

Suitable locations for a new restaurant in Vienna

Prepared by Myroslav Yatkevych

11-May-2020

1. Introduction

1.1 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.

1.2 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.

1.3 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.

2. Data acquisition and cleaning

2.1 Data sources

First, we need to identify what data is needed to use for a better understanding of the 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. Restaurants in each neighborhood of the city.

Data source: Foursquare 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

```
!pip install geopy
!pip install bs4
!pip install lxml
import numpy as np # Library to handle data in a vectorized manner
import pandas as pd # Library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
import json # Library to handle JSON files
!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API Lab
from geopy.geocoders import Nominatim # convert an address into Latitude and Longitude values
import requests # Library to handle requests
from pandas.io.json import json_normalize # transform JSON file into a pandas dataframe
# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
# import k-means from clustering stage
from sklearn.cluster import KMeans
from sklearn.datasets.samples_generator import make_blobs
!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API Lab
import folium # map rendering library
from bs4 import BeautifulSoup
import lxml
print('Libraries imported.')
```

2.2 Data cleaning

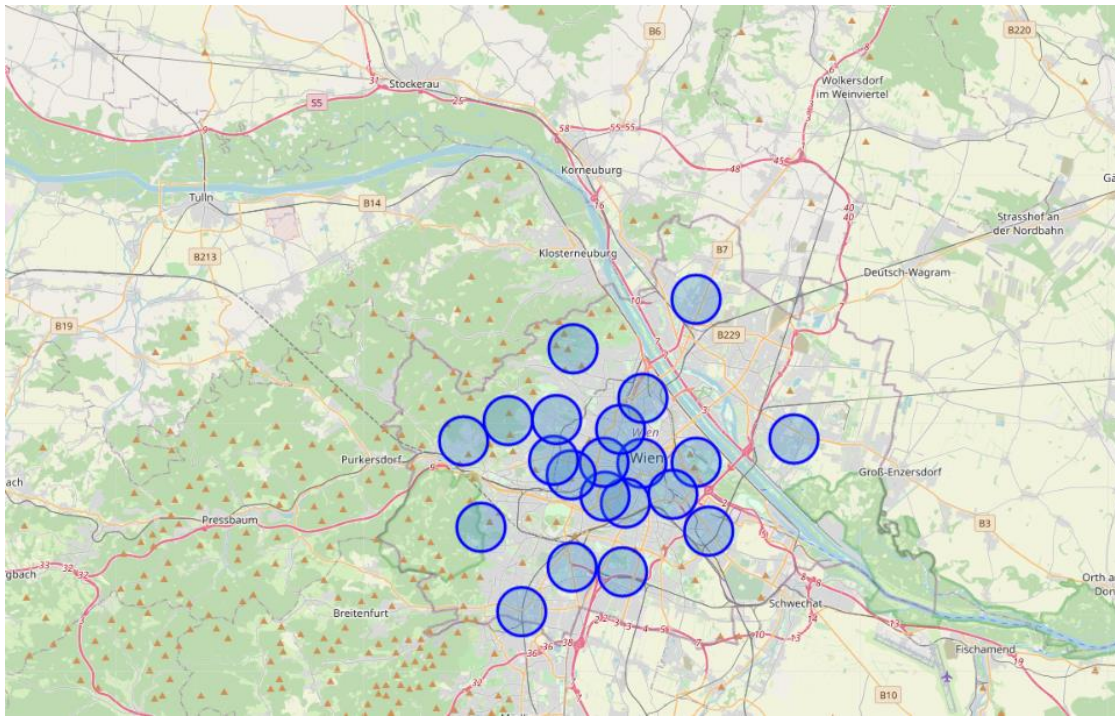
Data downloaded or scraped from multiple sources were combined into one table. We have found already cleaned data which we are going to use. Just one manipulation with rename of column.

3. 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.



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

```
CLIENT_ID = '5VSIIVDJ4LSBKW3RI45BBAVIC2MFNYMDAK5HCVWYS4GRG1QS' # your Foursquare ID
CLIENT_SECRET = 'L5PHIRYNMOBNB5INMT1TETZIA3MQE4BDYLW0I2HPR3R3XXNG' # your Foursquare Secret
VERSION = '20200512' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: 5VSIIVDJ4LSBKW3RI45BBAVIC2MFNYMDAK5HCVWYS4GRG1QS
CLIENT_SECRET: L5PHIRYNMOBNB5INMT1TETZIA3MQE4BDYLW0I2HPR3R3XXNG
```

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

Exploratory data analysis

Explore the first district in our dataframe to become familiar with the data

```
] : # Explore the first Neighborhood in our dataframe.
    # Get the Neighborhood's French name.

vienna.loc[0, 'Neighborhood']
vienna.loc[0, 'Neighborhood']

]: 'Innere Stadt'
```

The first Neighborhood is identified as "Innere Stadt"

```
] : # Get the Neighborhood's Latitude and Longitude values.

neighborhood_latitude = vienna.loc[0, 'Latitude'] # Neighborhood Latitude value
neighborhood_longitude = vienna.loc[0, 'Longitude'] # Neighborhood Longitude value

neighborhood_name = vienna.loc[0, 'Neighborhood'] # Neighborhood name

print('Latitude and longitude values of the neighborhood {} are {}, {}'.format(neighborhood_name,
                                                                              neighborhood_latitude,
                                                                              neighborhood_longitude))

Latitude and longitude values of the neighborhood Innere Stadt are 48.20583251, 16.36866519.
```

- 3) After identifying the first district we are going to get the top 100 venues that are in the neighborhood Innere Stadt within a radius of 500 meters

Get the top 100 venues that are in the neighborhood Innere Stadt within a radius of 500 meters

```
] : LIMIT = 100 # Limit of number of venues returned by Foursquare API
    radius = 500 # define radius

url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url # displays the URL

]: 'https://api.foursquare.com/v2/venues/explore?&client_id=5VSIIVD34LS8K03RI4588BAVIC2MFVWIDAKSHCVWYS4GRG1Q5&client_secret=L5PHIRYWM0BNBSIWT1TETZTIA3HQE48DYLM0I2HPR3R3XXNG&v=20200512&ll=48.20583251,16.36866519&radius=500&limit=100'
```

- 4)
5) Then we are going to structure the json file into a pandas dataframe

	name	categories	lat	lng
0	Albertina	Art Museum	48.204620	16.368256
1	1010 BAR CAFE	Bar	48.206445	16.370218
2	Prunksaal der Nationalbibliothek	Historic Site	48.206182	16.366775
3	Nightfly's Club American Bar	Cocktail Bar	48.207055	16.368883
4	The Guesthouse Vienna	Hotel	48.204784	16.368875
5	Hofburg	Palace	48.206629	16.365595
6	Albertinaplatz	Plaza	48.204319	16.368909
7	Hotel Sacher	Hotel	48.203798	16.369702
8	Dorotheum	Antique Shop	48.206340	16.368630
9	Michaelerplatz	Plaza	48.208007	16.366625
10	Burggarten	Park	48.204281	16.365519
11	Wiener Staatsoper	Opera House	48.203275	16.369050
12	Kärntner Straße	Pedestrian Plaza	48.207002	16.371795
13	Apple Kärntner Straße	Electronics Store	48.207045	16.371746
14	Ferrari Gelato	Ice Cream Shop	48.203908	16.371523
15	Ilona Stüberl	Hungarian Restaurant	48.208482	16.369517
16	Stephansplatz	Plaza	48.208299	16.371880
17	Kohlmarkt	Shopping Plaza	48.208330	16.366840
18	Graben	Pedestrian Plaza	48.208915	16.369379
19	Heidenplatz	Plaza	48.206247	16.363610

- 6) 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

Create a new dataframe called for the venues of Vienna called vienna-venues

```
#Create a new dataframe called vienna_venues.  
vienna_venues = getNearbyVenues(names=vienna['Neighborhood'],  
                                latitudes=vienna['Latitude'],  
                                longitudes=vienna['Longitude']  
                                )
```

```
Innere Stadt  
Leopoldstadt  
Landstrasse  
Wieden  
Mariahilf  
Josefstadt  
Alsergrund  
Favoriten  
Simmering  
Meidling  
Hietzing  
Penzing  
Rudolfsheim-Funfhaus  
Ottakring  
Hernals  
Wahring  
Dobling  
Brigittenau  
Floridsdorf  
Donaustadt  
Liesing
```

```
print(vienna_venues.shape)  
vienna_venues.head(250)
```

```
(380, 7)
```

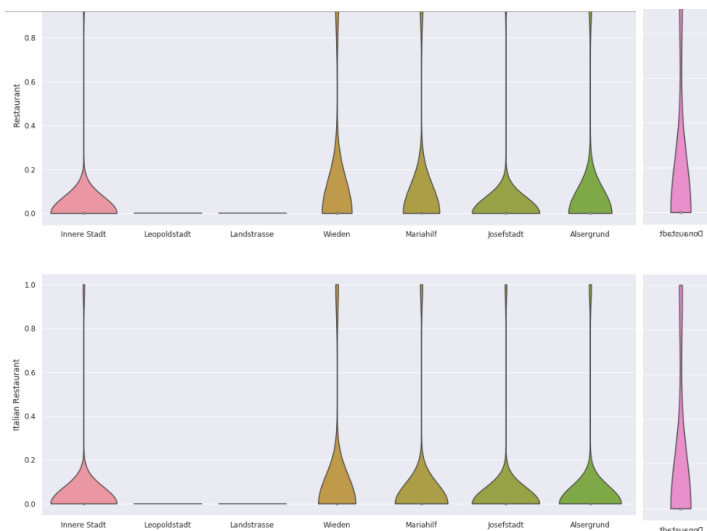
- 7) Analyze each of the Neighborhoods (Print each neighborhood with it's top 10 most common venues)
- 8) Show The top 10 venue categories for each neighborhood

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.

	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	10th Most Common Venue
0	Alsergrund	Café	Hotel	Restaurant	Tram Station	Asian Restaurant	Italian Restaurant	Park	Korean Restaurant	Bakery	Supermarket
1	Brigittenau	Supermarket	Pet Store	Gastropub	Plaza	Bus Stop	Gym Pool	Pharmacy	Scenic Lookout	Tram Station	Austrian Restaurant
2	Dobling	Restaurant	Vineyard	Farm	Bus Stop	Scenic Lookout	Drugstore	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
3	Donaustadt	Supermarket	Italian Restaurant	Restaurant	Plaza	Pizza Place	Breakfast Spot	Light Rail Station	Dessert Shop	Gym	Ethiopian Restaurant
4	Favoriten	Tram Station	Park	Plaza	Drugstore	Farm	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Wine Shop
5	Floridsdorf	Gastropub	Grocery Store	Wine Shop	Light Rail Station	River	Fast Food Restaurant	Supermarket	Wine Bar	Gym Pool	Farm
6	Hietzing	Restaurant	Park	Construction & Landscaping	Bakery	Dumpling Restaurant	Farm	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Drugstore
7	Innere Stadt	Plaza	Hotel	Park	Supermarket	Clothing Store	Museum	Pedestrian Plaza	Ice Cream Shop	Church	Cocktail Bar
8	Josefstadt	Supermarket	Mediterranean Restaurant	Asian Restaurant	American Restaurant	Organic Grocery	Restaurant	Clothing Store	Pub	Gym	Polish Restaurant
9	Landstrasse	Supermarket	Pizza Place	Hotel	Pub	Tram Station	Market	Korean Restaurant	Bar	Bakery	Plaza
10	Leopoldstadt	Sports Club	Gastropub	Dog Run	Park	Bus Stop	Snack Place	Sports Bar	Beer Garden	Baseball Field	Tennis Court
11	Liesing	Cosmetics Shop	Basketball Court	Dance Studio	Hotel	Coffee Shop	Clothing Store	Shopping Mall	Bus Stop	Plaza	Steakhouse
12	Mariahilf	Hotel	Restaurant	Austrian Restaurant	Vietnamese Restaurant	Supermarket	Tram Station	Bar	Italian Restaurant	Hostel	Korean Restaurant
13	Meidling	Supermarket	Asian Restaurant	Furniture / Home Store	Café	Athletics & Sports	Pharmacy	Doner Restaurant	Dog Run	Drugstore	Flower Shop
14	Ottakring	Supermarket	Bus Stop	Restaurant	Plaza	Tram Station	Bakery	Thrift / Vintage Store	Gym	Theater	Trattoria/Osteria
15	Penzing	Home Service	Wine Shop	Flower Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dog Run	Doner Restaurant	Drugstore	Dumpling Restaurant
16	Rudolfsheim-Funfhaus	Chinese Restaurant	Recreation Center	Bakery	Irish Pub	Jazz Club	Gym / Fitness Center	Grocery Store	Farmers Market	Electronics Store	Bus Stop
17	Simmering	Café	Tram Station	Hotel	Bus Stop	Supermarket	Gastropub	Shopping Mall	Shoe Store	Seafood Restaurant	Cafeteria
18	Wahring	Bus Stop	Gastropub	Pizza Place	Seafood Restaurant	Soccer Field	Wine Shop	Dumpling Restaurant	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
19	Wieden	Austrian Restaurant	Restaurant	Bar	Café	Plaza	Supermarket	Italian Restaurant	Bakery	Liquor Store	Eastern European Restaurant

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

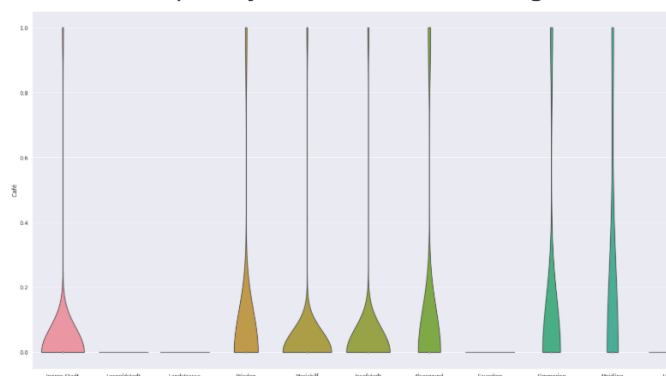
- Inner Stadt
- Wieden
- Mariahilf
- Josefstadt
- Alsergrund
- Donaustadt

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'

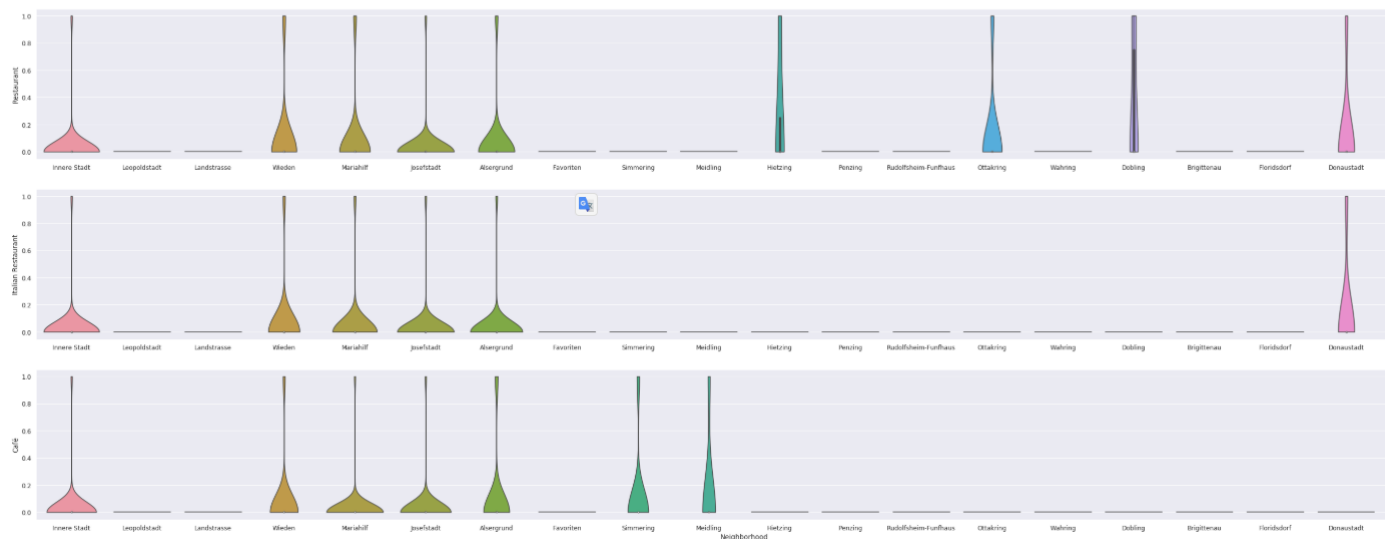
Frequency of Café for each neighborhood



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.

Frequency distribution for the top 3 venue categories for each neighborhood (includes Caf



4. 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

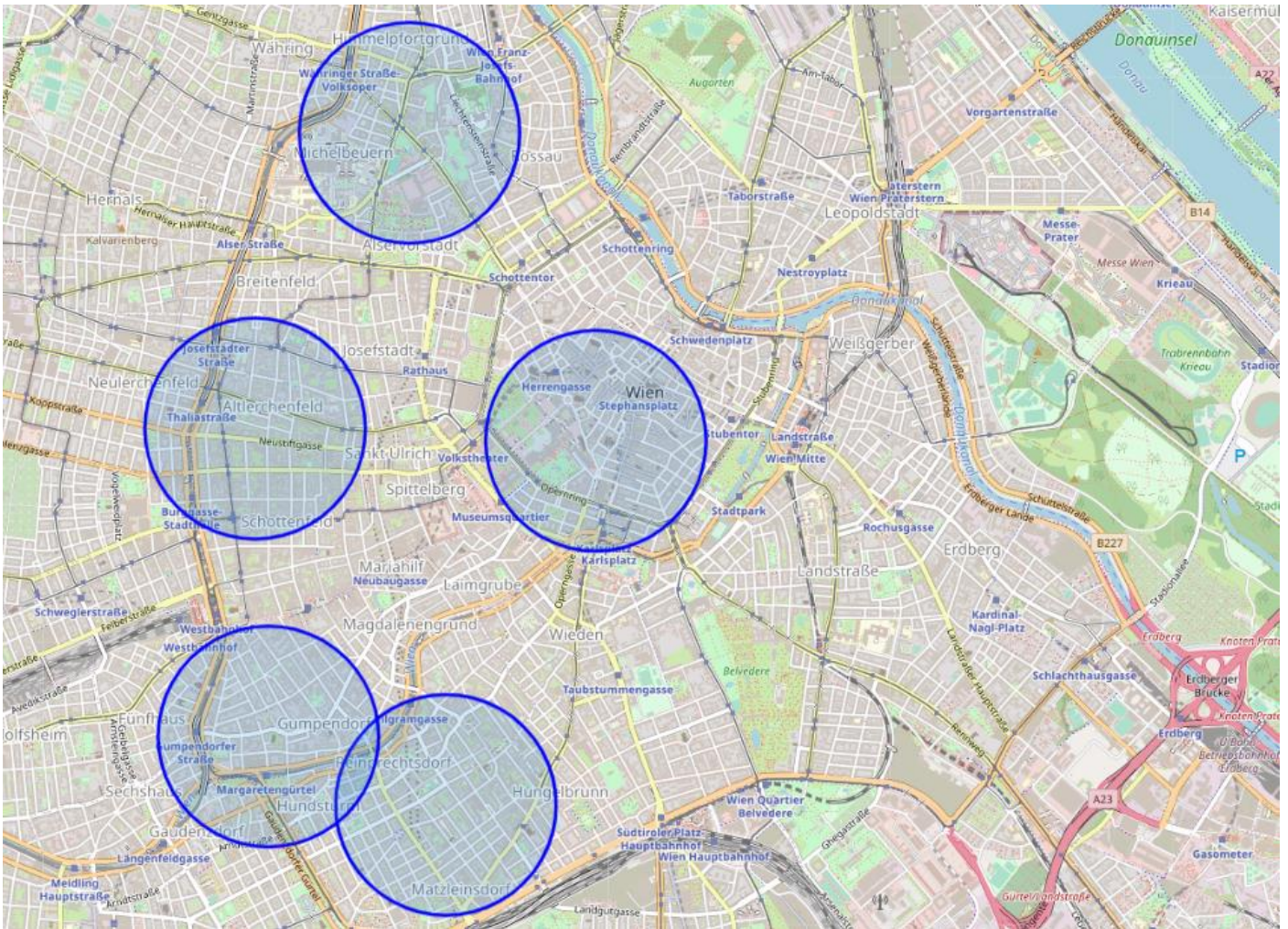
Wieden

Mariahilf

Josefstadt

Alsergrund

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



Observation

As we can see, all districts 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 the beginning of our research we had 21 districts. Without any data it would be extremely hard, time consuming and costly to do such research by visiting each of the districts. 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 a 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 focused 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.

