The Battle of Neighborhoods

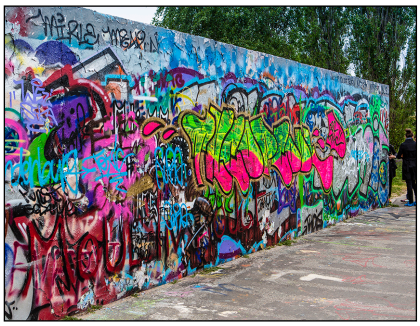
IBM Data Science Professional Certificate - Capstone Project

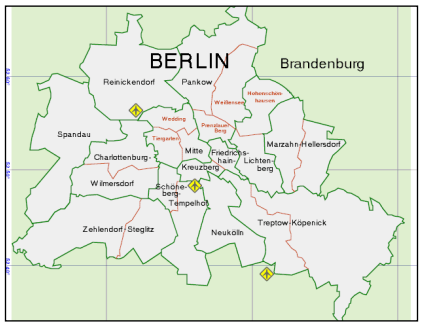
Opening a new Restaurant in Berlin, Germany

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 London. Their 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 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 a 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 a 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 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)