IBM Data Science Capstone Project

Sydney Johnson

*Determining a Location for a New Restaurant in Troy, AL*

Introduction

The city of Troy, AL is a small college town located in South-Eastern Alabama. Troy is divided by a major highway that leads directly to the beach. Because of this, Troy has several fast-food options for those passing through, but not many sit-down restaurant options for the residents and university students of Troy.

Business Problem

Levi Smith, a young entrepreneur, is looking to open a restraint that can fill this void in the community. He has a concept for a menu and bar that would attract the locals and college students alike. His main concern now is location. He would like his restaurant to stand apart from all of the fast food and chain restaurants in Troy. He also need the location to be somewhere that college students and local families can easily access. Mr. Smith believes that the best location would be somewhere outside of the main hub of food options, but not somewhere where it is the only restaurant around.

Data

For this project, I will be primarily using Four Square API. Four Square API is a database that provides tons of information about venue location. It will allow me to find all restaurant’s in the area and their locations. I will use only the information of restaurants located within city limits as Mr. Smith wants his restaurant to be accessible by the public and college students alike. The data provided by Four Square API will allow me to plot the area’s existing restaurants in order to determine a location for Mr. Smith’s restaurant that meets all of his criteria.

Methodology

I will begin my project by gathering the Zip Code information from Troy, Al. After I have researched and gathered this information, I will now need to get coordinates associated with these three zip codes. I will now use pgeocode to retrieve latitude and longitude points to combine into a dataframe. Now that my postal code information is loaded into a useable dataframe, I will now use folium to visualize a map of the city and plot the postal codes on the map.

Now that I have the city broken down into three regions, I will connect to the foursquare API to gather the relevant venue information. Since there are multiple categories for food and restaurants, I will pull all of the categories related to food. Next, I will merge my Zip Code data frame with my Restaurant data frame. This step gets me closer to seeing which postal code has the greatest number of restaurants. Next, I will use K-Means clustering to cluster the restaurants into the three zip code zones based on their longitude and latitude points. I will now use folium to create a map once again to further visualize this data.

Results

After clustering and visualizing my data, I can explore the clusters to see where most of the restaurants are located. The most restaurants are located in the cluster for the Zip Code 36081. The second greatest amount of restaurants are located in the Zip Code cluster of 36082. The zip code with the least amount of restaurants is 36079. With this knowledge, the new restaurant owner will be better equipped to make a decision on where his new restaurant might thrive best.