Opening a new Burger Restaurant in Hamburg, Germany

By: Simon Schuster

July 2021

Coursera Capstone Project

IBM Data Science

1. Introduction

A thick, juicy beef patty, cheese losing its hard edges as it melts overtop, a slathering of sauce, a spoonful of tangy relish and a couple of slices of smoky, salty bacon, a bit of onion, some tomato and crisp lettuce all tucked between two halves of a bun: it is all the things you could want.[1]
The fast-food trend is growing in Germany. As a result, burger restaurants become more and more popular in big cities like Hamburg. Opening a new restaurant requires serious consideration and is a lot more complicated than it seems. The location of a burger restaurant is one of the most important decisions that will determine whether it will be a success or a failure.

1.1 Business Problem

The objective of this capstone project is to find the best neighborhood in Hamburg to open a Burger Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide a solution to answer the business question: In the city of Hamburg, Germany, where would you recommend opening a new burger restaurant?

2. Data

To solve the problem, we need following data:

- List of neighborhoods in Hamburg
- The coordinates of the neighborhoods. This is required to plot the map and also get the venue data from the Foursquare API
- Venue data, particularly data related to food and restaurants. We will use this data to cluster the different neighborhoods. We will also be using population density data for clustering.

Source of the data and how to extract them:

This Wikipedia page (https://de.wikipedia.org/wiki/Liste_der_Bezirke_und_Stadtteile_Hamburgs) contains a list of neighborhoods in Hamburg. With the help of Python requests and beautifulsoup packages, we will use web scrapping techniques to get the data from the page. We store the data in a pandas dataframe. To get the coordinates of each neighborhood, we use a Python Geocoder package. This will give use the latitude and longitude. After this the dataframe looks like follows:

	Neighborhood	Latitude	Longitude	Borough	Population density
0	Hamburg-Altstadt, Hamburg	53.550468	9.994640	Hamburg-Mitte	979
1	HafenCity, Hamburg	53.542913	9.995835	Hamburg-Mitte	2239
2	Neustadt, Hamburg	53.549881	9.979048	Hamburg-Mitte	5549
3	St. Pauli, Hamburg	53.553935	9.959432	Hamburg-Mitte	8839
4	St. Georg, Hamburg	53.557149	10.014256	Hamburg-Mitte	4733

Then we will use the Foursquare API to get the venue data of the neighborhoods. Foursquare has one of the largest databases. It will provide many categories of the venue data, we are particularly interested in the food and restaurants.

This is a project that will make use many data science skills, from web scrapping, working with Foursquare API, data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization.

References:

[1]: https://calgaryherald.com/life/food/the-battle-of-the-burgers-an-introduction