Week5 - Report

July 2019

1 Introduction

Each year, the international consulting firm Mercer carries out a study to assess the quality of life in 231 cities around the world. In 2019 the Austrian federal capital was award the most livable city in the world for the tenth time in a row. Although Vienna is the most livable city in the world, people still would like to know more about which district in the city would suit them more, before moving to Vienna. For example, there are so many factors affect people to make decisions to start a new life in a city, such as house price, school rating, crime rating, neighborhood analysis etc. That calls for a search algorithm which normally gives the requested features.

2 Goal

This project is to help stakeholders or anyone who wants to explore Vienna, in order to make a better decision which district is the best area to start a new life. There are in total 23 districts in Vienna. The rental cost is quite different from one district to another. For example, the 1st district locates in the city center, which is of course the most convenient place to live. However, the renting price must be taken into consideration. According to the information we scraped from internet, the 1st district is the most expensive district. So the purpose of this project is to help stakeholders to find a district to rent an apartment, with a relatively cheaper rental cost and relative convenient living facilities such as cafes, restaurants, bakeries, gyms etc.

3 Datasets

3.1 Vienna Rental Price

In vienna, there are 23 districts in total. In order to find a balance between cheaper rental price and more convenient living facilities, we have to know the rental cost of each district. The link below provides the rental price of apartment with differece sizes. In this project, we only consider the average rental price ($Average/M^2$). However, the algorithm used in this project also can be applied

to small ($< 50M^2$)/medium ($50 - 129M^2$)/large ($> 130M^2$) apartments. The link is provided in the presentation.

3.2 Vienna Neighborhood Data - Districts Latitude and Longitude

The link for these data is provided in the presentation.

4 Steps

4.1 Data Wrangling

Numby and *Pandas* are used for this purpose. The postcode of each Vienna district needs to be added into the dataframe in order to read the data more clearly.

In this project, we only consider the average rental price (AVERAGE/M). So the other columns with the price of different apartment size are dropped.

dataset of Vienna neighborhood columns 'SHAPE', 'FEATURENAME', 'LOCALITYLIST' will be kept from the original data.

'SHAPE' contains the coordinates of each neighborhood. So the latitude and longitude need to be extracted from 'SHAPE' and make two new columns 'Latitude' and 'Longitude'.

'LOCALITYLIST' will be renamed to 'Postcode' and convert to the same format as the 'Postcode' column in Vienna rental price dataset. Some neighborhood are shared with more than 1 district, thus these neighborhoods has more than more postcode. These data will be dropped. We only consider the neighborhood uniquely belongs to one district.

Merge Vienna rental price dataset and Vienna neighborhood dataset based on the shared column 'Postcode'.

There are more than 4000 rows after merging the two datasets. For simplifying, 5 data points (neighborhoods) in each district are randomly selected to make a new dataframe.

4.2 Visualization

• Library seaborn is used for data visualization.
• Geopy library is used to get the latitude and longitude values of Vienna city.
• Folium library is used to visualize the neighborhoods of Vienna

4.3 Clustering

• Neighborhoods need to be clustered in this project. However, neighborhoods are categorical data. So one-hot enconding on categorical variables is being used

to convert categorical variables to numerical variables.
• Library Scikit learn is used for k means clustering.

5 Results

Vienna rental price has a huge difference among districts. As shown in the barplot below, the first district Innere Stadt has the highest rental price with almost 20 euro per M^2. However, the lowest rental price is in the 11th district Simmering, which only cost around 12 euro per M^2. This makes sense since Innere Stadt has more convenient living facilities.

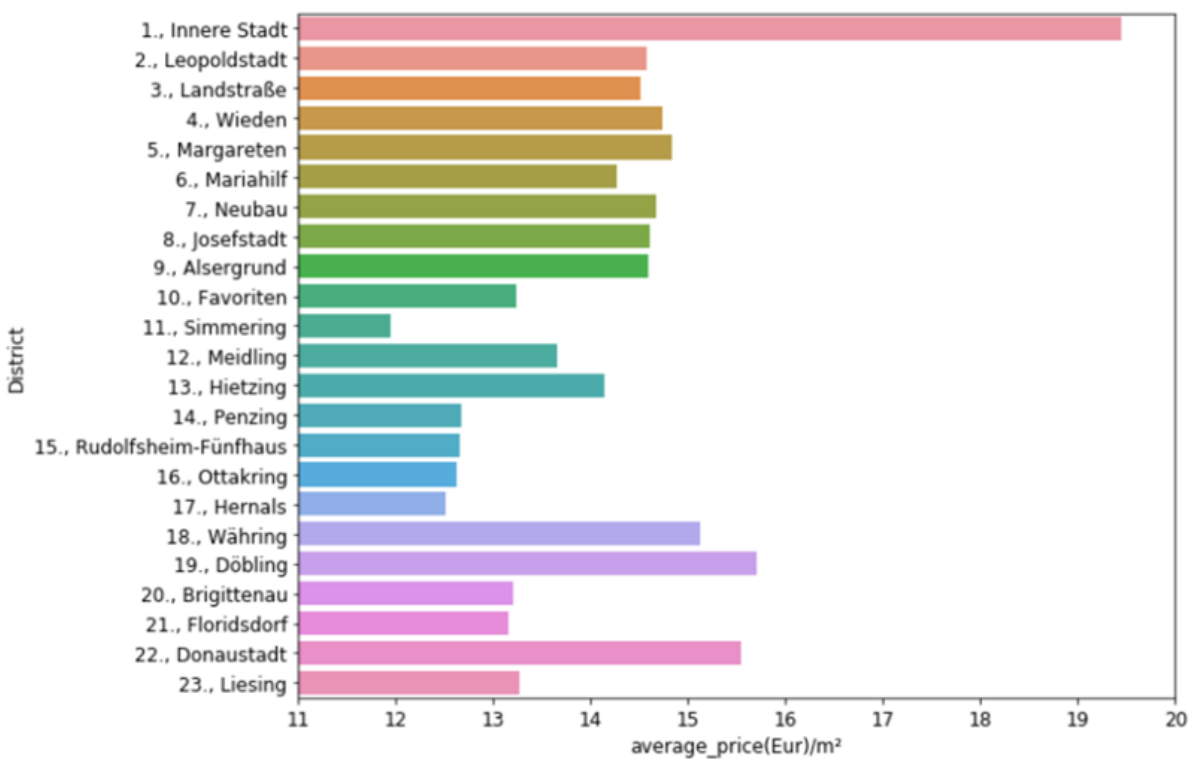

Figure 1: Rent for districts of Vienna

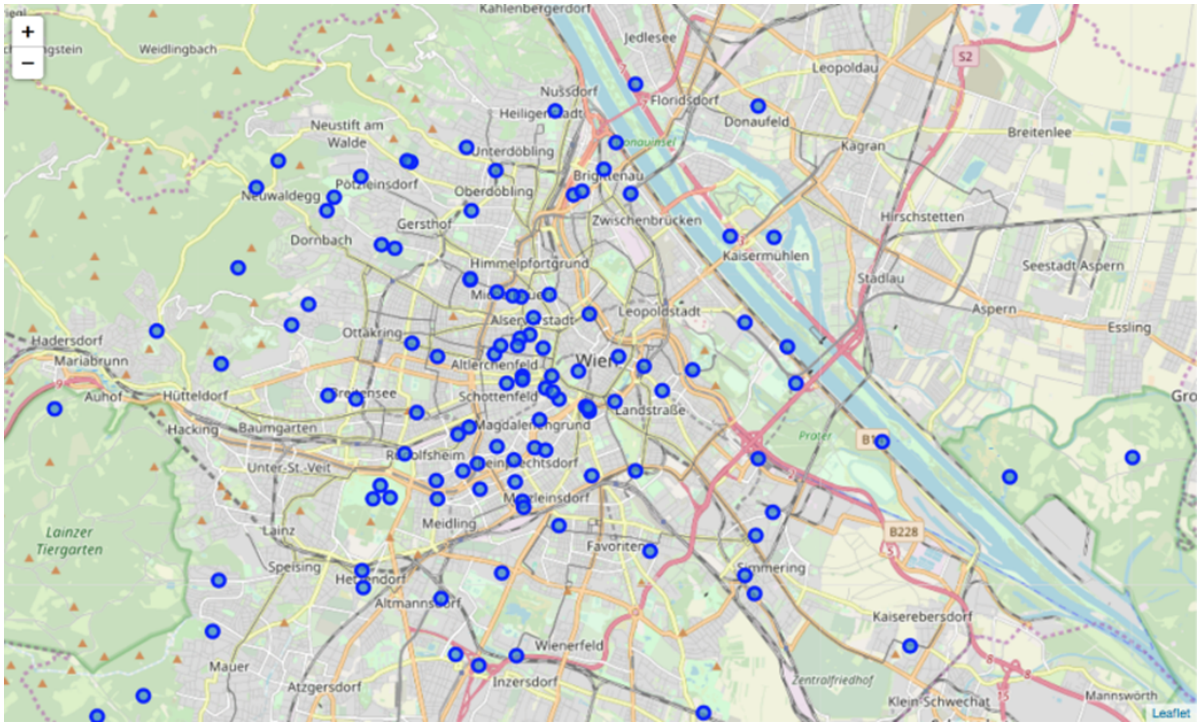

Figure 2: Map of Vienna

Venues

Here are the neighborhoods along with their common venuesThis is only partially visible due to the large size of table.

	Neighborhood	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
0	Am Bahndamm	Dog Run	Campground	Train Station	Restaurant	Zoo Exhibit	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaura
1	Amerling-Denkmal	Austrian Restaurant	Hotel	Italian Restaurant	Restaurant	Café	Bar	Wine Bar	Steakhouse	Bakery
2	Anschüttbrückl	Plaza	Zoo Exhibit	Fish Market	Exhibit	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field
3	Anton-Proksch-Wohnhausanlage	Italian Restaurant	Greek Restaurant	Wine Bar	Bus Stop	Market	History Museum	Gym Pool	Cocktail Bar	Austrian Restaura
4	Aspern Seestadt	Pizza Place	Park	Bakery	Austrian Restaurant	Pub	Fruit & Vegetable Store	Metro Station	Lake	Flower Shop

Figure 3: Top 10 venues in Vienna

6 Conclusion

In this project, we clustered the Vienna neighborhoods in order to find similar or more convenient living facility with relative cheaper price. Depends on which neighborhoods feature the stakeholders care about, they can easily focus on the cluster which they are interested in. Then it is more easier to make a decision according to rental price where to start a few life in Vienna.