Coursera.org

IBM Applied Data Science Capstone

Proposal for a new vegetarian / vegan restaurant in Berlin, Germany

By: Tatiane MP

January 2020

Introduction

In recent years, the German vegetarian/vegan population has rapidly grown. As of 2015, the vegetarian/vegan market was worth $520 million and saw a increase rate of 17%. This is the result of a population that is more and more concerned about animal welfare, the environment, and, especially, their health.

According to the European Vegetarian Union, 10% of German consumers (7.8 million individuals) are vegetarians, and 1.1% are vegans (900,000 individuals). These numbers have doubled since 2006 and many believe that these numbers will only continue to increase in the coming years.

For this reason, the opening of a vegetarian / vegan restaurant is a great business opportunity in a so eco – friendly city like Berlin. But of course, as with any business decision, opening a new venue requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the vegetarian / vegan restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best location in the city of Berlin, Germany to open a new vegetarian / vegan restaurant. For this purpose, it will be used data science methodology and machine learning techniques like clustering to find out which part of the city would be the ideal place to open a new vegetarian / vegan restaurant. This project aims to provide an answer to a very important question: If a property developer is looking to open a new vegetarian / vegan restaurant in Berlin, Germany, where would be the best recommend place to do it?

Target Audience

As with any restaurant, vegetarian and vegan businesses need a location where they have a strong chance of building positive buzz and developing a base of repeat customers. Entrepreneurs pointed out that new owners must consider questions like how accessible a spot is to likely patrons, the foot traffic in the area, plans for future development nearby and the affordability of

the lease. For a restaurant specializing in vegetarian cuisine, the best option is usually an relatively upscale area frequented by health-conscious people.

Therefore, this project is particularly useful to entrepreneurs, property developers and investors looking to open or invest in vegetarian / vegan restaurants in Berlin (DE). This project is timely as the city has a great potential to such a kind of business, because like the Culture Trip Internet Page[1] writes: "But the ever increasing number of vegetarians in Berlin has given rise to a vibrant meat-free dining scene: vegan currywurst, all-you-can-eat buffets, and Himalayan dumplings are just a few of the vegetarian offerings in the German capital."

A survey from "Forsa" also revealed that approximately 42 million people in Germany identify as flexitarians aka "part time vegetarians." Professionals at the German Official Agencies estimate that by 2020 over 20% of Germans will eat mostly vegetarian and this fact creates an excellent opportunity for those ones who would like to start a vegetarian / vegan restaurant. Besides that, many vegetarian / vegan restaurants also sell food products as well as vegetarian / vegan cookbooks, and this can be a terrific way to increase profits.

Data

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

- List of Berlin's districts . This defines the scope of this project which is confined to the city of Berlin, the capital of the Germany.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to vegetarian / vegan restaurants. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them:

- Wikipedia[2], that contains a list of all districts of Berlin , with a total of 12 districts. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the districts using a GitHub CSV file that's already there.
- Foursquare API: we will use Foursquare API to get the venue data for those districts. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API

[1] https://theculturetrip.com/europe/germany/articles/the-best-vegetarian-restaurants-in-berlin/
[2] https://de.wikipedia.org/wiki/Kategorie:Bezirk_von_Berlin

will provide many categories of the venue data, we are particularly interested in the Vegetarian / Vegan Restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).