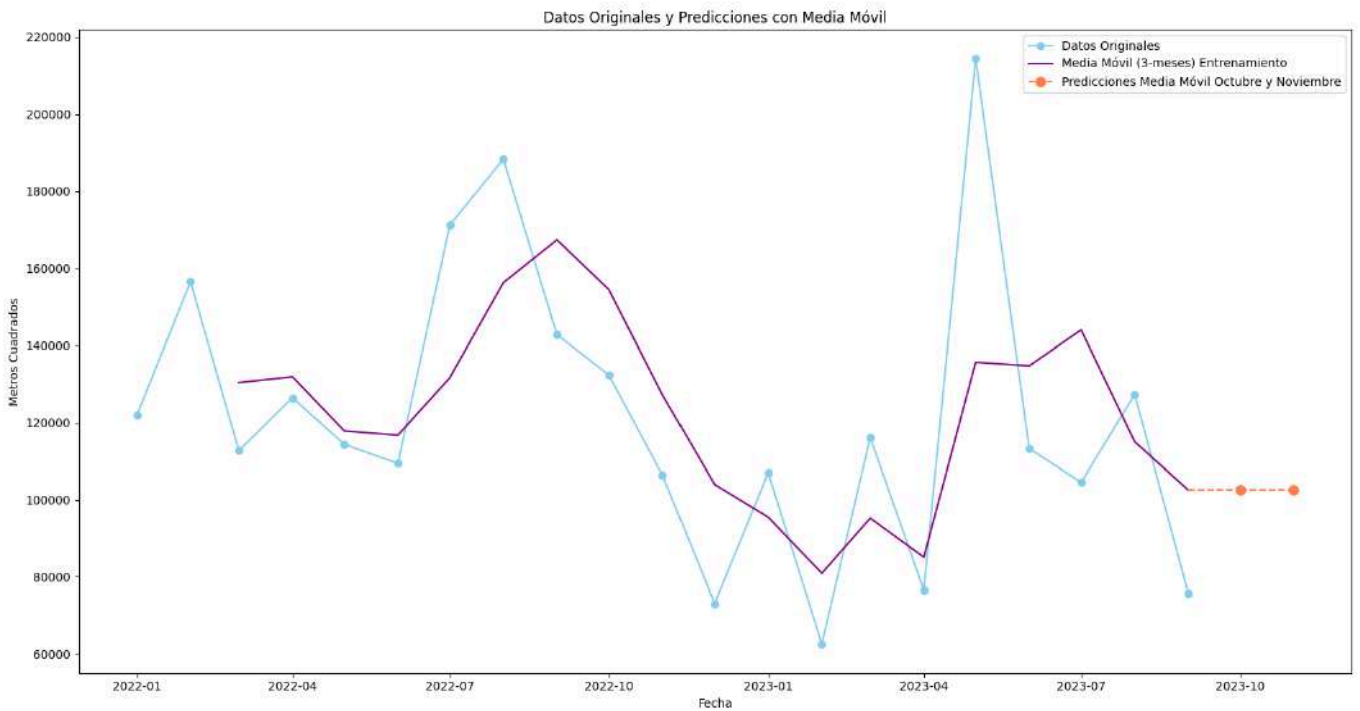


Figura 35: Procesos de Media Móvil, ventana = 3 meses. Fuente: Elaboración propia

Elección del modelo más adecuado:

Modelo	MAE	RMSE	R²	MAPE
<i>Media Móvil (Ventana de 3)</i>	1225020.63	1615453.73	0.9244	9.29%
<i>Suavizado Exponencial 1er Orden (Alpha=0.1)</i>	6788.47	8943.87	0.849	5.60%
<i>Suavizado Exponencial 2do Orden (Alpha=0.1, Beta=0.1)</i>	6807.56	8951.04	0.848	5.62%
<i>Suavizado Exponencial 3er Orden (Alpha=0.146, Beta=0.0001, Gamma=0.0001)</i>	1280120.53	1812747.45	0.969	5.82%

Comparando los modelos, el modelo de Suavizado Exponencial de Primer Orden ($\alpha = 0.1$) sería el más adecuado para estos datos, ya que tiene un RMSE y MAE significativamente más bajos en comparación con el modelo de Procesos de Media Móvil y el modelo de Suavizado Exponencial de Segundo y Tercer Orden.

Aunque el modelo de Suavizado Exponencial de Tercer Orden tiene un R^2 más alto, lo cual indica un mejor ajuste a los datos originales, el hecho de que tenga un RMSE y MAE mucho más alto sugiere que podría estar sobreajustándose y no generalizando bien a nuevos datos. El modelo de Procesos de Media Móvil también proporciona buenos resultados, pero aún así, el modelo de Suavizado Exponencial de Primer Orden tiene un desempeño más sólido en términos de las métricas.

10. Retrospectiva del proyecto

10.1. Estimaciones, objetivos y planificación

Factores como la disponibilidad del equipo, su experiencia, las etapas de investigación y las limitaciones técnicas que surgieron a lo largo del proceso afectaron la duración de algunas tareas, lo que resultó en retrasos.

En primera instancia, se establecieron objetivos primarios y secundarios que sirvieron de base para crear un plan de trabajo, delineando los roles y responsabilidades de los participantes. El siguiente paso consistió en definir el alcance del proyecto, determinando qué tareas eran esenciales y cuáles no lo eran.

Es importante destacar que la obtención de objetivos primarios y secundarios no fue un proceso sencillo. El equipo enfrentó varias dificultades desde el inicio. En primer lugar, el referente funcional necesitó clarificar los requerimientos antes de transmitirlos al equipo. Esta tarea de clarificación, en muchos casos, se llevó a cabo de manera colaborativa junto con el equipo de desarrollo. Posteriormente, los objetivos experimentaron múltiples ajustes en un proceso cambiante, influenciado por el contexto socioeconómico del país. Esta dinámica, combinada con el enfoque económico del referente funcional, desafió las habilidades de renegociación del equipo y en algunos casos, dificultó la estimación de tiempos y difuminó los límites de la fase final del proyecto. Una vez que los objetivos se estabilizaron, fue posible avanzar con la definición de un plan de trabajo.

La planificación inicial del proyecto fue afectada por el progreso de las tareas según el diagrama de Gantt presentado en el protocolo, logrando cumplir con ciertos hitos puntualmente. A lo largo del proyecto, el equipo adquirió más experiencia en gestión, mejorando la organización y permitiendo correcciones en estimaciones iniciales poco precisas. Sin embargo, algunas actividades planificadas con estimaciones consideradas precisas experimentaron aumentos en su duración debido a aspectos no considerados en el relevamiento inicial, necesitando clarificaciones con el referente funcional.

El desarrollo de consultas enfrentó desafíos de rendimiento, llevando a ajustes en el diseño y la introducción de una capa de caché en el frontend para mejorar el tiempo de respuesta. Estos cambios generaron un retraso en la etapa de pruebas y otras actividades, atribuido a limitaciones en la capacidad del equipo para manejar contratiempos y a la centralización de tareas.

En el modelado de series temporales, la comprensión de los modelos y la cantidad y calidad de datos complicaron el análisis, extendiendo el tiempo estimado de 40 a 60 horas. La implementación del modelo también se vio afectada, llevando 10 horas más de lo previsto. Finalmente, la documentación se vio interrumpida debido a complicaciones de tiempo por compromisos laborales, resultando en un aumento de 20 horas en su duración.

Este proyecto no solo marcó una diferencia en el crecimiento profesional del equipo, proporcionando valiosos conocimientos y habilidades interpersonales, sino que también destacó la importancia de la gestión efectiva en situaciones de cambios y limitaciones técnicas, más un desarrollo profesional notable al liderar un proyecto de esta magnitud.

La formación en ingeniería proporcionó la capacidad de adaptación al cambio, creatividad y responsabilidad.

10.2. Alcance de los objetivos

En cuanto a los objetivos generales y específicos, la creación de una aplicación web para la ingesta, procesamiento y visualización de datos con el añadido de un módulo de análisis de ventas se considera cumplida de manera satisfactoria. Sin embargo, al abordar la perspectiva futura de las ventas, cuando se identificó la necesidad de incorporar un modelo de predicción, el equipo se encontró con diferentes dificultades. Este modelo, que tiene en cuenta la facturación en pesos argentinos y la cantidad de metros cuadrados de material utilizado, enfrentó complicaciones debido a la escasa disponibilidad de datos, la cantidad de puntos aislados en una muestra pequeña y la calidad de los datos propiamente dicha, lo que afectó la robustez de las predicciones.

Este punto también fue relevante para la evaluación de la madurez de la empresa a lo largo del proyecto y se consideró como una amenaza en la matriz FODA, anticipando posibles problemas asociados con este aspecto. En resumen, el equipo no preveía un escenario ideal en términos de datos, pero los resultados fueron más desfavorables de lo inicialmente esperado. Nuevamente, y vinculado a la sección anterior, la extensión de ciertas tareas relacionadas con el modelado predictivo se hizo presente.

10.3. Trabajo en equipo, metodología y contacto con el cliente

El equipo enfrentó diversas dificultades durante el desarrollo del proyecto, destacando la importancia de la disponibilidad, un factor clave en cualquier planificación. La necesidad de mantener un *backlog* organizado y tareas granulares fue decisiva para la culminación exitosa de todas las actividades y del proyecto en su totalidad.

La colaboración en el equipo fue evolucionando, respaldada por una buena relación entre sus integrantes y una sinergia de trabajo, lo que permitió avanzar de manera efectiva en ciertos hitos. El uso de *brainstorming*, debates y el constante apoyo mutuo contribuyeron significativamente al desarrollo del proyecto, demostrando un alineamiento constante hacia un objetivo común.

A medida que las semanas avanzaban y el equipo ganaba experiencia, se logró retomar un ritmo de trabajo más eficiente y una mayor flexibilidad en la implementación de tareas. Sin embargo, esta flexibilidad conllevó ajustes y rediseños. Por lo tanto, el equipo destaca la importancia de una buena gestión y una metodología de trabajo bien delimitada para organizar eficazmente las actividades.

El respaldo del cliente fue fundamental para la finalización exitosa. Las reuniones periódicas de relevamiento y las revisiones de avances permitieron detectar errores de interpretación y enriquecer el proyecto, asegurando su funcionalidad óptima para el usuario final.

No obstante, para lograr el soporte y respaldo del cliente (y referente funcional), fueron necesarios diferentes procesos de renegociación, en donde se llevaron a cabo discusiones detalladas para ajustar y adaptar diversas funcionalidades del sistema según las expectativas y necesidades cambiantes del cliente. Estos procesos de renegociación se convirtieron en un componente esencial para asegurar la alineación entre las metas del proyecto y las expectativas del cliente a lo largo del desarrollo.

Profundizando, durante este proceso, se llevaron a cabo diversas sesiones en las cuales el referente funcional compartía aspectos de interés y requerimientos específicos. Sin embargo, al carecer de un enfoque informático o de desarrollo, fue necesario realizar explicaciones detalladas sobre cómo ciertas actividades impactarían significativamente en los plazos y la complejidad del proyecto. Se estableció una comunicación transparente para transmitir la complejidad técnica asociada con ciertos requerimientos, buscando un entendimiento mutuo sobre las implicaciones y *trade-offs* entre las funcionalidades deseadas y las limitaciones del desarrollo. Estas interacciones resultaron fundamentales para garantizar una alineación efectiva entre las expectativas del cliente y las capacidades del equipo.

11. Conclusiones

A pesar de los desafíos, la implementación exitosa de la aplicación web y del módulo de análisis de ventas se destacan como logros. Sin embargo, la incorporación de un modelo de predicción para las ventas futuras encontró dificultades debido a la limitada disponibilidad de datos, calidad y puntos aislados como ya se mencionó.

Al analizar los datos sobre la duración de las tareas, se observa que el tiempo asignado a la arquitectura y base de datos representa solo el 5% del total del proyecto. Es posible que si se hubiera dedicado más tiempo a esta etapa, se podría haber evitado el problema de rendimiento en las consultas más adelante. Esto quizás habría hecho innecesaria la implementación de una capa de caché para mejorar los tiempos de respuesta. La observación hecha conduce a una reflexión sobre las tareas fundamentales en el proyecto. Resulta evidente que un enfoque básico en el diseño inicial puede desencadenar problemas significativos en etapas posteriores. Es determinante reconocer la importancia de asignar tiempo y recursos adecuados a la planificación y ejecución de las fases iniciales del proyecto, para evitar complicaciones y optimizar la eficiencia en todo su ciclo de vida.

Al reflexionar sobre el proyecto, se identifican diversos desafíos vinculados a la disponibilidad del equipo, la experiencia acumulada, las fases de investigación, las limitaciones técnicas, el marcado perfil económico del referente funcional y el contexto económico y político, que condicionaron el presente trabajo y su estimación. Estos elementos impactaron en la extensión temporal de ciertas actividades, ocasionando demoras. A pesar de ello, se sostiene que los resultados representan una situación beneficiosa para ambas partes. El incremento notable en competencias de gestión, la capacidad de adaptación a cambios, el reconocimiento de fortalezas y debilidades, junto con la satisfacción del cliente, confirman su consideración de éxito en el proyecto. El equipo rescata un aprendizaje valioso de estos procesos de relevamiento, negociación y estimación de plazos. La experiencia obtenida se aplica actualmente en el entorno laboral.

Las actividades de investigación contribuyeron a aclarar y consolidar conceptos teóricos difusos que el equipo poseía. Un claro ejemplo es el de la definición de KPI y KRI.

Establecer una metodología de trabajo sólida y sinérgica entre los integrantes del equipo resultó ser un desafío considerable. La adaptación a los horarios de los demás integrantes, la asignación de tareas y, en definitiva, la autogestión del equipo, fueron aspectos que se fueron perfeccionando hasta lograr un equilibrio en la ejecución de las actividades.

Explorar la viabilidad de emplear un software de visualización del mercado en contraposición a un sistema desarrollado internamente contribuyó a considerar detenidamente los beneficios y

limitaciones de cada alternativa. Este análisis motivó a investigar más a fondo temas relacionados con la gestión de datos.

Por otro lado, realizar un análisis de madurez llevó al equipo a integrar la perspectiva ingenieril en el proyecto. Se llevó a cabo un análisis del escenario previo y posterior a la implementación del sistema, reconociendo logros y detectando nuevos puntos de dolor que surgieron de dicha evaluación.

Desde la perspectiva técnica, en la actualidad se han adquirido habilidades en nuevas herramientas y productos de software. Esto habilita la capacidad de realizar evaluaciones entre diversos productos, permitiendo una elección informada y consciente.

Además, trabajar en estrecha colaboración con un representante del negocio proporcionó una valiosa oportunidad para adquirir conocimiento detallado sobre las operaciones financieras y elementos clave de interés para la empresa. **Esta inmersión en el contexto empresarial permitió al equipo no solo comprender mejor las necesidades del referente funcional, sino también brindar un apoyo más efectivo al momento de definir y refinar los requerimientos.**

El proyecto no solo cumplió con sus objetivos, sino que también proporcionó valiosas lecciones y experiencias para el crecimiento profesional. Reforzó la importancia de la adaptabilidad, la colaboración y una comprensión profunda de la situación de la empresa, en sintonía con los conceptos de transformación digital y madurez organizacional previamente explorados.

Además, un indicador de éxito del proyecto radica en **la decisión de la empresa de seguir adelante con el mantenimiento evolutivo del sistema implementado.** Esto no solo demuestra que el producto ha cumplido con las expectativas, también indica que se ha logrado guiar a la empresa y a la capa gerencial a través de una transformación cultural basada en datos. **En la actualidad, la compañía experimenta una transformación en su modelo operativo, abarcando la gestión de operaciones y la toma de decisiones.** Este impacto va más allá de la implementación de un software; es el **valor añadido de identificar las necesidades del cliente, respaldado por una sólida formación en ingeniería y gestión.** Reconocer los datos como activos estratégicos marca un punto de inflexión en el desarrollo empresarial. En muchos casos se recopilan datos, pero no se les brinda el tratamiento adecuado para que puedan generar valor dentro de las organizaciones.

El equipo consultó con el referente funcional el significado de negocio de los indicadores desarrollados. Se describen a continuación en el lenguaje empleado por la empresa:

- “La vista de ventas permite identificar rápidamente qué material se vendió por sobre y por debajo del precio promedio. Esta vista se complementa con la de clientes, dado que

si determinada venta se realiza por debajo del precio promedio es posible que el cliente haya liquidado su pago dentro de los plazos esperados.”

- Continuando con la vista de clientes y en línea con el punto anterior: “permite revisar el comportamiento de pagos dentro de un plazo establecido para fijar los precios a los que se les vende determinado producto a clientes particulares. En esta vista, es importante destacar que pueden figurar clientes sin movimientos en el período seleccionado, pero igualmente se clasifican para mostrar los indicadores históricos. Este punto es algo que a la empresa le gustaría mejorar.”
- Con respecto a la vista de históricos: “esta funcionalidad brinda la capacidad de visualizar el desempeño en comparación con los últimos meses. También permite revisar los montos cobrados y facturados durante el último año, así como observar la evolución del negocio a lo largo del tiempo”.
- Las vistas de top cinco de cobranza y facturación: “de forma rápida se puede conocer quiénes son los clientes que más influyen en la facturación y la cobranza del último mes cerrado o para el filtro aplicado en particular.”
- Además, se sugiere incluir como parte del mantenimiento evolutivo un registro histórico de ventas y precios: “en el histórico de ventas se registraría el comportamiento de las ventas en pesos argentinos junto con la cantidad de metros vendidos, mientras que en otro gráfico se mostraría la evolución del precio promedio. También se anexaría una curva que muestre la variación inflacionaria como parámetro del período. Esta mejora es significativa ya que actualmente es posible revisar la variación del precio promedio de un mes a otro mediante filtros por meses consecutivos, pero tener toda esta información en una sola vista sería más conveniente.”

El último punto se destaca como un indicador clave del uso del sistema y cómo la empresa ha transformado su proceso de toma de decisiones. El inicio del relevamiento de esta nueva vista, junto con su desarrollo, se abordará en un futuro próximo.

Además, se le consultó específicamente al referente funcional qué fue lo que esta herramienta vino a solucionar, cuánto lo asistió en sus actividades a nivel gerencial y qué acciones le permite tomar. La respuesta, nuevamente en el lenguaje del referente es la siguiente:

“Anteriormente, teníamos las herramientas necesarias para realizar análisis de datos, como el sistema de origen y software de visualización, pero el proceso era extremadamente laborioso. Carecíamos de un esquema que transformara los datos y los preparara para la creación de indicadores, lo que implicaba realizar esta transformación manualmente. **Ahora, en poco tiempo, podemos obtener información sobre el aumento de los precios desde el inicio del año hasta la fecha de consulta, por ejemplo.** También podemos analizar cómo se distribuyeron las ventas en un mes específico en relación con los diferentes materiales (por

ejemplo, en cierto mes, el 36% de las ventas correspondían al material X). Incluso podemos identificar qué clientes compraron a precios por encima o por debajo del promedio.

Utilizando el histórico de datos podemos establecer objetivos comerciales, como aumentar la facturación en un 15%, por ejemplo. Con solo un clic, ahora podemos acceder a la misma información que antes nos llevaba dos días obtener. Este sistema nos permite obtener datos instantáneamente de manera muy sencilla. Además, a partir de un sistema externo que almacena información, podemos transformarla en datos financieros y económicos de manera accesible para cualquier persona. Solo necesitamos descargar un archivo y cargarlo en el sistema desde nuestra casa. Incluso el propietario de la empresa lo utiliza, ya que no es necesario ser un experto en datos para obtener información relevante sobre el negocio.”

Después de revisar las líneas expuestas en este apartado, el equipo llega a la conclusión de que el incremento en experiencia de gestión, liderazgo y conocimientos técnicos, junto con la transformación experimentada en Embalajes Marplatenses, es altamente significativa. Si los integrantes del equipo fueran inversionistas, sin duda optarían por invertir en la empresa que juntos han creado. Se establecieron múltiples objetivos al inicio de este proyecto que se lograron alcanzar tras el reajuste inicial en la planificación. Esto demuestra el potencial del equipo para cumplir con dichos objetivos.

12. Glosario

API: Interfaz de Programación de Aplicaciones. Un conjunto de reglas y definiciones que permite la comunicación entre diferentes aplicaciones. Proporciona funciones y protocolos para facilitar la interacción entre sistemas.

API REST: Estilo de arquitectura de software para sistemas distribuidos basado en estándares web. Utiliza métodos HTTP para operaciones CRUD (Crear, Leer, Actualizar, Eliminar) en recursos, siendo ampliamente adoptado en la construcción de servicios web.

Backend: La parte de una aplicación o sistema que gestiona la lógica de negocio, la manipulación de datos y la comunicación con la base de datos. El backend se ejecuta en el servidor y se encarga de procesar las solicitudes del frontend.

Backlog: En el contexto de metodologías ágiles como Scrum, un backlog es una lista prioritaria de tareas o funcionalidades que deben abordarse en un proyecto. Puede incluir elementos como características, mejoras y correcciones pendientes.

Base de datos: Un sistema organizado para recopilar, almacenar y gestionar datos. Proporciona un método estructurado para acceder, gestionar y actualizar información de manera eficiente.

Business Intelligence: Un conjunto de tecnologías, procesos y herramientas que ayudan a las empresas a recopilar, analizar y presentar información empresarial. El Business Intelligence facilita la toma de decisiones informadas al proporcionar una visión clara de los datos empresariales.

Caché: Un área de almacenamiento temporal que almacena datos para acelerar el acceso y mejorar el rendimiento. La caché se utiliza para almacenar copias de datos frecuentemente accedidos, reduciendo la necesidad de recuperarlos desde la fuente original.

Ceremonia: En el contexto de metodologías ágiles como Scrum, una ceremonia es un evento planificado que facilita la colaboración y la comunicación entre los miembros del equipo. Ejemplos de ceremonias en Scrum incluyen la planificación de sprint, la revisión de sprint y la retrospectiva.

Cliente: En el contexto de aplicaciones cliente-servidor, el cliente es la interfaz de usuario o la aplicación que solicita y consume servicios o recursos del servidor.

Contenedor: Un entorno ejecutable y aislado que contiene una aplicación específica y todas sus dependencias. Proporcionan una forma consistente de ejecutar aplicaciones en diferentes entornos, asegurando la consistencia y portabilidad del software.

Dashboard: Interfaz gráfica que proporciona una visualización rápida de datos clave y métricas para facilitar la toma de decisiones informadas. Utilizado para el monitoreo del rendimiento.

Data-Driven: Un enfoque que utiliza datos y análisis para guiar la toma de decisiones y estrategias. Ser data-driven implica confiar en evidencia cuantificable en lugar de suposiciones o intuiciones al tomar decisiones empresariales.

Data Analytics: El proceso de examinar, limpiar, transformar y modelar datos para descubrir información útil, llegar a conclusiones y apoyar la toma de decisiones. La analítica de datos utiliza diversas técnicas y herramientas para explorar patrones y tendencias en conjuntos de datos.

Docker: Una plataforma de código abierto que automatiza el despliegue de aplicaciones en contenedores. Docker permite empaquetar una aplicación y sus dependencias en un contenedor, asegurando su portabilidad y fácil implementación en cualquier entorno compatible.

Express: Framework de aplicación web para Node.js. Simplifica el desarrollo de aplicaciones web y APIs, proporcionando una capa de abstracción sobre Node.js para gestionar rutas, solicitudes y respuestas, así como la integración de middleware.

Frontend: La interfaz de usuario y la experiencia visual de una aplicación o sitio web. El frontend se encarga de la presentación y la interacción directa con el usuario.

Git: Sistema de control de versiones distribuido. Facilita el seguimiento de cambios en el código fuente durante el desarrollo de software y mantiene un historial completo de cambios.

Imagen: En el contexto de Docker, una imagen es un paquete ejecutable que incluye todo lo necesario para ejecutar una aplicación, incluyendo el código, bibliotecas, dependencias y configuraciones. Las imágenes de Docker son la base para crear contenedores.

Insights: En el contexto de análisis de datos, "insights" se refiere a percepciones o comprensiones significativas derivadas del examen de datos. Proporcionan una comprensión más profunda de los patrones y tendencias presentes en la información analizada.

JSON (JavaScript Object Notation): Un formato ligero de intercambio de datos que utiliza una sintaxis legible para humanos. JSON representa datos como pares clave-valor y estructuras de datos anidadas. Es ampliamente utilizado para transmitir datos entre un servidor y una aplicación web, y es fácilmente interpretable por las máquinas.

JWT (JSON Web Token): Un estándar abierto (RFC 7519) que define una manera compacta y segura de transmitir información entre partes como un objeto JSON. Los JWT se utilizan

comúnmente para autenticar usuarios y compartir información segura entre un servidor y un cliente.

KPIs (Indicadores Clave de Desempeño): Medidas cuantificables para evaluar el rendimiento empresarial. Ayudan a medir el progreso hacia objetivos y facilitan la toma de decisiones informadas.

Metodología Ágil: Un enfoque flexible y colaborativo para el desarrollo de software que prioriza la adaptabilidad a los cambios y la entrega continua. Las metodologías ágiles valoran la interacción constante con los clientes, la retroalimentación temprana y la entrega incremental.

Node: Entorno de ejecución para JavaScript del lado del servidor. Permite ejecutar código JavaScript en el servidor, facilitando la construcción de aplicaciones escalables y de alto rendimiento.

ORM: Mapeo Objeto-Relacional. Una técnica de programación que permite interactuar con bases de datos relacionales mediante objetos en lugar de consultas SQL directas, facilitando la integración entre el código y la estructura de la base de datos.

Passport: Librería de autenticación para Node.js. Simplifica la implementación de estrategias de autenticación, comúnmente utilizada en aplicaciones web para gestionar la autenticación de usuarios de manera eficiente y segura.

Product Owner: Rol en el desarrollo ágil de software que representa los intereses del cliente y define características y prioridades del producto. Colabora con el equipo de desarrollo para asegurar que el producto cumpla con requisitos y expectativas.

Rama: En el contexto de control de versiones, una rama es una línea independiente de desarrollo derivada de la línea principal del código. Permite trabajar en funcionalidades o correcciones sin afectar directamente el código principal.

React: Biblioteca de JavaScript para construir interfaces de usuario interactivas. Se centra en componentes reutilizables que gestionan su propio estado, simplificando el desarrollo de aplicaciones frontend robustas y eficientes.

Repositorio: Espacio de almacenamiento para guardar y gestionar el código fuente de un proyecto. Puede ser local o remoto y se utiliza comúnmente con sistemas de control de versiones como Git.

Responsive: Diseño que se adapta y funciona óptimamente en diferentes dispositivos y tamaños de pantalla. Garantiza una experiencia de usuario consistente y eficiente en diversos dispositivos.

Scrum: Un marco de trabajo ágil que organiza el desarrollo de software en iteraciones llamadas "sprints". Scrum define roles claros, como Scrum Master y Product Owner, y utiliza ceremonias como la planificación de sprint y la revisión de sprint para mantener un proceso transparente y colaborativo.

Servidor: En el contexto de aplicaciones cliente-servidor, el servidor es la parte que proporciona servicios, recursos o datos al cliente. Se encarga de procesar las solicitudes y gestionar la lógica de negocio.

SEO (Optimización para Motores de Búsqueda): Prácticas y técnicas para mejorar la visibilidad y clasificación en los resultados de búsqueda. SEO se centra en optimizar contenido y estructuras para aumentar la relevancia y atraer tráfico orgánico.

SPA (Single Page Application): Aplicación de Página Única. Un tipo de aplicación web que carga una sola página y actualiza dinámicamente el contenido mientras los usuarios interactúan. Esto mejora la velocidad y la experiencia del usuario al evitar recargas completas de la página.

Sprint: En el contexto de metodologías ágiles como Scrum, un sprint es un período de tiempo fijo y corto, generalmente de dos a cuatro semanas, durante el cual se realiza trabajo en una parte específica del producto.

SQL (Lenguaje de Consulta Estructurado): Un lenguaje estándar para gestionar y manipular bases de datos relacionales. SQL se utiliza para realizar consultas, actualizaciones, inserciones y eliminaciones de datos en bases de datos que siguen el modelo relacional.

Stakeholders: Las partes interesadas en un proyecto o proceso, como clientes, usuarios, patrocinadores y aquellos que se ven afectados por los resultados. Gestionar las expectativas y la participación de los stakeholders es esencial para el éxito de un proyecto.

XLS (Excel Spreadsheet): Un formato de archivo asociado comúnmente con hojas de cálculo de Microsoft Excel. Los archivos XLS almacenan datos tabulares y fórmulas que permiten realizar cálculos y análisis de datos de manera efectiva.

13. Bibliografía

- [1] Embalajes Santa Catalina. (s.f.). Nuestra Historia. Recuperado de <https://embalajesantacatalina.com/nuestra-historia/>
- [2] Pressman, R. S. (2010). Desarrollo Ágil. En "Ingeniería del Software: Un Enfoque Práctico" (7a ed., p. 69). University of Connecticut.
- [3] Pressman, R. S. (2010). Desarrollo Ágil. En "Ingeniería del Software: Un Enfoque Práctico" (7a ed., p. 70). University of Connecticut.
- [4] DataMiner Community. (s.f.). The Scrum Framework. Recuperado de <https://community.dataminer.services/the-scrum-framework/>
- [5] Atlassian. (s.f.). Gitflow Workflow. Recuperado de <https://www.atlassian.com/es/git/tutorials/comparing-workflows/gitflow-workflow>
- [6] Atlassian. (s.f.). Feature Branch Workflow. Recuperado de <https://www.atlassian.com/es/git/tutorials/comparing-workflows/feature-branch-workflow>
- [7] Docker. (s.f.). Docker Compose: Features and Uses. Recuperado de <https://docs.docker.com/compose/features-uses/#development-environments>
- [8] Heroku. (2023). Local Development with Docker Compose. Recuperado de <https://devcenter.heroku.com/articles/local-development-with-docker-compose>
- [9] Corporación Andina de Fomento. Vicepresidente de Conocimiento: Pablo Sanguinetti (2021). Experiencia: Datos e Inteligencia Artificial en el Sector Público. La versión digital de este libro se encuentra en: scioteca.caf.com. PARTE 1: Entender la inteligencia artificial y su potencial para el sector público, p. 48-53.
- [10] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Data Quality p. 55). Sebastopol, CA: O'Reilly Media, Inc.
- [11] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, p. 55-56). Sebastopol, CA: O'Reilly Media, Inc.
- [12] Corporación Andina de Fomento, Vicepresidente de Conocimiento: Pablo Sanguinetti, 2021, p. 57-58, PARTE 1: Entender la inteligencia artificial y su potencial para el sector público,

Capítulo 1. Conceptos fundamentales sobre inteligencia artificial en el sector público - ¿Qué son los datos?

[13] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 1. Data Engineering Described, p. 3). Sebastopol, CA: O'Reilly Media, Inc.

[14] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 1. Data Engineering Described, p. 4). Sebastopol, CA: O'Reilly Media, Inc.

[15] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 1. Data Engineering Described, The Data Engineering Lifecycle p. 5). Sebastopol, CA: O'Reilly Media, Inc.

[16] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Generation: Source Systems p. 35). Sebastopol, CA: O'Reilly Media, Inc.

[17] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Generation: Source Systems p. 35). Sebastopol, CA: O'Reilly Media, Inc.

[18] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Evaluating source systems: Key engineering considerations p. 36). Sebastopol, CA: O'Reilly Media, Inc.

[19] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Evaluating storage systems: Key engineering considerations p. 38). Sebastopol, CA: O'Reilly Media, Inc.

[20] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Ingestion p. 39). Sebastopol, CA: O'Reilly Media, Inc.

[21] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part II. The Data Engineering Lifecycle in Depth, Chapter 7 Ingestion p. 233). Sebastopol, CA: O'Reilly Media, Inc.

- [22] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Transformation p. 43). Sebastopol, CA: O'Reilly Media, Inc.
- [23] Wall Street Mojo. (s.f.). Data Transformation. Recuperado de <https://www.wallstreetmojo.com/data-transformation/>
- [24] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Serving Data p. 44). Sebastopol, CA: O'Reilly Media, Inc.
- [25] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Serving Data p. 45). Sebastopol, CA: O'Reilly Media, Inc.
- [26] Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Part I. Foundation and Building Blocks, Chapter 2 The Data Engineering Lifecycle, Serving Data p. 45). Sebastopol, CA: O'Reilly Media, Inc.
- [27] Amazon Web Services. (s.f.). Machine Learning. Recuperado de <https://aws.amazon.com/es/what-is/machine-learning/>
- [28] Parmenter, D. (2015). Key Performance Indicators: Developing, Implementing, and Using Winning KPIs (3ra ed., Part I: Setting the Scene, Chapter 1. The Great KPI Misunderstanding, p. 7). Wiley.
- [29] Parmenter, D. (2015). Key Performance Indicators: Developing, Implementing, and Using Winning KPIs (3ra ed., Part I: Setting the Scene, Chapter 1. The Great KPI Misunderstanding, p. 4). Wiley.
- [30] SCL Consulting. (2016). Conocimiento del Negocio Digital en Tiempo Real con SAP BW4HANA. Recuperado de <https://www.scl-consulting.com/conocimiento-del-negocio-digital-en-tiempo-real-con-sap-bw4hana/>
- [31] Joyanes Aguilar, L. (2019). Inteligencia de Negocios y Analítica de Datos: Una Visión Global de Business Intelligence & Analytics (Parte I: Visión Moderna de Inteligencia de Negocios y Analítica de Datos, Capítulo 1. Inteligencia de Negocios: Una Panorámica Global, p. 6). Bogotá: Alfaomega.

- [32] Joyanes Aguilar, L. (2019). Inteligencia de Negocios y Analítica de Datos: Una Visión Global de Business Intelligence & Analytics (Parte III: Analítica de Negocios y Analítica de Datos, Capítulo 10. Analítica de Datos (Big Data & Analytics), p. 305). Bogotá: Alfaomega.
- [33] Joyanes Aguilar, L. (2019). Inteligencia de Negocios y Analítica de Datos: Una Visión Global de Business Intelligence & Analytics (Parte I: Visión Moderna de Inteligencia de Negocios y Analítica de Datos, Capítulo 2. Analítica de Negocios (Business Analytics): Una Visión Global, p. 78). Bogotá: Alfaomega.
- [34] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting Familiar with Time Series, Chapter 1, Introducing Time Series, p. 4). Packt Publishing Ltd.
- [35] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting Familiar with Time Series, Chapter 3, Analyzing and Visualizing Time Series Data, p. 49). Packt Publishing Ltd.
- [36] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting Familiar with Time Series, Chapter 3, Analyzing and Visualizing Time Series Data, p. 72). Packt Publishing Ltd.
- [37] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting Familiar with Time Series, Chapter 4, Setting a Strong Baseline Forecast, p. 78). Packt Publishing Ltd.
- [38] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting Familiar with Time Series, Chapter 4, Setting a Strong Baseline Forecast, p. 79). Packt Publishing Ltd.
- [39] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting Familiar with Time Series, Chapter 4, Setting a Strong Baseline Forecast, p. 84). Packt Publishing Ltd.
- [40] Joseph, M. (2022). Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning (Part 1 – Getting

Familiar with Time Series, Chapter 4, Setting a Strong Baseline Forecast, p. 87). Packt Publishing Ltd.

[41] Joyanes Aguilar, L. (2019). Inteligencia de Negocios y Analítica de Datos: Una Visión Global de Business Intelligence & Analytics (Parte I: Visión Moderna de Inteligencia de Negocios y Analítica de Datos, Capítulo 3. Transformación Digital en Organizaciones y Empresas (Economía Colaborativa, Experiencia de Cliente y Blockchain), p. 98). Bogotá: Alfaomega.

[42] Joyanes Aguilar, L. (2019). Inteligencia de Negocios y Analítica de Datos: Una Visión Global de Business Intelligence & Analytics (Parte I: Visión Moderna de Inteligencia de Negocios y Analítica de Datos, Capítulo 3. Transformación Digital en Organizaciones y Empresas (Economía Colaborativa, Experiencia de Cliente y Blockchain), p. 99). Bogotá: Alfaomega.

[43] Joyanes Aguilar, L. (2019). Inteligencia de Negocios y Analítica de Datos: Una Visión Global de Business Intelligence & Analytics (Parte I: Visión Moderna de Inteligencia de Negocios y Analítica de Datos, Capítulo 3. Transformación Digital en Organizaciones y Empresas (Economía Colaborativa, Experiencia de Cliente y Blockchain), p. 100). Bogotá: Alfaomega.

[44] Press, G. (2015, 6 de diciembre). 6 Predictions About the Future of Digital Transformation. Recuperado de

<https://www.forbes.com/sites/gilpress/2015/12/06/6-predictions-about-the-future-of-digital-transformation/?sh=4a2c5cce1102>

[45] Newman, D. (2016, 30 de agosto). Top 10 Trends for Digital Transformation in 2017.

Forbes. Recuperado de

<https://www.forbes.com/sites/danielnewman/2016/08/30/top-10-trends-for-digital-transformation-in-2017/?sh=35fec54147a5>

[46] Grid.cl. (s.f.). La Transformación Digital: ¿Por qué es tan importante para tu empresa?

Recuperado de

<https://www.grid.cl/blog/la-transformacion-digital-por-que-es-tan-importante-para-tu-empresa/>

[47] Rogers, D. L. (2016). The Digital Transformation Playbook: Rethink Your Business for the Digital Age (Chapter 4: Turn Data Into Assets, p. 89). Columbia University Press.

[48] Rogers, D. L. (2016). The Digital Transformation Playbook: Rethink Your Business for the Digital Age (Chapter 4: Turn Data Into Assets, p. 91). Columbia University Press.

- [49] Rogers, D. L. (2016). *The Digital Transformation Playbook: Rethink Your Business for the Digital Age* (Chapter 4: Turn Data Into Assets, p. 91). Columbia University Press.
- [50] Rogers, D. L. (2016). *The Digital Transformation Playbook: Rethink Your Business for the Digital Age* (Chapter 4: Turn Data Into Assets, p. 117). Columbia University Press.
- [51] Rogers, D. L. (2016). *The Digital Transformation Playbook: Rethink Your Business for the Digital Age* (Chapter 4: Turn Data Into Assets, p. 118). Columbia University Press.
- [52] Rogers, D. L. (2016). *The Digital Transformation Playbook: Rethink Your Business for the Digital Age* (Chapter 4: Turn Data Into Assets, p. 118). Columbia University Press.
- [53] Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering: Plan and Build Robust Data Systems* (Part I. Foundation and Building Blocks, Chapter 1. Data Engineering Described, Data Maturity and the Data Engineer p. 13). Sebastopol, CA: O'Reilly Media, Inc.
- [54] Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering: Plan and Build Robust Data Systems* (Part I. Foundation and Building Blocks, Chapter 1. Data Engineering Described, Data Maturity and the Data Engineer p. 14). Sebastopol, CA: O'Reilly Media, Inc.
- [55] Prod-G2G Assets. (s.f.). Data Management Maturity Assessment Advice Note. Recuperado de https://prod-g2g-assets.s3.amazonaws.com/documents/Data_Management_Maturity_Assessment_Advice_Note.pdf