

Capstone Project Report

Opening an Italian Restaurant in Vienna



Author: Daniel Schnelldorfer
For IBM Applied Data Science Capstone

1.Introduction

A successful Italian restaurant chain with many locations spread all over Europe, and a stable financial equipment wants to use the current lockdown in Austria to expand their business. Therefore, they plan to open new restaurants in the biggest Austrian cities. The company is especially interested in the Austrian capital Vienna. What makes Vienna that interesting are the generally high loans, a continuous growth of the population as well as the high cultural diversity of the city and its many foreign cuisines.

So, the company wants to open a new restaurant in Vienna, but doesn't know exactly where.

The **business question** is: "Which district(s) of Vienna offers the most attractive conditions for opening a new restaurant?". To answer this particular question, the company hires an external data analyst in order to support decision making of the company's management.

The aim of the data scientists report will be getting an overview of the different districts and their key attributes like population density, competition density, as well as real estate prices. The main part of the explorative analysis will be a clustered map of Vienna's districts.

2. Data

The **data**, that is used to solve the above business problem, is extracted from the following different **sources**:

1. The table with the different district names, and their corresponding population density is scraped from the following Wikipedia page: https://de.wikipedia.org/wiki/Wiener_Gemeindebezirke
2. A table with the average real estate prices of each district in Vienna will be scraped from the following page: <https://de.statista.com/statistik/daten/studie/900881/umfrage/preise-fuer-eigentumswohnungen-in-wien-nach-bezirken/>
3. The geographical coordinates of each district are collected via the python geocoder package.
4. The competition density in each district is determined by the average number of venues for Italian restaurant in this area. We

get those venues by writing the according queries and using therefore the Foursquare API.

3.Methodology

At first, we need to get data regarding the city of Vienna and its districts. Therefore we scrape the table from our Wikipedia source in a data frame. In the next few steps we clean the data. Therefore we drop the columns, that we don't need for further calculation. Additionally we calculate the column with the values Residents per km² manually, because the values from the wikipedia column did not come through the download.

In the next steps we use the number of each district in our dataset, in order to calculate its Postal Code manually, by adding "1" at the beginning and "0" at the end of every district number. So now we can get both Latitude and Longitude coordinates for each Postal Code, using the python geocoder package.

In the next step we create a new data frame with the average real estate prices for each district and merge it with our other data frame on the column district, so that we have now one single big data frame, containing the columns: "District", "Price m²", "Postal Codes", "Residents per km²", "Latitude" and "Longitude" for each of the 23 districts.

Now we will use the Foursquare API to get all venues for the search term 'Italian Restaurant' within a radius of 1000 meters of each district. With the data, we can check, how many Italian Restaurants are in which district. Now we can add the number of Italian Restaurants as a new column to our data frame.

	District	Price m²	PostalCodes	Residents per km²	Latitude	Longitude	Italian Restaurants
0	Innere Stadt	17354	1010	5593	48.208750	16.370735	71.0
1	Josefstadt	9139	1080	22955	48.207793	16.352502	49.0
2	Wieden	8154	1040	18710	48.192806	16.369190	28.0
3	Döbling	7853	1190	2962	48.249367	16.338429	4.0
4	Währing	7513	1180	8113	48.231200	16.328649	4.0
5	Landstraße	7416	1030	12392	48.198756	16.393209	19.0
6	Mariahilf	7363	1060	21753	48.195262	16.350736	40.0
7	Alsergrund	7114	1090	14116	48.229912	16.351149	9.0
8	Neubau	7054	1070	19876	48.197257	16.342721	27.0
9	Leopoldstadt	6576	1020	5500	48.215485	16.380536	56.0
10	Hietzing	6420	1130	1432	48.175907	16.276513	6.0
11	Hernals	6342	1170	5006	48.227044	16.309270	2.0
12	Penzing	5982	1140	2773	48.207082	16.260296	4.0
13	Margareten	5846	1050	27397	48.186531	16.364161	17.0
14	Ottakring	5647	1160	11889	48.212891	16.313489	5.0
15	Donaustadt	5358	1220	1908	48.230605	16.479091	0.0
16	Rudolfsheim-Fünfhaus	5296	1150	19605	48.194818	16.327152	9.0
17	Floridsdorf	5082	1210	3779	48.267900	16.405105	2.0
18	Liesing	5052	1230	3445	48.144123	16.299105	2.0
19	Favoriten	5012	1100	6509	48.160150	16.381669	1.0
20	Meidling	4797	1120	11980	48.172302	16.323294	4.0
21	Brigittenau	4754	1200	15125	48.238962	16.387055	3.0
22	Simmering	4633	1110	4490	48.165578	16.436359	0.0

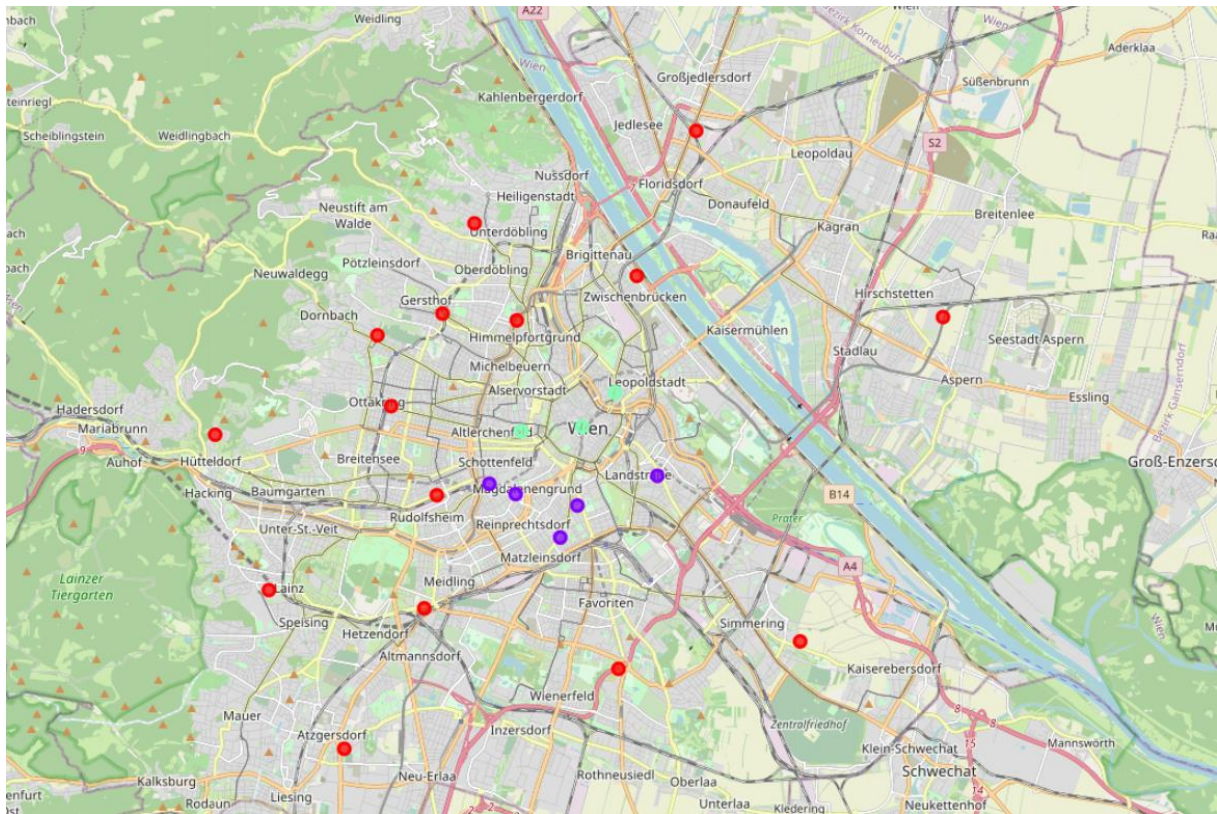
3.2 Exploratory data analysis

Now that we have all our data wrangled and cleaned, we can begin with the first steps of analysis. We visualize the distributions of population and the number of Italian Restaurants per district, to get an overview of potential demand.

We create the same plot again, to compare the distribution of Italian Restaurants as well as the real estate prices, to find a spot, where the competition and price-conditions are attractive.

In the last step, we want to Cluster the districts of the city of Vienna into three clusters. Then we visualize those clusters, to get a better overview of the clusters of the city.

4. Results



Our clustering algorithm has segmented Vienna's districts into three different clusters by the number of Italian Restaurants in each district:

- **red:** 0-9 Italian Restaurants
- **purple:** 17-40 Italian Restaurants
- **green:** 49-71 Italian Restaurants

The results show, that the ideal places to open a new restaurant are in the districts of the red and the purple markers. In those areas the demand for new restaurants is by far the highest. The green areas are too competitive in our case. The real estate prices, which have to be considered as well are especially high in the centered districts, but the difference between the red and purple districts are not really high. The population density is the highest in most of the purple markers, whereas the lowest population density can be found in the 'Innere Stadt' as well as in the outskirts.

5. Discussion

Data: The data regarding district information, real estate prices, as well as the geographical data, that has been used in this lab is of sufficient quality. On the other hand the venue data for the number of Italian Restaurants was not 100%

accurate, but still adequate to get a clue about the distribution of Italian Restaurants over Vienna's districts.

Methodology: The applied explorative analysis has delivered quite good insights in the dataset. The visualizations were necessary and easy to interpret. Although the clustering algorithm applied on this comparatively small dataset was quite an overkill, the results are still supporting our analysis and providing a good overview of the different districts.

Quality of results: Our results can provide helpful support in decision making for the company's management. But it's still important to note, that our dataset is not including all necessary information, which the management needs to make a 100% evidence-based decision. That would require a much higher quantity and quality of data. For example, we should not only cover the information for each district, but also for each neighborhood, perhaps even each single block has to be considered as a single unit. Additionally many other key aspects, such as individual financing plans and real estate prices for different locations, number of tourists, traffic density, income of the local residents, etc. have to be considered.

6. Conclusion

As we have seen in our results section, there are several districts in the city of Vienna, that seem very attractive for a new Italian restaurant. In those districts, we don't only have a very small number of existing Italian Restaurants, which means less competition. A high population density in those districts means a potentially unsatisfied demand, which offers great opportunity for a new restaurant.

In conclusion, the most attractive districts in Vienna to be considered are Rudolfsheim-Fünfhaus, Margareten, Meidling, Brigittenau and Währing. Those offer a high population density in combination with a moderate number of existing Italian restaurants.