Determining the best city to open a Pizza Delivery service

November 3, 2020

1. Introduction

1.1 Business Problem

A small business owner wants to open up a pizza delivery service to cater only to the surrounding pizza restaurants. It will speed up the time it takes to get pizza from restaurant to customer by only delivery from pizza restaurants. The business owner wants to make sure there are enough surrounding restaurants in the city that they open in, to ensure successful business. They want to open in a busy multicultural city in Pennsylvania.

1.2 Interest

The interest in this data will be this entrepreneur who wants to open in one of these two cities. Or anyone else wanting to open a pizza delivery service within Pennsylvania.

2. Data acquisition and cleaning

2.1 Data sources

All the necessary data for this project will be sourced from using the *Foursquare API*. By querying 'Pizza' restaurants within one city and building a DataFrame – and then querying the same but in the second city, we will be able to form our basis for evaluation.

2.2 Data cleaning

The data is a little messy when first querying from the *Foursquare API*. First, we'll want to view the raw json and we can better understand the data we are working with. Upon reviewing the json, we determine that we'll want to add the venues into a DataFrame and leave out the other headers.

venues = results['response']['venues']

dataframe = pd.json_normalize(venues)

The above code will be a great starting point in cleaning our data. Next, we will make sure to filter only the columns we want to view – everything dealing with "location", as well as the "name", "categories", and "id". Once we filter the columns, we can set this as a variable called "phil_df" -- short for Philadelphia DataFrame.

Now we can do the exact same process, but first querying Pittsburgh with the *Foursquare API*. Once we have cleaned this DataFrame, we can set the variable to be named "pitt_df" -- short for Pittsburgh DataFrame.

3. Exploratory Data Analysis

3.1 Plotting venues on map

Once we have both cities with their own separate DataFrame, we can plot these points on a map to better visualize where these venues exist. Using the folium library, and adding CircleMarkers to the map, we can see all our queried venues on the generated map. But still, we want to explore our data further.

3.2 Using k-Means cluster

Let's use k-Means cluster to separate our venues into groups. First, we will need to determine what would be the best k to use in this situation. Building a *for* loop and then generating the results onto a graph, we can find out what would be the best k to use.