



# Paris Housing

Data Science - 46295  
Profesor Jorge Ruiz  
Tutor Sergio Dagna

Maicol Pastor

***CODER HOUSE***

# Índice

- Introducción
  - Objetivo
  - Contexto comercial
  - Contexto analítico
  - Hipótesis
- Adquisición de datos
- EDA
  - Gráficos e insights
  - Matriz de correlación
  - Conclusiones
- Selección de modelos
- Conclusión final

# Introducción

Paris Housing  
Maicol Pastor

***CODER HOUSE***



# Objetivo

Dada la necesidad en la industria inmobiliaria de tener acceso a un estimativo de precio rápido y preciso de una propiedad, se presenta el problema para realizar un modelo que realice una predicción del precio.

Siendo el objetivo mejorar rotundamente el tiempo de negociación y saber la rentabilidad de la propiedad.



# Contexto comercial

La rapidez y precisión a la hora de negociar es un aspecto clave en la industria inmobiliaria, debido a esto el proyecto estará más enfocado a un ámbito de ventas. La habilidad de evaluar correctamente los precios de las diversas propiedades en zonas de gran concurrencia como lo es París es algo de suma importancia, las acciones que se toman en base a esto puede determinar las ganancias o pérdidas en las ventas o adquisiciones de la empresa.



# Contexto analítico

Predecir el precio que tendrá la propiedad permitiendo así tener un estimativo temprano del valor de la misma en base a los rasgos más importantes.

Se realizarán gráficos, análisis y procesamiento de los datos, seleccionar un modelo de machine learning adecuado a los tipos de datos y variable objetivo.



# Hipótesis

- El precio aumenta en base al tamaño de las propiedades.
- El precio varía positivamente según el año que se construyeron.
- Mientras menos dueños haya tenido la propiedad más alto es el precio.
- La relación entre precio y pisos es positiva.
- ¿Alguna otra variable influirá en el precio de la propiedad?
- ¿Cómo se distribuyen los precios de las casas?
- ¿En qué años se construyeron más casas?
- ¿Mientras más antiguas más baratas las casa, o es lo contrario?
- ¿La cantidad de pisos estará relacionada con los metros cuadrados de las propiedades?
- ¿Existen zonas con menor valor que otras?

# Adquisición de datos

Paris Housing  
Maicol Pastor

***CODER HOUSE***





# Adquisición de datos

El dataset a utilizar es 'ParisHousing' extraído de Kaggle, contiene 17 columnas y 10000 filas.

Las variables de este son:

squareMeters	Metros cuadrados totales de la propiedad
numberOfRooms	Número de habitaciones
hasYard	Si tiene o no patio
hasPool	Si tiene o no piscina
floors	Cantidad de pisos
cityCode	Código identificativo de cada propiedad
cityPartRange	Parte de la ciudad a la que pertenece
numPrevOwners	Cantidad de dueños previos



made	Año en que se hizo la propiedad
isNewBuilt	Si es una construcción nueva o no
hasStormProtector	Si tiene o no Protector para tormentas
basement	Tamaño del sótano
attic	Tamaño del ático
garage	Tamaño del garaje
hasStorageRoom	Si tiene o no almacén
hasGuestRoom	Si tiene o no habitación para visitas
price	Precio de la propiedad

# EDA (Exploratory Data Analysis)

Paris Housing  
Maicol Pastor

***CODER HOUSE***

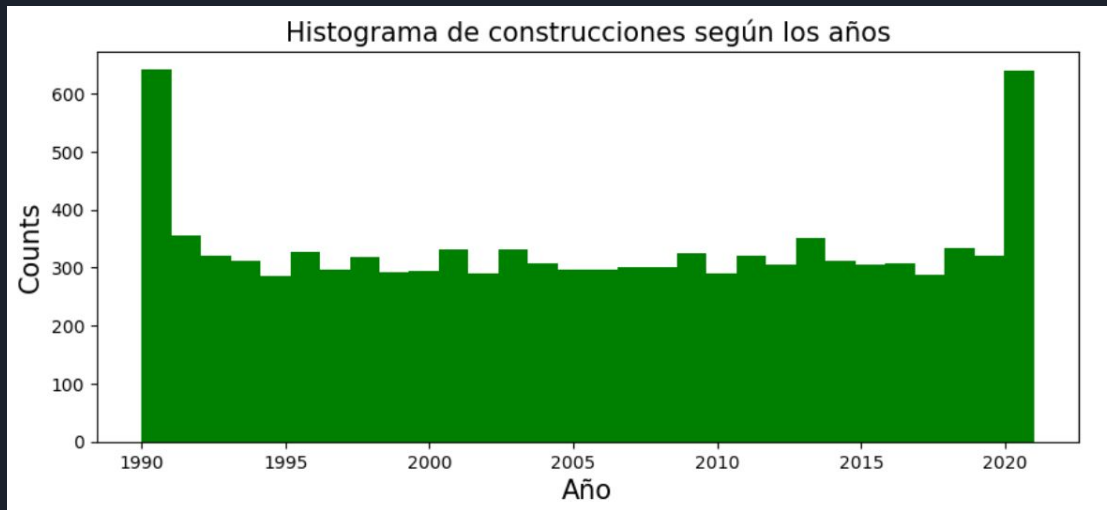
# Distribución de los precios

Se denota claramente una distribución multimodal con cuatro picos.



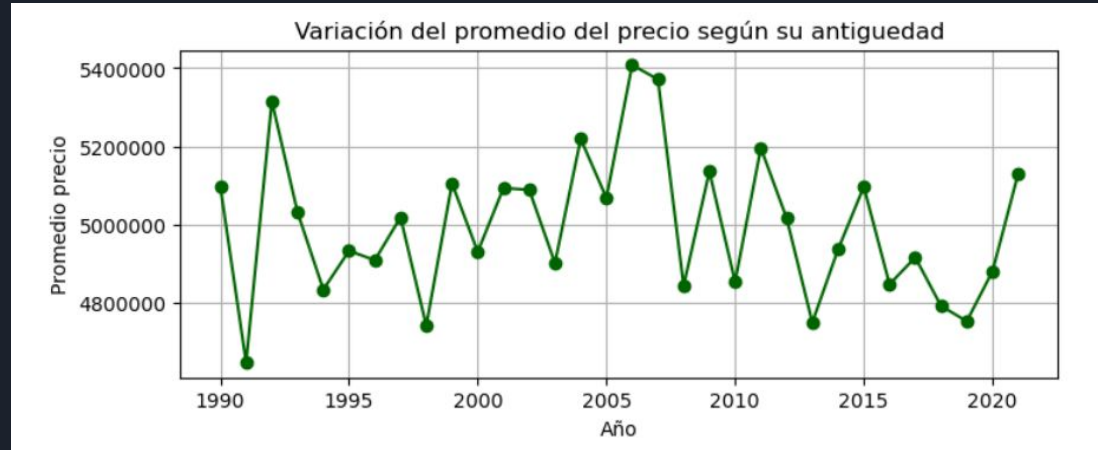
# Cantidad de construcciones por año

Se puede observar dos picos a lo largo del tiempo, tanto a mediados del 1990 como en el 2022, con una cantidad de más de 600 edificaciones.



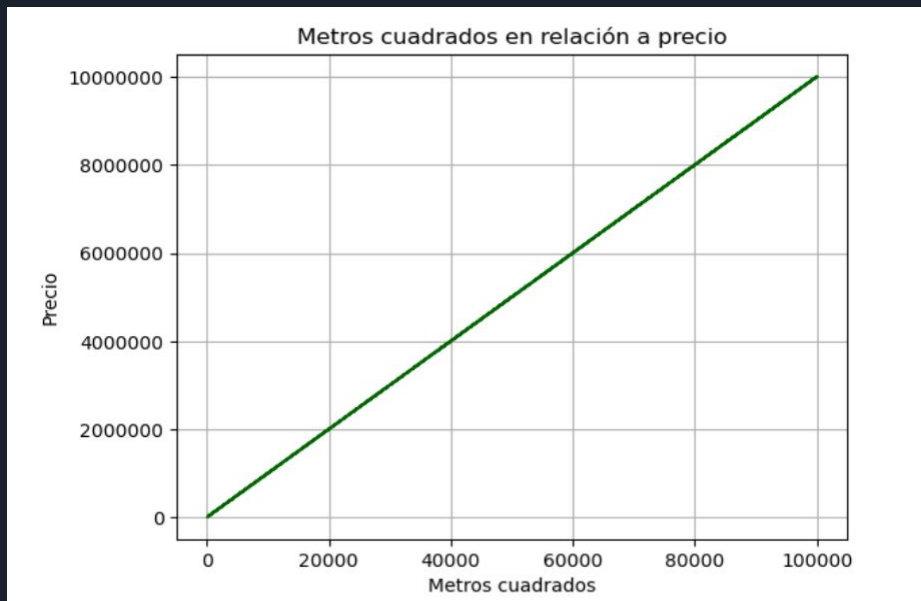
# Variación del promedio del precio según su antigüedad

En este gráfico de línea se visualiza la variación del precio según la antigüedad de las propiedades. Hay un pico de precios rondando el año 2006 y 2007.



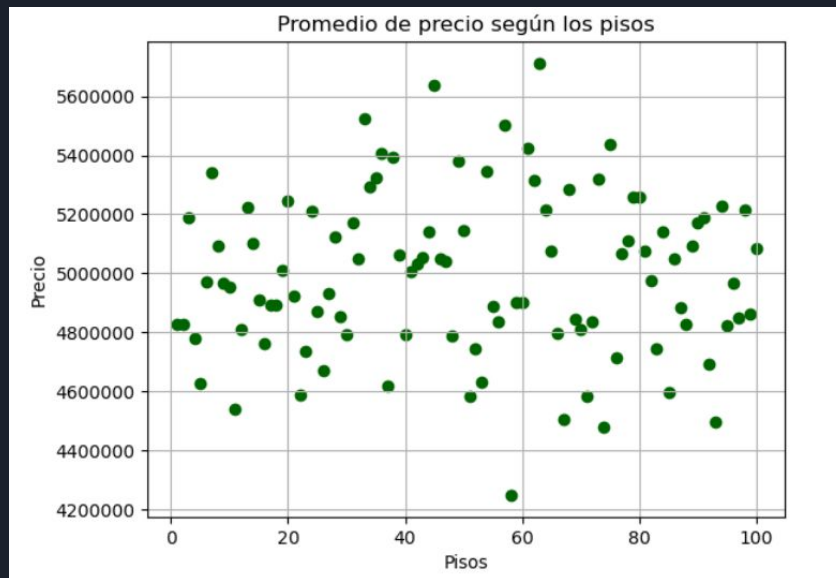
# Relación metros cuadrados y precio

Como primer gráfico se comprobó la relación entre precio y los metros cuadrados de las edificaciones, como resultado dió una correlación positiva tal como se había dicho en la hipótesis.



# Precio según los pisos

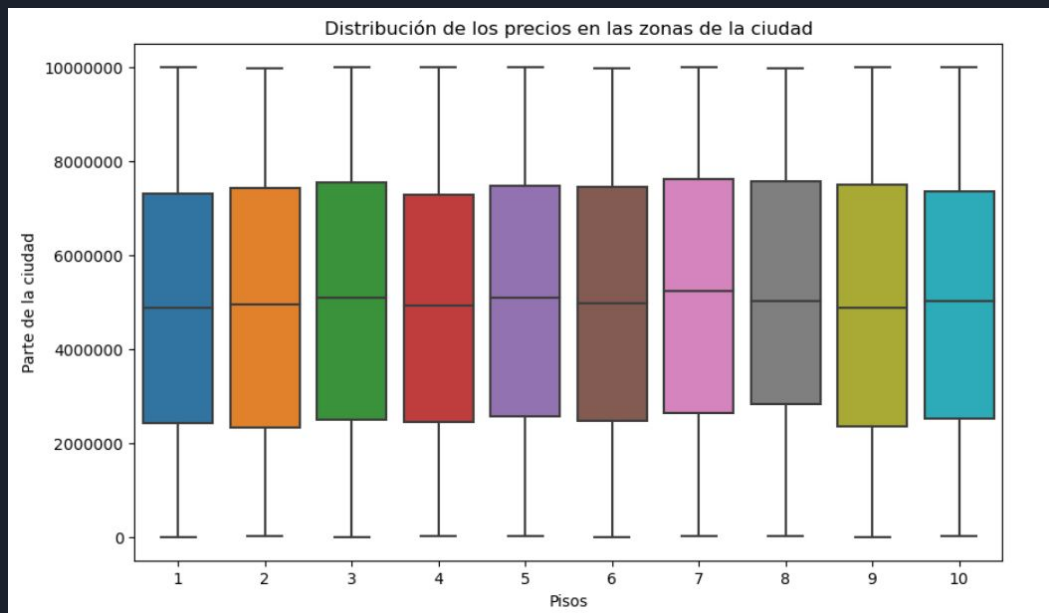
No existe relación alguna entre el precio y la cantidad de pisos.





# Precios según las zonas

No existe relación alguna entre el precio y la zona de la ciudad



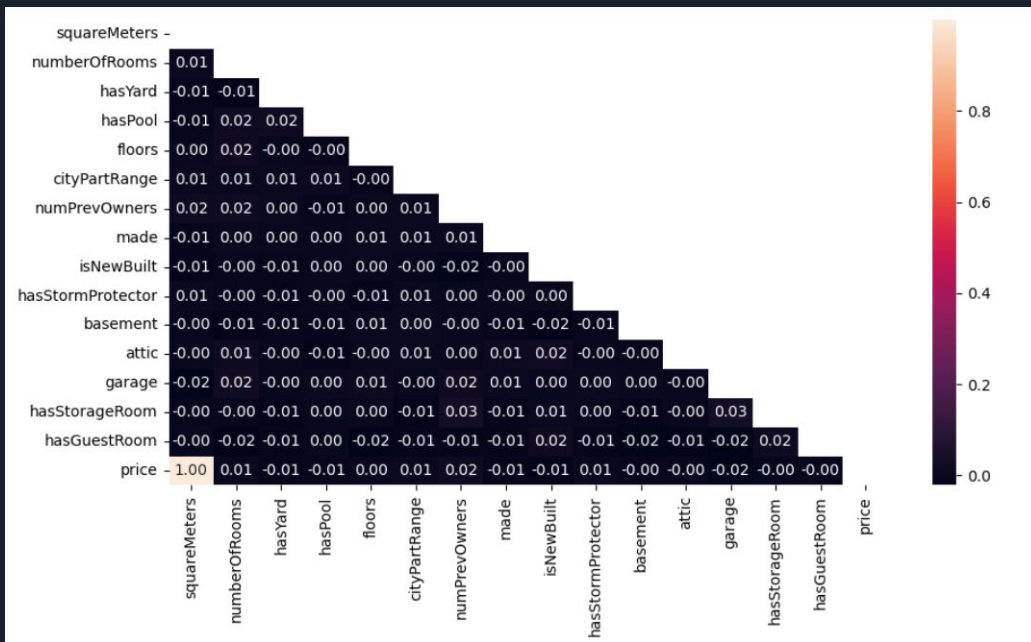
# Variación según el número de dueños anteriores

A pesar de ser muy variante, el precio suele aumentar mientras más dueños tuvo la propiedad.



# Matriz de correlación

En la matriz de correlación se aprecia la clara relación entre los metros cuadrados y el precio.





# Conclusiones

Como resultado del análisis, podemos comprobar que la única variable que se relaciona con 'price' es 'squareMeters', de forma lineal positiva, existe una fuerte correlación entre estas dos variables. Los datos son uniformes de tal manera que no hay distribuciones normales. Tanto la zona y la época, como lo pisos, no tienen relación con el precio. No se encontraron anomalías en los datos

# Selección de modelos

Paris Housing  
Maicol Pastor

***CODER HOUSE***



# Selección de modelos a mejorar

Se probaron varios modelos sin ajustes de hiperparámetros para evaluar rápidamente su rendimiento. Se utilizaron los siguientes modelos de regresión: Lasso, Linear Regression, Random Forest Regressor, K-Neighbors Regressor Y SVR (Support Vector Regressor).

De estos los mejores resultados fueron de Linear Regression, Random Forest Regressor y Lasso, con una precisión del 99% .



# Mejor modelo

Dado ese 99% de precisión en los modelos no era necesario realizar una búsqueda de hiperparámetros, pero igualmente se realizó la misma para probar los modelo más a fondo utilizando Cross Validation para evitar overfitting.

Los 3 modelos siguieron dando unos excelentes resultados.

# Conclusión final

Paris Housing  
Maicol Pastor

***CODER HOUSE***





# Conclusión final

Se puede observar una clara relación lineal, se intuye que la fórmula por la cual se calcula 'price' está directamente relacionada con 'squareMeters' y una constante que lo afecte, resultando en un gran porcentaje de precisión en los modelos.

El modelo que se utilizará será LinearRegressor. Este es un modelo mucho más simple, para este que es un caso de regresión lineal simple. Utilizar un modelo complejo para resolver este problema es inadecuado si hay un modelo más sencillo que logra cumplir con todas las expectativas sin utilizar un poder de cómputo tan grande.

# Muchas Gracias

Paris Housing  
Maicol Pastor



***CODER HOUSE***