

PROBABILIDAD Y ESTADÍSTICA

En la siguiente página baja un dataset, en este documento, me basaré para evaluar, adecua tus respuestas de acuerdo a tu contexto, busca un dataset que provenga de una tienda en línea o e-commerce:

<https://www.kaggle.com/datasets>

(puedes hacer web scraping o download de páginas que dispongan de la información)

1. El objetivo del proyecto es trabajar con un conjunto de datos reales o ficticios que contengan campos como:
 - a. ID o clave del cliente
 - b. Fecha de compra
 - c. Monto de la compra
 - d. Categoría de productos comprados
 - e. Frecuencia de compra del cliente
 - f. Calificación del servicio o estrellas
 - g. Ubicación del cliente
 - h. Precio de productos comprados
 - i. Marca
 - j. Modelo
 - k. Colores
 - l. Etcétera...
2. El objetivo será analizar los datos utilizando herramientas estadísticas y de probabilidad para:
 - Calcular probabilidades condicionales sobre comportamientos de compra.
 - Estimar la probabilidad de que un cliente vuelva a comprar.
3. Las tareas principales que deberás llevar a cabo son:
 - a. Exploración de los datos para identificar su integridad y consistencia
 - b. Identificar valores nulos o “raros” y de ser necesario aplicar métodos de programación Python o R para su limpieza.
 - c. Transformar datos si es necesario aplicando **NORMALIZACIÓN** EN PROGRAMACIÓN.
 - d. Generar medidas estadísticas y probabilísticas vistas en clase

[Escriba aquí]

- e. Visualizar las distribuciones y dispersiones gráficamente en R o Python
 - f. Analizar la probabilidad de recompra condicional según categorías de productos.
 - g. Determinar la probabilidad de recompra considerando el historial del cliente.
 - h. Extraer insights para mejorar la retención de clientes.
 - i. Proponer estrategias de marketing basadas en los resultados mostrando:
 - i. Dashboard o
 - ii. Reporte ejecutivo y visual en canva o pdf
4. Como metodología de trabajo y secuencia de la codificación se sugiere:
- a. Limpieza y preprocesamiento con librerías como pandas y numpy en Python
 - b. Análisis y visualización con matplotlib y seaborn
 - c. Modelado **(opcional por el momento)** con scikit-learn
5. Se espera que con su presentación a todo el grupo nos quede claro que factores o atributos influyen en la compra y sería excelente nos muestren los segmentos de clientes bien definidos, si logran analizar y reflexionar sobre los datos y su comportamiento.
6. Imaginen siempre que la presentación se hace a un cliente.
7. **Tu proyecto en R o Python debe ir en un ANEXO DEL ENTREGABLE PDF, documentado.**
8. **Dentro de los entregables una copia del dataset o un preview.**
9. Se recomienda analizar que columnas poseen datos cuantitativos, cualitativos, y a su vez discretos y continuos.
10. **Presenta en equipo de la manera más clara y precisa en las fechas indicadas en la asignación de TEAMS explica, muestra UN POCO DE código, gráficas y/o menciona los contratiempos que hayan surgido en su caso, resume esta explicación con lo más importante en 15 minutos. Con esta explicación y presentación podré identificar el que se hayan metido de lleno en el trabajo y que lo hayan comprendido, igualmente Si sus compañeros están atentos e interesados.**