# **Module-2-Real\_state\_EDA**

## **1. Introducción**

En el contexto de este trabajo, el ayuntamiento de Barcelona pide a **McEloy & Hervas, consultora de data analytics,** trabajar con un dataset para investigar sobre las siguientes preguntas de investigación y verificar las hipótesis del ayuntamiento sobre la realidad del precio en la ciudad condal de Barcelona.

1. **¿Cuáles son las zonas con pisos más caros?**

Hipótesis: El Distrito de Sarrià-Sant Gervasi es la zona con pisos más caros de Barcelona

1. **¿Hay una correlación entre la orientación del piso y el precio?**

Hipótesis: Sí hay correlación positiva entre la orientación del piso y el precio en los pisos de Barcelona

1. **¿De qué década del siglo XX son los pisos más caros?**

Hipótesis: Los pisos más caros de Barcelona son los de los años 80

1. **¿Hay correlación entre el coste del piso en Barcelona y la distancia a metro o calles emblemáticas como Diagonal?**

Hipótesis: Sí hay correlación positiva entre la distancia a estos servicios y la cercanía.

Para este caso se ha utilizado un archivo CSV que recoge más de 67.000 registros de pisos en Barcelona. En este dataset se ha seleccionado los pisos construidos durante el Siglo XX para poner foco en la investigación y eliminar outliers previamente detectados.

## **2. Objetivo y pasos del proyectos**

El objetivo principal es extraer insights y corroborar las hipótesis que el ayuntamiento de Barcelona tiena. Para ello se inicia un proceso de data wrangling y data visualization.

**2.1. Definir preguntas de investigación**

* **¿Cuáles son las zonas con pisos más caros?**
* **¿Hay una correlación positiva entre la orientación del piso y el precio?**
* **¿De qué década del siglo XX son los pisos más caros?**
* **¿Hay correlación positiva entre el coste del piso en Barcelona y la distancia a metro o calles emblemáticas como Diagonal?**

### **2.2 Data Wrangling**

En este proceso se seleccionaron las columnas con las que trabajar y se eliminaron las que no contribuían a cumplir nuestro objetivo objetivo principal, además de seguir todo el proceso estándar de Data Wrangling: Tratamiento de nulos, outliers, registros duplicados, formateo, agrupación y organización de los datos. Aquí puede ver el notebook: shark\_attack\_project.ipynb.

### **2.3. Data Visualization**

Una vez se finaliza el proceso de data wrangling se procede al proceso de visualización de datos para obtener insights y oportunidades. Para esto, hemos utilizado la librería seaborn.

**3. Tecnología utilizada**

* Lenguaje de programación
  + Python
* Librería
  + Numpy
  + Pandas
  + Seaborn
  + Matplotlib