Selección de DataSet

En https://www.kaggle.com/datasets encontramos el data set que utilizamos con nuestro modelo del TP. El mismo tiene muchas variables sobre casas como el año en la que se construyó, año en que se remodeló, el área del lote, cantidad de metros cuadrados, precio, entre otros. Nuestro objetivo será predecir el precio pero de modo que la variable sea categórica. Para ello, transformamos el variable de precio de la casa, en una que indique 1 si su precio es mayor a 2 mil y, 0 sino. Elegimos 200 mil dado que hicimos el cálculo del promedio de precios y nos daba aproximadamente 180 mil, por lo que el número 200.000 nos pareció que podría ser adecuado para que la cantidad de 0s y 1s fueran parejas.

El código que utilizamos para adaptar el dataset fue el siguiente:

In [ ]:
```
import pandas as pd
df = pd.read_excel(r"C:\Users\fedep\OneDrive\Documentos\TD VI\TP1\src\dat
df = df.dropna()
df['Mayor_2'] = 0

for i in range(len(df)):
     if df['SalePrice'].iloc[i] > 200000:
          df['Mayor_2'][i] = 1

df = df.drop(['SalePrice'],axis=1)

print(len(df))
df.to_csv(r"C:\Users\fedep\OneDrive\Documentos\TD VI\TP1\src\data\HousePr
```