



UNIVERSIDAD MAYOR DE SAN SIMÓN
FACULTAD DE CIENCIAS Y TECNOLOGÍA

DIRECCIÓN DE POSGRADO



DIPLOMADO CIENCIA DE DATOS

SEGUNDA VERSIÓN

DIGITALIZACIÓN Y ANÁLISIS PREDICTIVO PARA LA OPTIMIZACIÓN OPERATIVA DEL VIVERO “PARAÍSO DE HOJAS VERDES”

**PROYECTO PRESENTADO PARA OBTENER EL GRADO DE LICENCIATURA EN
INGENIERÍA DE SISTEMAS
MODALIDAD DOBLE TITULACIÓN**

POSTULANTE : NOELIA FELIPE RAMIREZ
TUTOR : ING. CAROLINA AGUILAR

Cochabamba – Bolivia
2025

DIGITALIZACIÓN Y ANÁLISIS PREDICTIVO PARA LA OPTIMIZACIÓN OPERATIVA DEL VIVERO “PARAÍSO DE HOJAS VERDES”

Por

Noelia Felipe Ramirez

El presente documento, Trabajo de Grado es presentado a la Dirección de Posgrado de la Facultad de Ciencias y Tecnología en cumplimiento parcial de los requisitos para la obtención del grado académico de Licenciatura en Ingeniería de sistemas, modalidad Doble Titulación, habiendo cursado el Diplomado “Estadística Aplicada a la Toma de Decisiones” propuesta por el Centro de Estadística Aplicada (CESA) en su tercera versión.

ASESOR/TUTOR

Ing. Carolina Aguilar

COMITÉ DE EVALUACIÓN

Ing. M.Sc. Ronald Edgar Patiño Tito. (Presidente)

Ing. M.Sc Guillen Salvador Roxana. (Coordinador)

Ing. M.Sc Samuel Roberto Achà Pérez (Tribunal)



DIRECCIÓN DE POSGRADO, FACULTAD DE CIENCIAS Y TECNOLOGIA
Cochabamba, Bolivia

Aclaración

Este documento describe el trabajo realizado como parte del programa de estudios de Diplomado “Ciencia de Datos” en el Centro de Estadística Aplicada CESA y la Dirección de Posgrado de la Facultad de Ciencias y Tecnología. Todos los puntos de vista y opiniones expresadas en el mismo son responsabilidad exclusiva del autor y no representan necesariamente las de la institución.

Resumen

El proyecto se desarrolló en el vivero **Paraíso de Hojas Verdes**, ubicado en Cochabamba, Bolivia, con el objetivo de mejorar la gestión de ventas e inventario mediante la digitalización de procesos. A pesar de utilizar redes sociales como Tik Tok y WhatsApp para promocionar y vender plantas ornamentales, el vivero gestionaba sus operaciones de manera manual, lo que dificultaba el control de ventas, inventario y atención al cliente.

Ante esta situación, se planteó como objetivo general el diseño e implementación de un sistema de registro y análisis de datos utilizando herramientas como Excel, Jupyter Notebook y Power BI, para optimizar la toma de decisiones estratégicas y facilitar la planificación basada en datos reales. Entre los objetivos específicos estuvieron la creación de una base de datos estructurada, el análisis de ventas y clientes, y la visualización de resultados para mejorar el control del negocio.

Durante el desarrollo del proyecto, se logró cargar exitosamente los datos en Excel, generando una base de datos con 116 filas y 10 columnas, suficiente para realizar análisis estadísticos representativos. Los resultados permitieron identificar patrones de compra, semanas de mayor demanda y clientes frecuentes, además de establecer una base sólida para futuras predicciones de ventas.

La implementación del proyecto en el vivero Jardín de Hojas Verdes permitió digitalizar y automatizar la gestión de ventas, inventario y preferencias de clientes mediante Excel, Jupyter Notebook y Power BI; se recolectaron y organizaron con éxito los datos necesarios en una base de Excel, se analizó el comportamiento de los clientes y se realizó una predicción de ventas para la temporada de verano, mientras que Power BI facilitó la toma de decisiones al revelar que las plantas de interior son las más demandadas y que La Paz lidera las compras en línea con Bs 6.903 en ventas.

El proyecto concluye destacando la importancia de incorporar tecnologías accesibles y efectivas en pequeños negocios, como el uso de Jupyter Notebook para análisis, Power BI para visualización, y Excel como base de datos, con el fin de fortalecer la competitividad, mejorar la experiencia del cliente y fomentar un crecimiento sostenible.

Palabras clave

Digitalización empresarial, gestión de inventario, análisis de datos, Power BI, ventas digitales, pequeños negocios.

*Dedico este proyecto con todo mi cariño y gratitud a mis padres, mis
hermanos que quienes me brindaron un apoyo incondicional a lo
largo de toda mi carrera profesional.*

*También quiero dedicar estas páginas a mi hermana que ahora está
en el cielo; aunque físicamente no esté conmigo, siento su presencia
cada día, guiándome y dándome fuerzas para seguir adelante y
cumplir mis metas.*

Agradecimientos

A la Ing. Carolina Aguilar por su valioso e incondicional apoyo, cuya guía a lo largo de este proyecto fue esencial. Su vasto conocimientos y constante orientación fueron pilar fundamentales para alcanzar el éxito de este proyecto.

A toda mi familia por brindarme todo el apoyo y la información necesaria para el inicio de este proyecto. Sin su respaldo incondicional, no habría sido posible llevar a cabo el desarrollo de esta iniciativa.

A todos mis compañeros de la universidad, con quienes compartí momentos de aprendizaje a lo largo de este trayecto de mi carrera profesional.

Tabla de contenidos

Contenido

UNIVERSIDAD MAYOR DE SAN SIMÓN.....	1
1. Introducción.....	1
1.1. Antecedentes	1
1.2. Justificación.....	3
1.3. Planteamiento del problema	4
1.4. Objetivo general	4
1.4.1. Objetivos específicos.....	4
2. Marco teórico	5
2.1. Ciencia de Datos: Concepto y Alcance	5
2.1.1. Componentes principales de la Ciencia de Datos:.....	5
2.1.2. Aplicaciones prácticas en el entorno empresarial:	5
2.2. Herramientas Digitales para el Registro y Análisis de Datos.....	6
2.2.1. Microsoft Excel	6
2.2.2. Power BI.....	6
2.2.3. Jupyter Notebook	7
2.3. Modelos predictivos y pronóstico de ventas	7
2.3.1. Modelos de series temporales.....	8
2.3.2. Prophet.....	8
2.3.3. Modelos de regresión lineal y múltiple	8
2.4. El Contexto de la Venta de Plantas Ornamentales.	9
3. Marco metodológico.....	10
3.1. Área de estudio.....	10
3.2. Flujograma metodológico.....	11
3.3. Fuentes de información	13
3.3.1. Fuente de información primaria	13
3.3.2. Fuente de información secundaria.....	13

3.4.	Adquirir datos.....	13
3.5.	Diseñar una base de datos	14
3.6.	Aplicar técnicas de análisis de datos con la herramienta de Jupyter notebook	15
3.7.	Evaluación y visualización de los resultados en Power BI	19
3.7.1.	Primera visualización	19
3.7.2.	Segunda visualización	20
3.7.3.	Tercera visualización.....	21
4.	Análisis de Resultados y Discusión.....	23
4.1	. Resultados de adquirir datos.....	23
4.2	. Resultados de diseñar una base de datos estructurada en Excel	23
4.3	. Resultados de aplicar técnicas de análisis de datos con la herramienta de Jupyter notebook	24
4.4	. Resultado de la evaluación y visualización de los resultados en Power BI.....	30
4.4.1.	Resultados de la Visualización 1	30
4.4.2.	Resultados de la Visualización 2	31
4.4.3.	Resultados de la Visualización 3	31
4.5.	Discusión de resultados	32
5.	Conclusiones	35
6.	Recomendaciones	36
	Referencias bibliográficas	37
	Bibliografía.....	37
	Anexos.....	39
Anexo 1.	Adquir datos	39
Anexo 2.	Código ejecutado en Jupyter notebook.....	40
Anexo 3.	Visualización de clientes con mayor compra de plantas de exterior	41
Anexo 4.	Contenido del CD.....	42
Anexo 5.	Repositorios	43

Lista de figuras

Figura 3-11Área de estudio Cercado-Cochabamba.....	10
Figura 3-2Fujograma metodológico.....	11
Figura 3-3Hoja de registro	14
Figura 3-4Importación de librerías.....	15
Figura 3-5Carga del archivo csv	15
Figura 3-6Verificacion de filas y columnas	15
Figura 3-7Verificación de datos nulos	15
Figura 3-8Verificación a detalle del dataset.....	15
Figura 3-9Análisis general del dataset	16
Figura 3-10Cantidad de gastos por departamento	16
Figura 3-11Evolución de ventas por fechas	16
Figura 3-12Verificacion de plantas preferidas para el cliente.....	17
Figura 3-13Clientes con más compras de plantas	17
Figura 3-14Clientes con mayor gasto en plantas.....	18
Figura 3-15Compra de plantas mas preferidas por destino	18
Figura 3-16Predicción de ventas	19
Figura 3-17Selecion de columnas “Nombre de cliente y precio total de compra”.....	20
Figura 3-18Selección de la columna "Nombre de cliente".....	20
Figura 3-19Seleccion de la columna "Fecha, año, mes, día y suma total de plantas (Interior y exterior)"	21
Figura 3-20Selección de la columna Destino.....	21
Figura 3-21Seleccion de la columna "Fecha de compra y total precio de compra (Año y mes)"	22
Figura 3-22Seleccion de la columna "Fecha de compra (mes y día)"	22
Figura 3-23Selección de "Fecha de compra(mes)"	22
Figura 4-1Cantidad de filas y columnas del dataset.....	24
Figura 4-2Verificacion datos nulos	25
Figura 4-3Verificación de tipo de datos	25
Figura 4-4Visualización de ventas por destino	26
Figura 4-5Visualizacion de la evolución de ventas en el tiempo	27
Figura 4-6Visualización de las preferencias de los clientes al momento de adquirir plantas	27
Figura 4-7Visualizacion de los clientes con mayor volumen de compras de plantas	28
Figura 4-8Visualizacion de 5 clientes con el mayor monto acumulado en compras.....	28
Figura 4-9Visualización de plantas de interior y exterior más adquiridas por los clientes según destino final.....	29
Figura 4-10Visualizacion de la proyección de ventas de plantas hasta fin de año.....	29
Figura 4-11Visualizacion del gasto realizado por cada cliente	30
Figura 4-12Visualizacion de las plantas mas demandadas según su detino: interior o exterior.....	31
Figura 4-13Visualizacion del total de ventas a lo largo del tiempo según fechas registradas.....	32

Figura 4-14Gráfico de porcentajes acerca de la tienda virtual	33
Figura 4-15Visualizacion de la proyección de ventas de plantas hasta fin de año.....	34

Lista de tablas

Tabla 1Planilla en Excel con datos transcritos	14
Tabla 2Transcripción de datos en Excel.....	24
Tabla 3Análisis estadístico general	26
Tabla 4Verificación de resultado según cliente.....	30
Tabla 5Resultado sobre el uso de tiendas virtuales	33

1. Introducción

En la actualidad, vivimos en una era digital donde las redes sociales se han consolidado como herramientas fundamentales tanto en la vida cotidiana como en el ámbito empresarial. Plataformas como Tik Tok y WhatsApp no solo facilitan la comunicación entre personas, sino que también se han convertido en canales clave para promocionar productos, fortalecer la relación con los clientes y concretar ventas. Según. (PACHECO-DIAZ, 2024), las redes sociales actúan como incubadoras de grandes negocios debido a la gran variedad de contenidos, productos y servicios que se consumen a través de diversas páginas de entretenimiento.

Este proyecto se desarrolla en el vivero **Paraíso de Hojas Verdes**, una tienda virtual ubicada en la ciudad de Cochabamba, Bolivia, dedicada a la venta de plantas ornamentales de interior y exterior. La empresa utiliza activamente redes sociales como Tik Tok y WhatsApp para promocionar sus productos, compartir promociones y mantener comunicación directa con sus clientes. Sin embargo, a pesar de aprovechar estas plataformas digitales, el vivero enfrenta importantes limitaciones en la gestión de sus operaciones, ya que actualmente lleva un control manual de su inventario y ventas, utilizando cuadernos y hojas sueltas. Esta práctica no solo dificulta el seguimiento eficiente de las ventas y del stock de productos, sino que también representa un riesgo ante posibles reclamos, al no contar con un respaldo digital de la información relacionada con sus clientes (SANCHEZ BARAHONA LUIS ALEXANDER, 2024).

En este contexto, resulta fundamental que los pequeños negocios, como el vivero **Paraíso de Hojas Verdes**, adopten tecnologías que optimicen su organización y faciliten una toma de decisiones más informada. Herramientas como Excel, Power BI y entornos de análisis como Jupyter Notebook ofrecen soluciones prácticas y accesibles para digitalizar procesos clave, realizar análisis de datos y generar reportes que revelen patrones en el comportamiento de ventas, identifiquen a los clientes más frecuentes y determinen los periodos de mayor demanda. A partir de estos análisis, también es posible realizar predicciones sobre la evolución futura de las ventas, lo cual permite anticiparse a la demanda y planificar de manera estratégica. La implementación de estas tecnologías representa una ventaja significativa para el crecimiento sostenible y la competitividad del negocio en el entorno digital actual.

1.1. Antecedentes

La comercialización de plantas ornamentales ha tenido un crecimiento constante en los últimos años, impulsado por el creciente interés en la jardinería doméstica, la preocupación por el bienestar emocional, y la tendencia global de incorporar espacios verdes en hogares, oficinas y entornos comerciales. Diversos estudios coinciden en que el contacto con las plantas mejora la salud mental, reduce el estrés y fomenta un entorno más saludable, lo que ha incrementado significativamente la demanda de especies ornamentales en zonas urbanas (MedicalNewsToday, 2021).

Como consecuencia, numerosos viveros y tiendas especializadas han adoptado nuevas estrategias para llegar a sus clientes, en particular mediante el uso intensivo de plataformas digitales y redes sociales. En

América Latina, se observa una migración progresiva hacia la venta en línea, donde aplicaciones como WhatsApp, Facebook, Instagram y más recientemente, Tik Tok, se han convertido en canales clave para promocionar productos, generar contenido visual atractivo y mantener un contacto directo con los consumidores (SANCHEZ BARAHONA LUIS ALEXANDER, 2024). Estas herramientas digitales han facilitado a los pequeños negocios del rubro ampliar su alcance geográfico, aumentar su visibilidad comercial y captar nuevos segmentos del mercado, especialmente entre los públicos más jóvenes y urbanos (BIMATICO, 2012).

Sin embargo, este crecimiento acelerado también ha puesto en evidencia limitaciones estructurales dentro del sector, particularmente en lo que respecta a la gestión interna del negocio. Muchos viveros siguen operando con métodos manuales para el control de stock, registro de ventas y seguimiento de clientes, lo que genera errores, pérdida de información y falta de trazabilidad (CEPAL, 1948). Ante este escenario, se vuelve cada vez más necesario incorporar herramientas tecnológicas como sistemas de inventario automatizados, hojas de cálculo organizadas y plataformas de análisis de datos como Power BI o entornos de trabajo como Anaconda que permitan tomar decisiones estratégicas basadas en información real y oportuna (Fundacion Alternativa, 2021).

Diversos estudios y experiencias han demostrado que la falta de digitalización en procesos clave como el registro de ventas, el manejo del inventario y la gestión de información del cliente representa un obstáculo serio para el desarrollo de los pequeños negocios. Esta carencia afecta directamente su crecimiento operativo y financiero, ya que limita su capacidad para tomar decisiones informadas, adaptarse a la demanda del mercado y optimizar sus recursos.

De forma complementaria, un estudio del Banco Interamericano de Desarrollo (BIMATICO, 2012), en su programa *Digitalización para MIPYMES*, demostró que los pequeños comercios que adoptaron herramientas simples de control digital como Microsoft Excel para gestionar ventas e inventario, y Power BI para el análisis de datos, lograron una reducción del 30% en pérdidas por errores de inventario y un incremento del 25% en la frecuencia de compra de clientes habituales, gracias al conocimiento más preciso de sus patrones de consumo.

En el sector específico de los viveros y negocios agrícolas, estas limitaciones también están presentes. De hecho, el programa *AL-Invest 5.0* en Bolivia documentó mejoras claras tras la implementación de herramientas digitales en viveros locales: mayor capacidad de respuesta ante la demanda, menos casos de desabastecimiento y una mejor rotación de productos, lo que se traduce en un crecimiento económico más sólido y sostenible (Félix Zenón Limachi Mamani, 2020).

En respuesta a estas carencias, cada vez más emprendedores del sector han comenzado a utilizar herramientas digitales como Excel, Power BI y entornos de análisis como la herramienta de Jupyter Notebook. Estas soluciones permiten registrar las ventas de forma ordenada, visualizar el comportamiento del negocio y realizar análisis predictivos para tomar decisiones estratégicas basadas en datos reales. Sin embargo, en muchos viveros tradicionales aún persisten serias limitaciones (OpenWebinars, 2024).

Entre las principales deficiencias que enfrentan los viveros que operan sin digitalización se encuentran:

- **Falta de trazabilidad:** sin registros digitales, no se puede saber con exactitud qué plantas se venden más, en qué fechas ni a qué tipo de cliente.

- **Errores en el control de inventario:** el conteo manual es propenso a fallas, generando desabastecimientos o excesos de stock.
- **Desorganización de la información del cliente:** no se lleva un seguimiento de quién compra, con qué frecuencia ni qué tipo de productos prefiere.
- **Pérdida de datos físicos:** los registros en papel pueden extraviarse o deteriorarse, lo que dificulta auditar el historial comercial del vivero.

Un diagnóstico del programa (Pro Mujeres, 2025) reveló que el 72% de los viveros de Cochabamba no utilizan ningún sistema digital para el control de ventas o inventario, y que el 81% lleva sus registros a mano o en libretas. Esto dificulta, por ejemplo, responder con rapidez a picos de demanda en fechas especiales como ferias o festividades, planificar compras y reposiciones, o identificar qué productos son los más rentables.

Por otro lado, el estudio de la (Fundación Alternativa, 2021), titulado *“La digitalización como motor de las microempresas rurales”*, evidenció que los viveros que migraron de registros manuales a sistemas básicos como Excel lograron:

- Reducir en un 40% los errores de inventario.
- Incrementar en un 20% la identificación de clientes frecuentes, gracias al orden en sus bases de datos.
- Aumentar hasta un 25% las ventas durante los primeros tres meses, al detectar con mayor rapidez qué productos tienen mayor demanda.

Uno de los beneficios más importantes de la digitalización es la posibilidad de realizar predicciones de ventas a futuro. Utilizando herramientas como Power BI, Prophet o ARIMA dentro de Jupyter Notebook, los viveros pueden modelar el comportamiento histórico de sus ventas y proyectar cómo evolucionarán en los próximos meses. Esto no solo permite anticiparse a los periodos de alta demanda y preparar el inventario con mayor precisión, sino también planificar campañas promocionales, ajustar precios estratégicamente y administrar mejor los recursos financieros. La predicción de ventas se convierte así en un pilar fundamental para asegurar la sostenibilidad del negocio, evitar pérdidas y aprovechar al máximo las oportunidades del mercado (ARIMA, 2024).

1.2. Justificación

La necesidad de modernizar los procesos de registro y gestión en el vivero "Paraíso de hojas verdes" surge a raíz de las dificultades que enfrenta al utilizar métodos manuales para controlar su inventario y registrar sus ventas. Esta situación no solo representa una pérdida de tiempo y recursos, sino que también limita la capacidad del negocio para analizar su rendimiento, conocer los hábitos de consumo de sus clientes y planificar su crecimiento de manera estructurada.

Implementar un sistema informático con una base de datos permitirá automatizar el control de ventas, obtener reportes detallados, y visualizar con claridad información relevante, como qué semana generó mayores ingresos, cuáles son los productos más vendidos y qué clientes realizaron más compras. Además, integrar este sistema con canales de venta como WhatsApp y Tik Tok, fortalecerá la atención al cliente y mejorará la experiencia de compra, generando confianza, fidelización y nuevas oportunidades de crecimiento para el negocio.

1.3. Planteamiento del problema

La necesidad de modernizar los procesos de registro y gestión en el vivero "Paraíso de hojas verdes" surge ante las limitaciones que implica el uso de métodos manuales para controlar el inventario y registrar las ventas. Esta situación genera pérdidas de tiempo y recursos, dificulta el análisis del rendimiento del negocio y limita la comprensión de los hábitos de consumo de los clientes, lo que impide una planificación estructurada y estratégica para su crecimiento.

La implementación de un sistema informático con una base de datos permitirá automatizar el control de ventas, generar reportes detallados y visualizar información clave, como las semanas con mayores ingresos, los productos más vendidos y los clientes más frecuentes. Además, se podrá incorporar un modelo de predicción de ventas que facilitará la toma de decisiones acertadas, con el objetivo de maximizar las ganancias y diseñar estrategias que impulsen las ventas.

Asimismo, el sistema fortalecerá la atención al cliente mediante el uso de plataformas como WhatsApp y Tik Tok, mejorando la experiencia de compra, generando confianza, fidelización y abriendo nuevas oportunidades de crecimiento para el negocio.

1.4. Objetivo general

Diseñar e implementar un sistema de registro y análisis de datos utilizando herramientas digitales como Jupyter Notebook, Excel y Power BI, que permita al vivero **Paraíso de Hojas Verdes** mejorar la gestión de sus ventas, controlar su inventario y optimizar la toma de decisiones estratégicas. Este sistema incluirá la aplicación de modelos de predicción de ventas a futuro, lo que permitirá anticipar la demanda, planificar el abastecimiento con mayor precisión y contribuir al crecimiento sostenible del negocio.

1.4.1. Objetivos específicos

- **Adquirir datos** una recolección de datos que almacene información clave.
- **Diseñar una base de datos estructurada en Excel** una base de datos estructurada que almacene información clave sobre productos, ventas y clientes
- **Aplicar técnicas de análisis de datos con la herramienta Jupyter Notebook** con el fin de identificar las semanas con mayores ventas, los clientes con mayor volumen de compra y realizar predicciones que permitan mejorar la gestión de ventas en el futuro
- **Evaluación y visualización de los resultados en Power BI** con el propósito de obtener una mayor visibilidad de las ventas y de los clientes, permitiendo así un análisis detallado que facilite la toma de decisiones estratégicas.

2. Marco teórico

2.1. Ciencia de Datos: Concepto y Alcance

La **Ciencia de Datos** es una disciplina que ha emergido con fuerza en la era digital debido al crecimiento exponencial de los datos disponibles en todos los ámbitos, desde el comercio y la salud hasta la agricultura y la industria. Se define como un campo interdisciplinario que combina estadística, informática, análisis de datos y conocimiento del dominio para extraer información útil a partir de grandes volúmenes de datos estructurados (como hojas de cálculo o bases relacionales) y no estructurados (como textos, imágenes o publicaciones en redes sociales) (Provost & Fawcett, 2013).

En esencia, la Ciencia de Datos busca transformar datos en información valiosa para la toma de decisiones, lo cual se logra mediante un ciclo que incluye **la recolección, procesamiento, análisis, visualización y modelado predictivo**. Este ciclo permite pasar del simple almacenamiento de datos a la generación de conocimiento accionable, una capacidad estratégica clave para cualquier organización.

2.1.1. Componentes principales de la Ciencia de Datos:

1. **Estadística:** permite describir y resumir los datos, además de formular inferencias.
2. **Aprendizaje automático (Machine Learning):** facilita la creación de modelos predictivos y prescriptivos a partir de patrones observados.
3. **Ingeniería de datos:** involucra la recolección, limpieza y transformación de datos en formatos útiles para su análisis.
4. **Visualización de datos:** ayuda a comunicar los hallazgos de forma clara y comprensible mediante gráficos, dashboards y reportes interactivos.
5. **Conocimiento del dominio:** permite contextualizar los datos y validar los resultados obtenidos.

2.1.2. Aplicaciones prácticas en el entorno empresarial:

En el ámbito corporativo, la Ciencia de Datos se utiliza para resolver problemas de diversa índole, como:

- Analizar el comportamiento de los clientes.
- Prever ventas y gestionar la cadena de suministro.
- Detectar fraudes o anomalías en sistemas financieros.
- Optimizar campañas de marketing.
- Automatizar procesos rutinarios mediante algoritmos inteligentes.

Sin embargo, su uso no se limita a grandes empresas con equipos tecnológicos robustos. Gracias al desarrollo de herramientas accesibles como Python, Excel avanzado, Power BI y plataformas de código abierto, la Ciencia de Datos está al alcance de emprendedores, microempresas y PYMES que deseen adoptar un enfoque de gestión más estratégico y eficiente.

2.2. Herramientas Digitales para el Registro y Análisis de Datos.

En un entorno cada vez más digitalizado, el uso de herramientas tecnológicas para la gestión de datos se vuelve crucial para los pequeños negocios. En el caso de los viveros, estas herramientas permiten optimizar procesos clave como el registro de ventas, el control del inventario, el análisis del comportamiento del cliente y la toma de decisiones estratégicas. A continuación, se describen tres herramientas fundamentales que se utilizan en este proyecto: Microsoft Excel, Power BI y Jupyter Notebook (ISDI, 2024).

2.2.1. Microsoft Excel

Microsoft Excel es una de las herramientas más conocidas y utilizadas a nivel mundial para la gestión de datos. Su interfaz amigable y su amplia gama de funcionalidades lo convierten en un recurso accesible incluso para usuarios con conocimientos básicos en informática (IBM, 2007).

En el contexto de un vivero, Excel ofrece múltiples aplicaciones prácticas:

- **Registro de ventas** por fecha, tipo de planta, cantidad y cliente.
- **Gestión del inventario**, permitiendo llevar un control detallado del stock disponible y de los productos agotados.
- **Elaboración de reportes de desempeño**, como ingresos mensuales, productos más vendidos o frecuencia de compras.
- **Creación y mantenimiento de una base de datos de clientes**, que puede incluir información de contacto, historial de compras y preferencias.

Además, Excel permite automatizar tareas mediante el uso de fórmulas, funciones, tablas dinámicas y macros. Su bajo costo (incluso gratuito en versiones en línea como Excel Online) (Excel, 2016) y su curva de aprendizaje accesible lo convierten en una opción ideal para negocios pequeños en proceso de digitalización.

2.2.2. Power BI

Power BI es una herramienta avanzada de visualización de datos que permite transformar información compleja en gráficos interactivos, paneles de control (dashboards) y reportes dinámicos. Desarrollada por Microsoft, esta plataforma se integra fácilmente con Excel, lo que permite importar datos ya organizados para su análisis avanzado (BIMATICO, 2012).

Entre los beneficios de Power BI para un vivero se incluyen:

- **Visualización de productos más vendidos** por semana, mes o temporada.
- **Identificación de clientes frecuentes** y segmentación por tipo de compra.

- **Detección de patrones de demanda** según fechas clave (festividades, promociones, fines de semana).
- **Seguimiento del rendimiento de ventas** en tiempo real y de manera visual.
- **Análisis comparativo entre diferentes periodos**, permitiendo evaluar el crecimiento o decrecimiento de las ventas.

El uso de Power BI no solo mejora la toma de decisiones, sino que también permite compartir reportes de manera eficiente con otros miembros del negocio, incluso a través de dispositivos móviles.

2.2.3. Jupyter Notebook

Jupyter Notebook es un entorno de desarrollo interactivo que permite escribir y ejecutar código en múltiples lenguajes de programación, siendo Python el más utilizado. Esta herramienta es especialmente útil para proyectos de análisis de datos, ya que combina código, visualizaciones y documentación en un mismo archivo (Coursera, 2023).

En el presente proyecto, Jupyter Notebook se utilizará con fines como:

- **Limpieza y transformación de los datos** del vivero, eliminando errores, duplicados o inconsistencias.
- **Desarrollo de modelos predictivos simples**, como la estimación de la demanda futura de ciertas plantas o productos.
- **Aplicación de análisis estadísticos**, como la distribución de ventas, tendencias por categoría o análisis de frecuencia de compra.
- **Automatización de tareas analíticas repetitivas**, reduciendo el tiempo necesario para generar reportes.

2.3. Modelos predictivos y pronóstico de ventas

El **pronóstico de ventas** constituye una herramienta estratégica fundamental en la ingeniería de sistemas aplicada a la gestión empresarial, ya que permite anticipar el comportamiento futuro del mercado, optimizar recursos y reducir incertidumbres en la toma de decisiones. Esta práctica se basa en el análisis de datos históricos de ventas mediante técnicas estadísticas y computacionales que generan modelos capaces de realizar predicciones confiables sobre la evolución de la demanda (Predik Dat Driven, 2023).

Desde un enfoque ingenieril, los modelos predictivos se apoyan en algoritmos estructurados que procesan grandes volúmenes de datos (Big Data), detectan patrones y generan proyecciones. Su implementación se realiza en entornos de análisis como Jupyter Notebook, utilizando herramientas de

procesamiento como **Python** junto con bibliotecas específicas como pandas, numpy, matplotlib, scikit-learn, statsmodels y prophet.

2.3.1. Modelos de series temporales

Estos modelos son fundamentales cuando se analiza una variable que depende del tiempo, como es el caso de las ventas. En este contexto, los más comunes son:

- **ARIMA (Autoregressive Integrated Moving Average):** permite modelar series temporales no estacionarias y es útil para datos con tendencia. Se compone de tres parámetros (p , d , q) que representan la autoregresión, la diferenciación y el promedio móvil.
- **SARIMA (Seasonal ARIMA):** una extensión de ARIMA que incluye componentes estacionales, ideal para datos que presentan comportamiento cíclico durante periodos específicos como semanas o meses.

Estos modelos se implementan habitualmente mediante la biblioteca statsmodels, que permite ajustar y validar la precisión del modelo mediante métricas como MAE, RMSE y AIC (ARIMA, 2024).

2.3.2. Prophet

Desarrollado por Facebook, **Prophet** es un modelo aditivo robusto para el pronóstico de series temporales con fuerte componente estacional y efectos de días festivos. Es especialmente útil para pequeñas empresas por su facilidad de uso, tolerancia a datos faltantes y capacidad de incorporar múltiples eventos especiales (Medium, 2025).

Prophet modela la serie temporal como una suma de tres componentes:

- **Tendencia no lineal** (con saturación logística o crecimiento lineal).
- **Estacionalidad múltiple** (diaria, semanal, anual).
- **Eventos especiales** (ferias, festividades, campañas).

La implementación se realiza con la librería prophet de Python, lo que permite generar pronósticos visuales e interpretables para usuarios no expertos.

2.3.3. Modelos de regresión lineal y múltiple

Estos modelos se enfocan en explicar las ventas como una función lineal de una o más variables independientes (por ejemplo, clima, descuentos, días festivos). A través de la regresión múltiple, se pueden analizar factores combinados que influyen en las ventas y ajustar campañas de marketing o producción con base en los resultados (Ciencia de Datos, 2016).

2.4. El Contexto de la Venta de Plantas Ornamentales.

El sector de las plantas ornamentales ha experimentado un notable crecimiento en los últimos años, impulsado por una serie de factores sociales, culturales y ambientales. La jardinería urbana y la decoración con plantas en interiores se han convertido en tendencias populares en diversos países, debido a su impacto positivo en el bienestar físico y emocional. Estudios de la (FAO, 2020) destacan que el contacto con la naturaleza y la inclusión de elementos verdes en el entorno contribuyen a reducir el estrés, aumentar la productividad y mejorar el estado de ánimo de las personas. Estas condiciones han favorecido un incremento en la demanda de plantas ornamentales tanto en espacios residenciales como comerciales.

En el contexto latinoamericano, este auge ha abierto nuevas oportunidades para pequeños emprendedores, viveros familiares y negocios independientes, que han encontrado en las redes sociales y plataformas digitales un canal accesible para comercializar sus productos. Aplicaciones como WhatsApp, Facebook, Instagram y Tik Tok se han convertido en vitrinas virtuales donde los viveros no solo promocionan sus plantas, sino también comparten consejos, generan contenido educativo y se conectan directamente con sus clientes (El Foro Economico Mundial, 2015).

A pesar de este avance en la promoción y difusión digital, muchas de estas microempresas siguen operando con métodos manuales para el control de inventario, registro de ventas y seguimiento de clientes. Esta falta de digitalización representa una importante limitación para el crecimiento y sostenibilidad del negocio. Según un informe de la Comisión Económica para América Latina y el Caribe (CEPAL, 1948), más del 70% de los pequeños negocios agrícolas en la región no cuentan con sistemas informáticos básicos para la gestión operativa, lo que genera errores frecuentes, pérdida de información y dificultades para tomar decisiones estratégicas basadas en datos reales.

La digitalización en viveros no solo implica modernizar la forma en que se almacenan los datos (ScienceDirect, 2023), sino también permite mejorar la trazabilidad de los productos, planificar compras de manera más eficiente, identificar a los clientes frecuentes, conocer los patrones de demanda por temporada y elaborar reportes que faciliten la toma de decisiones. Un estudio del Banco Interamericano de Desarrollo (BIMATICO, 2012) indica que los viveros que adoptaron herramientas como Excel para la organización de datos y Power BI para el análisis de ventas reportaron una mejora del 30% en la rotación del inventario y un aumento del 25% en la fidelización de clientes.

Además, el contexto actual de competencia comercial exige a los pequeños negocios diferenciarse no solo por la calidad de sus productos, sino también por la eficiencia de su gestión. En este sentido, el uso de herramientas tecnológicas accesibles y de bajo costo puede marcar la diferencia entre un vivero que sobrevive y uno que crece sostenidamente. Al adoptar tecnologías como hojas de cálculo organizadas, dashboards interactivos o análisis predictivos, los viveros tienen la posibilidad de transformar su información dispersa en inteligencia de negocio (fastercapital logo, 2025).

3. Marco metodológico

En el presente estudio se detalla la metodología empleada para abordar los objetivos del vivero **Paraíso de Hojas Verdes** mediante el uso de tecnologías digitales y análisis predictivos. El proceso inició con la recolección de datos, seguida del diseño de una base de datos estructurada en Excel. Posteriormente, se realizó el tratamiento de datos nulos y se procedió al análisis utilizando herramientas digitales como Jupyter Notebook. Finalmente, se llevó a cabo la evaluación y visualización de los datos relacionados con clientes y ventas mediante Power BI, con el objetivo de optimizar la toma de decisiones

3.1. Área de estudio

El área de estudio se centra en el municipio de Cercado zona sur se encuentra en el departamento de Cochabamba-Bolivia.



Figura 3-11 Área de estudio Cercado-Cochabamba

Fuente: Google Maps, (2025)

3.2. Flujograma metodológico

A lo largo del desarrollo del proyecto se logró implementar la metodología CRISP-DM, estructurando el trabajo en torno a cuatro objetivos principales. Esta organización permitió abordar el problema de manera integral, desde su comprensión inicial hasta la propuesta de soluciones basadas en el análisis de datos.

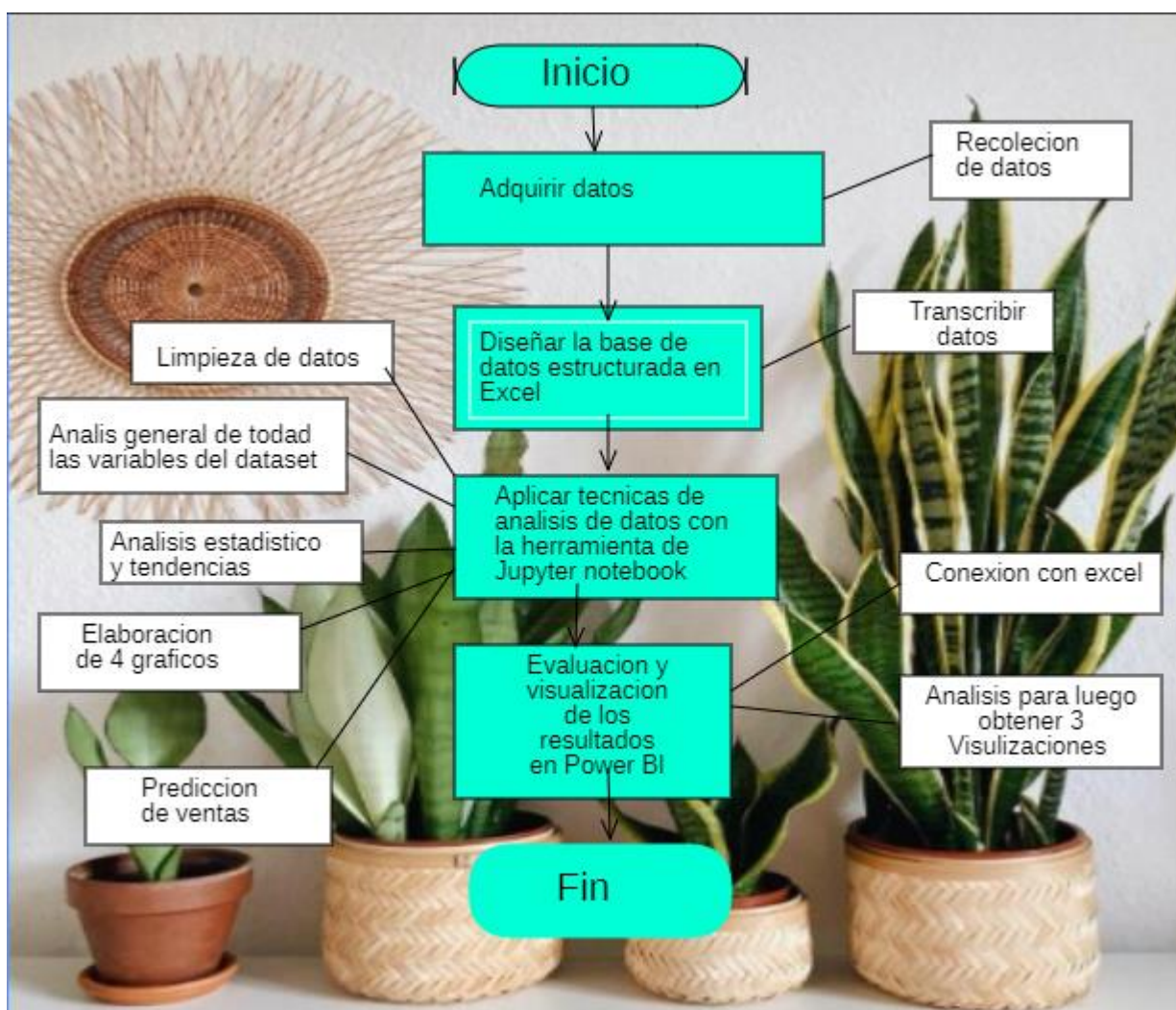


Figura 3-2 Flujograma metodológico

Fuente: Elaboración propia (2025)

Adquirir datos: Se realizó la recolección de todos los datos de vivero "Paraíso de hojas verdes"

Diseñar la base de datos estructurada en Excel: Se elaboró previamente una base de datos en Excel con una estructura clara y ordenada, con el objetivo de facilitar la organización y el análisis de la información recolectada.

Transcribir los datos: Una vez diseñada la base de datos, se procedió a transcribir todos los datos registrados en hojas sueltas. La información fue ingresada cuidadosamente en la base de datos, asegurando su correcta organización y evitando la pérdida de datos durante el proceso.

Aplicar técnicas de análisis de datos con la herramienta Jupyter notebook: Para llevar a cabo el análisis, en primer lugar se realizó la carga del archivo en formato CSV dentro del entorno Jupyter Notebook.

Limpeza de datos: Una vez cargado el archivo, se procedió al tratamiento de valores nulos. Tras la revisión, se verificó que no existían datos faltantes en el dataset.

Análisis general de dataset: Se efectuó un análisis exploratorio para identificar las variables categóricas y comprender mejor la estructura y el contenido del conjunto de datos. Este paso permitió obtener información clave sobre la naturaleza de las variables y su distribución.

Análisis estadístico y tendencia: Seguidamente, se realizó un análisis estadístico descriptivo del dataset, calculando métricas como la media, el valor mínimo y máximo, así como los percentiles (25%, 50% y 75%). Esto permitió detectar posibles tendencias y patrones útiles para la toma de decisiones.

Realización de gráficos con el lenguaje de programación Python: Se elaboraron seis gráficos utilizando Python, los cuales resultaron fundamentales para el análisis del comportamiento de los clientes y las tendencias de ventas. Estas representaciones visuales permitieron identificar patrones, facilitar la interpretación de los datos y respaldar la toma de decisiones basada en evidencia.

Predicción de ventas: Se logró generar un gráfico de predicción de ventas que permitió visualizar la proyección del comportamiento de las ventas hasta fin de año, facilitando la planificación estratégica.

Evaluación y visualización de los resultados en Power BI: En esta etapa se evaluaron diferentes tipos de visualizaciones, identificando cuáles representaban mejor los datos y ofrecían mayor claridad para el análisis.

Conexión con Excel: Se estableció con éxito la conexión entre Power BI y el archivo de Excel, lo que permitió trabajar directamente con los datos estructurados previamente.

Análisis para luego obtener 3 visualizaciones: A partir del análisis de los datos, se desarrollaron tres visualizaciones clave que permitieron identificar:

1. Qué clientes realizaron mayores compras.
2. Qué departamento generó más ventas.
3. En qué mes se registró el mayor volumen de ventas.

3.3. Fuentes de información

3.3.1. Fuente de información primaria

Son datos propios, obtenidos por el autor del proyecto. En este caso, se incluye la observación directa del funcionamiento diario del vivero, el registro manual de ventas e inventario, y el análisis personal de las necesidades del negocio. Dado que el autor es también el propietario del vivero, se tiene un conocimiento profundo y detallado del contexto operativo, lo que ha permitido identificar con precisión los problemas existentes y las áreas susceptibles de mejora

3.3.2. Fuente de información secundaria

Además de la experiencia directa del autor, se realizaron entrevistas a miembros de su familia, quienes también forman parte activa en la administración del vivero. Estas entrevistas permitieron recopilar diferentes perspectivas sobre los principales desafíos que enfrenta el negocio, especialmente en lo referente al control de ventas, la organización del inventario y la atención al cliente.

3.4. Adquirir datos

Se llevó a cabo un proceso de recolección de datos a partir de los registros existentes en cuadernos y hojas sueltas, que documentaba las ventas realizadas por el vivero. Además, se recopilaron todas las facturas de envío emitidas por diferentes empresas transportadoras, con el fin de contar con un historial completo de las transacciones y entregas efectuadas. Esta información sirvió como base para estructurar una base de datos más organizada y confiable.

Fecha de compra	Nombre de Cliente	Celular	Planta de Interior	Planta de Exterior	Cantidad de plantas Compradas	Total precio de compra	Transportadora	Destino	Costo de envío
13/03/2025	Coro Vedia Maria Fernanda	78742439	1	1	2	50	Bolivar	Potosi	15
13/03/2025	Quispe Veronica	76985321	7	1	9	800	Illimani	Illimani	50
13/03/2025	Cordova Mercedes	72451845	8	3	11	346	Bolivar	Sucre	20
13/03/2025	Yovana Mollo	65372345	4	2	6	234	Omar	Uyuni	50
13/03/2025	Mancilla Keila	76542314	4	3	7	45	Lupjansa	La Paz	20
13/03/2025	Quispe Mamani Adriana	65481544	6	1	7	90	Lupjansa	El Alto	20
13/03/2025	Condori Huanca Alejandro	71123429	3	2	5	70	Lupjansa	La Paz	15
13/03/2025	De los Angeles Flores Maria	71123429	2	4	6	100	Lupjansa	La Paz	20
13/03/2025	Porco Quispe Hans	64982344	1	1	2	72	Lupjansa	El Alto	20
13/03/2025	Fernandez Joselin	69031081	1	1	2	50	Lupjansa	Santa Cruz	20
13/03/2025	Hinojosa Sandra	73900043	2	5	7	120	Lupjansa	La Paz	20
13/03/2025	Choque Maria	62349990	6	2	8	130	Lupjansa	La Paz	20

Figura 3-3Hoja de registro

Fuente: Elaboración propia (2025)

3.5. Diseñar una base de datos

Con base en toda la información y los datos recopilados, se procedió a transcribir y organizar los registros en hojas de cálculo de Excel, con el objetivo de estructurar la información de manera clara, ordenada y accesible para su posterior análisis.

Tabla 1Planilla en Excel con datos transcritos

	B	C	D	E	F	G	H	I	J
1	Nombre de cliente	Celular	Planta de interior	Planta de exterior	Cantidad de plantas compradas	Total precio de compra	Transportadora de envío	Destino	Costo de envío
2	Orellana Pao	73731155	7	2	9	300	trans Lupjansa	La Paz	50
3	Coro Vedia Maria Fernanda	78712439	1	1	2	50	trans Bolivar	Potosi	15
4	Quispe Veronica	76985321	7	1	8	800	Trans Illimani	Illimani	50
5	Cordova Mercedes	72451845	8	3	11	346	tran Bolivar	Sucre	20
6	Yovana Mollo	65372345	4	2	6	234	trans omar	Uyuni	50
7	Mancilla Keila	76542314	4	3	7	45	trans Lupjansa	La Paz	15
8	Quispe Mamani Adriana	65481544	6	1	7	90	trans Lupjansa	El Alto	20
9	Condori Huanca Alejandro	71123429	3	2	5	70	trans Lupjansa	La Paz	15
10	De Los angeles Flores Maria	67088942	2	4	6	100	trans Lupjansa	La Paz	20
11	Porco Quispe Hans	64982344	1	1	2	72	trans Lupjansa	El Alto	20
12	Fernandez Joselin	69031081	1	1	2	50	trans Lupjansa	Santa Cruz	15
13	Hinojosa Rodriguez Sandra	73900043	2	5	7	120	trans Lupjansa	La Paz	20
14	Choque Aguilar Maria	62349990	6	2	8	130	trans Lupjansa	La Paz	20
15	Huasco Giovana	70631065	7	3	10	170	trans Lupjansa	La Paz	45
16	Torrez Corina	65893333	5	1	6	95	trans Lupjansa	La Paz	20
17	Condori Alejandra	71000098	4	1	5	180	trans Lupjansa	La Paz	20
18	Sanjines Eva	75291269	10	5	15	380	trans Lupjansa	El Alto	40
19	Quispe Pablo	73460086	6	2	8	230	trans Lupjansa	El Alto	20
20	Mollo Magali	67483455	4	1	5	180	trans Lupjansa	Oruro	20
21	Laura Eliana	72265723	5	2	7	150	trans Lupjansa	El Alto	20

Fuente: Elaboración propia (2025)

3.6. Aplicar técnicas de análisis de datos con la herramienta de Jupyter notebook

Realizamos las importaciones de las librerías

```
import numpy as np
import pandas as pd
import os
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Figura 3-4 Importación de librerías

Fuente: Elaboración propia (2025)

Se realizó la carga del archivo csv.

```
datos = pd.read_csv('PARAISO_DE_HOJAS_VERDES_1.csv', sep=';')
datos
```

Figura 3-5 Carga del archivo csv

Fuente: Elaboración propia (2025)

Se verificó cuantas columnas y filas existen en el archivo csv.

```
datos.shape
```

Figura 3-6 Verificación de filas y columnas

Fuente: Elaboración propia (2025)

Una vez identificadas y comprendidas las variables categóricas, se procedió a realizar un análisis más detallado, comenzando con el tratamiento de los valores nulos para garantizar la calidad y consistencia de los datos.

```
datos.isnull().sum()
```

Figura 3-7 Verificación de datos nulos

Fuente: Elaboración propia (2025)

Se verificó a detalle la información del dataset

```
datos.info()
```

Figura 3-8 Verificación a detalle del dataset

Fuente: Elaboración propia (2025)

Se realizó un análisis general de todas las variables del dataset.

```
datos.describe()
```

Figura 3-9 Análisis general del dataset

Fuente: Elaboración propia (2025)

Posteriormente se realizó 6 gráficos para obtener resultados y conclusiones.

Gráfico 1 Muestra el precio total de precio por cada departamento.

```
# Agrupar los datos
df_ventas_por_destino = datos.groupby('Destino')['Total precio de compra'].sum().reset_index()

# Crear gráfico
plt.figure(figsize=(10,6))
grafico = sns.barplot(data=df_ventas_por_destino, x='Destino', y='Total precio de compra', palette='viridis')

# Rotar etiquetas del eje X
plt.xticks(rotation=45)

# Título y etiquetas con "Bs" en lugar de "USD"
plt.title('Total de ventas por departamento de destino')
plt.ylabel('Total precio de compra (Bs)')
plt.xlabel('Destino')

# Añadir etiquetas numéricas encima de cada barra con "Bs"
for index, row in df_ventas_por_destino.iterrows():
    grafico.text(index, row['Total precio de compra'] + 1, f"Bs {row['Total precio de compra']:.2f}",
                ha='center', va='bottom', fontsize=10)

# Ajustar Layout
plt.tight_layout()
plt.show()
```

Figura 3-10 Cantidad de gastos por departamento

Fuente: Elaboración propia (2025)

Gráfico 2 Ventas durante el periodo comprendido.

```
datos['Fecha de compra'] = pd.to_datetime(datos['Fecha de compra'], dayfirst=True, errors='coerce')

datos['Fecha de compra'] = pd.to_datetime(datos['Fecha de compra'])

# Agrupar por fecha
ventas_diarias = datos.groupby('Fecha de compra')['Total precio de compra'].sum().reset_index()

# Graficar línea de tiempo
plt.figure(figsize=(14, 6))
sns.lineplot(data=ventas_diarias, x='Fecha de compra', y='Total precio de compra', marker='o')
plt.title('Evolución del Total de Ventas en el Tiempo', fontsize=16)
plt.xlabel('Fecha de compra')
plt.ylabel('Total vendido (USD)')
plt.grid(True)
plt.tight_layout()
plt.show()
```

Figura 3-11 Evolución de ventas por fechas

Fuente: Elaboración propia (2025)

Gráfico 3 Ventas con más preferencia

```
# Crear una nueva columna para indicar qué tipo de planta compró más cada cliente
datos['Tipo de planta preferida'] = datos.apply(
    lambda row: 'Interior' if row['Planta de interior'] > row['Planta de exterior']
    else ('Exterior' if row['Planta de exterior'] > row['Planta de interior'] else 'Igual'),
    axis=1
)

# Contar cuántos clientes prefieren cada tipo
preferencias = datos['Tipo de planta preferida'].value_counts().reset_index()
preferencias.columns = ['Tipo de planta', 'Cantidad de clientes']

# Calcular el total para convertir a porcentaje
total_clientes = preferencias['Cantidad de clientes'].sum()
preferencias['Porcentaje'] = (preferencias['Cantidad de clientes'] / total_clientes) * 100

# Graficar
plt.figure(figsize=(8,6))
grafico = sns.barplot(data=preferencias, x='Tipo de planta', y='Cantidad de clientes', palette='Set3')

# Agregar etiquetas con porcentajes en cada barra
for i, row in preferencias.iterrows():
    grafico.text(i, row['Cantidad de clientes'] + 0.5, f"{row['Porcentaje']:.1f}%", ha='center', fontsize=11)

plt.title('Preferencia de clientes: ¿Plantas de interior o exterior?', fontsize=14)
plt.xlabel('Tipo de planta comprada en mayor cantidad', fontsize=12)
plt.ylabel('Cantidad de clientes', fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Figura 3-12 Verificación de plantas preferidas para el cliente

Fuente: Elaboración propia (2025)

Gráfico 4 Clientes con mayor compra de plantas

```
# Agrupar y ordenar por cantidad total de plantas compradas
top_clientes = (datos.groupby('Nombre de cliente')['Cantidad de plantas compradas']
                .sum()
                .sort_values(ascending=False)
                .head(5)
                .reset_index())

# Crear el gráfico de barras
plt.figure(figsize=(10,6))
ax = sns.barplot(data=top_clientes, x='Nombre de cliente', y='Cantidad de plantas compradas', palette='Set2')

# Agregar etiquetas con la cantidad exacta encima de cada barra
for i, row in top_clientes.iterrows():
    ax.text(i, row['Cantidad de plantas compradas'] + 0.5, int(row['Cantidad de plantas compradas']),
           ha='center', va='bottom')

# Ajustes de presentación
plt.xticks(rotation=45, ha='right')
plt.title('Top 5 Clientes por Cantidad de Plantas Compradas')
plt.xlabel('Nombre del Cliente')
plt.ylabel('Cantidad de Plantas Compradas')
plt.tight_layout()
plt.show()
```

Figura 3-13 Clientes con más compras de plantas

Fuente: Elaboración propia (2025)

Gráfico 5 Clientes con mayor gasto en compra de plantas

```
# Agrupar y ordenar por total de precio de compra
top_clientes_precio = (datos.groupby('Nombre de cliente')['Total precio de compra']
                        .sum()
                        .sort_values(ascending=False)
                        .head(5)
                        .reset_index())

# Crear el gráfico de barras
plt.figure(figsize=(10,6))
ax = sns.barplot(data=top_clientes_precio, x='Nombre de cliente', y='Total precio de compra', palette='viridis')

# Etiquetas exactas del total en cada barra con "Bs" en lugar de "USD"
for i, row in top_clientes_precio.iterrows():
    ax.text(i, row['Total precio de compra'] + 1, f"Bs {row['Total precio de compra']:.2f}",
            ha='center', va='bottom', fontsize=10)

# Ajustes visuales
plt.xticks(rotation=45, ha='right')
plt.title('Top 5 Clientes por Total de Compra')
plt.xlabel('Nombre del Cliente')
plt.ylabel('Total Precio de Compra (Bs)')
plt.tight_layout()
plt.show()
```

Figura 3-14 Clientes con mayor gasto en plantas

Fuente: Elaboración propia (2025)

Gráfico 6 Cantidad de plantas compradas por departamento

```
# Analistas
datos['Planta de interior'] = pd.to_numeric(datos['Planta de interior'], errors='coerce')
datos['Planta de exterior'] = pd.to_numeric(datos['Planta de exterior'], errors='coerce')

# Agrupar por destino y sumar la cantidad de plantas interiores y exteriores
preferencias = datos.groupby('Destino')[['Planta de interior', 'Planta de exterior']].sum().reset_index()

# Crear gráfico de barras agrupadas
plt.figure(figsize=(12,6))
bar_width = 0.4
x = range(len(preferencias))

# Barras de plantas de interior
bars1 = plt.bar(x, preferencias['Planta de interior'], width=bar_width, label='Planta de Interior', color='seagreen')

# Barras de plantas de exterior
bars2 = plt.bar([p + bar_width for p in x], preferencias['Planta de exterior'], width=bar_width, label='Planta de Exterior', color='sandybrown')

# Etiquetas de cantidad exacta encima de cada barra
for i, value in enumerate(preferencias['Planta de interior']):
    plt.text(i, value + 0.5, int(value), ha='center', va='bottom', fontsize=9)

for i, value in enumerate(preferencias['Planta de exterior']):
    plt.text(i + bar_width, value + 0.5, int(value), ha='center', va='bottom', fontsize=9)

# Etiquetas
plt.xlabel('Destino', fontsize=12)
plt.ylabel('Cantidad de Plantas Compradas', fontsize=12)
plt.title('Preferencia de Plantas por Destino', fontsize=14)
plt.xticks([p + bar_width/2 for p in x], preferencias['Destino'], rotation=45, ha='right')
plt.legend()
plt.tight_layout()
plt.grid(axis='y', linestyle='--', alpha=0.6)
plt.show()
```

Figura 3-15 Compra de plantas mas preferidas por destino

Fuente: Elaboración propia (2025)

Gráfico 7 Predicción hasta fecha de 31 de diciembre 2025

```
pip install prophet
```

```
from prophet import Prophet
import pandas as pd
import matplotlib.pyplot as plt

# Asegurar formato correcto para Prophet
df = ventas_diarias.rename(columns={'Fecha de compra': 'ds', 'Total precio de compra': 'y'})

# Estimar un valor tope razonable para las ventas (capacidad máxima esperada)
capacidad_maxima = df['y'].max() * 1.5 # Puedes ajustar este valor según tu criterio
df['cap'] = capacidad_maxima

# Crear el modelo con crecimiento logístico
modelo = Prophet(growth='logistic', daily_seasonality=True)
modelo.fit(df)

# Crear fechas futuras
fecha_final = pd.to_datetime('2025-12-31')
dias_a_predecir = (fecha_final - df['ds'].max()).days
futuro = modelo.make_future_dataframe(periods=dias_a_predecir)

# Asignar también la capacidad a los datos futuros
futuro['cap'] = capacidad_maxima

# Realizar la predicción
prediccion = modelo.predict(futuro)

# Graficar la predicción
modelo.plot(prediccion)
plt.title('Predicción de la Evolución de Ventas (con crecimiento proyectado) hasta el 31 de diciembre de 2025')
plt.xlabel('Fecha')
plt.ylabel('Total vendido (Bs)')
plt.grid(True)
plt.show()
```

Figura 3-16 Predicción de ventas

Fuente: Elaboración propia (2025)

3.7. Evaluación y visualización de los resultados en Power BI

Se realizó la carga de del archivo Excel a la herramienta Power BI

3.7.1. Primera visualización

Para la visualización en el gráfico, se seleccionaron las columnas "**Nombre de Cliente**" y "**Total Precio de Compra**" con el fin de analizar el monto total invertido por cada comprador. Esta representación permite identificar fácilmente a los clientes que han realizado mayores inversiones en el vivero, facilitando la segmentación del mercado, la evaluación del valor de cada cliente y la planificación de estrategias comerciales personalizadas, como promociones exclusivas o programas de fidelización.

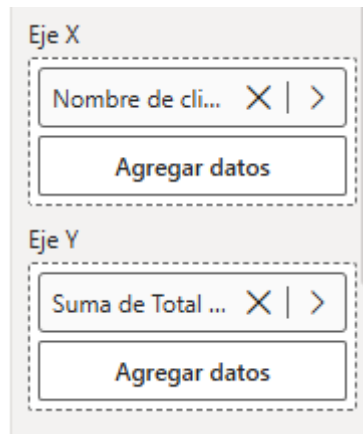


Figura 3-17 Selección de columnas “Nombre de cliente y precio total de compra”

Fuente: Elaboración propia (2025)

Se seleccionó la columna “Nombre de Cliente” con el objetivo de generar una visualización específica que muestre únicamente los nombres de las personas que realizaron compras. Esta representación permite identificar con mayor claridad quiénes son los clientes registrados en la base de datos, lo cual es útil para analizar el comportamiento individual, detectar patrones de compra frecuentes, y fortalecer la gestión de relaciones con los clientes mediante un seguimiento personalizado.

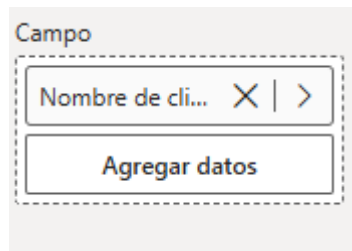


Figura 3-18 Selección de la columna "Nombre de cliente"

Fuente: Elaboración propia (2025)

3.7.2. Segunda visualización

Para la visualización en la tabla y en el gráfico se seleccionó la columna 'Fecha de compra', desglosándola en año, mes y día, junto con las columnas 'Plantas de interior' y 'Plantas de exterior', con el objetivo de analizar las tendencias de compra a lo largo del tiempo y por tipo de planta.



Figura 3-19 Selección de la columna "Fecha, año, mes, día y suma total de plantas (Interior y exterior)"

Fuente: Elaboración propia (2025)

Se seleccionó la columna 'Destino' para visualizar cuántas plantas de interior y cuántas plantas de exterior fueron compradas en cada mes por ubicación."

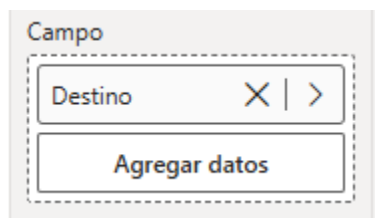


Figura 3-20 Selección de la columna Destino

Fuente: Elaboración propia (2025)

3.7.3. Tercera visualización

Para la visualización en la tabla, se seleccionó la columna 'Fecha de compra', desglosada en año y mes, junto con la columna 'Total precio de compra', con el objetivo de analizar el monto total de compras realizadas por mes

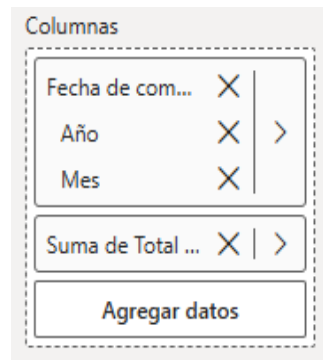


Figura 3-21 Selección de la columna "Fecha de compra y total precio de compra (Año y mes)"

Fuente: Elaboración propia (2025)

Para el gráfico, se seleccionó la columna 'Fecha de compra', desglosada en mes y día, junto con la columna 'Total precio de compra', con el fin de visualizar la variación diaria del monto total de compras a lo largo del mes

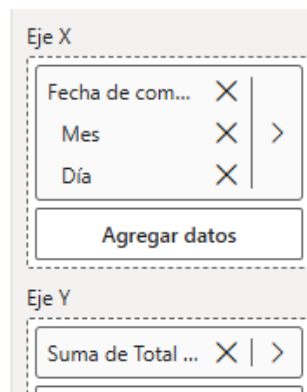


Figura 3-22 Selección de la columna "Fecha de compra (mes y día)"

Fuente: Elaboración propia (2025)

Para el segmento de datos, se seleccionó la columna 'Fecha de compra' con el objetivo de permitir la filtración dinámica mediante un control de calendario o botones, lo que facilita visualizar el total del precio de compra según el mes o la fecha específica que se desee analizar

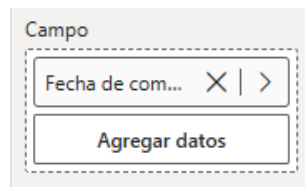


Figura 3-23 Selección de "Fecha de compra(mes)"

Fuente: Elaboración propia (2025)

4. Análisis de Resultados y Discusión

4.1. Resultados de adquirir datos

A pesar de que el vivero "**Jardín de Hojas Verdes**" no contaba con una gran cantidad de datos ni con un historial extenso de operaciones, ya que comenzó a realizar sus ventas a través de plataformas digitales como Tik Tok y WhatsApp a partir del **8 de enero de 2025**, se logró obtener toda la información necesaria para el desarrollo del proyecto.

Durante el proceso de recolección de datos, se identificaron y organizaron diversos registros relacionados con las ventas, los productos ofrecidos, la interacción con los clientes y las estrategias utilizadas en redes sociales para la promoción de las plantas ornamentales. Aunque el volumen de datos inicial era limitado, la información obtenida resultó ser **muy valiosa y relevante** para cumplir con los objetivos del proyecto.

Gracias al análisis de estos datos, fue posible comprender el comportamiento de los clientes, identificar las semanas de mayor demanda, los productos más solicitados y los clientes con mayor volumen de compra. Esta información sirvió como base para **diseñar una solución digital personalizada** que permite al vivero llevar un mejor control de su inventario, registrar las ventas de manera estructurada y tomar decisiones estratégicas fundamentadas en evidencia (ver Figura 3-19).

4.2. Resultados de diseñar una base de datos estructurada en Excel

Con el objetivo de contar con una base de datos organizada y funcional, fue necesario realizar un proceso minucioso de transcripción y digitalización de la información disponible. Muchos de los registros del vivero "**Jardín de Hojas Verdes**" se encontraban dispersos en hojas sueltas, cuadernos y apuntes físicos, lo que dificultaba su análisis y uso eficiente.

Por ello, se procedió a ingresar toda la información relevante en una hoja de cálculo de Microsoft Excel, priorizando el orden cronológico de las ventas, es decir, organizando los datos por fecha de compra. Esta estructura permitió no solo tener una visión más clara de las actividades comerciales del vivero, sino también facilitar su posterior análisis.

El archivo de Excel se convirtió en la fuente principal de datos para el desarrollo del proyecto, ya que gracias a su formato digital estructurado fue posible cargarlo en el entorno de Jupyter Notebook, donde se llevaron a cabo diversos procesos de análisis y visualización utilizando herramientas como Python, pandas y matplotlib. Sin esta transformación inicial de los datos físicos a un formato digital adecuado, habría sido imposible realizar los análisis necesarios ni aplicar herramientas tecnológicas avanzadas.

Tabla 2 Transcripción de datos en Excel

	Fecha de compra	Nombre de cliente	Celular	Planta de interior	Planta de exterior	Cantidad de plantas compradas	Total precio de compra	Transportadora de envío	Destino	Costo de envío
0	8/1/2025	Orellana Pao	73731155	7	2	9	300	trans Lupjansa	La Paz	50
1	13/1/2025	Coro Vedia Maria Fernanda	78712439	1	1	2	50	trans Bolivar	Potosi	15
2	13/1/2025	Quispe Veronica	76985321	7	1	8	800	Trans Illimani	Illimani	50
3	13/1/2025	Cordova Mercedes	72451845	8	3	11	346	tran Bolivar	Sucre	20
4	13/1/2025	Yovana Mollo	65372345	4	2	6	234	trans omar	Uyuni	50
...
111	23/4/2025	Torrez Claudia	72832131	6	6	12	300	trans Lupjansa	La Paz	25
112	23/4/2025	Apaza Belen	63874468	9	2	11	350	trans Lupjansa	Oruro	25
113	23/4/2025	Rodriguez Imana	71559280	3	3	6	90	trans Lupjansa	La Paz	20
114	23/4/2025	Machicado Javier	65874890	8	5	13	190	trans Lupjansa	La Paz	25
115	23/4/2025	Huanca Alex	63649367	7	3	10	100	trans Lupjansa	Oruro	25

Fuente: Elaboración propia (2025)

4.3. Resultados de aplicar técnicas de análisis de datos con la herramienta de Jupyter notebook

Se procedió a realizar la carga de los datos en Excel de manera exitosa, garantizando que toda la información recopilada estuviera correctamente organizada, estructurada y lista para su posterior análisis. Esta etapa fue clave para asegurar la integridad y utilidad de los datos, permitiendo su manipulación eficiente en herramientas tecnológicas como Jupyter Notebook.

Otro de los resultados obtenidos durante el proceso de digitalización y carga de los datos fue que el conjunto de datos final contó con 116 filas y 10 columnas, lo cual resultó suficiente para llevar a cabo diversos análisis estadísticos y exploratorios. A pesar de no tratarse de una base de datos extensa, la información disponible fue lo bastante representativa para identificar patrones de comportamiento, evaluar el desempeño de las ventas y extraer conclusiones relevantes que contribuyeron significativamente al desarrollo del proyecto

(116, 10)

Figura 4-1 Cantidad de filas y columnas del dataset

Fuente: Elaboración propia (2025)

También se realizó un tratamiento de datos nulos, pero no se encontró ninguno el cual favorece

```

Fecha de compra          0
Nombre de cliente        0
Celular                  0
Planta de interior       0
Planta de exterior       0
Cantidad de plantas compradas 0
Total precio de compra    0
Transportadora de envío  0
Destino                  0
Costo de envío           0
dtype: int64

```

Figura 4-2 Verificación de datos nulos

Fuente: Elaboración propia (2025)

A continuación, se verificaron los tipos de datos presentes en el dataset para asegurar su correcta interpretación y procesamiento

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 116 entries, 0 to 115
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Fecha de compra                       116 non-null   object
1   Nombre de cliente                     116 non-null   object
2   Celular                               116 non-null   int64
3   Planta de interior                    116 non-null   int64
4   Planta de exterior                    116 non-null   int64
5   Cantidad de plantas compradas         116 non-null   int64
6   Total precio de compra                 116 non-null   int64
7   Transportadora de envío                116 non-null   object
8   Destino                               116 non-null   object
9   Costo de envío                        116 non-null   int64
dtypes: int64(6), object(4)
memory usage: 9.2+ KB

```

Figura 4-3 Verificación de tipo de datos

Fuente: Elaboración propia (2025)

Posteriormente, se realizó un análisis estadístico general del dataset con el fin de obtener una visión global de las principales características de los datos

Tabla 3 Análisis estadístico general

	Celular	Planta de interior	Planta de exterior	Cantidad de plantas compradas	Total precio de compra	Costo de envío
count	1.160000e+02	116.000000	116.000000	116.000000	116.000000	116.000000
mean	7.165612e+07	4.905172	2.784483	7.689655	135.051724	24.655172
std	4.886400e+06	2.268800	1.826406	3.071365	91.319396	8.615561
min	6.232752e+07	1.000000	1.000000	2.000000	35.000000	15.000000
25%	6.744674e+07	3.000000	1.000000	5.000000	83.750000	20.000000
50%	7.283213e+07	5.000000	2.000000	8.000000	115.000000	20.000000
75%	7.549663e+07	7.000000	3.250000	10.000000	170.000000	25.000000
max	7.959006e+07	10.000000	8.000000	16.000000	800.000000	60.000000

Fuente: Elaboración propia (2025)

Resultados del gráfico 1

El gráfico muestra el total del precio de compra por cada departamento o región, reflejando la cantidad de dinero invertido en la adquisición de plantas. Se observa que el departamento de La Paz registra el mayor gasto, con un total de 6.903 Bs, mientras que el menor gasto corresponde al departamento de Trinidad, con solo 120 Bs

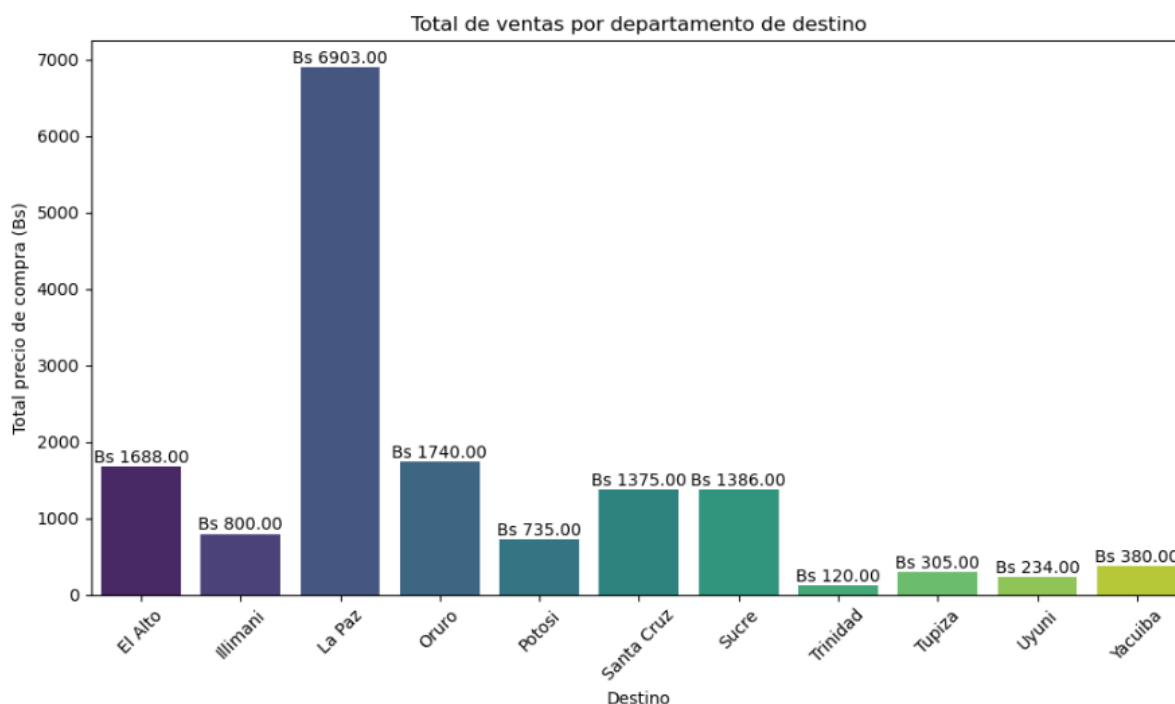


Figura 4-4 Visualización de ventas por destino

Fuente: Elaboración propia (2025)

Resultados del gráfico 2

Este gráfico permite observar el comportamiento de las ventas durante el periodo comprendido entre el 8 de enero y el 23 de abril. Se destaca que el mayor volumen de ventas se registró en enero. Posteriormente, las ventas fueron disminuyendo **progresivamente, alcanzando su punto** más bajo el 4 de abril. Sin embargo, el 23 de abril se evidenció un repunte significativo, lo que muestra una recuperación en la actividad comercial del negocio de plantas hasta la fecha

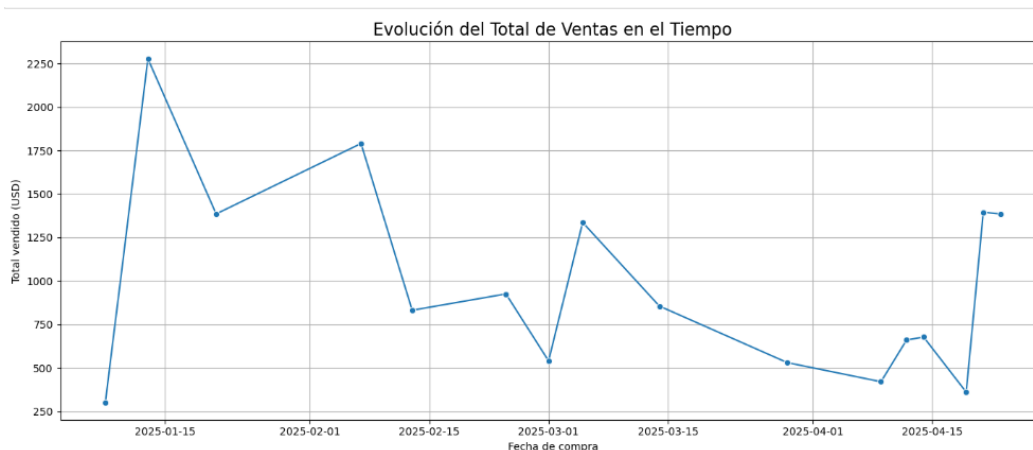


Figura 4-5 Visualización de la evolución de ventas en el tiempo

Fuente: Elaboración propia (2025)

Resultados del gráfico 3

Este gráfico muestra el porcentaje de clientes según su preferencia de compra: plantas de interior, plantas de exterior o ambas. Se observa claramente que el 66,4 % de los clientes prefieren comprar plantas de interior, mientras que el 13,8 % opta por plantas de exterior. Además, un 19,8 % de los clientes prefiere adquirir ambos tipos de plantas.

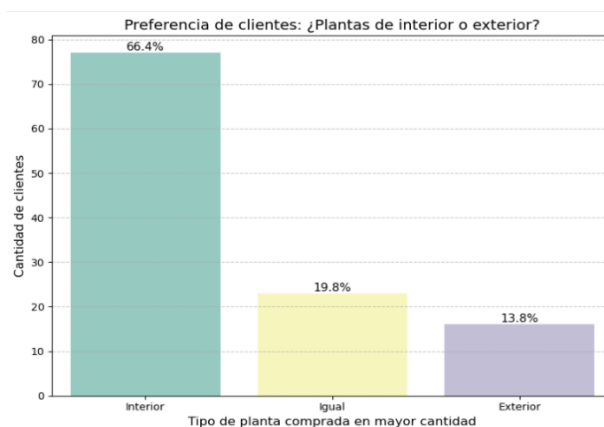


Figura 4-6 Visualización de las preferencias de los clientes al momento de adquirir plantas

Fuente: Elaboración propia (2025)

Resultados del gráfico 4

Este gráfico permite evidenciar a los cinco mejores clientes con mayor cantidad de compras realizadas entre el 8 de enero y el 23 de abril. La cliente con la mayor cantidad de plantas adquiridas es Herrera Aydee, con un total de 37 plantas. Por otro lado, entre estos cinco clientes, quien registró la menor cantidad de compras fue Eva Sanciones, con 25 plantas en total

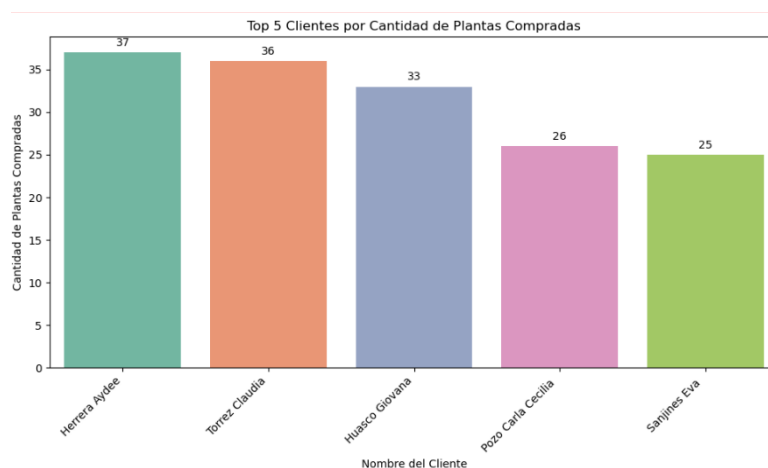


Figura 4-7 Visualización de los clientes con mayor volumen de compras de plantas

Fuente: Elaboración propia (2025)

Resultados del gráfico 5

Este gráfico presenta a los cinco mejores clientes según el monto total gastado en la compra de plantas. La cliente con el mayor gasto es Verónica Quiere, con un total de 800 Bs. En contraste, entre estos cinco clientes, quien realizó el menor gasto fue Giovana Huasco, con 430 Bs.

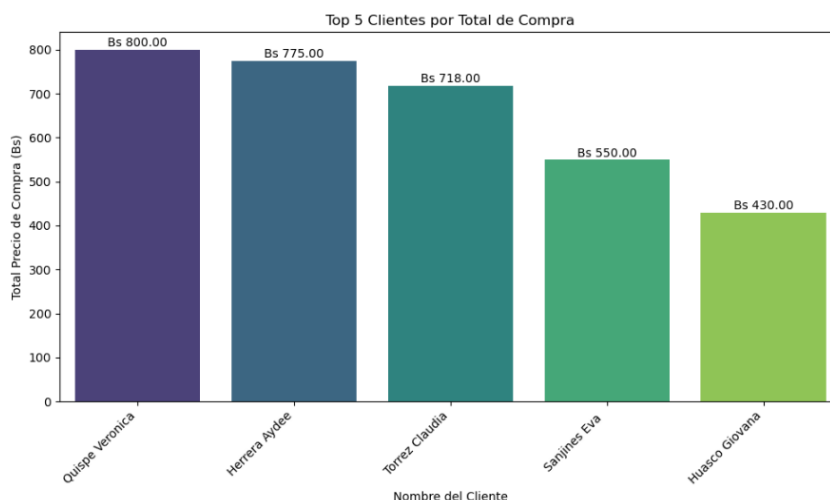


Figura 4-8 Visualización de 5 clientes con el mayor monto acumulado en compras

Fuente: Elaboración propia (2025)

Resultados del gráfico 6

Este gráfico muestra la cantidad de plantas compradas por cada departamento o región, diferenciando entre plantas de interior y exterior. Se puede observar claramente que el departamento de La Paz registra la mayor cantidad de compras, con 258 plantas de interior y 154 plantas de exterior, lo que evidencia una alta preferencia de los clientes por ambos tipos de plantas en esta región.”

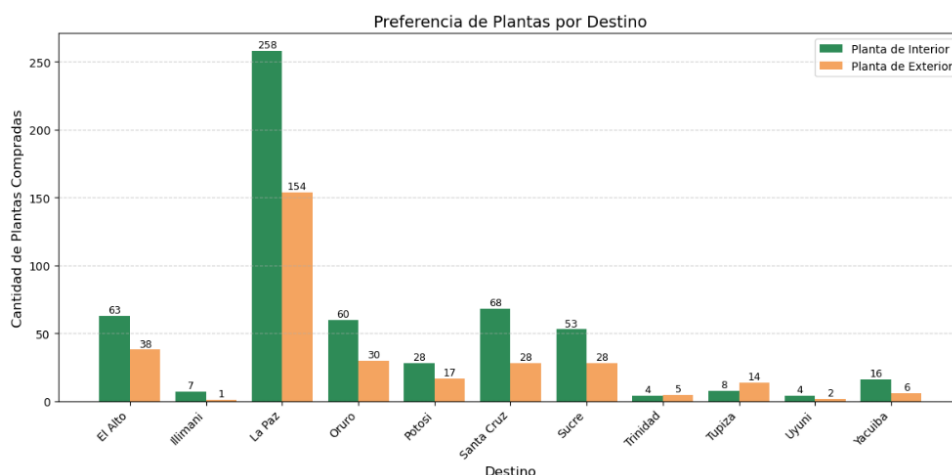


Figura 4-9 Visualización de plantas de interior y exterior más adquiridas por los clientes según destino final

Fuente: Elaboración propia (2025)

Resultados del gráfico 7

Se predice un aumento sostenido y saludable de las ventas diarias hasta diciembre de 2025, lo cual sugiere una proyección optimista que podría deberse a una estrategia comercial efectiva, a una temporada alta prevista (como las festividades de fin de año) o al crecimiento esperado del vivero. Además, la coherencia en los resultados, reflejando un comportamiento de ventas plausible y alineado con tendencias positivas.

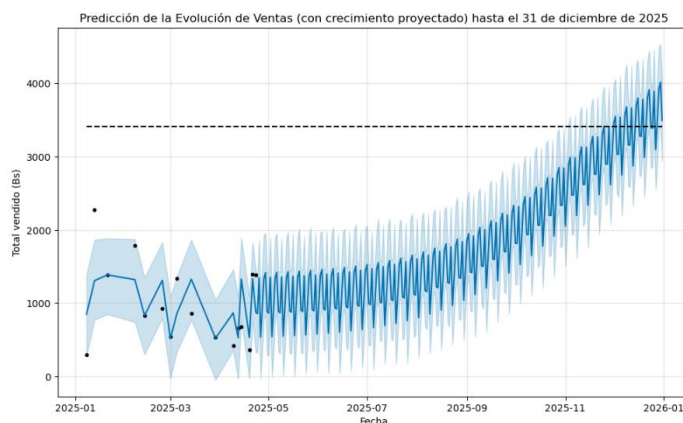


Figura 4-10 Visualización de la proyección de ventas de plantas hasta fin de año

Fuente: Elaboración propia (2025)

4.4. Resultado de la evaluación y visualización de los resultados en Power BI

4.4.1. Resultados de la Visualización 1

La visualización muestra, mediante un segmento en forma de tabla, todos los nombres de los clientes que realizaron compras. Al seleccionar el nombre de Américo Oropeza, se actualiza el gráfico para mostrar que el total gastado por este cliente en todas sus compras es de 185 Bs.

Tabla 4 Verificación de resultado según cliente

B	C	D	E	F	G	H
Pinto Evelin	73704540	2	2	4	105	trans Lupjansa
Perez Veronica	76254929	9	3	12	190	trans Lupjansa
Talavera Soto Flor	79031763	3	3	6	65	trans Lupjansa
Usmayo Lucy	76220042	7	3	10	115	trans Lupjansa
Mendizabel Camila	78928174	6	4	10	150	trans Lupjansa
Garcia Jose	76265315	6	3	9	180	trans Lupjansa
Mamani Carla	77728766	3	3	6	140	trans Lupjansa
Cuellar Jose	69206338	9	1	10	180	trans Lupjansa
Antequera Omar	77755950	5	2	7	100	trans Lupjansa
De Los angeles Flores Maria	67088942	7	3	10	130	trans Lupjansa
Gutierrez Luis	77513377	5	5	10	180	trans Lupjansa
Americo Oropeza	70534574	4	7	11	185	Trans Illimani
Mercado Daniela	75836812	1	1	2	35	trans Lupjansa
Herrera Aydee	63239394	4	1	5	90	trans Lupjansa
Pozo Carla Cecilia	73581998	8	3	11	165	trans Lupjansa
Aguilar Mirian	70516859	6	3	9	100	trans Lupjansa
Torrez Claudia	72832131	6	6	12	300	trans Lupjansa
Apaza Belen	63874468	9	2	11	350	rans Lupjansa
Rodriguez Imana	71559280	3	3	6	90	trans Lupjansa
Machicado Javier	65874890	8	5	13	190	trans Lupjansa
Huanca Alex	63649367	7	3	10	100	trans Lupjansa

Fuente: Elaboración propia (2025)

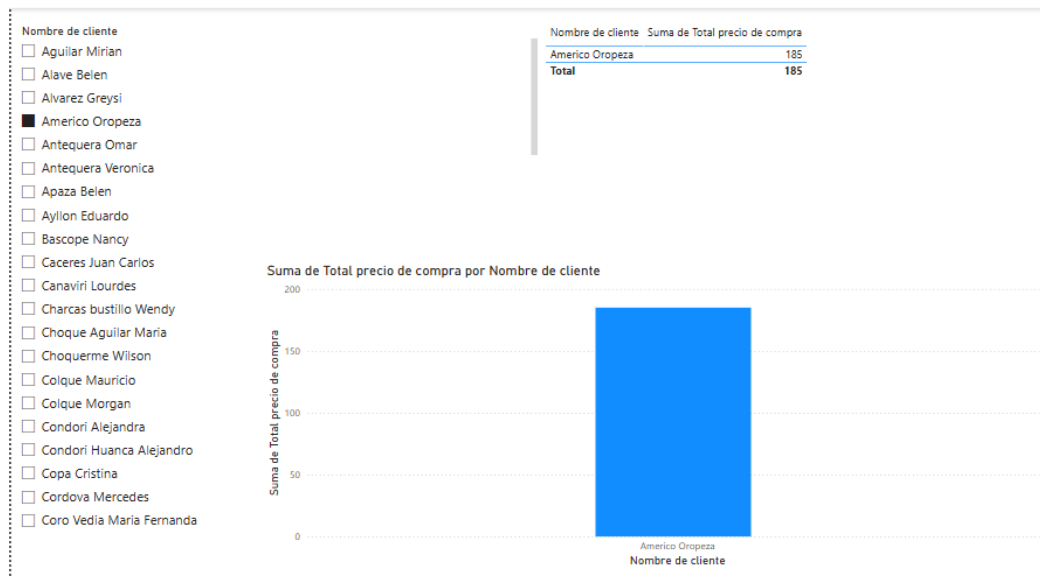


Figura 4-11 Visualización del gasto realizado por cada cliente

Fuente: Elaboración propia (2025)

4.4.2. Resultados de la Visualización 2

Se puede observar que, al filtrar los datos por el departamento de La Paz, los resultados muestran que entre el 8 de enero y el 21 de abril, los clientes de este departamento prefirieron en mayor medida la compra de plantas de interior, con un total de 258 Bs, en comparación con 154 Bs en plantas de exterior.

Además, el gráfico evidencia que, mes a mes, la preferencia por plantas de interior se mantuvo constante en La Paz, lo que indica una tendencia clara en los hábitos de consumo de los clientes de esta región.

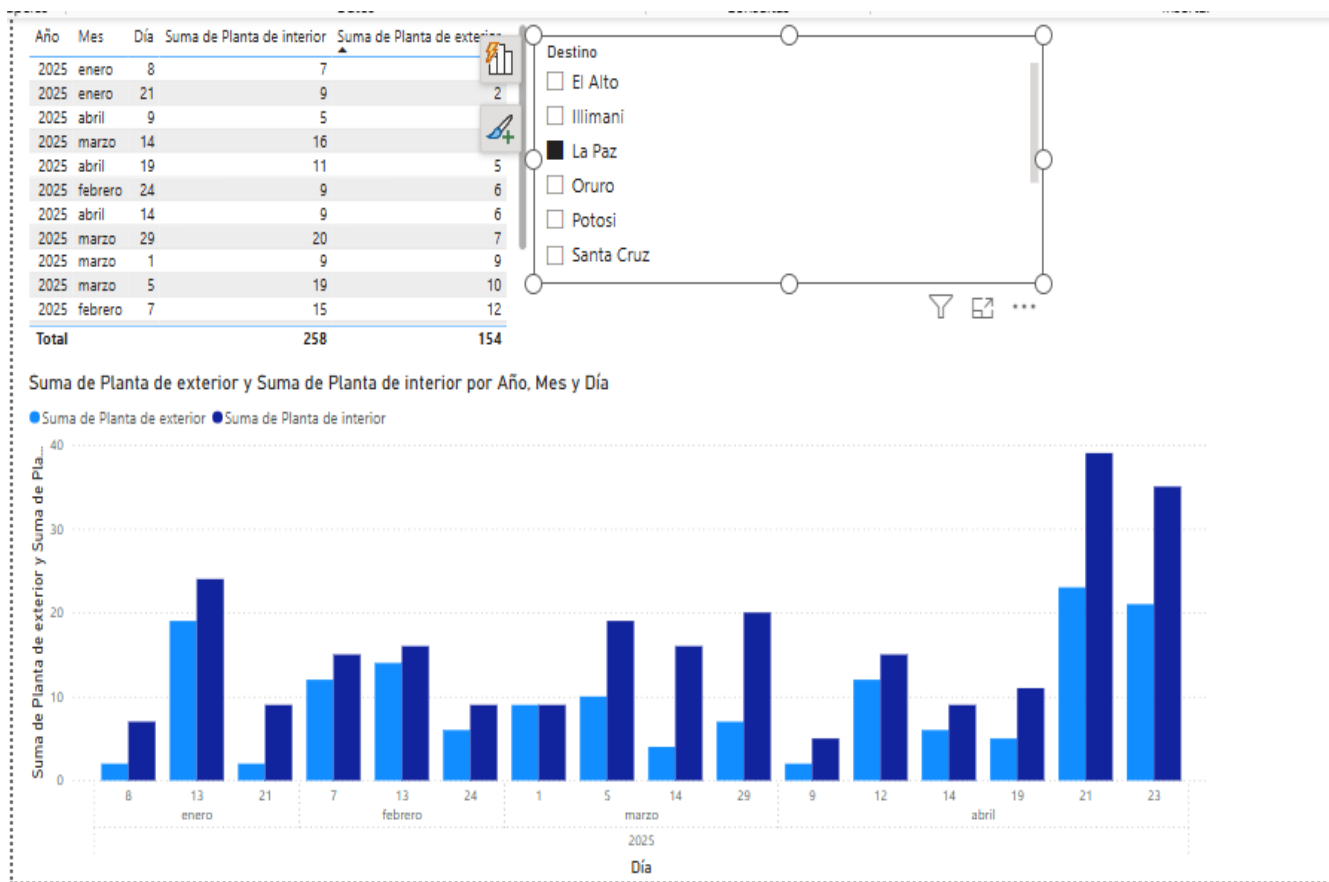


Figura 4-12 Visualización de las plantas más demandadas según su destino: interior o exterior

Fuente: Elaboración propia (2025)

4.4.3. Resultados de la Visualización 3

En esta visualización se ha aplicado un segmento de datos que filtra los resultados a partir del 26 de enero. Esto permite mostrar el valor exacto del total del precio de compra por mes: en febrero, el total es de 3.546; en marzo, 3.261; y en abril, 4.897. La suma total de los tres meses asciende a 11.704

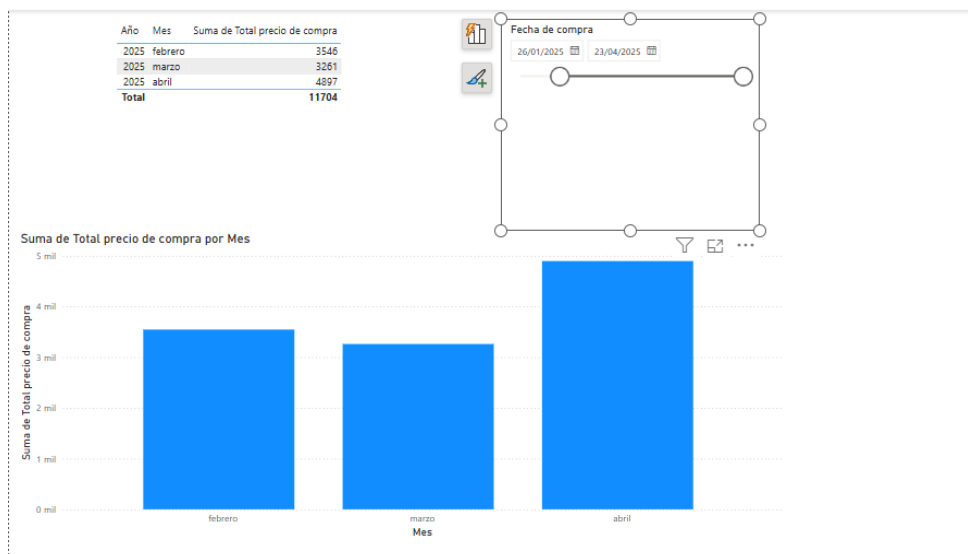


Figura 4-13 Visualización del total de ventas a lo largo del tiempo según fechas registradas

Fuente: Elaboración propia (2025)

4.5. Discusión de resultados

El proyecto se centró en optimizar las operaciones del Vivero Danielito mediante la implementación de un sistema web con herramientas open source para registrar la producción y ventas. Se abordaron limitaciones como la falta de una base de datos centralizada y la dependencia de transacciones presenciales. Utilizando la metodología XP, se desarrollaron fases de planeación, diseño, codificación, pruebas e implementación, apoyadas en técnicas de recolección de datos como observación, entrevistas y encuestas, y se modelaron procesos con diagramas UML. El sistema fue desarrollado con PHP, MVC y MySQL, superando las expectativas al incorporar una tienda virtual y mejorar la experiencia del cliente. En nuestro caso, se utilizaron diversas herramientas para realizar un análisis detallado sobre los clientes y las ventas de plantas efectuadas a través de plataformas digitales como Tik Tok y WhatsApp. Mediante el uso de **Jupyter Notebook** y **Power BI**, fue posible profundizar en el comportamiento de compra de los clientes, evaluando cada transacción en detalle.

Este enfoque analítico coincide con el proyecto (SANCHEZ BARAHONA LUIS ALEXANDER, 2024) del Vivero Danielito, el cual también destacó la importancia de las ventas en línea. En uno de sus gráficos, por ejemplo, se evidenció que **el 80 % de las compras se realizan por internet**, lo cual refuerza la tendencia hacia la digitalización de las ventas en viveros y la necesidad de aprovechar plataformas digitales para mejorar la experiencia del cliente y aumentar las ventas.

Recomendación de tienda virtual

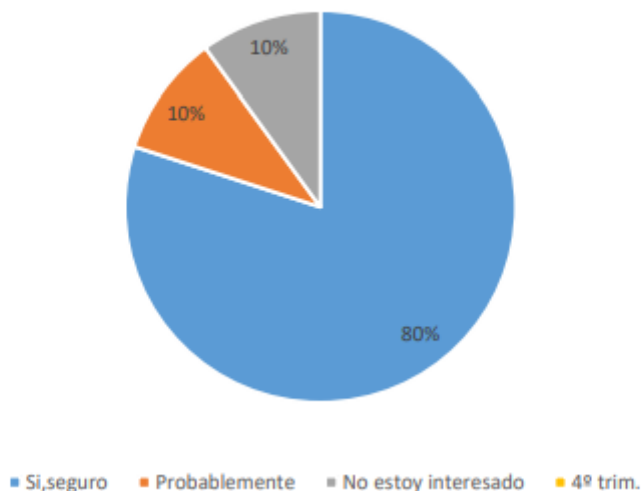


Figura 4-14 Gráfico de porcentajes acerca de la tienda virtual

Fuente: SANCHEZ BARAHONA LUIS ALEXANDER, 2024)

l

Tabla 5 Resultado sobre el uso de tiendas virtuales

Opciones	Respuestas	Porcentaje
Si, seguro	20	80%
Probablemente	5	10%
No estoy interesado	5	10%
Total	30	100%

Fuente: SANCHEZ BARAHONA LUIS ALEXANDER, 2024)

En la figura 4-10 se muestra la proyección de la evolución de ventas del vivero *Jardín de Hojas Verdes* hasta finales de este año. Los datos reflejan una **tendencia positiva**, especialmente en lo que respecta a las ventas realizadas por internet. Esta proyección evidencia el impacto favorable de la digitalización y el uso de plataformas como WhatsApp y Tik Tok, lo cual representa una oportunidad clave para seguir fortaleciendo las estrategias de venta en línea.

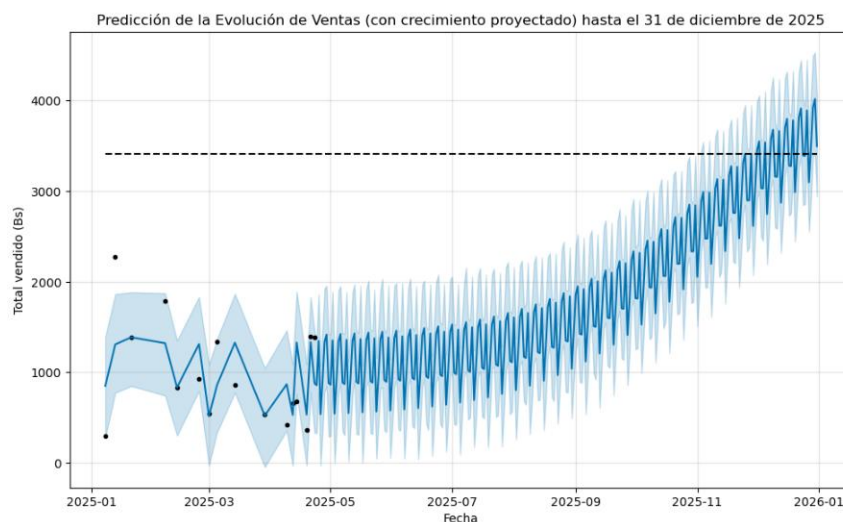


Figura 4-15 Visualización de la proyección de ventas de plantas hasta fin de año

Fuente: Elaboración propia (2025)

Para finalizar tanto la implementación de sistemas digitales como el análisis de datos sobre ventas en línea han demostrado ser estrategias eficaces para mejorar la gestión y aumentar la rentabilidad en viveros. La coincidencia entre los hallazgos del proyecto del Vivero Danielito y nuestro propio análisis evidencia que la transformación digital, especialmente a través del comercio en plataformas sociales y herramientas de análisis, es un camino clave para el crecimiento sostenible del sector.

5. Conclusiones

La implementación del proyecto en el vivero Jardín de Hojas Verdes demostró cómo la incorporación de herramientas digitales puede transformar profundamente la gestión operativa y estratégica de un pequeño negocio dedicado a la venta de plantas. Al integrar soluciones tecnológicas como Microsoft Excel, Jupyter Notebook y, especialmente, Power BI, fue posible digitalizar y automatizar procesos clave como el registro de ventas, el control de inventario y el seguimiento de las preferencias de los clientes.

La recolección de datos fue exitosa, ya que se logró obtener toda la información necesaria para llevar a cabo el análisis. Esta etapa resultó ser clave para el inicio del proyecto, ya que sin estos datos no habría sido posible comenzar con el desarrollo del análisis y la toma de decisiones.

Se logró diseñar con éxito una base de datos en Excel, en la cual se almacenó toda la información recolectada sobre los clientes y las ventas. Esta base de datos constituye un recurso fundamental para la organización y análisis de la información recopilada.

Mediante el uso de herramientas como Jupyter Notebook, se logró analizar exitosamente el comportamiento de los clientes durante los meses de enero, febrero, marzo y abril. Además, se realizó una predicción de futuras ventas, la cual será de gran utilidad para la planificación y gestión de la temporada de verano en el futuro.

El uso de la herramienta de Power BI permitió obtener información valiosa de forma clara, comprensible y accesible que fortalecen la toma de decisiones informadas: por ejemplo, se identificó que las plantas de interior son las más demandadas, lo que orienta la producción y las campañas publicitarias; además, se detectó que el departamento de La Paz es el mercado líder en compras en línea, con un total de Bs 6.903 en ventas, superando notablemente a otros departamentos.

6. Recomendaciones

Se recomienda implementar un **sistema** de gestión de inventario digital que permita registrar en tiempo real los ingresos y salidas de productos, optimizando así el control de existencias y facilitando la toma de decisiones. Sin embargo, en este proyecto no se logró desarrollar dicho sistema web debido a las limitaciones de tiempo establecidas para la ejecución del trabajo. Aunque se reconoció su importancia, el alcance temporal del proyecto no fue suficiente para abordar el diseño, desarrollo e implementación completa de esta funcionalidad, por lo que se sugiere considerarlo como una fase futura de mejora.

Se recomienda aumentar el stock de plantas de interior, dado que representan la categoría de mayor demanda, lo que podría traducirse en un incremento directo de las ventas si se logra satisfacer esta preferencia del cliente de forma constante. Asimismo, es recomendable desarrollar campañas de marketing digital dirigidas específicamente a los habitantes del departamento de La Paz, ya que representan el segmento con mayor poder de compra en línea. Estas campañas podrían incluir promociones exclusivas, envíos gratuitos o descuentos en productos seleccionados para reforzar la fidelización.

Además, se aconseja implementar una planificación de inventario basada en la estacionalidad y en los datos de comportamiento de compra, lo cual permitirá optimizar la disponibilidad de productos en función de la demanda esperada, evitando tanto el exceso como la falta de stock.

También sería útil ampliar el análisis geográfico a otros departamentos, con el fin de identificar nuevas oportunidades de mercado o regiones con potencial de crecimiento.

Finalmente, se recomienda mantener el uso continuo de herramientas como Power BI para el monitoreo y evaluación constante de las métricas clave, ya que esto permitirá ajustar las estrategias en tiempo real y sostener una toma de decisiones basada en evidencia.

Referencias bibliográficas

Bibliografía

- ARIMA. (2024). <https://www.ibm.com/es-es/think/topics/arima-model>.
- BID. (2012).
- BIMATICO. (2012). Obtenido de <https://www.bimatico.com/es/bi-news/para-que-se-utiliza-power-bi-que-es-capaz-de-hacer-y-donde-descargarlo>.
- CEPAL. (1948).
https://www.google.com/search?gs_ssp=eJzj4tTP1TcwrqhIT1JgNGB0YPBiTU4tSMwBAEKsBeE&q=cepal&oq=cepal&gs_lcrp=EgZjaHJvbWUqEAgBEC4YxwEYsQMY0QMYgAQyDQgAEAAY4wIYsQMYgAQyEAgBEC4YxwEYsQMY0QMYgAQyBggCEEUYOzIHCAMQABiABDIHCAQ_QABiABDIHCAUQABiABDIHCAYQABiABDIHCACQABiAB.
- Ciencia de Datos. (2016). https://cienciadedatos.net/documentos/25_regresion_lineal_multiple.
- Coursera. (2023). <https://www.coursera.org/mx/articles/what-is-python-used-for-a-beginners-guide-to-using-python>.
- De la Cruz & Ramírez. (2021).
- DIAZ, P. (2024).
- El Foro Economico Mundial. (2015). <https://es.weforum.org/stories/2015/01/los-negocios-que-nacieron-en-las-redes-sociales/>.
- ESTEFANIA, G. T. (2022). GUAYAQUIL – ECUADOR .
- Excel. (2016). <https://chromewebstore.google.com/detail/excel-online/iljnkagajgfdmfnnidjijobjlffjfgnb?hl=es&pli=1>.
- FAO. (2020). <https://www.fao.org/news/archive/news-by-date/2020/es/>.
- fastercapital logo. (2025). <https://fastercapital.com/es/contenido/Analisis-competitivo-de-viveros--Estrategias-para-el-analisis-de-mercado-en-el-negocio-de-viveros.html>.
- Félix Zenón Limachi Mamani. (2020). “SISTEMA DE REGISTRO GEOLOCALIZACION DE VIVEROS EN.
- Fundacion Alternativa. (2021). <https://fundacionalternativas.org/>.
- IBM. (2007). <https://www.ibm.com/docs/es/spss-statistics/beta>.
- ISDI. (2024). <https://www.isdi.education/es/blog/herramientas-de-analisis-de-datos>.
- MedicalNewsToday. (2021). Obtenido de <https://www.medicalnewstoday.com/articles/es/hierbas-para-la-ansiedad>.
- Medium. (2025). <https://medium.com/@tarangds/traditional-prediction-models-prophet-arima-83bc8b980ec4>.
- OpenWebinars. (2024). <https://openwebinars.net/blog/jupyter-notebook/>.

PACHECO-DIAZ. (2024). Obtenido de

<https://cia.uagraria.edu.ec/Archivos/SANCHEZ%20BARAHONA%20LUIS%20ALEXANDER%20Y%20VASQUEZ%20PINEDA%20GISSELA%20DAYANARA.pdf>.

Predik Dat Driven. (2023). <https://predikdata.com/es/que-son-y-para-que-se-usan-los-modelos-predictivos/>.

Pro Mujeres. (2025). <https://promujer.org.bo/es/>.

Provost & Fawcett. (2013).

https://www.researchgate.net/publication/256439081_Data_Science_and_Its_Relationship_to_Big_Data_and_Data-Driven_Decision_Making.

SANCHEZ BARAHONA LUIS ALEXANDER. (2024). Obtenido de

<https://cia.uagraria.edu.ec/Archivos/SANCHEZ%20BARAHONA%20LUIS%20ALEXANDER%20Y%20VASQUEZ%20PINEDA%20GISSELA%20DAYANARA.pdf>.

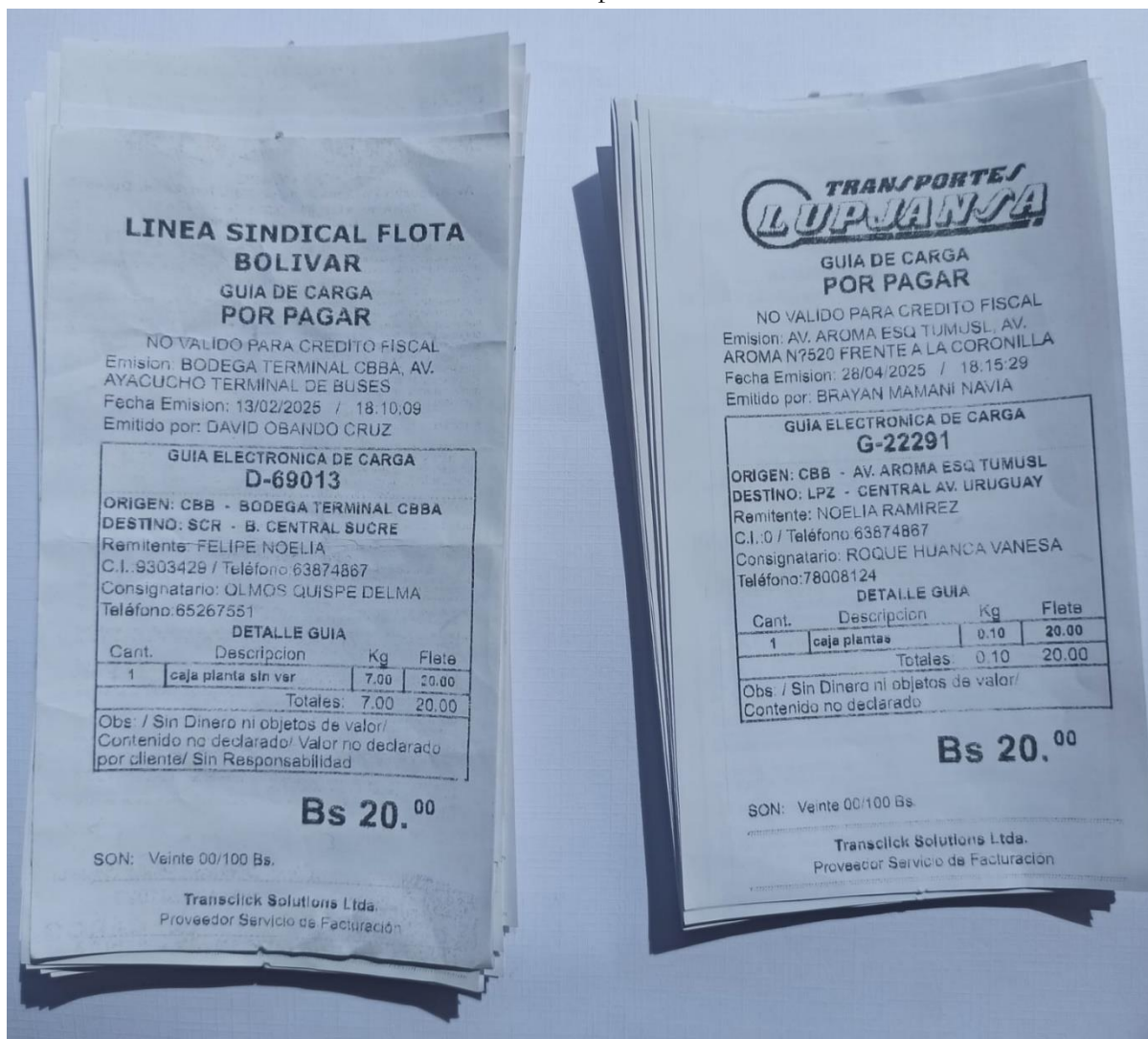
ScienceDirect. (2023). <https://www.sciencedirect.com/topics/social-sciences/digital-transformation>.

STEFANIA, P. D. (2024). Babahoyo-Los Ríos-Ecuador .

Anexos

Anexo 1. Adquir datos

Se recolectaron datos del costo de envío de las transportadoras a través de baucher



Ubicación: CD: Imagen de facturas/Proyecto final-Noelia Felipe Ramirez

Anexo 2. Código ejecutado en Jupyter notebook

```
# Agrupar y ordenar por total de precio de compra
top_clientes_precio = (datos.groupby('Nombre de cliente')['Total precio de compra']
                        .sum()
                        .sort_values(ascending=False)
                        .head(5)
                        .reset_index())

# Crear el gráfico de barras
plt.figure(figsize=(10,6))
ax = sns.barplot(data=top_clientes_precio, x='Nombre de cliente', y='Total precio de compra', palette='viridis')

# Etiquetas exactas del total en cada barra con "Bs" en lugar de "USD"
for i, row in top_clientes_precio.iterrows():
    ax.text(i, row['Total precio de compra'] + 1, f"Bs {row['Total precio de compra']:.2f}",
            ha='center', va='bottom', fontsize=10)

# Ajustes visuales
plt.xticks(rotation=45, ha='right')
plt.title('Top 5 Clientes por Total de Compra')
plt.xlabel('Nombre del Cliente')
plt.ylabel('Total Precio de Compra (Bs)')
plt.tight_layout()
plt.show()

datos['Fecha de compra'] = pd.to_datetime(datos['Fecha de compra'], dayfirst=True, errors='coerce')

datos['Fecha de compra'] = pd.to_datetime(datos['Fecha de compra'])

# Agrupar por fecha
ventas_diarias = datos.groupby('Fecha de compra')['Total precio de compra'].sum().reset_index()

# Graficar Línea de tiempo
plt.figure(figsize=(14, 6))
sns.lineplot(data=ventas_diarias, x='Fecha de compra', y='Total precio de compra', marker='o')
plt.title('Evolución del Total de Ventas en el Tiempo', fontsize=16)
plt.xlabel('Fecha de compra')
plt.ylabel('Total vendido (USD)')
plt.grid(True)
plt.tight_layout()
plt.show()

# Agrupar y ordenar por cantidad total de plantas compradas
top_clientes = (datos.groupby('Nombre de cliente')['Cantidad de plantas compradas']
                .sum()
                .sort_values(ascending=False)
                .head(5)
                .reset_index())

# Crear el gráfico de barras
plt.figure(figsize=(10,6))
ax = sns.barplot(data=top_clientes, x='Nombre de cliente', y='Cantidad de plantas compradas', palette='Set2')

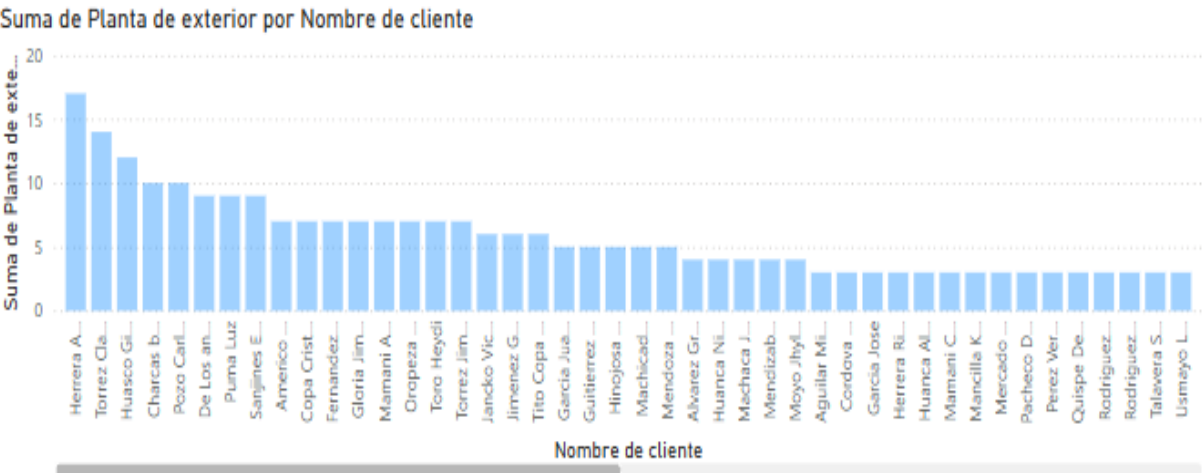
# Agregar etiquetas con la cantidad exacta encima de cada barra
for i, row in top_clientes.iterrows():
    ax.text(i, row['Cantidad de plantas compradas'] + 0.5, int(row['Cantidad de plantas compradas']),
            ha='center', va='bottom')

# Ajustes de presentación
plt.xticks(rotation=45, ha='right')
plt.title('Top 5 Clientes por Cantidad de Plantas Compradas')
plt.xlabel('Nombre del Cliente')
plt.ylabel('Cantidad de Plantas Compradas')
plt.tight_layout()
plt.show()
```

Ubicación:CD: Notebook/Proyecto final-Noelia Felipe Ramirez

Anexo 3. Visualización de clientes con mayor compra de plantas de exterior

Año	Mes	Día
2025	enero	1
2025	enero	2
2025	enero	3
2025	enero	4
2025	enero	5
2025	enero	6
2025	enero	7
2025	enero	8
2025	enero	9
2025	enero	10
2025	enero	11
2025	enero	12
2025	enero	13
2025	enero	14
2025	enero	15
2025	enero	16



Ubicación:CD: Visualización en Power BI/Proyecto final-Noelia Felipe Ramirez

Anexo 4. Contenido del CD

Este apartado presenta la organización del contenido del CD, el cual incluye todos los archivos y carpetas requeridos para el desarrollo del proyecto.

Excel

Reporte de venta de plantas.xlsx.

Imagenes

Adquirir datos.ipg

Código ejecutado en Jupyter notebook.jpg
--

Visualización de clientes con mayor compra de plantas.ipg

Power BI

Visualización 1.pptx.

Visualización 2.pptx

Visualización 3.pptx.

Codigo fuente

Gráfica 1.ipynb.

Gráfica 2.ipynb.

Gráfica 3.ipynb.

Gráfica 4.ipynb.

Gráfica 5.ipynb.

Gráfica 6.ipynb.

Gráfica 7.ipynb.

proyecto pdf

ProyectoCienciaDatosDiplomado.pdf.

Gráficas de análisis.csv

predicción.csv.

Anexo 5. Repositorios

GitHub: <https://github.com/NoeliaFelipeRamirez/PROYECTO-FINAL-DIPLOMADO-CIENCIA-DE-DATOS.git>

