A picture containing sky, outdoor, city, shore

Description automatically generated

Housing Sale Prices and Venues Data Analysis of Jeddah City

Data Acquisition and Cleaning

2.1 Data Source

2.2 Data Cleaning

2.3 Feature Selections

1. **Data Acquisition and Cleaning**

As the data is not available in readymade format, so we collect it from different sources spread on the internet and other data repositories.

**2.1 Data Source**

* As Jeddah have 140 administrator units so we need to have the name of each district and it location coordinates e.g. latitudes and longitudes.

We collect the list of Dist. from the url below and scrap it

[Jeddah - Wikipedia](https://en.wikipedia.org/wiki/Jeddah)

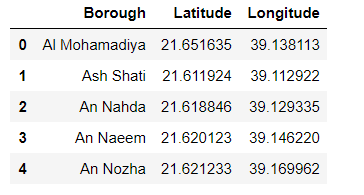
The scraped list dist is stored in into a data frame. As we just have the list of Dist. names, we also need the latitudes and longitudes of the those dist. So we collect latitude and longitude through google.

* Foursquare API is used to collect the venues in the nearby Borough.
* We go through and manually search the house prices in each Borough
  1. **Data Cleaning**

We have 140 Dist in the list. We short list this to 45.

* 1. **Feature Selections**

For our purpose we need name of Borough, Latitude, longitude. Here is the list.



**References**

1. <https://www.jeddah.gov.sa/>
2. [Jeddah - Wikipedia](https://en.wikipedia.org/wiki/Jeddah)
3. Google maps