Máster Universitario en *Data Analytics for Business*

Análisis exploratorio de datos

Alexandra Abós

upf. BARCELONA SCHOOL OF MANAGEMENT

Noviembre-2025

Proyectos

1.

Proyectos

1. Introducción a los proyectos
2. Estructura de los informes
3. Datasets

upf. BARCELONA SCHOOL OF MANAGEMENT

1

Introducción a los proyectos

La asignatura de **Análisis exploratorio de datos** se evalúa con un trabajo en grupo.

- 3 personas por grupo y, en total, 8-9 grupos.
- El trabajo consiste en un memoria escrita (50% de la nota) y una presentación oral (50% de la nota).
- La memoria escrita se tiene que entregar en formato informe (en pdf) y en notebook (código).
- La presentación y la entrega de la memoria serán el último día de clase.

2

Estructura del informe

El informe tendrá que seguir los pasos del análisis exploratorio de datos que veremos en la asignatura:

1. Introducción al dataset (resumen y explicación de los datos)

2. Detección y corrección de valores faltantes (missings) y outliers

3. Exploración de variables numéricas y categóricas

4. Relaciones y Correlaciones entre Variables

2

Estructura del informe

5. Análisis de tendencias temporales

6. Interpretación de los datos

7. Conclusiones y posibles siguientes pasos

8. Opcional:

- Usar otros datasets para enriquecer/mejorar el análisis
- Construcción de modelos estadísticos o de machine learning con los datos trabajados

* Es importante usar distintos métodos de visualización en las diferentes etapas.

2

Datasets

Puntos comunes en todos los datasets

- Análisis y corrección de duplicados/missings y outliers. Decidid la mejor manera de tratarlos y razonad vuestra elección.

- Exploración de variables numéricas y categóricas.

- Tratamiento de la información temporal (fecha).
  - Corrección del formato para poner las fechas en formato estándar (Y/m/d)
  - Sacar información del año y del mes por separado

- Análisis de tendencias temporales. ¿Podemos observar tendencias positivas o negativas en nuestras variables?

- Análisis de correlación entre variables.

2

Datasets

Proyecto 1: Gym

- 2 datasets con información sobre usuarios de gym y su historial de visitas.
- Podéis juntar los datasets o trabajar con ellos por separado.
- Cread una columna con el número de mes de la fecha de check-in.
- Ejemplo de preguntas a responder:
  - Hay diferencias entre los usuarios de distintos gyms/tarifas?
  - ¿Qué actividades son las más realizadas?
  - ¿Cuál es el tiempo medio de entreno? Hay diferencias según el perfil del usuario?
  - ¿Hay diferencias entre meses del año?

Datasets

Proyecto 2: Los Angeles Crime Data 2020-2023

- 1 dataset con información de crímenes en Los Angeles entre 2020-2023.
- Cread una columna con el número de mes de la fecha del crimen.
- Ejemplo de preguntas a responder:
  - Analizad las características de las víctimas. Hay alguna edad/género más frecuente según la categoría de crimen?
  - Analizad los patrones temporales de los delitos en diferentes zonas de la ciudad de Los Ángeles. Hay zonas con tendencia creciente? Y decreciente?
  - Hay algún mes donde haya frecuentemente más crímenes?

* Información de los datos en: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data

Datasets

Proyecto 3: TV shows

- 1 dataset con información de series de TMDB.
- Cread una columna con el número de mes de la fecha de estreno de la serie.
- En las columnas number_of_seasons, number_of_episodes, vote_count, vote_average, popularity y episode_run_time, los missings se codifican con un 0, os parece una manera correcta?
- Ejemplo de preguntas a responder:
  - Explorad las tendencias en la popularidad de los programas de televisión según el recuento de votos y el promedio
  - Hay alguna relación entre el número de temporadas/episodios de las series y sus valoraciones (ratings)?
  - Investigad las tendencias de producción de programas de televisión en diferentes países y plataformas.

Datasets

Proyecto 4: Películas

- 1 dataset con información de películas de TMDB.
- Cread una columna con el número de mes de la fecha del estreno.
- Ejemplo de preguntas a responder:
  - Identificad tendencias en las fechas de estreno de películas. ¿Hay algún mes que destaque por encima de los otros a nivel de estrenos?
  - Analizad la relación entre presupuesto, ingresos y popularidad para determinar los factores que contribuyen al éxito de una película.
  - ¿Qué genero de películas es el más popular? Hay algún género que tenga una tendencia negativa respecto a la popularidad?

Datasets

Proyecto 5: Barcelona AirBnB

- 1 dataset con información sobre pisos de AirBnB en Barcelona.
- Cread una columna que sea el número de meses entre la 1ª y la última reseña.
- Ejemplo de preguntas a responder:
  - ¿En qué barrios hay más pisos de AirBnB? ¿Y dónde tienen la mejor puntuación?
  - ¿Qué perfil tienen los pisos según el barrio? ¿Hay diferencias?
  - ¿Hay pisos que llevan mucho tiempo en AirBnB? ¿O la mayoría son nuevos?
  - ¿Hay propietarios que tienen más de 1 piso en AirBnB? ¿Cuál es el número máximo de pisos?
  - ¿Qué tipo de pisos tiene mejores puntuaciones?

Datasets

Proyecto 6: Steam Games

- 1 dataset con información de juegos de Steam.
- Cread una columna con el número de mes de la fecha de estreno.
- Ejemplo de preguntas a responder:
  - Analizad los tipos de juegos. ¿Qué género tiene más juegos? ¿Cuál tiene la mejor puntuación en la columna Metacritic? Los juegos más caros son los que tienen mejor puntuación?
  - ¿Qué juego es el más recomendado? ¿Y a nivel de género?
  - ¿En qué plataforma hay más juegos disponibles?
  - Analizad las tendencias de estrenos de juegos a lo largo del tiempo e identificad cualquier patrón de estacionalidad. ¿Hay algún mes que destaque por sus estrenos?

Datasets

Proyecto 7: Spotify weekly

- 1 dataset con información semanal de Spotify.
- Cread una columna con el número de mes de la fecha de estreno de la canción.
- Ejemplo de preguntas a responder:
  - ¿Qué artista/grupo tiene más canciones en las listas de más escuchados?
  - Hay alguna relación entre las características de las canciones y su popularidad?
  - Analizad las tendencias semanales de las listas más escuchadas. Explorad la tendencia de canciones por país. ¿Hay países donde las listas son más estables?
  - Hay algún artista/grupo que aparezca consistentemente en el top10 durante el último mes?

2

Datasets

Proyecto 8: Juegos Olímpicos Paris 2024

- 2 datasets con información de los atletas y las medallas de los juegos olímpicos de París 2024
- Calculad el número de medallas por país teniendo en cuenta el tipo de medalla.
- También está disponible la información por deporte (si se quiere usar).
- Ejemplo de preguntas a responder:
  - ¿Qué país ha ganado más medallas de oro? Y de plata? Y de bronce?
  - ¿Hay algún país que destaque en alguna disciplina?
  - Hay algún atleta que tenga medallas en distintas disciplinas?
  - ¿Los anfitriones mejoraron su posición (en medallas) comparado con años anteriores?
  - ¿Los atletas ganadores tienen características parecidas?

Datasets

Proyecto 9 Marketing Campaign

- 1 dataset con información sobre clientes de una campaña de marketing.
- Cread una columna con el número de mes de la fecha de compra.

- Ejemplo de preguntas a responder:

  - ¿Qué perfil tienen los clientes de esta campaña de marketing?
  - ¿Qué tipo de producto se compra más según el perfil?
  - ¿Qué tipo de plataforma se usa más?
  - ¿Qué tipo de cliente aceptó la campaña (respuesta positiva)?
  - ¿Hay alguna relación entre el salario y el número de productos comprados?

¿PREGUNTAS?

Alexandra Abós

alexandra.abos@bsm.upf.edu

¡GRACIAS!

BARCELONA SCHOOL OF MANAGEMENT

Alexandra Abós

alexandra.abos@bsm.upf.edu