Project Proposal
By Fatimah Alrashed

One of the most Machine Learning problems is predicting house prices. We have different questions about the USA_Housing prices:

1- Is there a relationship between the age of houses and the price?

2- Does the number of rooms has an effect on the house price?

3- Does the number of bedrooms has an effect on the house price?

4- Does the population of the area has an influence on the house price?

We are going to use the USA_Housing dataset, and this is a regression problem. USA_Housing dataset contains 5000 rows. The dataset has 6 columns, and the prices column is the target value. There is no non-null in USA_Housing dataset. Also, there is no duplicated rows in USA_Housing dataset.

The data contains the following columns:

• 'Avg. Area Income': This Colum describes the average income of house residents located in the same city.

• 'Avg. Area House Age': This column characterizes the average age of houses in the same city.

• 'Avg. Area Number of Rooms': This column contains the average number of rooms for houses in the same city.

• 'Avg. Area Number of Bedrooms': This Colum describes average number of bedrooms for houses in the same city.

• 'Area Population': This column describes population of the city the house is located in.

• 'Price': Price that the house sold at.

• 'Address': Address of the house.

The data info is shown in the picture bellow:

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 7 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Avg. Area Income              5000 non-null   float64
 1   Avg. Area House Age           5000 non-null   float64
 2   Avg. Area Number of Rooms     5000 non-null   float64
 3   Avg. Area Number of Bedrooms  5000 non-null   float64
 4   Area Population               5000 non-null   float64
 5   Price                         5000 non-null   float64
 6   Address                       5000 non-null   object 
dtypes: float64(6), object(1)
memory usage: 273.6+ KB
```

Tools:

We will use Python and Jupyter notebooks for this project. We will use Pandas , Scikit Learn , NumPy , Seaborn and Matplotlib libraries. Pandas library use for handling structured data. NumPy library use for linear algebra and mathematics. Scikit Learn library use for machine learning. Seaborn and Matplotlib use for data visualization.

Dataset Web Site:

https://www.kaggle.com/faressayah/linear-regression-house-price-prediction/notebook