**Linear regression of Vacation Home Rental in Saudi Arabia**



Raghad Althunayan

23 November 2021

**Introduction**

In this project, I will do scraping for Gathern website. Gathern is the first platform for shared housing (peer-to-peer rental) in Saudi Arabia View through the platform of private vacation homes (such as villas and private apartments, farms, chalets, buildings, camps, other vacation homes)

**Problem statement:**

Many seller and buyers are confused about the price and what feature makes the price high.

**Value of the Platform and recommendations.**

The primarily goal of project to answer the following questions/needs:

As a Buyer :

I need to know what makes price higher

As a seller:

I need know which features means a lot to buyer and they may pay more for it

**Dataset**

**Datasets with description:**

This project is based on the data available on the Gathern website:

https://gathern.co/

**Scope of the work**

**Sample size:**

All cities in KSA and all types of shared housing

**Description of scrapped data:**

The dataset represents the price of shared housing in different cities in KSA

Number of features: 6 features/Columns

Number of rows: Approx.: 2000 rows

Names of columns with description and type:

|  |  |
| --- | --- |
| Field Name | Description |
| property | Represents name of property of housing |
| city | Represents the name of city |
| area | Represents the area of housing |
| category | Represents the category of housing |
| rating | Represents the rating of housing |
| price | Represents the price of housing per night |

**Algorithms**

I will start by using web scraping taqnieq to get the data then cleaning data and pre-processing, I will start by deleting the duplicate records and check if there are any null values then drop them. I will visualize the result and try to find any correlations. After that I will do Linear regression to predict price.

**Tools**

* Technologies: Python, Jupyter notebook, HTML/CSS
* Libraries: NumPy,Pandas, Matplotlib, Seaborn, BeautifulSoup and selenium, os, SKlearn, PLotly