The Battle of Neighborhoods

# Applied Data Science Capstone Project

Nikola Kostic

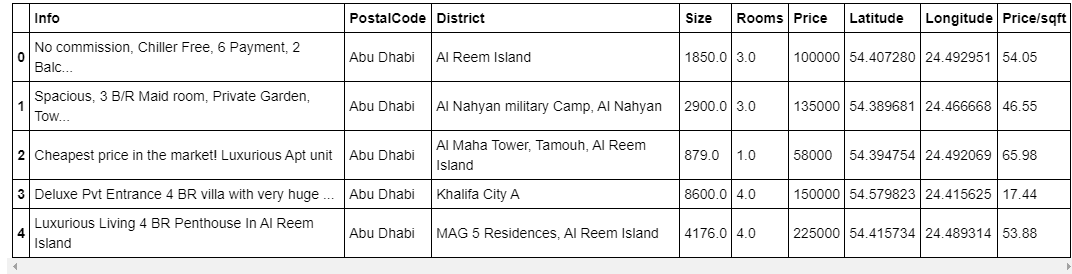
# Introduction

In this project, we will help people who are looking for renting an apartment in Abu Dhabi. If they are looking to move to Abu Dhabi or changing there residence they can see:

* Which district has cheaper rent or to have an idea on how the prices effected per district or number of rooms,
* It was planned to use choropleth maps on order to visualize the data but I couldn’t find district Geojson for Abu Dhabi

# Data

The data on apartments: size, number of rooms, address, and the price is collected by scraping a local website with apartment listings (dubizzle.com). We clean up the values and calculate the price (K)/sqft by dividing the price by the size column. The data is pre-processed and we get our first data frame:

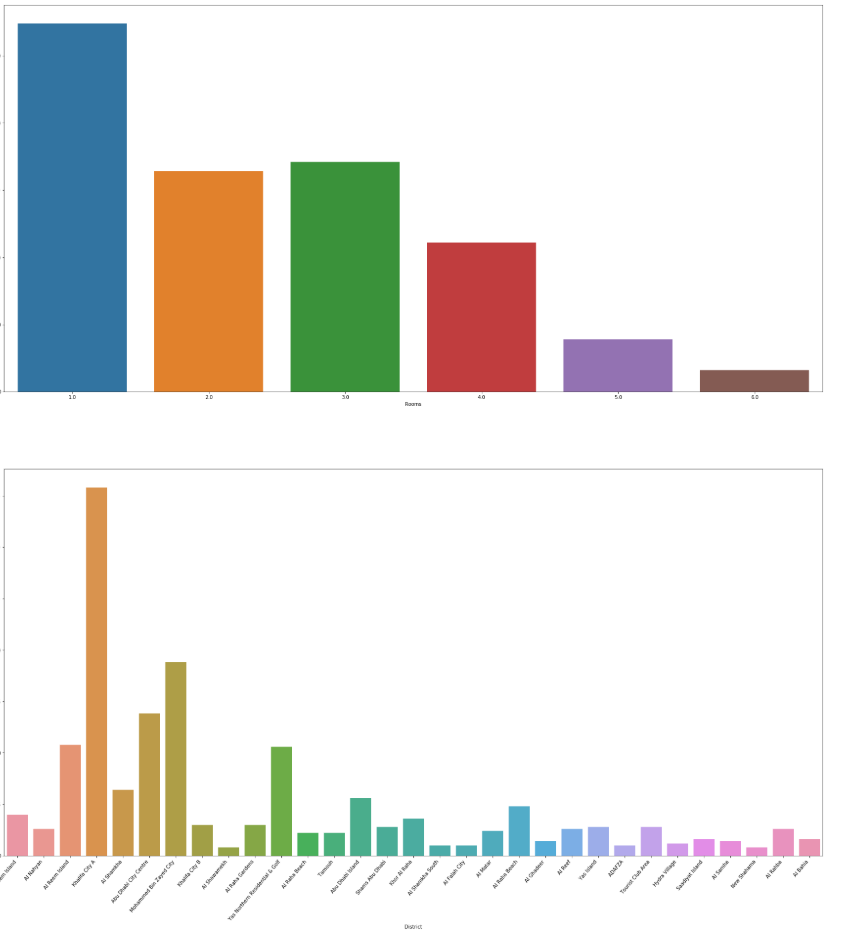


It is further cleaned by removing NA values and outliers, which results in a dataset. From this dataset we extract the rows with a unique Postal Code and then using geopy we find the coordinates for each district.

Using Foursquare we collect the closest venues (supermarket, restaurant, park, etc.) and select the top 10 venues for each district. After the data collection, we can run k-means clustering to cluster the districts into residential and commercial areas and visualize all the data.

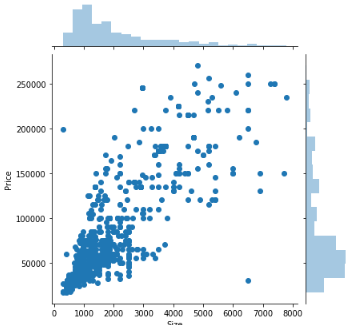
# Methodology

We check the data set for how many apartments per room number we have and how many apartments we have in each district. After removing any outliers we can plot the following charts:

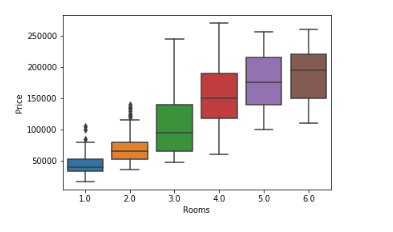


From this we can see that 1 and 3 room apartments are the most common.

We can also see if there is a correlation between the price and apartment size:

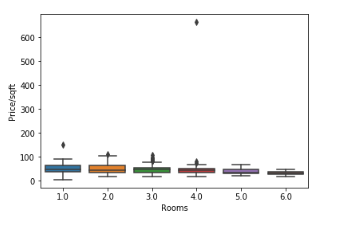


And as expected, the larger the apartment, the higher the yearly rent. Another thing we can look at is the rent per number of rooms:

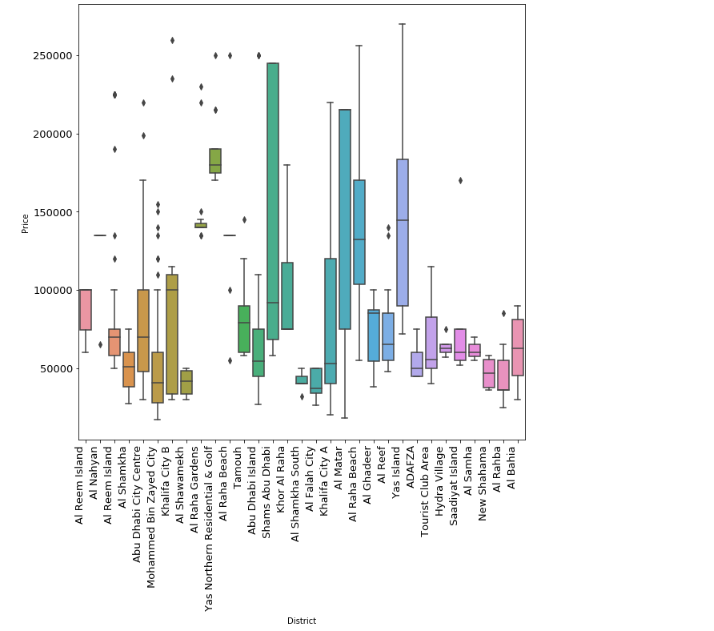


As expect the value of properties to go up as the number of rooms increases. The interesting aspect in this boxplot is that 4,5 and 6 rooms apartments are competing in the same price range.

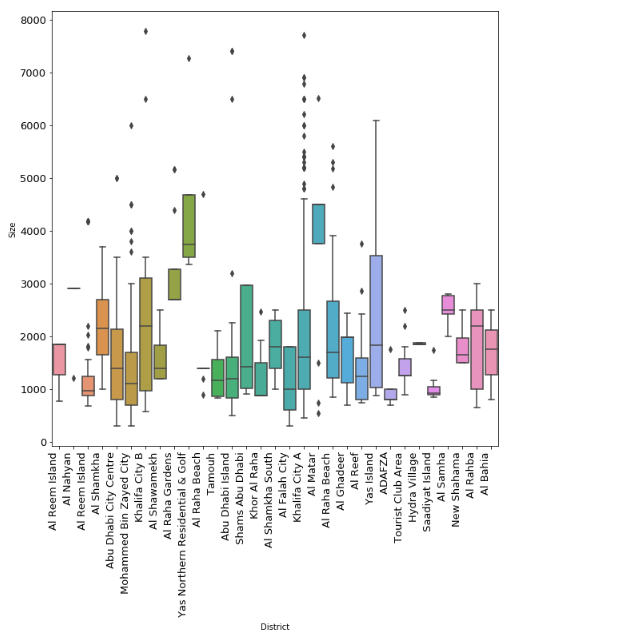
In the following plot we can see that the price/sqft is in the same price range for all size apartments besides for single room apartments where the price/sqft goes even higher.



Using the apartment data and the we can visualize the average price ranges for each district in Abu Dhabi.



Also we can notice that Al Raha and Al Matar districts have the biggest apartments sizes



# Results

Once Foursquare data is collected. We find the most common venues (supermarket, restaurant, park, etc.) and select the top 10 venues for each district. After the data collection we can run k-means clustering to cluster the districts. By analyzing the clusters we can see that cluster 1 is more residential since it contains lots of parks and supermarkets.

# Discussion & Conclusion

With this our analysis, one could determine for example that the 1st district is the best district to live in, however by clustering we determined that there are several more similar districts where the price/sqft is significantly lower.