



Parcial I

Objetivo general

Aplicar las herramientas vistas en clase (Pandas, Seaborn, Matplotlib) para realizar un análisis exploratorio de datos (EDA) sobre un conjunto de datos reales elegido por el estudiante. Se evaluará la preparación de datos, la construcción de visualizaciones y la comunicación de hallazgos.

1) Elección y descripción del dataset

Se deberá seleccionar un dataset real, público y descargable (p. ej., datos.gob.ar, INDEC, Kaggle). El dataset deberá contener al menos tres variables numéricas continuas, una variable temporal y una categórica. El dataset puede ser obtenido de una o varias fuentes.

En el informe se incluirá:

- Fuente del dataset (enlace o referencia precisa).
- Breve descripción del tema (1–2 oraciones).

2) Preparación de los datos

- Carga en Pandas y estandarización de nombres de columnas.
- Conversión de la columna temporal a tipo fecha y creación de columnas auxiliares (mes, año), si aplica.
- Auditoría de calidad: detección y tratamiento de nulos, duplicados y tipos incorrectos.
- Derivación de métricas: por ejemplo, variaciones porcentuales mensuales.

3) Análisis descriptivo

- Medidas de tendencia central (media, mediana).
- Medidas de dispersión (desviación estándar, rango).
- Medidas de posición (mínimo, máximo, cuartiles cuando corresponda).
- Interpretación breve en lenguaje claro (2–4 oraciones).

Introducción al Análisis de Datos

Tecnatura Universitaria en Programación

4) Visualización

- Histograma simple de una variable numérica.
- Histograma superpuesto de dos variables comparables.
- Boxplot por categoría o mes.
- Scatterplot bivariado con una dimensión adicional (color o tamaño).
- Heatmap de correlaciones entre 3 o más variables numéricas.
- Opcional: matriz de dispersión (pairplot) para un subconjunto de variables.
- Cada gráfico deberá acompañarse con una interpretación concisa (1–2 oraciones).

5) Conclusiones

- Síntesis de los hallazgos principales.
- Relación con la pregunta o motivación del análisis.
- Limitaciones del dataset o del enfoque aplicado.
- Preguntas que podrían investigarse a futuro.

6) Entregables

- Archivo .py con el código completo y comentarios.
- Informe breve (1–3 páginas) en PDF o Word, con links a las fuentes de datos, gráficos insertados y texto explicativo.

7) Criterios de evaluación

Criterio	Descripción	Peso
Selección y comprensión del dataset	Pertinencia de la temática y claridad en la descripción de la fuente.	15%
Preparación de los datos	Limpieza, tipos y creación de columnas auxiliares/derivadas.	20%
Gráficos e interpretación	Claridad visual, elección adecuada de gráficos y lectura coherente de resultados.	40%
Presentación del informe	Estructura, redacción y síntesis; calidad de figuras y relato.	15%
Creatividad y autonomía	Valor agregado en preguntas, enfoques o herramientas empleadas.	10%

Introducción al Análisis de Datos

Tecnicatura Universitaria en Programación

8) Formato y presentación

- Se recomienda el uso de títulos informativos en todos los gráficos.
- Se sugiere incluir comentarios en el código para facilitar la corrección.
- Los ejes deberán estar etiquetados con unidades cuando corresponda.
- Se deberá mantener un orden lógico en el archivo .py (secciones claras).

9) Entrega y plazos

La fecha y hora límite serán informadas en el aula virtual. Se deberán cargar el archivo .py y el informe (PDF o Word) en el mismo espacio.

10) Sugerencias de fuentes de datos

- Portal de Datos Abiertos (datos.gob.ar).
- INDEC (series e informes).
- Kaggle (datasets públicos con fines educativos).

 Nota: El parcial puede ser desarrollado en grupos de 1 a 4 estudiantes.