

PARCIAL 1

El objetivo de este parcial es evaluar las capacidades del estudiante para realizar análisis de datos utilizando técnicas estadísticas y de visualización, que le permitan entender un conjunto de datos, descubrir *insights* y sugerir accionables al negocio.

Contexto de negocio: Recomendación de compra de vehículos usados.

Como analista de un marketplace de vehículos usados, su tarea consiste en asesorar a un cliente respecto al “mejor carro a comprar” según sus necesidades o perfil de comprador. Estas necesidades pueden estar dadas en términos de precio, kilometraje, modelo, marca, tipo de combustible, transmisión y cualquier otro atributo que se tenga disponible.

Más específicamente, usted debe entender y preparar los datos, extraer patrones y relaciones útiles para la tarea que se le ha encomendado, ejecutar pruebas estadísticas básicas para sustentar sus afirmaciones y proponer una recomendación clara y defendible sobre qué carro comprar.

Para este fin se le ha proporcionado un *snapshot* de la base de datos disponible [aquí](#), en el que encontrará para los diferentes carros listados atributos como precio, modelo, tipo de combustibles, kilometraje, tipo de carrocería, transmisión, entre otros. El diccionario detallado de los datos lo puede consultar [aquí](#).

ACTIVIDADES Y ENTREGABLES

1. **[15%] Entendimiento inicial:** Genere un reporte de exploración inicial de datos en donde destaque aspectos claves como la dimensión del dataset y los tipos de datos. Para el top 5 de los atributos que considere más importantes, incluya análisis univariados apropiados que permitan determinar su comportamiento o distribución. Destaque cualquier insight que sea relevante para las actividades posteriores. *Evite utilizar herramientas automáticas como Pandas Profiling.*
2. **[20%] Análisis de calidad y preparación de datos:** Identifique potenciales problemas en los datos y clasifíquelos de acuerdo a la dimensión de calidad correspondiente. En la medida de lo posible, implemente alguna estrategia para la

corrección de dicho problema de calidad de datos. Justifique brevemente cada estrategia.

3. **[20%] Relaciones entre variables:** Aplique técnicas de exploración de datos (gráficas o no gráficas) que le permitan entender la relación que existe entre los atributos que considera más importantes para el análisis. Asegúrese de aplicar la técnica correcta (o la más efectiva) dependiendo de los atributos a analizar. Destaque cualquier insight que sea relevante para las actividades posteriores.
4. **[20%] Pruebas de hipótesis:** Corra una o dos pruebas de hipótesis sobre atributos clave para su análisis. Asegúrese de realizar el proceso siguiendo los pasos recomendados: formulación de las hipótesis nula y alternativa, validación de supuestos, seleccionar la prueba adecuada, hacer los cálculos correspondientes e interpretar los resultados.
5. **(Bono) [20%] Modelos multivariados:** Utilice alguna técnica de análisis multivariado que le permita identificar la relación que existe entre más de dos variables simultáneamente, como por ejemplo regresión lineal, análisis de componentes principales, entre otros. Explique por qué la técnica aplicada es útil en el contexto de su análisis e interprete los resultados (es decir, los coeficientes de la regresión o los componentes principales).
6. **[20%] Generación de la recomendación:** Defina los criterios explícitos que utilizaría para recomendar el “mejor carro a comprar” a partir de un perfil de comprador hipotético, como por ejemplo, relación precio/año, precio/km, tipo de combustible, transmisión. Puede proponer un *score* compuesto que le permita establecer un ranking sobre todos los carros disponibles. Justifique apropiadamente la estrategia de recomendación y muestre la lista de carros recomendados.
7. **[5%] Transparencia en el uso de IA:** Si usa alguna IA, incluya los prompts exactos y describa brevemente en qué parte del trabajo influyó (2-4 bullets). En caso contrario, justifique la razón para no utilizar ninguna.

INSTRUCCIONES DE ENTREGA

- La entrega debe hacerse por Bloque Neón antes de la hora estipulada. **Solo es posible realizar un envío.**
- Adjunte el **notebook ejecutado** con el desarrollo de cada una de las actividades planteadas de forma secuencial, clara, ordenada y en el formato original .ipynb.

Ingeniería de Sistemas y Computación

Escuela de Posgrado

MINE-4101: Ciencia de Datos Aplicada

Semestre: 2025-20

Horario: Jueves de 6:00 a 9:00 p.m.

Sábado de 9:00 a 12:00 a.m.

Escuela de Posgrado
Departamento de Ingeniería
de Sistemas y Computación

- Utilice bloques tipo markdown para demarcar el inicio y fin de cada actividad, así como las diferentes explicaciones y/o interpretaciones solicitadas. Se recomienda utilizar la plantilla proporcionada.
- Solo es posible utilizar el conjunto de datos proporcionado.