Project Proposal

-Problem statement:

Due to the information gap about house prices and the difference in house prices from one neighborhood to another, the individual may face a problem estimating house prices and may be subject to deception by real estate owners.

Therefore, we will build a model that predicts house prices in various neighborhoods in Saudi Arabia's capital Riyadh based on several features.

-Data Description:

We did a web scraping on "Aqar" website, which is a site specialized in Saudi real estate. We collected data for villas offered for sale in Riyadh. We scraped a thousand pages on the site to get 21979 observations and five features.

Aqar website**:** **https://sa.aqar.fm/**

Columns:

Price (target variable) → Price, float, continuous variable
Size → Area in square meters, float, continuous variable.
n_rooms → Number of bedrooms ,int, discrete variable.
n_bathroom → Number of bathrooms, int, discrete variable.
Districts → Neighborhood name, string, nominal variable.

We are going to use dummy variables to deal with Districts because it is a categorical vari able.
(21979, 5)

-Tools:

- Programs: Jupyter Notebook
- Libraries: numpy, pandas, matplotlib, seaborn, csv, itertools, requests, BeautifulSoup, mean_squared_error, r2_score, LinearRegression, statsmodels, sklearn, mean_squared_error, mean_absolute_error
- Functions: shape, value_counts, strip, replace, info, drop, astype, groupby, mean, get, content, find_all, zip_longest, writer, writerow, fit, predict, OLS, summary,
- Plots: pointplot, barplot, histogram, scatter, distplot, Pie Chart

-MVP Goal:

The goal of this project is to predict the prices of houses in different neighborhoods of Riyadh city based on the number of rooms, bathrooms, area and neighborhood using regression algorithms.