Vienna Cafés District Map

Finding the best location for opening a new coffee place in Vienna

Coursera IBM Data Science Capstone Project

Katerina Mincheva, April 2019

1.   Introduction

Vienna is known for its strong coffee culture, which can be also recognized by the solid number of cafés and roasteries in the city, as well as the good quality of the coffee, that the Viennese highly value. It is a renowned cultural trait as well as a habit of the people to go out and enjoy fresh brewed coffee in local cafés. Furthermore, the vibrant city atmosphere, the big number of students, as well as that of tourists that are looking to experience the coffee culture, secure the constant demand for such places. This is the reason why both established café owners seek to expand their businesses in new locations, as well as new entrants see the market as an attractive and never too saturated to enter.

Thus, the aim of this report is to create an overview of the 'Café map' of Vienna that will have many uses. It will provide a cluster analysis and a valuable overview of the amount of coffee places in each of Vienna's so-called districts. This could, on one hand, be used by entrepreneurs looking for opening a new café in the right location. However, it is also a tool that could help students or Vienna's visitors to choose, for example, an accommodation in an area with bigger saturation of renowned coffee places.

In a business perspective, we will look for an answer to the question: **Which districts of Vienna offer the most attractive location for opening a new café?**

2.   Data

The data to be used in the scope of the project comes from few different sources:

i.   A Wikipedia article containing the needed information about the 23 different districts in Vienna, called 'Bezirke' (German): https://de.wikipedia.org/wiki/Wiener_Gemeindebezirke. In order to extract the required information - the number of the district and its name, that is contained in a table, web scrapping technique will be applied.

ii.   Location data - the coordinates of each district, will be then collected through the Python Geocoder. Having each district's coordinates will then be used in the next step.

iii.   Foursquare - after having each district's coordinates, analysis of all Café venues will be done by querying data and gathering all relevant information through the Foursquare API. Foursquare provides information related to venues in a specific area, such as location, venue category, reviews and tips. Data will be extracted by using the latitude and longitude coordinates of each district and saved in a json object, which will then be transformed in a data frame with the corresponding venue information. Based on that, a clustering of the districts will be done.

iv.   In addition, from Statistics Austria we will extract information regarding the population at the end of 2018 of each district, to later see if there are districts with bigger population, but smaller amount of coffee places: http://statistik.at/web_en/statistics/index.html, a potentially attractive location for opening up new cafés.

3.   Methodology

4.   Results

5.   Discussion

6.   Conclusion