1. **Análisis de Datos**

**1.1. Limpieza y preparación de datos**

Tienes un conjunto de datos con las siguientes columnas:

* **ID\_cliente**
* **Nombre**
* **Fecha\_de\_compra**
* **Monto\_compra**
* **Producto**

Al analizar el dataset, detectas valores nulos, duplicados y algunos registros donde el formato de las fechas no es consistente.

**Instrucciones:**

* Describe el proceso que seguirías para limpiar estos datos utilizando Python o R.
* Implementa una función en el lenguaje de tu preferencia que realice las siguientes acciones:
  + Eliminar duplicados.
  + Corregir el formato de la columna de fecha.
  + Eliminar los registros con valores nulos en **Monto\_compra**.

**1.2. Análisis descriptivo de datos**

Tienes los siguientes datos de ventas:

* **Ventas\_por\_mes**: [200, 150, 180, 220, 210, 190, 230, 220, 240, 230, 210, 250]

**Instrucciones:**

* Calcula el promedio, la mediana y la desviación estándar de las ventas.
* Realiza una visualización en la que compares las ventas mensuales con el promedio general.

1. **Creación de Modelos Predictivos  
   2.1. Regresión lineal**

Tienes un dataset con las siguientes columnas:

* **Horas\_estudio**
* **Notas\_examen**

**Instrucciones:**

* Entrena un modelo de regresión lineal para predecir las **Notas\_examen** en función de las **Horas\_estudio**.
* Muestra el coeficiente de determinación R2R^2R2 del modelo y realiza una predicción para un estudiante que ha estudiado 5 horas.

**2.2. Clasificación con árbol de decisión**

Dispones de un dataset de clientes con las siguientes columnas:

* **Edad**
* **Ingresos**
* **Compra\_realizada** (0: No, 1: Sí)

**Instrucciones:**

* Entrena un árbol de decisión para predecir si un cliente realizará una compra en función de su edad e ingresos.
* Muestra la importancia de las características y el **accuracy** del modelo.

1. **Estrategia y Liderazgo en Analítica**

**3.1. Definición de una estrategia de analítica**

Imagina que trabajas en una empresa que está comenzando su transformación digital y te han pedido definir una estrategia de analítica de datos.

**Instrucciones:**

* Describe los pasos clave que seguirías para implementar una estrategia de analítica en una organización que nunca ha trabajado con grandes volúmenes de datos.
* ¿Qué factores considerarías para priorizar los proyectos de analítica?

**3.2. Gestión de Stakeholders en un proyecto de analítica**

¿Cómo gestionarías la relación con los stakeholders no técnicos en un proyecto de analítica para asegurar que los resultados sean comprendidos y aplicados correctamente?

1. **Visualización de Datos**
   1. **Creación de un dashboard**

Te han pedido crear un dashboard en Power BI, Qlik o Tableau para visualizar los siguientes indicadores de un negocio de retail:

* **Ventas totales por región**
* **Margen de ganancia por producto**
* **Comparación de ventas mensuales del último año**

**Instrucciones:**

* Describe cómo estructurarías el dashboard para asegurar que sea claro y útil para los stakeholders.
* Adjunta una captura de pantalla de cómo quedaría el dashboard.

1. **Caso Práctico**

Como Científico de Datos usted debe comprender el comportamiento de los clientes, segmentarlos, estar en capacidad de presentarle a la alta gerencia de la compañía sus hallazgos y emplear modelos analíticos para determinar qué campañas se deberán hacer para mejorar la rentabilidad de los clientes.

Para el siguiente caso práctico, usted deberá trabajar con datos reales de unas campañas se telemercadeo de un banco portugués.

El objetivo del ejercicio es analizar la base de clientes, estimar un modelo de predicción de clientes que deben ser seleccionados en cada campaña y con base en este proponer la estrategia sobre la cual se deben realizar las campañas de este producto.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamenteInterfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamenteLa base de datos la puede descargar de <http://archive.ics.uci.edu/dataset/222/bank+marketing> usar el conjunto de datos **“bank-full.csv”.** de la carpeta **bank.zip**

Ilustración 2Bank.zip

1Contenido

Este conjunto de datos está descrito en el artículo académico: Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology.

**El entregable de este ejercicio es una presentación que cubra los siguientes aspectos:**

**Capítulo técnico** (Asuma que lo presentaría ante un público técnico con amplios conocimientos estadísticos, de ciencia de datos y machine learning):

* 1. Análisis descriptivo de la población
  2. Modelo de predicción seleccionado.

1. Justificación de la metodología de modelamiento seleccionada
2. Presentar métricas de clasificación (Puede usar el criterio de área bajo la curva ROC o los criterios que considere convenientes)

**Capítulo de Negocio** (Asuma que lo presentará ante gerentes de negocio, con un buen entendimiento de los clientes, pero bajo conocimiento técnico):

1. Principales hallazgos y conclusiones
2. Propuesta de estrategia para seleccionar los clientes buscando el mayor impacto con el menor costo posible.

**Nota**: Puede trabajar con las técnicas de análisis, algoritmos y programas que considere convenientes, conozca o use regularmente.