

Coursera Capstone

IBM Applied Data Science Capstone

Opening a Falafel fast food restaurant in Berlin, Germany

1- Introduction

Vegetarian food becomes a habit for plenty of people, oriental cuisine is also a favorite one since it serves many vegetarian dishes. For many people, Falafel sandwich or snack is one of the favorite oriental vegetarian food in European countries like Germany.

In recent years, the country has embraced vegetarian and vegan cuisine, and today it is almost always possible to find several delicious vegetarian options on every restaurant menu. se, as with any business decision, opening a new fast-food restaurant – even as small business - requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurant is one of the most important decisions that will determine whether it will be a success or a failure.(source: <https://fearlessfemaletravels.com/eating-vegetarian-in-germany/>)

"It's certainly fashionable. If you really want to be cool in Berlin, you've got to be vegan," Sebastian Joy of Vebu, Germany's largest vegetarian and vegan organization, told DW. (source: <https://www.dw.com/en/berlin-vegan-capital-of-the-world/a-35951064-0>)

So in Germany and especially in the multicultural Metropolis Berlin such restaurants will be a good and profitable low cost business.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Berlin, Germany to open a new Falafel restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: **In the city of Berlin, if an investor is looking to open a new Falafel fast-food restaurant or even a series of restaurant with a special brand, where would you recommend that they open it?**

Target Audience of this project

This project is particularly useful to middle and small investors looking to open or invest in new Falafel fast-food restaurant in the capital city of Germany i.e. Berlin. This project is timely as the city is currently suffering from oversupply of restaurants but not vegan ones!

According to "organic-market.info": The vegan trend has now reached the gastronomy sector. According to the GV-Barometer 2017, a rich variety of vegetarian food is one of the

most important trends in gastronomy of the future. More and more guests want to see healthy vegetarian dishes on the menu, and 58 percent of entrepreneurs in gastronomy assume that the importance of vegetarian food will continue to grow. In big and medium-size cities in Germany there are currently at least 616 purely vegetarian eateries, which is nearly twice as many as last year. Berlin is clearly in the lead with its 193 wholly vegetarian restaurants and cafés but in other cities too you find plenty of places where you can try out the vegetarian offer. (source: <https://organic-market.info/news-in-brief-and-reports-article/germany-9-3-million-vegetarians-and-vegans.html>)

2- Data

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

- _List of neighborhoods in Berlin. This defines the scope of this project which is confined to the city of Berlin, the capital city of the country of 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 restaurants. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin) contains a list of neighborhoods in Berlin, with a total of 12 boroughs. 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 neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 150,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the restaurants 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). Next, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.