PREDCTING REAL-ESTATE PRICE IN RIYADH

PROPOSED BY

Hayat Aldhahri
Muneera Alshunaifi

DATE PROPOSED

5 Oct 2021

Table of Contents

List of Figures

List of Tables

Overview

Saudi Arabia's residential real estate market, which is majorly concentrated across major cities such as Riyadh, Jeddah, Makkah, and Dammam Metropolitan Area (DMA), is expected to grow further in the coming years on the back of rising demand for housing units. Also, it is a non-stable market that attempts to grow or fall periodically. Therefore, the real-estate prices change dynamically in which it is hard for real-estate owners to measure the pricing criteria well. In order to facilitate the pricing process for the real-estate owners, we decided to develop a solution that aims to predict the real-estate prices based on a machine learning algorithm; specifically, we will be using a linear regression model. To do so, we will use a scraped date from the AQAR website using comprehensive analysis tools. AQAR is a website that targets people willing to sell/rent their property and people searching for properties.

To accomplish the prediction, The working flow will be as the following:

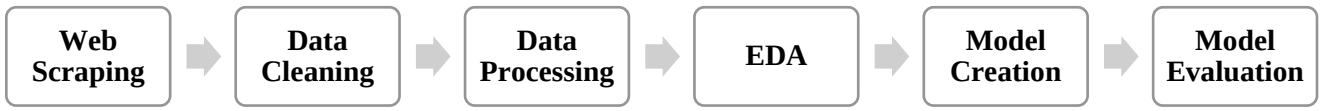

Figure 1 Model workflow

Data Description

The data set for the model will be scraped from AQAR website: https://sa.aqar.fm/ which targets people willing to sell/rent their real-estate. The below table outlines the features that will be used to develop the prediction model.

Table 0.1 Features description

Feature	Description	Data Type
Price	The price of the villa	Integer
#BEDROOMS	The number of bedroom for each villa	Integer
#BATHROOMS	The number of bathrooms for each villa	Integer
#LIVINGROOMS	The number of living rooms for each villa	Integer
SIZE(m^2)	The total building area in squared meters of each villa	Integer
DISTRICT	The name of the district where the villa is located	Object
STREETWIDTH	The width of the street on which the villa is located	Integer

Tools

To explore and analyze the data, I will be using Jupyter notebook to use python language and Python libraries, such as:

- Pandas and NumPy packages for data manipulation.
- Matplotlib, seaborn library for data visualization.
- BeautifulSoup library for web scraping.
- LinearRegression from the sklearn.linear_model package for applying the linear regression model

In addition, we will use Instant Data Scraper tool which is an automated data extraction tool for any website. It uses AI to predict which data is most relevant on a HTML page and allows saving it to Excel or CSV file.

MVP Goal

The MVP Goal would be looking at the correlations between the features outlined and the selling price for a real-estate and which component have the most significant impact on prices. MVP will be submitted on Monday 11th October to get the feedback and progress on the analysis.

Conclusion

Predicting the real-estate prices will give real-estate owners a reasonable estimation of the pricing criteria; this can be carried by using a prediction model, specifically a linear regression model. The expected outcomes from this model will assist real estate owners in setting the selling price using the prediction model.