**Suitable locations for a new restaurant in Vienna**

Prepared by Myroslav Yatkevych

11-May-2020

1. **Introduction**
   1. **Background**

The City of Vienna is the most populated city in Austria. It is diverse and is known as the financial capital of Austria. It also provides a lot of business opportunities and has a business-friendly environment. Vienna is a global hub of business and commerce in its region. The city is a major center for banking and finance, transportation, tourism, legal services, accountancy, insurance, theater, fashion, and the arts in Europe.

This also means that the market is highly competitive. As it is a highly developed city, therefore, the cost of doing business is also one of the highest in Europe. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from the analysis will give a good understanding of the business environment which helps in strategically targeting the market. This will help in the reduction of risk. And the Return on Investment will be reasonable. Therefore, the best way of understanding the market is to collect data. In our case we are going to analyze the restaurant niche.

* 1. **Problem**

A restaurant business is one of the most difficult businesses, as there are too many factors that influence the results. However, the profit margin is also quite high.

In the city of Vienna, we can observe a huge variety of restaurants, café, bars, night clubs, etc. Therefore, it is extremely important to have a strategic plan to survive in such a competitive market. Several key factors need to be studied in order to decide what location is better to choose or what kind of restaurant to open. For example:

* The population of the city of Vienna;
* Demographic statistics of the city of Vienna;
* Segmentation of the boroughs of the city of Vienna;
* Types of food services: restaurant, café, bar, etc.;
* Competitors in the respective boroughs of the city of Vienna;
* The most crowded places, such as business centers, historical places, entertainment zones, parks, etc.;
* Other factors that have an impact on decision-making which location to choose.
  1. **Interest**

Obviously, the stockholders, investors who are planning to open a new restaurant would be extremely interested in this research. The detailed analysis gives a full picture of the restaurant market in the city. It would save time and money. Such a report will provide the first step for investors in the new city.

1. **Data acquisition and cleaning**

**2.1 Data sources**

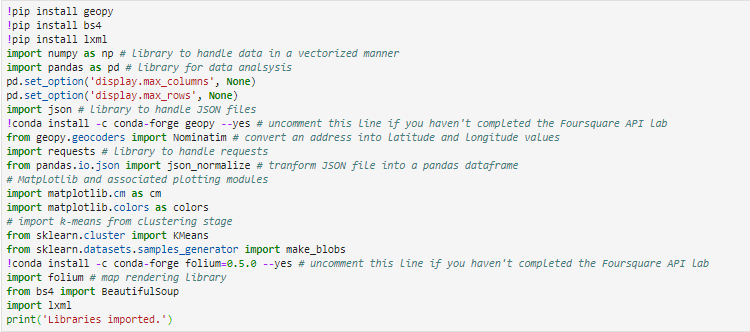
First, we need to identified what data is need to use for a better understanding of market. We will use information from an open source, such as: “Districts and population in Vienna”.  This data set contains the required information. And we will use this data set to explore various neighborhoods of Vienna. Restaurant in each neighborhood of the city.

Data source: Fousquare API Description: By using this API we will get all the venues in each neighborhood. We can filter these venues to get all restaurants.

GeoSpace data: By using this geo space data we will get the Vienna district’s boundaries that will help us visualize choropleth map.

**Required Libraries**

* pandas and numpy: library to handle data in a vectorized manner
* request library to handle requests.
* geopy to get co-ordinates of Vienna and convert an address into latitude and longitude values
* folium to visualize the results on a map



**2.2 Data cleaning**

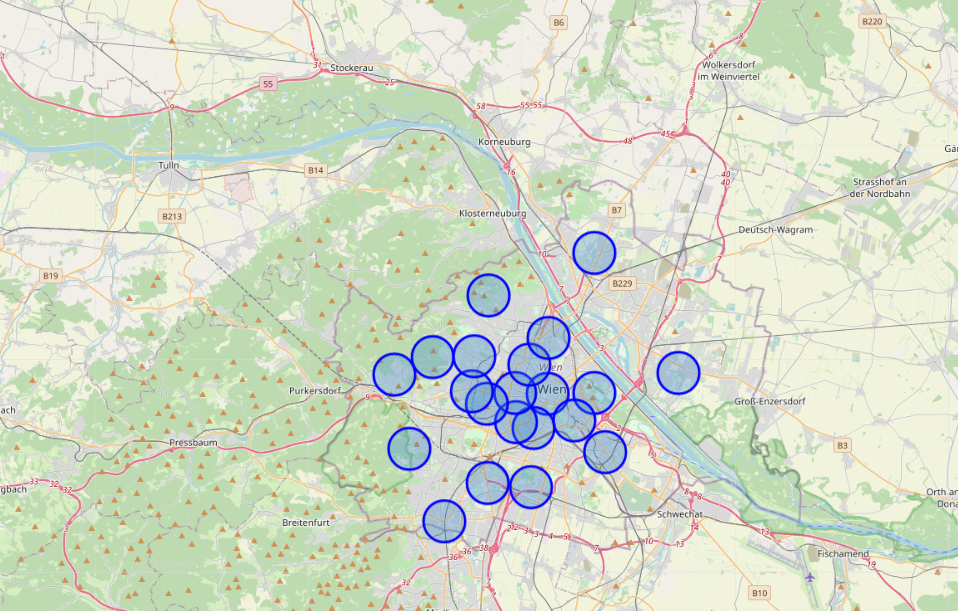
Data downloaded or scraped from multiple sources were combined into one table. We have found already cleaned data which we are going to use. Just one manipulation with rename of column.

1. **Methodology and Exploratory Data Analysis**

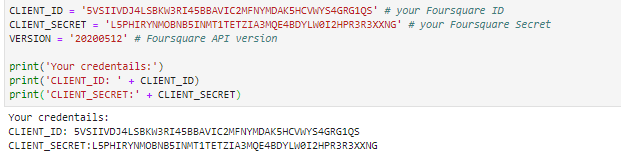
**3.1 Data Analysis and Location Data:**

* Foursquare location data will be leveraged to explore or compare districts around Vienna.
* Data manipulation and analysis to derive subsets of the initial data.
* Identifying the high traffic areas using data visualization and statistical analysis.

1. We are using the geopy library to get the latitude and longitude values of Vienna. Then we using Folium map for visualization of all districts in Vienna city.

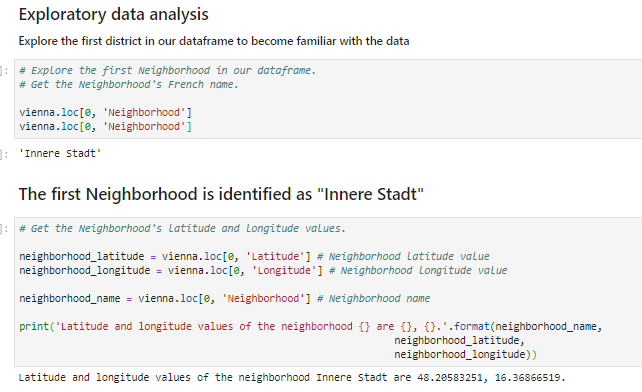


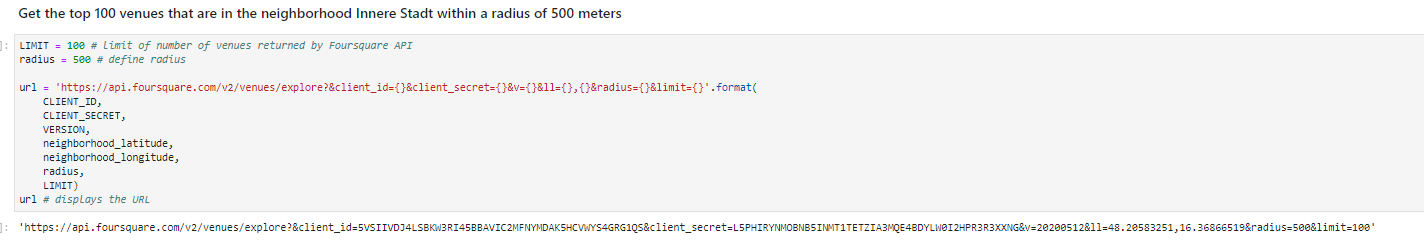
1. We are using the Foursquare API to explore the districts of Vienna.



**3.2 Exploratory data analysis**

1. Explore the first district in our dataframe to become familiar with the data
2. For example, we will explore the first Neighborhood in our datafram



1. After identifying the first disctrict we are going to get the top 100 venues that are in the neighborhood Innere Stadt within a radius of 500 meters
2. 
3. Then we are going to structure the json file into a pandas dataframe



1. We are going to create a nearby venues function for all the neighborhoods in Vienna and check how many venues were returned for each neighborhood



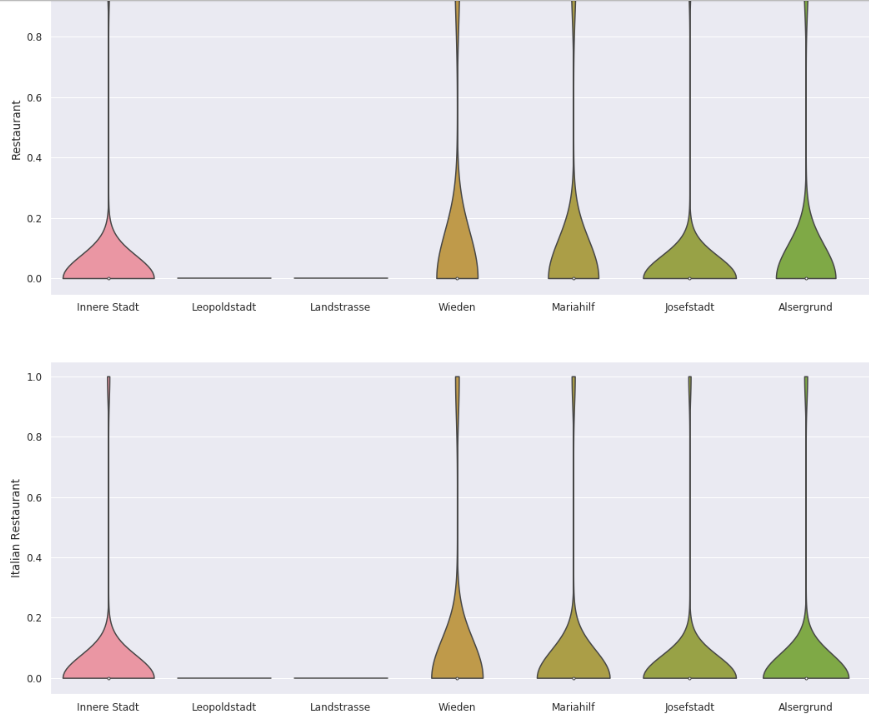
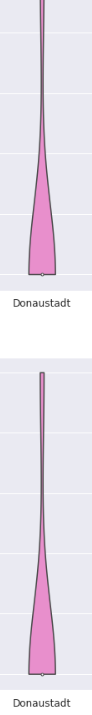
1. Analyze each of the Neighborhoods (Print each neighborhood with it's top 10 most common venues)
2. Show The top 10 venue categories for each neighborhood

This is a very useful results table that can provide at a glance information for all of the districts. Even once any conclusions are drawn further into the data workflow, we can refer back to this table for meaningful insights about the top categories of businesses in all the neighborhoods. Even without actual counts and numbers, it makes a great reference table for the client.



The business types criteria were specified by the stockholders. Two main category will be presented: Restaurant and Italian Restaurant.

Let's look at their frequency of occurrence for all the Vienna neighborhoods, isolating the categorical venues These are the types of places that the client wants to have in abundance in order to gain their market share. I've used a violin plot from the seaborn library - it is a great way to visualize frequency distribution datasets, they display a density estimation of the underlying distribution.

So as we can see from the analysis there are 6 neighborhoods to open new restaurant - according to the criteria that they have the 2 specified venues in a great frequency (Italian Restaurant, Restaurants. They are as follows:

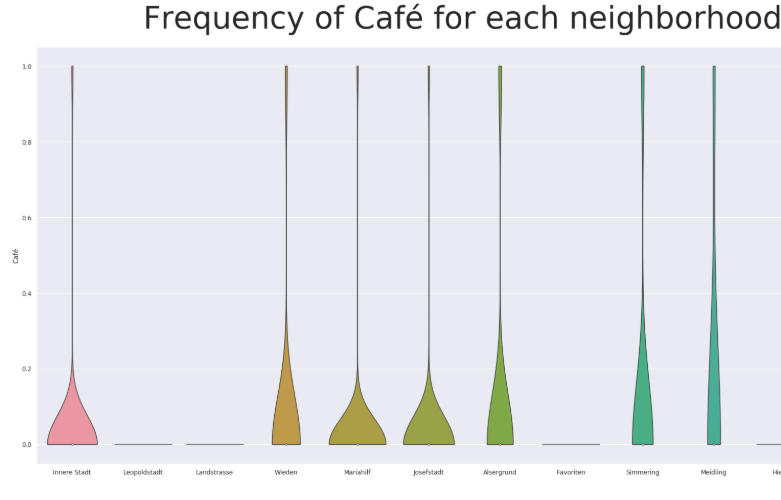
Neighborhoods

* Inner Stadt
* Wieden
* Mariahilf
* Josefstadt
* Alsergrund
* Donaustadt

Let's take this further with some exploration and Inferential Analysis

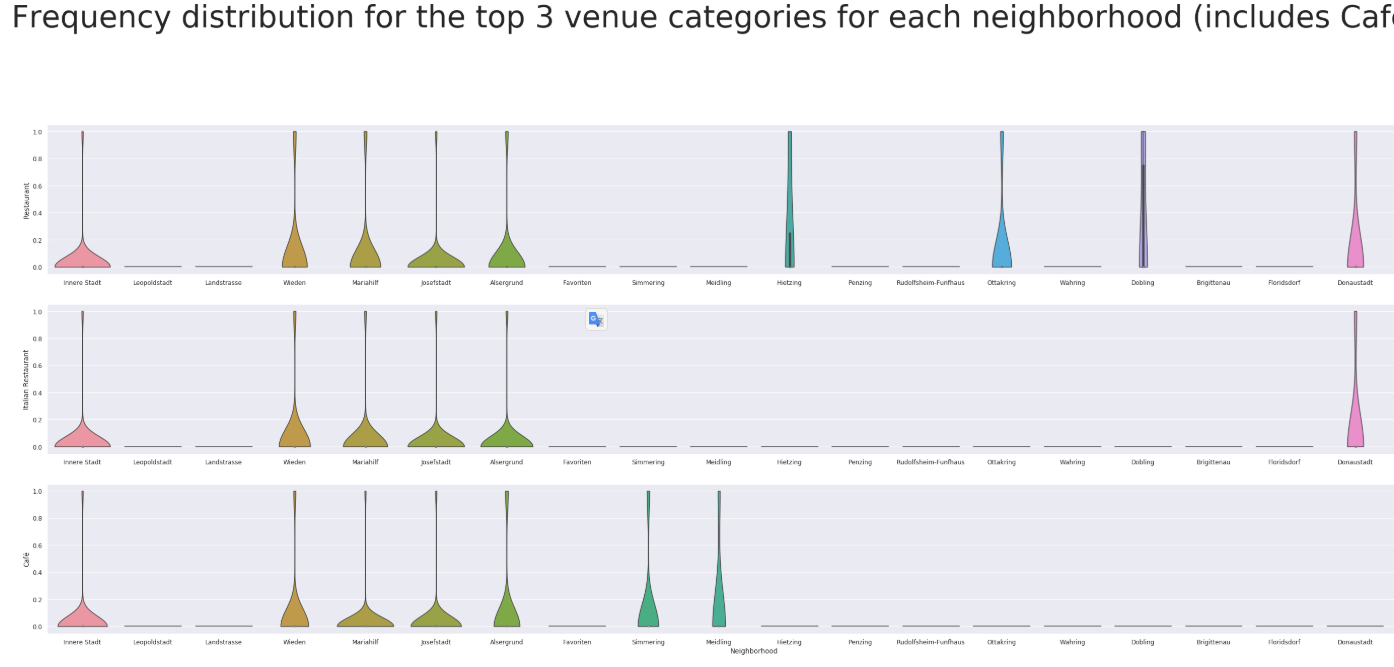
We have the 6 neighborhoods that all include the venue category criteria. But if we included the 'Café" venue category into the analysis, then we might be able to make some inferences based on the data, and domain knowledge of marketing and the industry, to focus the list.

Let's look at the venue category - 'Café'



So there are 7 neighborhoods that have a significant frequency density of Café.

Let's add this to the analysis with the other 3 specified categories as below.



1. **Inferences and Discussion**

**Chosen Neighborhoods - Results**

Inferential analysis using the data, as well as domain knowledge of retail and marketing, allow the list to be focused to just 5 neighborhoods from the previous 6.

The reasoning being that if the 3 criteria have been met - identifying neighborhoods that are lively with Restaurants and Italian Restaurants - adding Cafe into the mix of food services in the area is a significant bonus. Huge verity of restaurant in the same area - give us insight that:

* a lot of people usually spent time here
* people loves to be here
* high concentration of people
* a competitive market.

So the final 5 prospective neighborhoods for new restaurant are where 3 criteria are met:

Inner Stadt

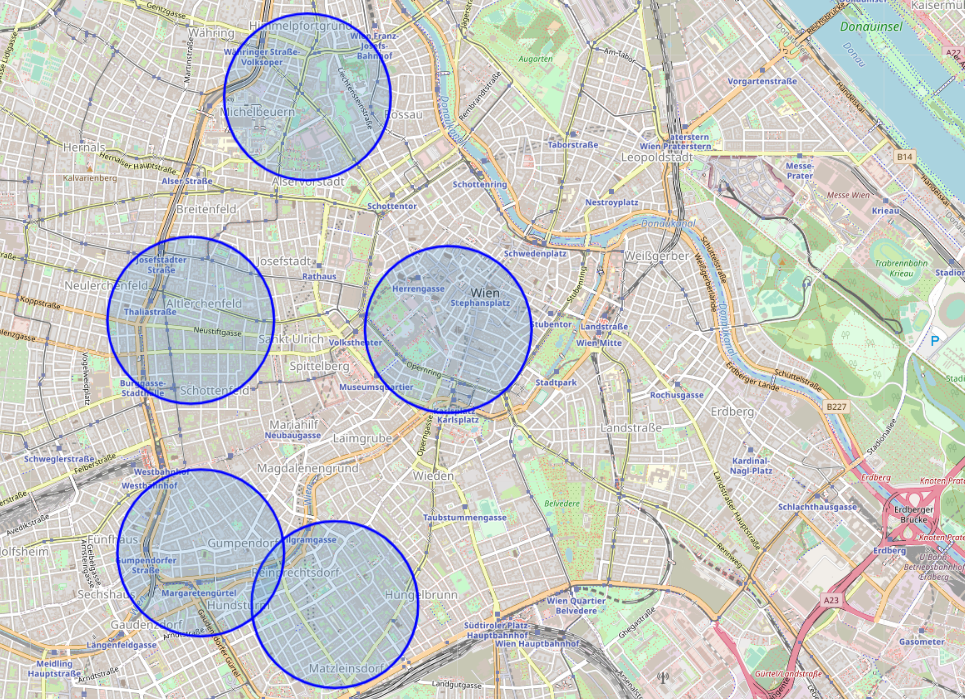
Wieden

Mariahilf

Josefstadt

Alsergrund

Let's look at the 5 districts on a Vienna map



**Observation**

As we can see, all district located in the central part of Vienna. Locations fitting the criteria for popular venues would normally be in central locations in many cities of the world. From beginning of our research we had 21 district. Without any data it would be extremely hard, time consuming and costly to do such research by visiting each of district. In our case we were able to collect the proper data and narrowed the search area down significantly from 21 potential districts to 3 that should suit the client's retail business.

**Inferences**

We have made inferences from the data in making the location recommendations, but that is exactly the point. There is no right or wrong answer or conclusion for the task at hand. The job of data analysis here is to advise the best location. For sure, if we were able to use more specific data, such report will be more valuable. However, as we mentioned at the beginning of our case, this report will be more suitable for investors who just want to understand the market without any knowledge.

**Conclusions**

There are many ways this analysis could have been performed based on different methodologies and different data sources. In this study, I analyzed the districts in Vienna to understand which location is better for a new restaurant. I identified districts, Latitude, Longitude. After I used the folium map for building a visualization of districts. Then we use a foursquare API to check how many venues were returned for each neighborhood. I originally intended to use the clustering algorithm to cluster the data, but as it progressed it became obvious that this only complicated the task at hand. The analysis and results are not an endpoint, but rather a starting point that will guide the next part of the process to find a better place for a new restaurant. Without leveraging data to make focussed decisions, the process could have been drawn out and resulted in sub-standard areas for our stakeholders. Data has helped to provide a better strategy and way forward, these data-driven decisions will lead to a better solution in the end.