**DATOS DEL ESTUDIANTE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Apellidos y Nombres: | 1. Clemente Ramos Jhordan Michael 2. Calsin Muñoz Kevin Jhojaner 3. Alaya Ventocilla Leonardo Anderson 4. Infantes Hernández | ID: | 1. 1541227 2. 1559805 3. 1566201 4. 1573765 | |
| Dirección Zonal/CFP: | Virtual | | | |
| Carrera: | Ingeniería De Software Con Inteligencia Artificial | Semestre: | | IV |
| Curso/ Mód. Formativo: | Módulos y paquetes para machine learning con Python | | | |
| Tema de Trabajo Final: | Análisis de datos y visualización con pandas, numpy y matplotlib | | | |

1. **INFORMACIÓN**

* **Identifica la problemática del caso práctico propuesto.**

Es fundamental contar con un sistema eficiente para evaluar las tendencias de ventas y tomar decisiones estratégicas basadas en datos. Sin embargo, al analizar el conjunto de datos proporcionado, se pueden identificar ciertos desafíos clave:

1. **Variabilidad en las ventas:** Existen fluctuaciones en los montos de venta a lo largo del tiempo, lo que dificulta la previsión de ingresos y la planificación de inventario.
2. **Segmentación y rendimiento por categoría:** No todas las categorías de productos generan el mismo nivel de ingresos, lo que plantea la necesidad de evaluar cuáles son las más rentables y cuáles requieren estrategias de mejora.
3. **Clientes y patrones de compra:** Determinar cuáles son los clientes más valiosos y cómo se comportan sus compras a lo largo del tiempo puede permitir mejorar la personalización de ofertas y fidelización.
4. **Distribución de ventas:** Identificar la distribución de montos de venta y si existen valores atípicos o inconsistencias que puedan afectar la interpretación de los datos.
5. **Correlaciones entre variables:** Analizar si hay relaciones significativas entre diferentes factores (como la categoría del producto y el total de ventas) que permitan optimizar estrategias comerciales.

* **Identifica propuesta de solución y evidencias.**

**Propuesta de Solución**

Para mejorar la gestión y optimización de las ventas en el negocio analizado, se proponen las siguientes soluciones basadas en el análisis de datos realizado:

1. **Optimización del Inventario y Gestión de Productos**
   * A partir del análisis de ventas por categoría, se identificó que ciertas categorías tienen un rendimiento superior a otras. Se recomienda reasignar recursos hacia los productos más vendidos y evaluar estrategias de promoción para aquellos con menor desempeño.
2. **Segmentación de Clientes y Estrategias de Fidelización**
   * Mediante la identificación de los clientes con mayores compras, se pueden diseñar estrategias personalizadas como descuentos, programas de lealtad o promociones exclusivas para fomentar su permanencia y aumentar la frecuencia de compra.
3. **Análisis de Patrones Temporales para Optimizar Estrategias de Venta**
   * El gráfico de evolución de ventas en el tiempo muestra fluctuaciones en los ingresos. Esto sugiere la necesidad de estrategias de marketing dirigidas en períodos de baja demanda para equilibrar el flujo de ingresos.
4. **Reducción de Inconsistencias y Datos Faltantes**
   * Se identificaron valores nulos en algunas columnas del dataset. Para evitar problemas en el análisis y toma de decisiones, se recomienda implementar un sistema de validación de datos al momento de ingresar registros en el sistema de ventas.
5. **Uso de Modelos Predictivos para Pronóstico de Ventas**
   * A partir de la correlación entre variables numéricas, se pueden desarrollar modelos predictivos que ayuden a prever tendencias futuras y mejorar la planificación de recursos.

**Evidencias del Análisis**

A continuación, se presentan las principales evidencias obtenidas del análisis de datos:

* **Estadísticas descriptivas:** Permiten conocer la media, mediana y desviación estándar de las ventas, facilitando la identificación de variaciones.
* **Mapa de calor de correlaciones:** Muestra qué variables tienen mayor relación entre sí, lo que puede ayudar a descubrir factores que influyen en el rendimiento de ventas.
* **Gráficos de tendencias y distribuciones:** Se evidencian patrones de comportamiento de las ventas a lo largo del tiempo, así como la distribución de los montos de venta y la participación de cada categoría en el total de ingresos.
* **Segmentación por cliente y categoría:** Se identifican los clientes y categorías más relevantes, permitiendo focalizar estrategias de negocio.
* **Respuestas a preguntas guía**

**Durante el análisis y estudio del caso práctico, debes obtener las respuestas a las interrogantes:**

|  |  |
| --- | --- |
| Pregunta 01: | ¿Cómo se pueden optimizar las estructuras de datos utilizando Pandas y NumPy? |
| * **Tipo de datos adecuado:** Especificar tipos de datos ejemplo loat32 en lugar de float64 para reducir el uso de memoria. * **Uso de índices:** Usar índices adecuados en DataFrames para acelerar búsquedas y operaciones. * **Métodos vectorizados:** Utilizar operaciones vectorizadas en NumPy en lugar de bucles para mejorar la eficiencia. | |
| Pregunta 02: | ¿Cómo se puede extraer información de un archivo CSV con Pandas y qué funciones son clave para este proceso? |
| Para extraer información de un archivo CSV con Pandas, se utiliza la función pd.read\_csv(),  **Funciones clave para el proceso incluyen:**   * df.head(): Muestra las primeras filas. * df.info(): Muestra información sobre el DataFrame. * df.describe(): Proporciona estadísticas descriptivas. | |
| Pregunta 03: | ¿Cuál es la diferencia entre un gráfico de líneas, un gráfico de barras y un gráfico de torta, y en qué situaciones se recomienda cada uno? |
| **Diferencia:**   * **Gráfico de líneas:** Ideal para mostrar tendencias a lo largo del tiempo ejemplo series temporales. * **Gráfico de barras:** Útil para comparar categorías discretas ejemplo ventas por producto). * **Gráfico de torta:** Bueno para mostrar proporciones de un todo ejemplo distribución de mercado, aunque se recomienda usarlo con pocas categorías. | |
| Pregunta 04: | ¿Cómo se pueden manejar valores nulos en un DataFrame de Pandas y qué estrategias existen para tratarlos? |
| Para manejar valores nulos en un dataframe se puede usar varias estrategias   * **Eliminar:** Usar df.dropna() para eliminar filas o columnas con valores nulos. * **Rellenar:** Usar df.fillna() para sustituir valores nulos con un valor específico (media, mediana, etc.). * **Interpolar:** Utilizar df.interpolate() para estimar valores nulos basados en otros datos. | |
| Pregunta 05: | ¿Qué ventajas ofrece el uso de NumPy en comparación con las listas nativas de Python para el manejo de grandes volúmenes de datos? |
| Bueno sobre las ventajas de NumPy frente a listas nativas de Python incluyen:   * Eficiencia en memoria: Arrays de NumPy son más compactos que listas. * Velocidad: Operaciones en arrays son más rápidas debido a su implementación en C. | |

**2. PLANIFICACIÓN DEL TRABAJO**

* **Cronograma de actividades:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **N°** | **ACTIVIDADES** | **CRONOGRAMA** | | | | | |
| 28/03 | 31/03 | 01/04 | 02/04 | 04/04 |  |
| 1° | Desarrollo del código (script) | X | X |  |  |  |  |
| 2° | Elaboración de informe |  |  | X |  |  |  |
| 3° | Desarrollo de la presentación y subida de trabajos al blackboard |  |  |  | X |  |  |
| 4° | Exposición de los proyectos |  |  |  |  | X |  |
| 5° |  |  |  |  |  |  |  |
| 6° |  |  |  |  |  |  |  |

* **Lista de recursos necesarios:**

|  |  |
| --- | --- |
| 1. **MÁQUINAS Y EQUIPOS** | |
| **Descripción** | **Cantidad** |
| computadora | 4 |
| celular | 4 |
| Disco duro | 4 |
|  |  |

|  |  |
| --- | --- |
| 1. **HERRAMIENTAS E INSTRUMENTOS** | |
| **Descripción** | **Cantidad** |
| Editor de código (vs code) | 1 |
| Herramientas de visualización (matplotlib) | 1 |
|  |  |
|  |  |

|  |  |
| --- | --- |
| 1. **MATERIALES E INSUMOS** | |
| **Descripción** | **Cantidad** |
| Archivo.csv | 1 |
| Documento de referencia | 1 |
|  |  |
|  |  |

**3. DECIDIR PROPUESTA**

* **Describe la propuesta determinada para la solución del caso práctico**

|  |
| --- |
| **PROPUESTA DE SOLUCIÓN** |
| El caso práctico consiste en analizar un conjunto de datos de ventas, que contenía valores no numéricos. La solución propuesta aborda la limpieza de datos, el análisis estadístico y la visualización de los resultados para obtener información valiosa.   * **Limpieza de datos:** Se convierte la columna "Total Venta" a valores numéricos utilizando pd.to\_numeric() y eliminamos las filas con valores NaN con dropna() * **Análisis estadístico:** Se calcula suma, promedio, mediana y desviación estándar de las ventas utilizando NumPy. Estas métricas proporcionan una visión general de cómo se comporta las ventas * **Visualización de datos:** Se generan gráficos   + Mapa de calor de correlaciones: Muestra las relaciones entre variables numéricas.   + Gráfica de evolución temporal de ventas: Representa cómo han cambiado las ventas a lo largo del tiempo.   + Gráfico de barras: Compara las ventas por categoría.   + Gráfico de torta: Visualiza la distribución de ventas por categoría. * **Segmentación de datos:** Se aplican filtros para analizar las ventas de productos por categoría o cliente específico * **Ordenación temporal:** Los datos se ordenan por fecha para facilitar el análisis de las tendencias de ventas a lo largo del tiempo.   En conjunto, estas soluciones permiten un análisis completo de los datos de ventas, lo que facilita la toma de decisiones estratégicas para optimizar las ventas y mejorar el rendimiento del negocio. |

**4. EJECUTAR**

* **Resolver el caso práctico, utilizando como referencia el problema propuesto y las preguntas guía proporcionadas para orientar el desarrollo.**
* **Fundamentar sus propuestas en los conocimientos adquiridos a lo largo del curso, aplicando lo aprendido en las tareas y operaciones descritas en los contenidos curriculares.**

**INSTRUCCIONES:** Ser lo más explícito posible. Los gráficos ayudan a transmitir mejor las ideas. Tomar en cuenta los aspectos de calidad, medio ambiente y SHI.

|  |
| --- |
| **OPERACIONES / PASOS / SUBPASOS** |
| Se importo las librerías (pandas numpy matplotlib.pyplot y seaborn) |
| Leemos el archivo con pd.read\_csv() |
| Se obtiene y muestra información básica sobre el DataFrame:  df.info() | df.shape() | df.dtypes | df.head() | df.tail() |
| Manejamos los valores nulos en cada columna |
| Corregimos valores no numéricos en “Total Venta” y los convertimos en valores  numéricos con pd.to\_numeric() con el parámetro errors='coerce' que reemplaza  cualquier valor no numérico con NaN |
| Usamos cálculos estadísticos básicos  suma, promedio, mediana, desviación estándar y imprimimos el resultado |
| Se seleccionan las columnas numéricas del DataFrame usando  df.select\_dtypes(include=[np.number]) y calculamos la matriz de correlaciones  entre variables con .corr() |
| Se utiliza sns.heatmap() para generar un mapa de calor que muestra las  correlaciones entre las variables numéricas.  Se ajusta el tamaño de la figura con plt.figure(figsize=(10, 6))  Se agrega un título con plt.title() y se visualiza con plt.show(). |
| Agrupamos los datos por la columna "Categoria" y se suman los valores de  "Total Venta" y se crea un grafico de barras usando plot(kind="bar") y se  visualiza el grafico con plt.show() |
| Agrupamos datos por categoría y se calcula las ventas totales por categoría y  generamos un grafico de torta usando plt.pie() y visualizamos con plt.show() |
| Se genera un histograma usando plt.hist() para mostrar la frecuencia de los  montos de ventas ajustamos el titulo y etiquetas y visualizamos con plt.show() |
| Filtramos las filas donde "Total Venta" es mayor a 1000 e imprimimos |
| Filtramos las filas donde Cliente es igual a "Toni Sanchez |
| Se filtran las filas donde "Categoria" es igual a "Electronica" |
| Se convierte la columna "Fecha de Venta" a tipo de dato datetime usando  pd.to\_datetime() |
| Se ordenan las filas del DataFrame por "Fecha de Venta" usando  df.sort\_values() |

|  |  |  |
| --- | --- | --- |
|  | | |
| C:\Unidad_D\Nuevo Logo\SENATI_FF-01 Modificado.jpg | [NOMBRE DEL TEMA DEL TRABAJO FINAL] | |
| [APELLIDOS Y NOMBRES] | [ESCALA] |

**DIBUJO / ESQUEMA / DIAGRAMA DE PROPUESTA**

**(Adicionar las páginas que sean necesarias)**

**5. CONTROLAR**

* **Verificar el cumplimiento de los procesos desarrollados en la propuesta de solución del caso práctico.**

|  |  |  |
| --- | --- | --- |
| **EVIDENCIAS** | **CUMPLE** | **NO CUMPLE** |
| * ¿Se identificó claramente la problemática del caso práctico? |  |  |
| * ¿Se desarrolló las condiciones de los requerimientos solicitados? |  |  |
| * ¿Se formularon respuestas claras y fundamentadas a todas las preguntas guía? |  |  |
| * ¿Se elaboró ​​un cronograma claro de actividades a ejecutar? |  |  |
| * ¿Se identificaron y listaron los recursos (máquinas, equipos, herramientas, materiales) necesarios para ejecutar la propuesta? |  |  |
| * ¿Se ejecutó la propuesta de acuerdo con la planificación y cronograma establecidos? |  |  |
| * ¿Se describieron todas las operaciones y pasos seguidos para garantizar la correcta ejecución? |  |  |
| * ¿La propuesta es pertinente con los requerimientos solicitados? |  |  |
| * ¿Se evaluó la viabilidad de la propuesta para un contexto real? |  |  |

**6. VALORAR**

* **Califica el impacto que representa la propuesta de solución ante la situación planteada en el caso práctico.**

|  |  |  |  |
| --- | --- | --- | --- |
| **CRITERIO DE EVALUACIÓN** | **DESCRIPCIÓN DEL CRITERIO** | **PUNTUACIÓN MÁXIMA** | **PUNTAJE**  **CALIFICADO POR EL ESTUDIANTE** |
| Identificación del problema | Claridad en la identificación del problema planteado. | 3 |  |
| Relevancia de la propuesta de solución | La propuesta responde adecuadamente al problema planteado y es relevante para el contexto del caso práctico. | 8 |  |
| Viabilidad técnica | La solución es técnicamente factible, tomando en cuenta los recursos y conocimientos disponibles. | 6 |  |
| Cumplimiento de Normas | La solución cumple con todas las normas técnicas de seguridad, higiene y medio ambiente. | 3 |  |
| **PUNTAJE TOTAL** | | **20** |  |

