The Battle of Neighborhoods

Introduction

Background

This project is for those who are planning to move in to a new apartment in the city of Vienna. It provides a suggestion on what would be the best district in Vienna to get an apartment which has already many good apartments. Vienna is Austria's most populous city, with about 1.9 million inhabitants (2.6 million within the metropolitan area, nearly one third of the country's population), and its cultural, economic, and political center. It is the 6th-largest city by population within city limits in the European Union.

Business Problem

This project focusses on the issue of finding an apartment to rent a city like Vienna which is the largest city in Austria. Potential stakeholders for this project would be people looking to rent an apartment. Business problem would be providing people with information like

- Which district has cheaper rent
- Choose to live in residential or commercial areas and can see for example which residential districts is best Or

If they already live in one of the 23 districts in Vienna, they will be able to see:

- If they are paying more than the average price for their apartment
- If there are similar districts to theirs with lower rents

Data

Two sources are identified as potential platforms of data we need for exploratory and descriptive analysis. One source is a popular apartment listings website 'willhaben.at' and the latter is 'Foursquare'.

We use web scraping to derive from the website and use Beautiful Soup library to parse and extract required field of interest like size of the apartment, number of rooms, address, price. Data is derived from Foursquare by making API calls to endpoints exposed by the data provider. Features of interest from this provider would by closet venues like supermarket, restaurant, theater, park etc.

After data wrangling procedures are performed on the collected data, it is taken to next step to cluster the districts into residential and commercial areas and visualize all the data on a single choropleth map using k-means clustering unsupervised machine learning algorithm.

After data wrangling and cleaning up the values and calculating the price/m^2 by dividing the price by the size column, we can view our first pandas dataframe as follows:

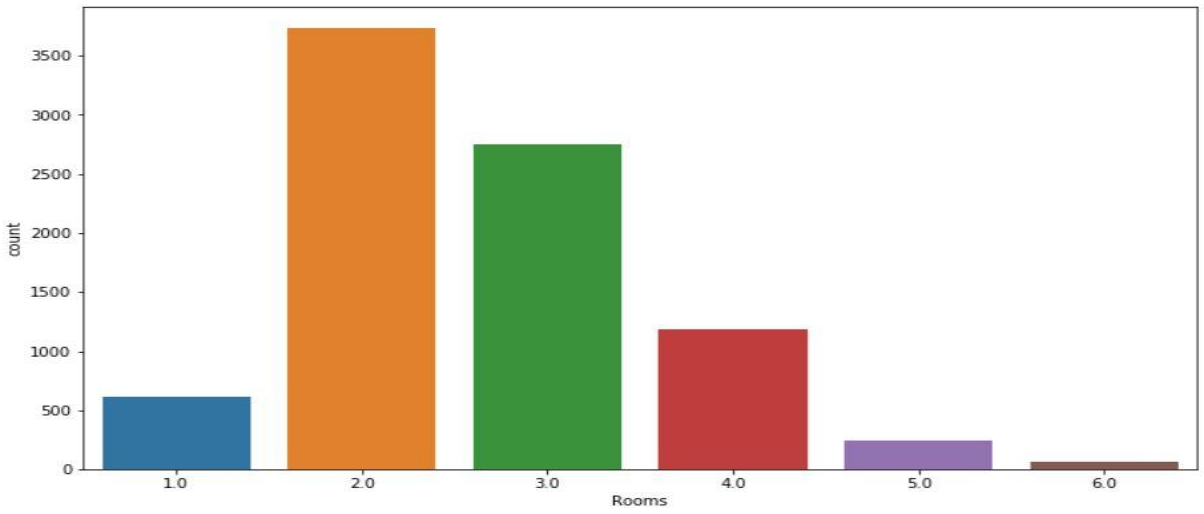
	PostalCode	District	Size	Rooms	Price	Price/m2
0	1150	Rudolfsheim-Fünfhaus	74.0	3.0	887.25	11.99
1	1220	Donaustadt	110.0	4.0	1500.00	13.64
2	1100	Favoriten	48.0	2.0	800.00	16.67
3	1120	Meidling	133.0	3.0	1931.16	14.52
4	1030	Landstraße	50.0	2.0	898.94	17.98

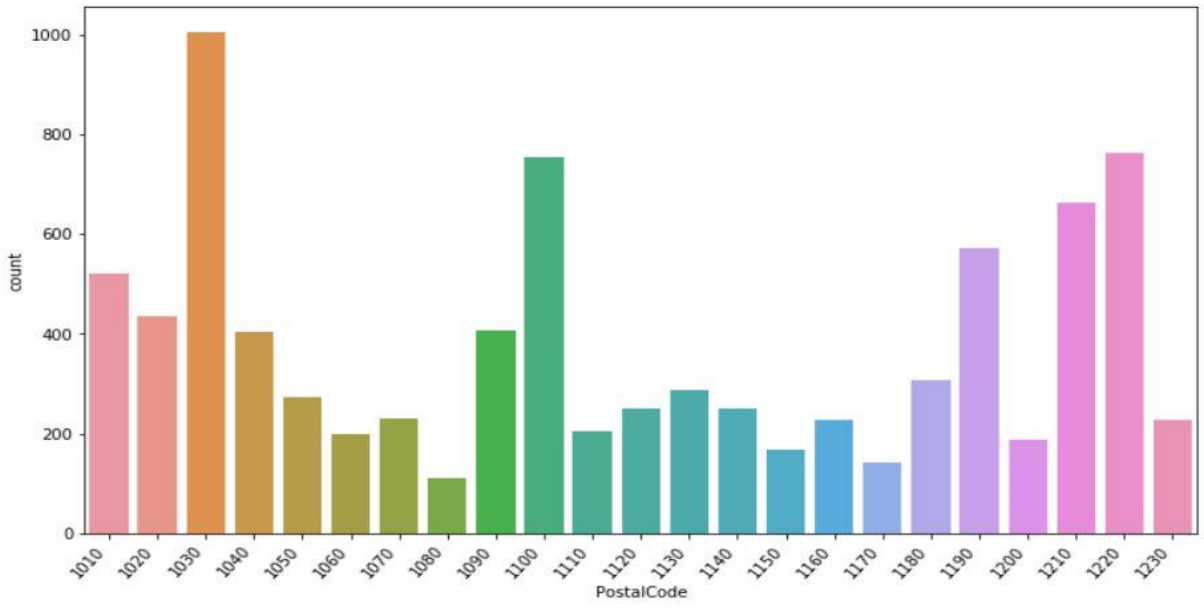
Further the dataframe is cleaned by removing NA values and outliers, which results in a dataset of 9275 apartments. From the latest dataset rows with a unique Postal Code are extracted and use geopy to locate the coordinates for each of district in Vienna.

We then employ Foursquare website and make API calls to the exposed endpoints that we are interested in and collect the closest venues (supermarket, restaurant, shopping, park, etc.). We select the top 10 venues for each district. After the data collection we run k-means clustering machine learning algorithm to cluster the districts into residential and commercial areas. After clustering process is completed, we visualize the results and all the data on a single choropleth map.

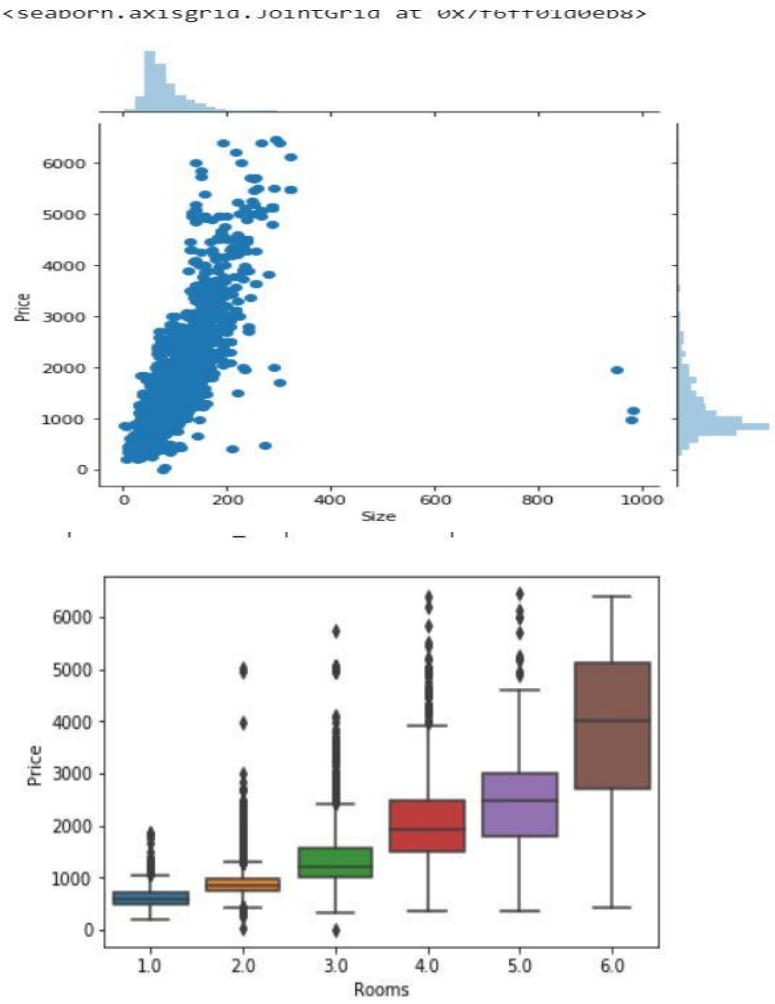
Methodology

Processed data set is analyzed to see the count of apartments which falls in to particular number of room (2, 3, 4 rooms...). We could employ histogram to visualize the findings. After removing any outliers, plotting histogram against data would yield the following charts:

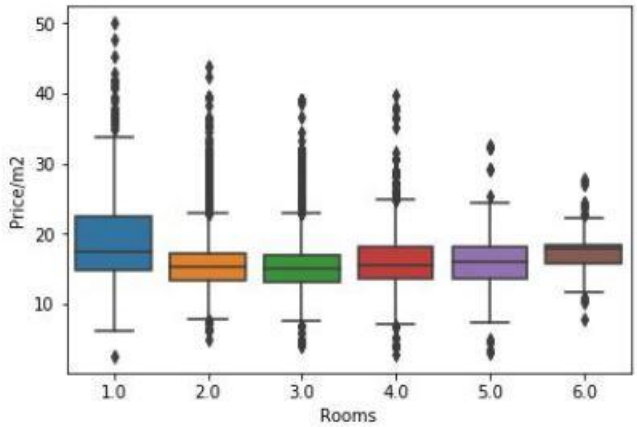




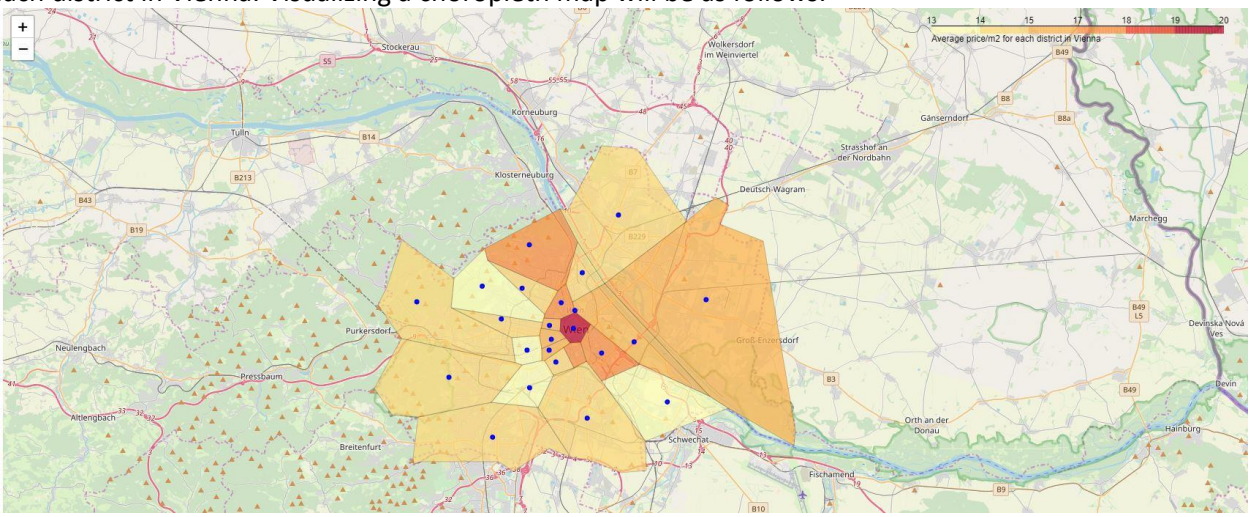
We can infer from the plots that frequency of 2 and 3 room apartments are the most common.
We check if the price is increasing or decreasing based on change in other features like apartment size.



Generally, the price of a property goes up as the number of rooms increases. In Contrary to what we might expect 1 and 2 room apartments are competing in the similar price range as seen in boxplot.
In the following plot we can infer that the price/m2 is in the same price range for all size apartments besides for single room apartments where the price/m2 goes even higher.



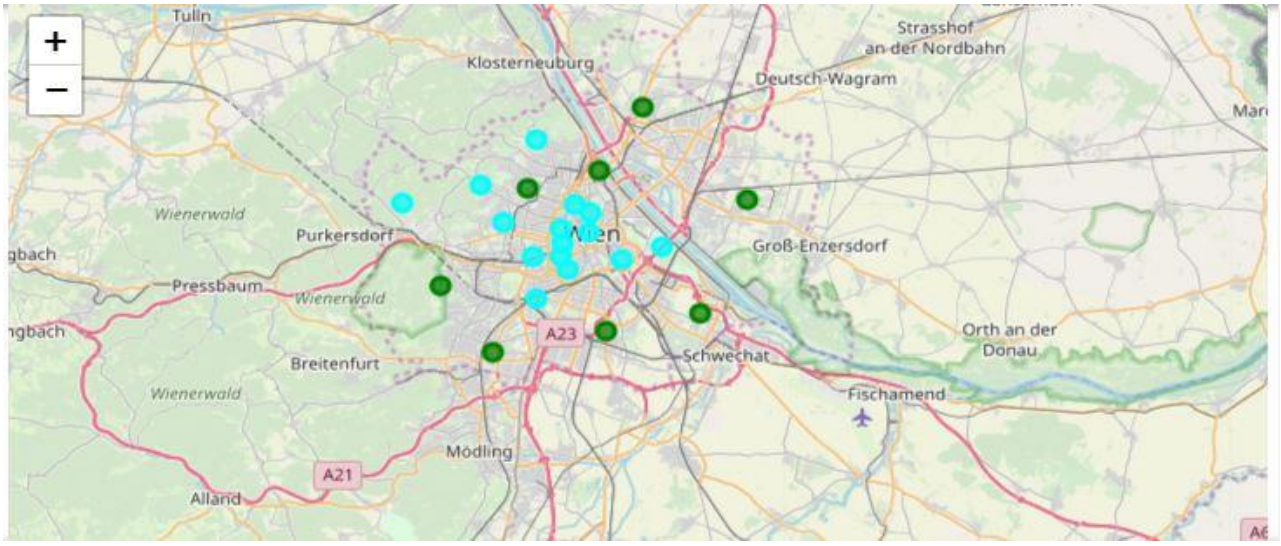
Using the data about apartment and coordinates of districts we could visualize the average price/m ^{2} for each district in Vienna. Visualizing a choropleth map will be as follows:



Data from Foursquare is also collected which features near by venues. Venues are grouped based on the categories and we get 10 most common venues in each district. The data would look below

	PostalCode	District	Latitude	Longitude	Price/m2	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1010	Innere Stadt	48.209023	16.369851	20.344605	1	Hotel	Austrian Restaurant	Plaza	Restaurant	Italian Restaurant	Café	Bar	Ice Cream Shop	Boutique	Jazz Club
1	1020	Leopoldstadt	48.200638	16.426895	16.125701	1	Restaurant	Supermarket	Hotel	Bakery	Gastropub	Beer Garden	Hot Dog Joint	Plaza	Gym / Fitness Center	Austrian Restaurant
2	1030	Landstraße	48.193644	16.396286	16.823695	1	Hotel	Restaurant	Austrian Restaurant	Pizza Place	Concert Hall	Café	Asian Restaurant	Plaza	Italian Restaurant	Steakhouse
3	1040	Wieden	48.220210	16.371216	16.717916	1	Restaurant	Hotel	Austrian Restaurant	Café	Plaza	Coffee Shop	Cocktail Bar	Pizza Place	Park	Lounge
4	1050	Margareten	48.188073	16.353386	15.157500	1	Austrian Restaurant	Bar	Hotel	Asian Restaurant	Café	Ice Cream Shop	Restaurant	Burger Joint	Indian Restaurant	Middle Eastern Restaurant
5	1060	Mariahilf	48.195475	16.347023	16.433182	1	Hotel	Austrian Restaurant	Ice Cream Shop	Pizza Place	Clothing Store	Hostel	Coffee Shop	Café	Japanese Restaurant	Seafood Restaurant
6	1070	Neubau	48.202264	16.349123	15.853609	1	Hotel	Plaza	Austrian Restaurant	Café	Park	Ice Cream Shop	Gym / Fitness Center	Theater	Pizza Place	Bar
7	1080	Josefstadt	48.210852	16.347360	15.737890	1	Hotel	Plaza	Italian Restaurant	Café	Coffee Shop	Park	Pizza Place	Ice Cream Shop	Gym	Theater

K means clustering is performed on the data. It yields us clusters as follows:
The cyan markers represent the commercial districts and the green markers represent the residential districts.

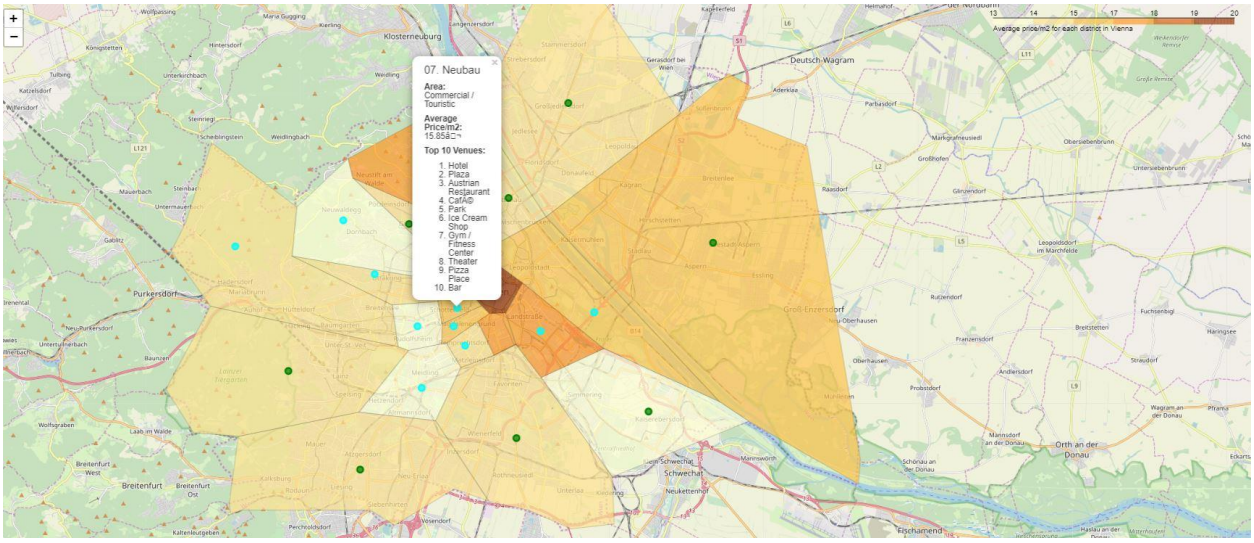


Cluster 1: green, Cluster 2: Cyan

Results

With all the gathered data creating a choropleth map displays the average price/m2 for each district as well as display information about the area type and top 10 locations for each district when clicked on marker labels. Analyzing the clusters conclude that

- cluster 1 is more residential since it contains lots of parks and supermarkets
- cluster 2 is more commercial / touristic and contains many hotels and restaurants.



Discussion

With the above map, one could understand for example that the 1st district is the most expensive district to live in, however by clustering we concluded that there are several more similar districts where the price/m2 is significantly lower.

Therefore, if someone is looking to rent an apartment but cannot afford to live in the 1st district, they could look for apartments' in other districts like 12 or 15 district which have similar venues but comparatively lower price.

Conclusion

In conclusion, analysis projected in the current project would definitely be beneficial to those who is looking for apartment recommendations with type of neighborhood they are wishing for be it commercial or residential or apartment within a particular price range and so on. This project is also beneficial to realtors, landlords and apartment listing websites to gain understanding on the current scenarios in the rental preferences. Similar analysis could be carried out on other cities of interest to derive suggestions.