IBM Applied Data Science Capstone

Opening a New Restaurant in Minneapolis, Minnesota, US

By: Shaoqing Liu

Dec 2019

Data

To solve the problem, we need the following data:

• List of suburbs in Minneapolis. This defines the scope of this project which is confined to the city of Minneapolis, Minnesota

• Latitude and longitude coordinates of suburbs. 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 suburbs.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Hennepin_County,_Minnesota) contains a list of suburbs in Minneapolis, with a total of 45 suburbs. We will use web scraping 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 suburbs using Python Geocoder package which will give us the coordinates of the suburbs.

After that, we will use Foursquare API to get the venue data for those suburbs. Foursquare has one of the largest database of 100 million places and is used by over 100 thousands developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the 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, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, 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.