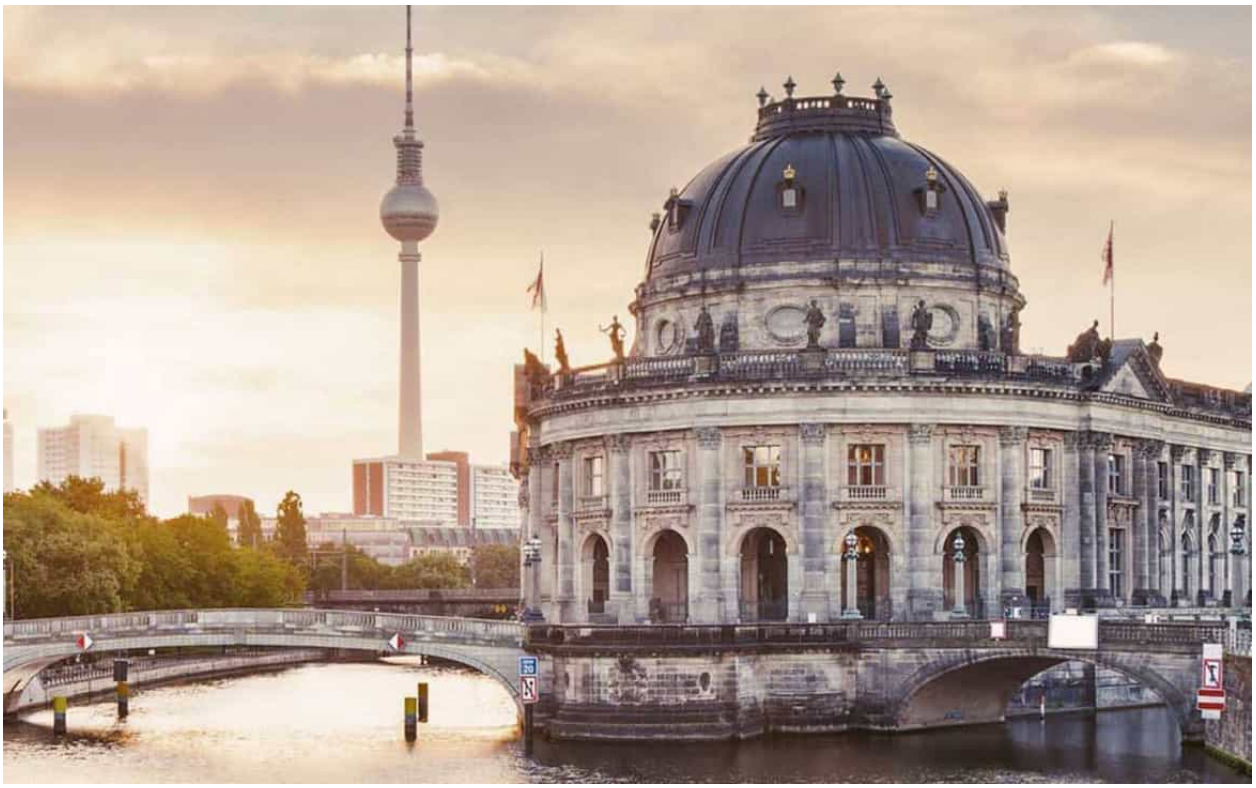
Coursera Capstone

IBM Applied Data Science Capstone

Opening a Mexican restaurant in Berlin, Germany



By: David Briones

Introduction

Everyone loves restaurants. They are great for social interaction and obviously, they are great for getting a kind of food that you cannot cook at home. Nonetheless, it turns out that restaurants are also good for urban development. They can revive an abandoned neighborhood and turn it into a hot location. They can attract young professionals to rundown areas and turn them into beautiful neighborhoods. They can bring other parts of the world closer to you. Restaurants are not merely a place to eat, but a place to congregate, interact, and enjoy life.

The Problem

Opening a restaurant is a challenging venture for many reasons. It requires a hefty capital investment, a great menu, a reliable supply chain, a food safety system, and a great location. However, in this report we will focus on one of these challenges: location. Location is incredibly important because a great restaurant placed in the incorrect area cannot possibly succeed. The problem is that choosing the right location appears to be almost impossible for people. How can you consider every aspect? This is where Data Science can come in. An entrepreneur can use machine learning algorithms to aid in the search for the right location. In this project we will consider the fictional Mexican restaurant Avo-Cado and their search for a great location in Berlin. By using a K-Means algorithm and Python programming we will help Avo-Cado's team to choose the best location.

Data

In order to solve this problem, we obviously need data, and this is what we will need:

- A list of neighborhoods or localities in Berlin. This will allow us to limit the scope of the project to restaurant in Berlin.
- Latitude and longitude coordinates for these localities. This will allow us to properly create clusters and then map them in order to display the data visually.
- Restaurant data regarding the type of restaurant. This will allow us to adequately create clusters of localities and determine which ones have a lot of restaurants but lack Mexican food.

Sources and Methods of Extraction

A list of localities in Berlin can be found in Wikipedia at https://en.wikipedia.org/wiki/Category:Localities_of_Berlin. Using the BeautifulSoup Python package we will scrape this website and extract the necessary data. Then we will use the Geocoder Python package to find the geographical coordinates for these localities.

Once we have the necessary data, we will use the Foursquare API to get locational data regarding the restaurants in these localities. Once we have done this we will use a K-Means Clustering algorithm to determine which one these neighborhoods are a prime location for opening a Mexican restaurant.