

## PROYECTO 1

### **Análisis exploratorio y modelado predictivo de precios de viviendas en Barcelona usando Python y SQL**

**(Fecha límite de entrega: 4 de noviembre del 2024 hasta las 23:59)**

#### **Objetivo:**

Desarrollar un análisis completo y un modelo predictivo para los precios de viviendas en Barcelona, utilizando datos extraídos del portal Fotocasa. El objetivo es aplicar técnicas de extracción, manipulación y análisis de datos, así como algoritmos de Machine Learning, para predecir los precios de las viviendas en función de diversas características.

#### **Dataset:**

[https://www.kaggle.com/datasets/imanollaconcha/barcelona-fotocasa-housingprices?select=Barcelona\\_Fotocasa\\_HousingPrices.csv](https://www.kaggle.com/datasets/imanollaconcha/barcelona-fotocasa-housingprices?select=Barcelona_Fotocasa_HousingPrices.csv)

#### **Requisitos:**

1. **SQL y Python:** Debéis ser capaces de extraer, transformar y cargar datos (ETL) utilizando tanto SQL como Python. El uso de bases de datos relacionales es clave para gestionar grandes volúmenes de datos de viviendas.
2. **Uso de GitHub:** Toda la documentación, el código y las versiones del proyecto deben estar gestionados en GitHub. Esto implica el uso de control de versiones y colaboración efectiva a través de esta plataforma.
3. **Ética de la IA y Gobernanza de Datos:** El proyecto debe considerar la privacidad de los datos y el impacto ético de las predicciones. Se requiere que el equipo analice la gobernanza de los datos utilizados, asegurando que el tratamiento de la información cumple con las normativas vigentes (e.g., GDPR).
4. **Exploratory Data Analysis (EDA):** Se debe realizar un EDA exhaustivo utilizando Python, que incluya la visualización de las distribuciones de

precios, identificación de patrones en los datos (por ejemplo, según la ubicación o características del inmueble) y detección de posibles outliers.

5. **Modelos de Machine Learning:** Se requiere aplicar y evaluar diferentes algoritmos de ML, tales como:

- **Regresión Lineal:** Como punto de partida para predecir el precio.
- **Algoritmos avanzados (si se justifica):** Podrían utilizarse árboles de decisión, bosques aleatorios o cualquier otro modelo que ofrezca mejores resultados predictivos.

**Pasos del Proyecto:**

1. **Extracción y preparación de datos:**

- **Uso de SQL y Python:** Extraer los datos de precios de viviendas desde el portal Fotocasa. Se requiere que los datos sean almacenados en una base de datos y manipulados usando SQL para su posterior análisis.
- **Formulación del problema de negocio:** El equipo debe definir claramente el problema de predicción, identificando las variables relevantes como la ubicación (código postal o coordenadas), tamaño del inmueble, número de habitaciones, antigüedad, entre otras.
- **Preparación del dataset:** A través de Python y SQL, se debe limpiar y transformar el dataset para eliminar inconsistencias, manejar valores nulos y realizar una correcta codificación de las variables categóricas.

2. **Análisis Exploratorio de Datos (EDA):**

- Utilizar herramientas como **Jupyter Notebook** o **Google Colab** para realizar un análisis exploratorio detallado, visualizando los patrones en los datos y generando insights clave que sirvan como base para la construcción del modelo.
- **Visualización de datos:** Crear gráficos que representen la distribución de precios, tendencias por zona geográfica, correlaciones entre las variables, etc. El objetivo es entender qué variables influyen más en los precios.

3. **Modelado Predictivo:**

- **Desarrollo del modelo:** Utilizar los datos preparados para construir un modelo de predicción de precios de viviendas utilizando algoritmos de **Machine Learning** (e.g., regresión lineal).

- **Validación del modelo:** Evaluar el rendimiento del modelo utilizando métricas de precisión y ajustar el modelo según sea necesario. Se debe realizar una comparación entre diferentes algoritmos para seleccionar el más adecuado.

#### **4. Entrega y documentación:**

- **Integración con GitHub:** Debes subir todo el código, la documentación y el trabajo final a una cuenta de GitHub, asegurándose de hacer un seguimiento de los cambios mediante control de versiones.
- **Documentación del EDA y modelado:** Redactar un informe que detalle el análisis exploratorio, los resultados de la modelización y las recomendaciones finales basadas en los datos obtenidos.

#### **Resultados Esperados:**

1. **Análisis Exploratorio Completo:** Informe detallado con visualizaciones y conclusiones clave sobre el mercado de vivienda en Barcelona.
2. **Modelo Predictivo:** Modelo capaz de predecir el precio de una vivienda basándose en las características del inmueble con un margen de error aceptable.
3. **Código en GitHub:** Repositorio completo con el código utilizado, explicaciones y documentos asociados.