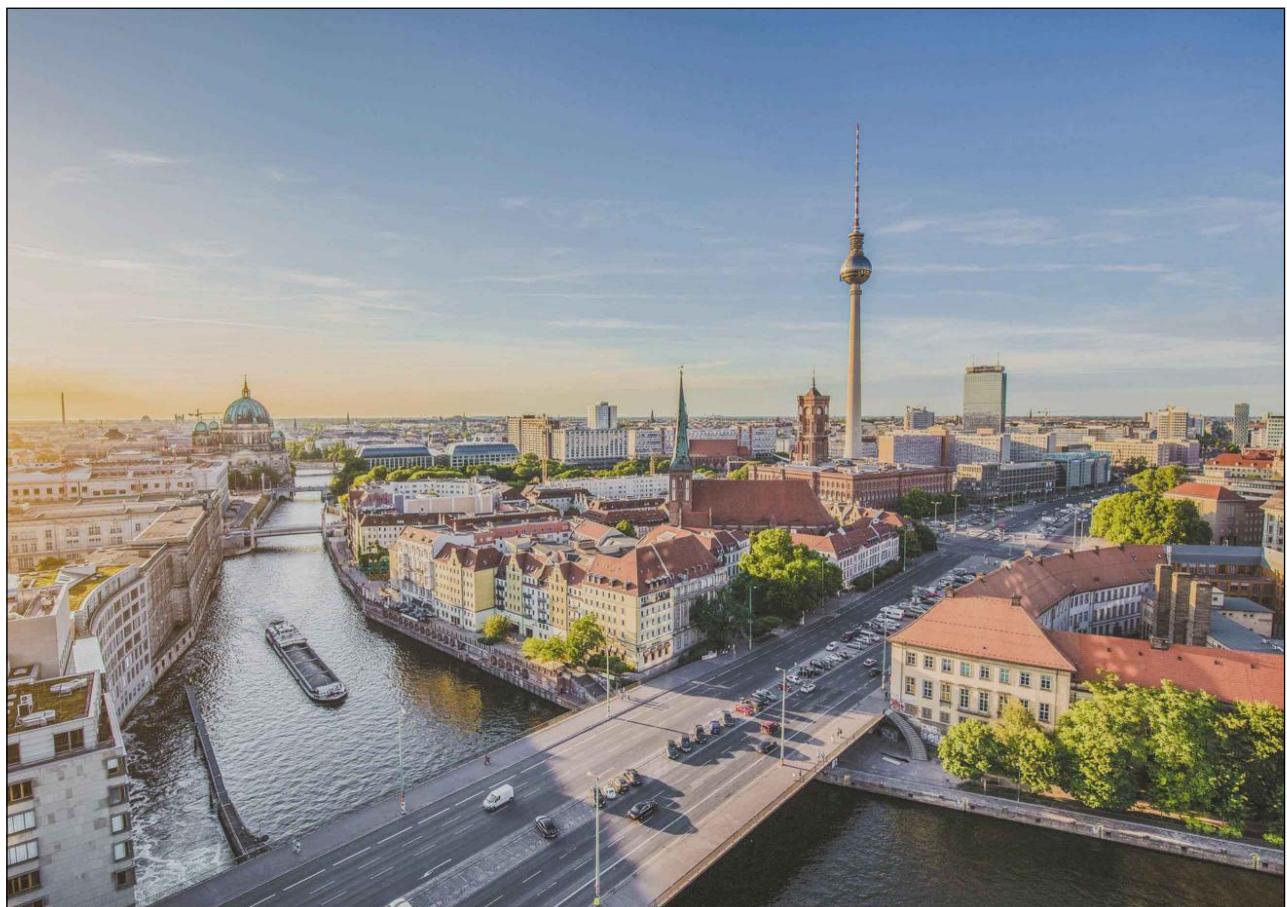

The Battle of Neighborhoods

IBM Data Science Professional Certificate - Capstone Project
Opening a new Restaurant in Berlin, Germany

Utkarsh Sharma - 4 June 2019



Introduction

One of the most culturally and historically rich city in Europe, Berlin is famous for its food scene, coffee houses and good beer. Our hypothetical client, *XYZ Foods*, is a restaurant chain that runs a successful restaurant in Milan. Their Italian restaurant, *XYZ Fine Dining* is a hit among the people and they want to start another branch of *XYZ Fine Dining* in Berlin.



Berlin: Rich in culture and food

Business Problem: The client is unable to decide which neighbourhood of Berlin would be best suited for the new chain of their Italian restaurant. The objective of this Capstone project is to analyse the various Neighborhoods of Berlin, Germany and select the best location for opening a new restaurant. In simple words, the business question this project aims at solving is: If someone is looking to open an Italian restaurant in the city of Berlin, where would you recommend that they open it?

Target Audience: This project is particularly useful for restaurant chain owners and/or investors looking to open and/or invest in an Italian restaurant in the capital of Germany.

Data Requirements and Extraction

To solve this problem, we will need the following data:

- List of Neighborhoods in Berlin
- Latitude and Longitude coordinates of all the Neighborhoods
- Venue data (powered by Foursquare), particularly related to Italian Restaurants

Sources of Data and Extraction Procedure: The Wikipedia page of Boroughs and neighborhoods of Berlin list all the neighborhoods of the city. Using Web Scraping techniques, we will extract the required data, with the help of Python Requests and BeautifulSoup packages. The Python Geocoder package would be essential in finding the location data of these neighborhoods. At last, the Foursquare API would be used to extract the data regarding existing restaurants in the city.

In general, this project would be encompassing a series of Data Science techniques, including, but not limited to, Web Scraping (using BeautifulSoup and Requests), Data Cleaning, Data Wrangling and Machine Learning (K-Means clustering algorithm)

Methodology

To get a list of neighborhoods in the city of Berlin. This data is extracted from the Wikipedia page titled, 'Localities of Berlin' (https://en.wikipedia.org/wiki/Category:Localities_of_Berlin). Using the BeautifulSoup and Requests packages of Python, the required data is scraped from the webpage. In order to use the Foursquare API, we fetch the location data of all these neighbourhoods from the Python Geocoder package. Next, the Foursquare API is used to get the top 100 venues in a radius of 2000 meters of a particular neighborhoods. To prepare the data for K-means clustering, we group the data frame by neighborhoods. Lastly, K-means clustering is performed on this data set to return 4 clusters, or categories of neighborhoods in terms of number of Italian Restaurants.

Results

The output from the clustering algorithm shows the division of all the neighborhoods into 4 groups in terms of number of Italian restaurants present in that area.

- Cluster 0: Neighborhoods with very low number of Italian restaurants
- Cluster 1: Neighborhoods with low number of Italian restaurants
- Cluster 2: Neighborhoods with a significant number of Italian restaurants
- Cluster 3: Neighborhoods with heavy competition of Italian restaurants

The results are visualised as follows:

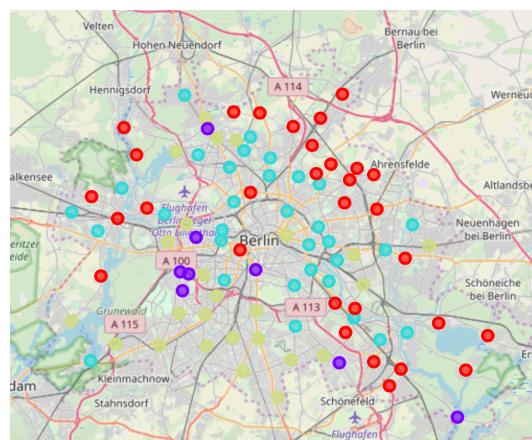
Here,

Cluster 0 = Red

Cluster 1 = Purple

Cluster 2 = Light Blue

Cluster 3 = Yellow



Discussion

As observed from the clusters, opening a restaurant in Berlin could be tricky business. Going for Restaurants in Category 0 can lead to less footfall, but also non-existence of competition, hence, this is not recommended. On the other hand, Category 3 suffers from

an intense competition and this is also not recommended. Italian restaurants with a USP like innovative menu, dirt cheap prices etc. can go for Category 1 but Category 2 is recommended more than Category 1 due to a significantly higher expected target audience. Please check the Appendix for the complete list of neighborhoods divided into these categories.

Limitations and Suggestions

Since Berlin isn't an Italian city, nor are Italians a major ethnic group, tourism is a major part of the business of Italian Restaurants in the city. Due to unavailability of competent neighbourhood-based travel data, this factor couldn't be taken into account in this project. Also, the clustering results highly depend on the accuracy of the Foursquare API. Lastly, since the API calls to the Foursquare API were made from a free Sandbox account, the results were limited. Data Scientists with a premium Foursquare account can overcome this bottleneck.

Appendix

Category 1: Neighborhoods with very low number of restaurants

Alt-Hohenschönhausen	Lübars	Wartenberg (Berlin)
Altglienicke	Malchow (Berlin)	Friedrichshagen
Baumschulenweg	Marzahn	Gatow
Blankenburg (Berlin)	Müggelheim	Gesundbrunnen
Blankenfelde	Neu-Hohenschönhausen	Grünau (Berlin)
Bohnsdorf	Oberschöneweide	Haselhorst
Buch (Berlin)	Rahnsdorf	Heiligensee
Falkenberg (Berlin)	Spandau (locality)	Johannisthal (Berlin)
Falkenhagener Feld	Stadtrandsiedlung Malchow	Karow (Berlin)
Französisch Buchholz	Tiergarten, Berlin	
Kaulsdorf (Berlin)	Konradshöhe	

Category 2: Neighborhoods with low number of restaurants

Charlottenburg	Rudow
Grunewald	Schmargendorf
Halensee	Schmöckwitz
Kreuzberg	Waidmannslust

Category 3: Neighborhoods with a significant number of restaurants

Adlershof	Hellersdorf	Rosenthal
Alt-Treptow	Köpenick	Rummelsburg
Biesdorf	Lichtenberg	Schöneberg

Borsigwalde	Moabit	Siemensstadt
Britz	Neukölln	Staaken
Fennpfuhl	Niederschöneweide	Wannsee
Friedrichshain	Niederschönhausen	Wedding
Frohnau	Pankow	Weissensee
Hakenfelde	Plänterwald	Karow (Berlin)
Hansaviertel	Prenzlauer Berg	Wilhelmstadt
Heinersdorf	Reinickendorf	

Category 4: Neighborhoods crowded with restaurants

Buckow	Lankwitz	Tegel
Charlottenburg-Nord	Lichtenrade	Tempelhof
Dahlem	Lichterfelde	Westend
Friedenau	Mahlsdorf	Wilmersdorf
Friedrichsfelde	Mariendorf	Wittenau
Gropiusstadt	Märkisches Viertel	Zehlendorf
Hermsdorf	Nikolassee	Kladow
Karlshorst	Steglitz	Mitte