Opening a New Pizza Place in Washington, D.C

IBM Applied Data Science Capstone

Uday Kiran Bonthu

Table of contents

Introduction

Washington D.C. is well-known for the diverse restaurants. There are restaurants of various cuisines like American, Indian, Thai etc. Additionally, there are many fast food joints there are becoming increasingly popular in this area. My friend, Alex wants to open a Pizza Place in one of the neighborhoods in Washington, D.C. Opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the place is one of the most important decisions that will determine whether the pizza place will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best neighborhood in the city of Washington, D.C to open a new pizza place. Using data science methodology and data collected from numerous sources, this project aims to provide solutions to answer the business question: In the city of Washington, D.C if someone is looking to open a new pizza, which would be the best neighborhood to open it?

Target Audience of this project

This project is particularly useful to investors looking to open or invest in new pizza places in Washington, DC. Other investors who are looking to open any restaurant in any place can use this project as a basis of their analysis.

Data Description

To solve the problem, we will need the following data:

- List of neighborhoods in Washington, D.C. This defines the scope of this project which is confined to the city of Washington D.C.

- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.

- Venue data, particularly data related to pizza places, nearby universities, high schools and offices.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Neighborhoods_in_Washington,_D.C.) contains a list of neighborhoods in Washington, D.C, with a total of 131 neighborhoods. We will use web scraping techniques 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 neighborhoods using Open Street Map API which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Pizza places, universities, high schools and offices categories 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 (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, and map visualization (Folium).

Methodology

Firstly, we need to get the list of neighborhoods in the city of Washington, D.C. Fortunately, the list is available in Wikipedia. We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighborhoods data. Following is the sample of the list of neighborhoods.

|	Neighbourhoods
0	Adams Morgan
1	Columbia Heights
2	Howard University
3	Kalorama
4	LeDroit Park

Next, we need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Nomintim package in the Open Street Map API that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas dataframe and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Washington, D.C.

Coursera Capstone Project

Following is a sample rows of the resulting dataframe

	Neighbourhoods	Latitude	Longitude
0	Adams Morgan	38.921500	-77.042199
1	Columbia Heights	38.925672	-77.029429
2	Howard University	38.921525	-77.019536
3	Kalorama	38.916778	-77.052477
4	LeDroit Park	38.915907	-77.015721

Below screenshot shows all the neighborhoods plotted on the map using Folium
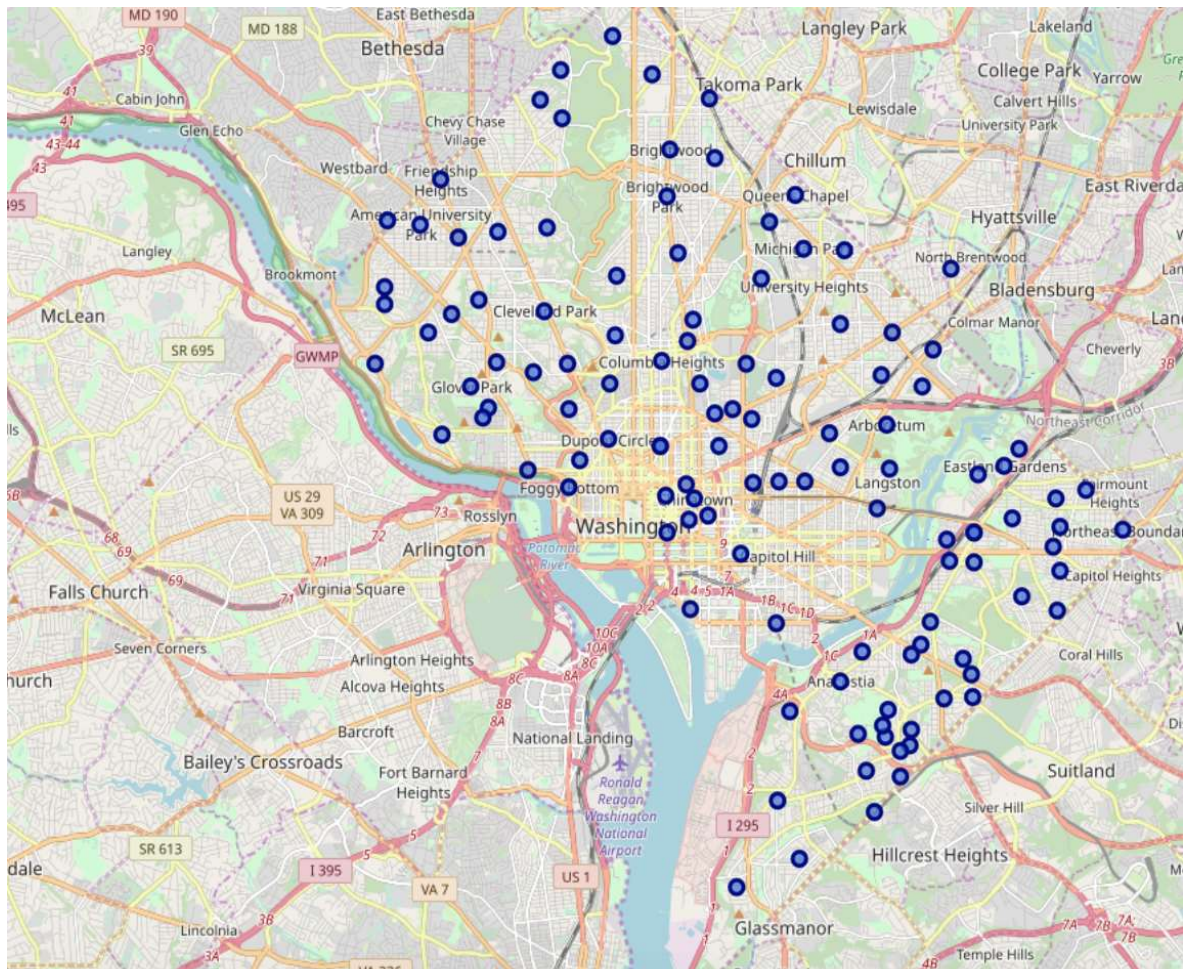
Next, we will use Foursquare API to get the top 250 venues that are within a radius of 500 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues.

FourSquare Venue locations of nearby Pizza places

	Neighbourhoods	Latitude	Longitude	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category
0	Adams Morgan	38.9215	-77.042199	Mellow Mushroom	38.921605	-77.042495	Pizza Place
1	Adams Morgan	38.9215	-77.042199	Jumbo Slice Pizza	38.920594	-77.041958	Pizza Place
2	Adams Morgan	38.9215	-77.042199	Pizza Mart	38.921607	-77.042400	Pizza Place
3	Adams Morgan	38.9215	-77.042199	&pizza	38.922140	-77.042170	Pizza Place
4	Adams Morgan	38.9215	-77.042199	Pizza Boli's	38.922018	-77.042394	Pizza Place

**Plot of the pizza venues on the map using Folium**

Coursera Capstone Project
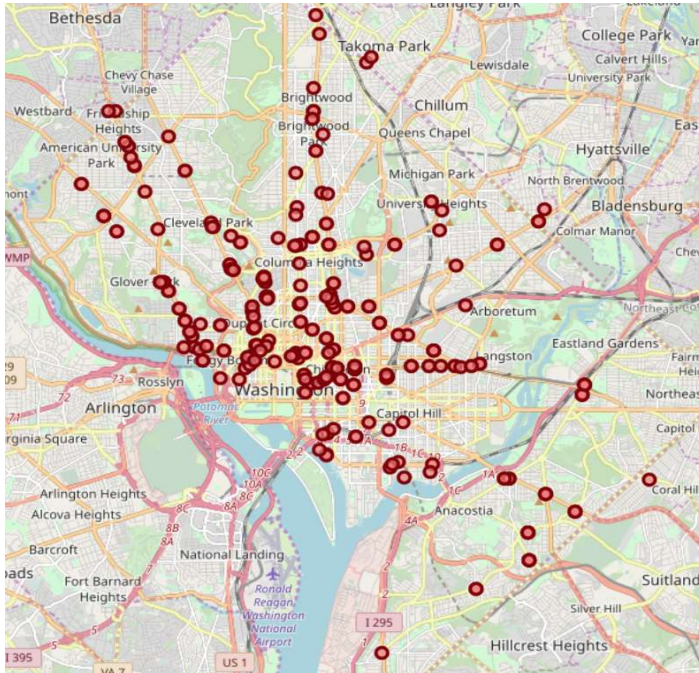
FourSquare Venue locations of nearby high schools

	Neighbourhoods	Latitude	Longitude	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category
0	Columbia Heights	38.925672	-77.029429	Carlos Rosario Int'l Public Charter School	38.926513	-77.027777	High School
1	Columbia Heights	38.925672	-77.029429	Cardozo High School	38.921728	-77.028846	High School
2	Columbia Heights	38.925672	-77.029429	Benjamin Banneker Academic High School	38.923689	-77.023762	High School
3	Columbia Heights	38.925672	-77.029429	Bell Multicultural High School	38.929328	-77.036166	High School
4	Columbia Heights	38.925672	-77.029429	Youth Build - LAYC	38.927639	-77.032693	High School

**Plot the high schools venues on the map using Folium**

Coursera Capstone Project
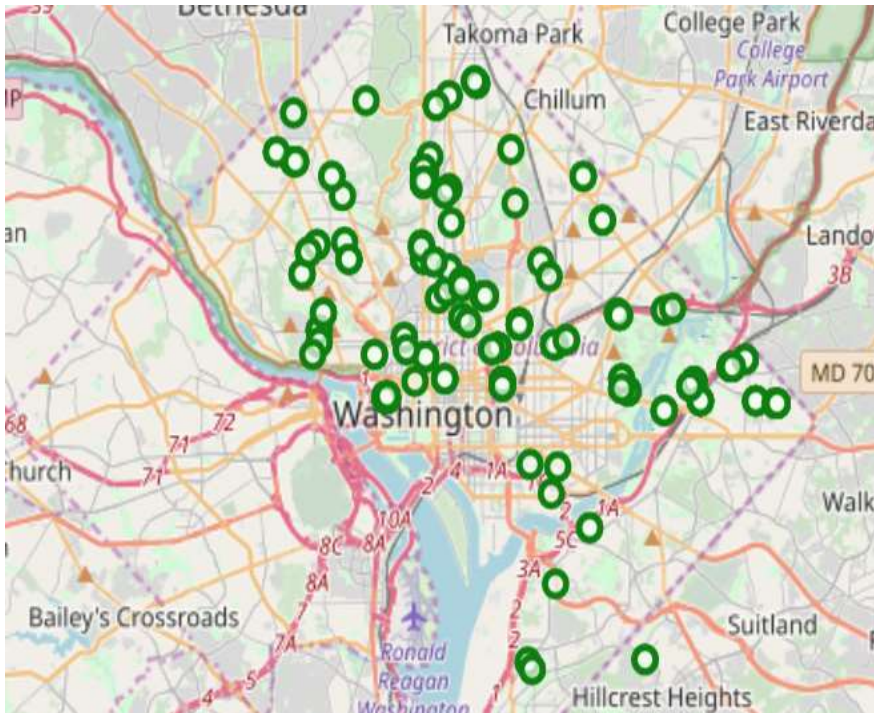
FourSquare Venue locations of nearby universities

	Neighbourhoods	Latitude	Longitude	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category
0	Adams Morgan	38.921500	-77.042199	Arizona State University Washington Center	38.915488	-77.046366	College Administrative Building
1	Howard University	38.921525	-77.019536	Howard University	38.921143	-77.020954	University
2	Howard University	38.921525	-77.019536	Howard University School Of Education	38.923174	-77.018431	University
3	Howard University	38.921525	-77.019536	Howard University Department Of Psychology	38.920852	-77.020021	University
4	Kalorama	38.916778	-77.052477	Arizona State University Washington Center	38.915488	-77.046366	College Administrative Building

**Plot the universities on the map using Folium**

Coursera Capstone Project
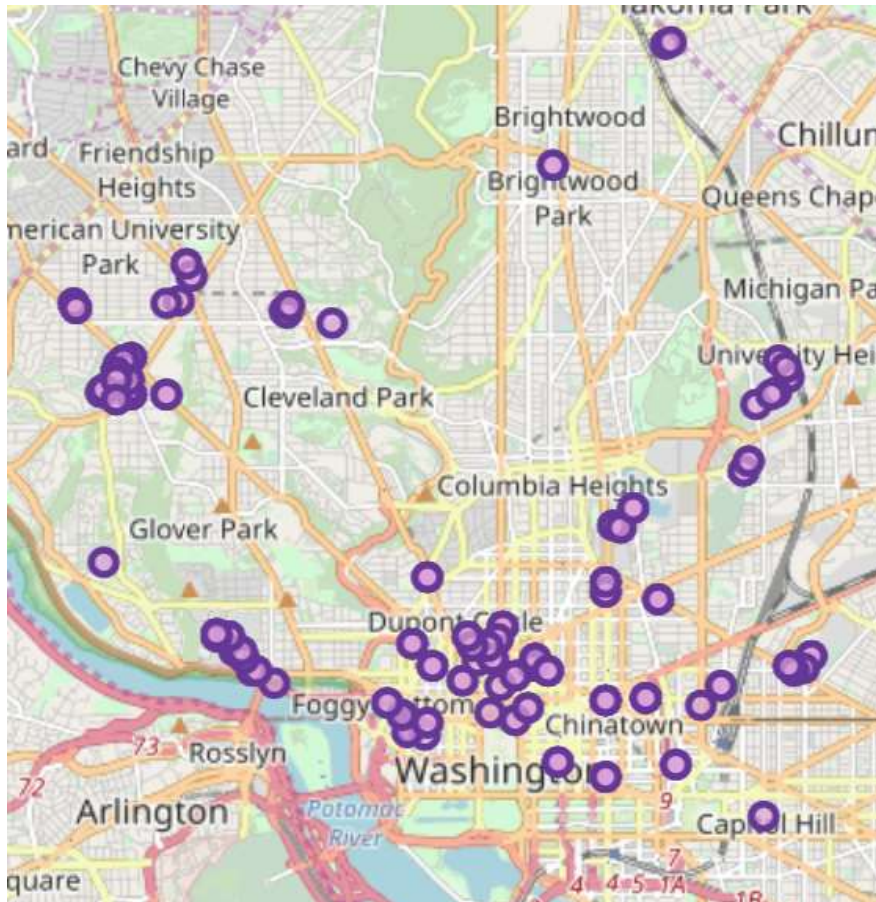
FourSquare Venue locations of nearby offices

	Neighbourhoods	Latitude	Longitude	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category
0	Adams Morgan	38.9215	-77.042199	Housing Counseling Services	38.922014	-77.039566	Office
1	Adams Morgan	38.9215	-77.042199	Washington Area Bicyclist Association	38.924284	-77.040286	Office
2	Adams Morgan	38.9215	-77.042199	Maga Design Inc.	38.921628	-77.043156	Print Shop
3	Adams Morgan	38.9215	-77.042199	Counter Culture Coffee DC Training Center	38.917591	-77.040799	Corporate Coffee Shop
4	Adams Morgan	38.9215	-77.042199	Environmental Working Group	38.916593	-77.033911	Office

Plot the offices on the map using Folium

Coursera Capstone Project
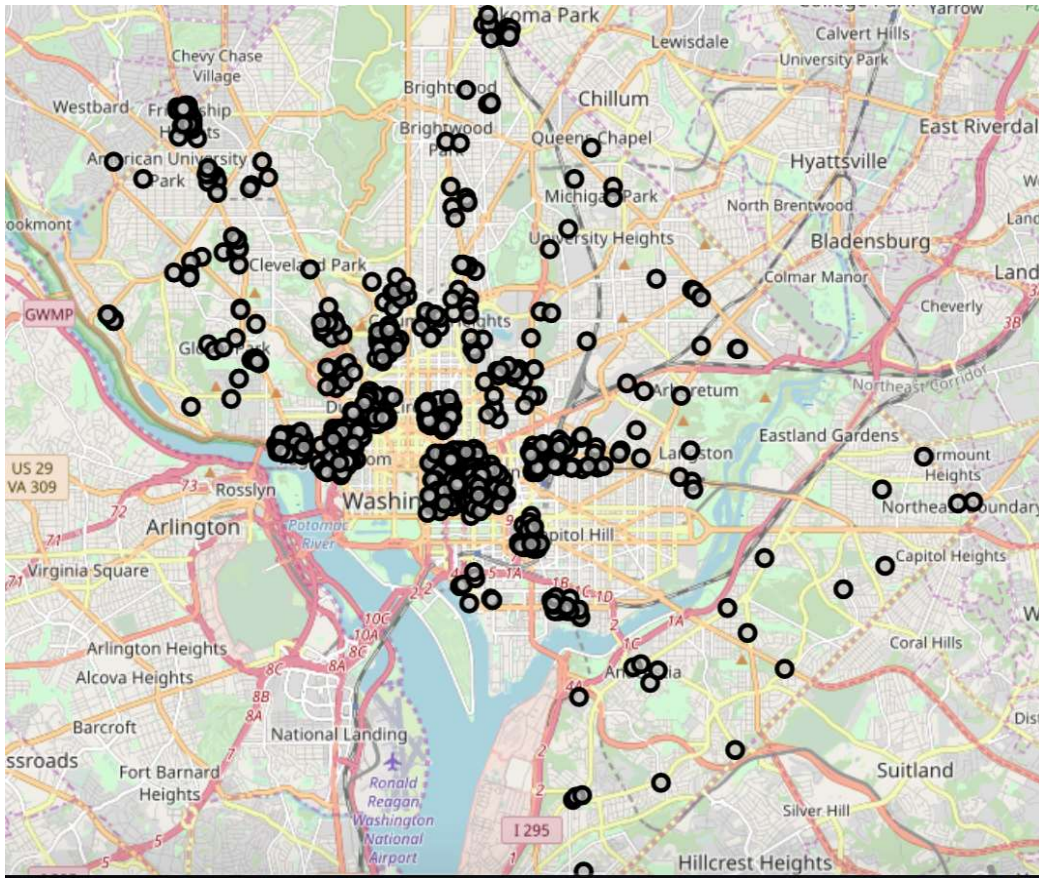
Then we merged all the venue dataframes into a single dataframe.

|	Neighbourhoods	Latitude	Longitude	Pizza	High_Schools	Universities	Offices
0	Adams Morgan	38.921500	-77.042199	11.0	0.0	1.0	50.0
1	Columbia Heights	38.925672	-77.029429	6.0	9.0	0.0	47.0
2	Howard University	38.921525	-77.019536	5.0	6.0	3.0	38.0
3	Kalorama	38.916778	-77.052477	3.0	0.0	1.0	50.0
4	LeDroit Park	38.915907	-77.015721	8.0	7.0	6.0	46.0

Lastly, we will assign weights to each of the categories and calculate the weighted sum for each neighborhood to calculate the ideal neighborhood to open a pizza place.

11

Results

The results from the weighted sum shows that Foggy Bottom is the best place to open a Pizza place.

	Neighbourhoods	Score
15	Foggy Bottom	117.5
8	Shaw	110.0
17	Logan Circle	110.0
13	Dupont Circle	107.0
24	Cathedral Heights	106.5
84	NoMa	103.0
21	West End	103.0
83	Near Northeast	102.5
63	Eckington	102.5
78	Capitol Hill	100.0
4	LeDroit Park	100.0
11	Chinatown	98.5
16	Georgetown	98.5
3	Kalorama	98.5
12	Downtown	98.0
79	Judiciary Square	98.0
10	Burleith	97.5

Plotting all the venues near Foggy Bottom

.
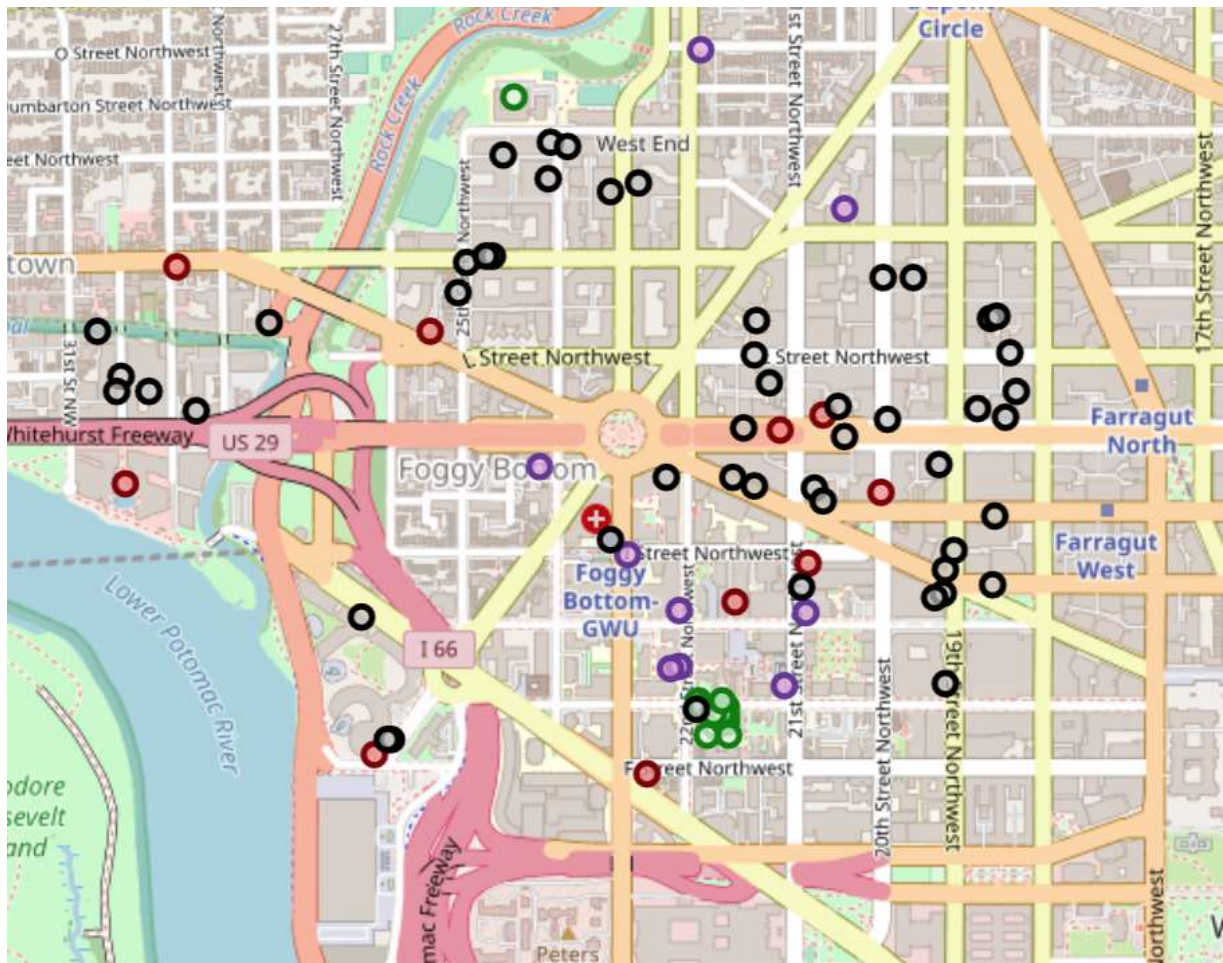
Discussion

As observations noted from the map in the Results section, most of the offices, universities and high schools are concentrated in the Foggy Bottom neighborhood with less pizza places in that neighborhood. This represents a great opportunity and high potential area to open new pizza place as there is very little competition from existing pizza places.

Limitations and Suggestions for Future Research

In this project, we only considered nearby offices, high schools, universities and pizza places.There are other factors such as population, other fast food restaurants that could influence the location decision of a new pizza place. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new pizza place. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results such as tips and comments from customers about the restaurants in the neighboorhoods.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, and lastly providing recommendations to Alex and other relevant investors regarding the best locations to open a new pizza place. To answer the business question that was raised in the introduction section, the answer proposed by this project is: Foggy Bottom the most preferred location to open a new pizza place. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations in their decisions to open a new pizza place.

Coursera Capstone Project

References

Neighborhoods in Washington D.C retrieved from the Wikipedia

page:https://en.wikipedia.org/wiki/Neighborhoods_in_Washington,_D.C.