

FINAL ASSIGNMENT DA BBDD

El famoso portal inmobiliario idealista, con actividad en España, Andorra, Italia y Portugal, reveló hace poco un dataset con información de 2018.

Este dataset cuenta con información relativa a las coordenadas aproximadas de los inmuebles (latitud y longitud), precios de venta de cada vivienda listada y diversas variables de características interiores. Los listados se enriquecieron con información oficial del catastro español (por ejemplo, calidad de los materiales de construcción) y otras características geográficas relevantes, como la distancia a puntos de interés urbano.

En este caso nos vamos a centrar en el dataset con información sobre la ciudad de **Madrid**.

Aquí tienes una descripción de las columnas:

ASSETID: Un identificador único para cada propiedad o activo en el conjunto de datos. Es una variable categórica con múltiples niveles.

PERIOD: Periodo de tiempo en el que se registraron los datos, normalmente representado como un número entero de seis dígitos (AAAAMM) que indica el año y el mes.

PRICE: El precio total de la propiedad en la moneda local (€). Se trata de un valor numérico.

UNITPRICE: El precio por unidad de superficie (metro cuadrado) del inmueble. También es un valor numérico.

CONSTRUCTEDAREA: La superficie total construida de la propiedad en metros cuadrados. Es un valor entero que indica el tamaño de la propiedad.

ROOMNUMBER: Número de habitaciones del inmueble. Este valor entero incluye todos los tipos de habitaciones (por ejemplo, dormitorios, salas de estar).

BATHNUMBER: Número de cuartos de baño de la vivienda. Se trata de un valor entero.

HASTERRACE: Indicador binario (0/1) en el que 1 indica que el inmueble tiene terraza y 0 que no.

HASLIFT: Un indicador binario (0/1) en el que 1 indica que el inmueble dispone de ascensor y 0 que no.

HASAIRCONDITIONING: Indicador binario (0/1) en el que 1 indica que el inmueble dispone de aire acondicionado y 0 que no.

HASPARKINGSPACE: Un indicador binario (0/1) donde 1 indica que la propiedad incluye una plaza de aparcamiento y 0 indica que no.

ISPARKINGSPACEINCLUDEDINPRICE: Indicador binario (0/1) en el que 1 indica que la plaza de aparcamiento está incluida en el precio de la vivienda y 0 que no lo está.

HASNORTHORIENTATION: Indicador binario (0/1) en el que 1 indica que la propiedad tiene orientación norte y 0 que no.

HASSOUTHORIENTATION: Indicador binario (0/1) en el que 1 indica que la propiedad tiene orientación sur y 0 que no.

HASEASTORIENTATION: Indicador binario (0/1) en el que 1 indica que la propiedad tiene orientación este y 0 que no.

HASWESTORIENTATION: Indicador binario (0/1) en el que 1 indica que la propiedad tiene orientación oeste y 0 que no.

HASBOXROOM: Un indicador binario (0/1) en el que 1 indica que la propiedad incluye un trastero (una pequeña habitación utilizada normalmente como almacén) y 0 indica que no.

HASWARDROBE: Un indicador binario (0/1) donde 1 indica que la propiedad incluye un armario empotrado y 0 indica que no.

HASSWIMMINGPOOL: Un indicador binario (0/1) donde 1 indica que la propiedad incluye una piscina y 0 indica que no.

HASDOORMAN: Indicador binario (0/1) en el que 1 indica que el inmueble dispone de portero o conserje y 0 que no.

HASGARDEN: Indicador binario (0/1) en el que 1 indica que la propiedad tiene jardín y 0 que no lo tiene.

ISDUPLEX: Un indicador binario (0/1) en el que 1 indica que la propiedad es un dúplex (una unidad de dos plantas) y 0 indica que no lo es.

ISSTUDIO: Un indicador binario (0/1) en el que 1 indica que la propiedad es un estudio (un apartamento de una sola habitación con zonas de estar y dormitorio combinadas) y 0 indica que no lo es.

ISINTOPFLOOR: Indicador binario (0/1) en el que 1 indica que el inmueble está situado en la última planta del edificio y 0 que no lo está.

CONSTRUCTIONYEAR: Año de construcción del inmueble. Se trata de un valor entero, aunque pueden faltar algunas entradas (NA).

CADCONSTRUCTIONYEAR: Año de construcción del inmueble según los registros catastrales. Se trata de un valor entero.

CADMAXBUILDINGFLOOR: Número máximo de plantas del edificio según los registros catastrales. Se trata de un valor entero.

CADDWELLINGCOUNT: Número de unidades de vivienda del edificio según los registros catastrales. Se trata de un valor entero.

CADASTRALQUALITYID: Número entero que representa la calidad o clasificación catastral del edificio, probablemente basado en normas legales o gubernamentales.

BUILTTYPEID_1: Indicador binario (0/1) en el que 1 indica que la propiedad es de nueva construcción (sin propietarios anteriores) y 0 que no lo es.

BUILTTYPEID_2: Un indicador binario (0/1) en el que 1 indica que la propiedad es de segunda mano para restaurar y 0 indica que no lo es.

BUILTTYPEID_3: Indicador binario (0/1) en el que 1 indica que el inmueble es de segunda mano y está en buen estado, y 0 que no lo es.

DISTANCE_TO_CITY_CENTER: Valor numérico que representa la distancia (kilómetros) entre la propiedad y el centro de la ciudad.

DISTANCE_TO_METRO: Valor numérico que representa la distancia (kilómetros) entre el inmueble y la estación de metro más cercana.

DISTANCE_TO_CASTELLANA: Valor numérico que representa la distancia (kilómetros) de la vivienda a la avenida principal de cada ciudad, en Madrid es Paseo de la Castellana o Castellana.

LONGITUDE: Longitud geográfica de la ubicación de la propiedad. Se trata de un valor numérico que indica la posición este-oeste.

LATITUDE: La latitud geográfica de la ubicación de la propiedad. Se trata de un valor numérico que indica la posición norte-sur.

PRIMERA PARTE (8,5 ptos)

0. Crea una nueva base de datos llamada 'idealista' y sube la tabla que se te proporciona. (0,5 ptos)

Utiliza tu conocimiento de SQL para responder a las siguientes preguntas (8 ptos):

1. Muestra las columnas ASSETID, PRICE y CONSTRUCTEDAREA de todas las propiedades del dataset.
2. ¿Qué meses están informados en el dataset?
3. ¿Cuáles son las propiedades con precio mayor a 500.000€? Muestra solo las columnas mencionadas en la pregunta anterior.
4. Muestra el ASSETID, PRICE y DISTANCE_TO_CITY_CENTER de las propiedades, ordenadas por la distancia al centro de Madrid, de menor a mayor.
5. Selecciona todas las propiedades que tengan más de 3 habitaciones y un precio mayor a 300000€. Muestra el ASSETID, ROOMNUMBER y PRICE.
6. ¿Cuáles son las propiedades que tienen al menos 2 baños, un ascensor y una plaza de aparcamiento? Selecciona el ASSETID, BATHNUMBER, HASLIFT, HASPARKINGSPACE y PRICE
7. ¿Cuál es el precio promedio de las propiedades con terraza?
8. ¿Cuántas propiedades hay según el número de habitaciones? ¿Cuál es el precio medio de los apartamentos según el número de habitaciones?
9. ¿Cuáles son las propiedades cuyo precio es mayor que el precio promedio de todas las propiedades en el dataset? Muestra el ASSETID, PRICE y CONSTRUCTEDAREA
10. ¿Cuáles son las propiedades que están a menos de 2 kilómetros de Castellana? Selecciona el ASSETID, PRICE y DISTANCE_TO_CASTELLANA.
11. Selecciona todas las propiedades que son dúplex y tienen más de 100 metros cuadrados de superficie construida. Muestra el ASSETID, PRICE, CONSTRUCTEDAREA y ROOMNUMBER.

12. ¿Hay alguna tendencia en el valor medio de las propiedades en los distintos meses del año 2018? ¿Y en el valor máximo o mínimo?
13. Supón que quieres aumentar en un 5% el precio de todas las propiedades que tienen aire acondicionado. Haz una actualización del campo PRICE para reflejar este cambio.
14. Elimina todas las propiedades que tengan menos de 30 metros cuadrados de superficie construida.
15. Crea una vista que muestre solo las propiedades con terraza y piscina, incluyendo los campos ASSETID, PRICE, ROOMNUMBER y BATHNUMBER.
16. Calcula el precio promedio de las propiedades nuevas y de las de segunda mano en buen estado, y muestra ambos precios.
17. Calcula el número total de propiedades que tienen terraza y piscina, y cuántas de ellas también tienen ascensor. **Utiliza la función CASE** para contar las propiedades con estas combinaciones.

SEGUNDA PARTE (1,5 ptos)

Supongamos que tienes una tabla adicional que contiene información sobre las agencias inmobiliarias que listan las propiedades.

Tabla imaginaria: agencias

Esta tabla contiene información sobre las agencias inmobiliarias que gestionan las propiedades. Aquí tienes una descripción de las columnas:

- AGENCY_ID: Un identificador único para cada agencia inmobiliaria. (clave primaria)
- AGENCY_NAME: Nombre de la agencia inmobiliaria.
- RATING: Puntuación promedio que han dado los usuarios a la agencia (valor entre 1 y 5).

Ejemplo de Datos en la tabla agencias

AGENCY_ID	AGENCY_NAME	RATING
1	Agencia Madrid Plus	4.5
2	Madrid Viviendas	4.2
3	Madrid Propiedades	4.8
4	Madrid Casas Deluxe	4.7
5	Inmo Casa Madrid	4.1

En el dataset original, añadiríamos una columna adicional (no es necesario que lo hagas):

- AGENCY_ID: Identificador que hace referencia a la agencia inmobiliaria que gestiona la propiedad (se relaciona con la tabla agencias).

Responde a las siguientes preguntas (simplemente escribe la query en tu notebook, se valorará la capacidad de abstracción para relacionar tablas):

1. Muestra las propiedades junto con la agencia que los gestiona.
2. ¿Cuáles son todas las propiedades gestionadas por una agencia con puntuación mayor a 4.5? Muestra el ASSETID, AGENCY_NAME, PRICE y RATING.
3. ¿Cuál es el precio medio de las propiedades listadas por cada agencia?