

# Battle of the Neighborhoods Capstone Project

## 1. Introduction

### 1.1 Problem

To find the best neighborhood in Prince William County for a general location of a fast food restaurant.

### 1.2 Background

Prince William County is located in Virginia in an area known as Northern Virginia. Prince William County has had numerous developments appearing in many different neighborhoods. In some neighborhoods there are many fast food restaurants that are clumped close together.

### 1.3 Interested

The ones that will be interested in this project is any one that wants to have a fast food restaurant franchise and Prince William Board of Supervisors.

## 2. Data

### 2.1 Data Sources

I will be creating a dataframe for the neighborhoods of Prince William County and this dataframe will contain the neighborhood names, and the general latitude and longitude for each neighborhood. I will be making calls to FourSquare API to locate fast food restaurants in each neighborhood. From the calls made to FourSquare API, I will create tables that list each fast food in each neighborhood. From the tables I will make a decision on the best neighborhood for a fast food restaurant.

## 3. Methodology

### 3.1 Creating Dataframe

#### 3.1.1 Prince William County Dataframe.

I started by creating a Jupyter notebook which I imported libraries that I will need for the project. I created a dataframe of the neighborhoods in Prince William County by using a website that I found that had the names of the neighborhoods as well as the latitudes and longitudes for each neighborhood. Then I needed a dataframe of the fast food restaurants in the Prince William County area which I got my credentials for FourSquare API and the latitude and longitude of Prince William County.

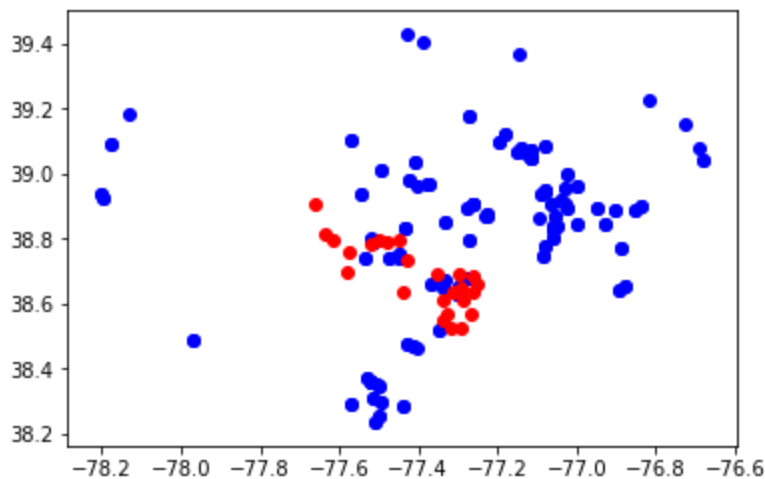
#### 3.1.2 Fast Food Dataframe

After creating the data frame of Neighborhoods in Prince William County, I looked up the general latitude and longitude of Prince William county because I will need it to make calls to FourSquare API. Then I got my FourSquare API credentials that I would need to help with making calls. I selected a 50 mile radius for my search and I had to convert it to meters because that is what was required as a parameter. Lastly I limited myself to the top 150 fast food restaurants in the radius

I created two urls and requested to get the JSON files. The reason is to divide Prince William County into two parts, west and east. I used the most known cities in Prince William County as my center points for the search: Manassas for the west and Woodbridge for the east. I got the latitude and longitude for each city and passed it into the urls. After I got the JSON files, I converted them into data frames and then combine them into one dataframe.

### 3.2 First Visual Inspection of the data

I decided to see how the data looks in a scatter plot. I took the longitude and latitude of the fast food restaurants as well as the longitude and latitude of the neighborhoods. I plotted this on a scatter plot using the latitude as the y-axis and the longitude as the x-axis. Below is an image of the scatter plot where the red dots represent the neighborhoods and the blue dots represent the fast food restaurants.



From this I could tell that a lot of the blue dots were away from the red dots. I tried to see if I could find a JSON map file of Prince William County to help better understand the data, but I couldn't find one. Also the red dots represent the centers of each neighborhood in Prince William County. So the next thing is to start clustering the Fast Food Restaurants into a neighborhood.

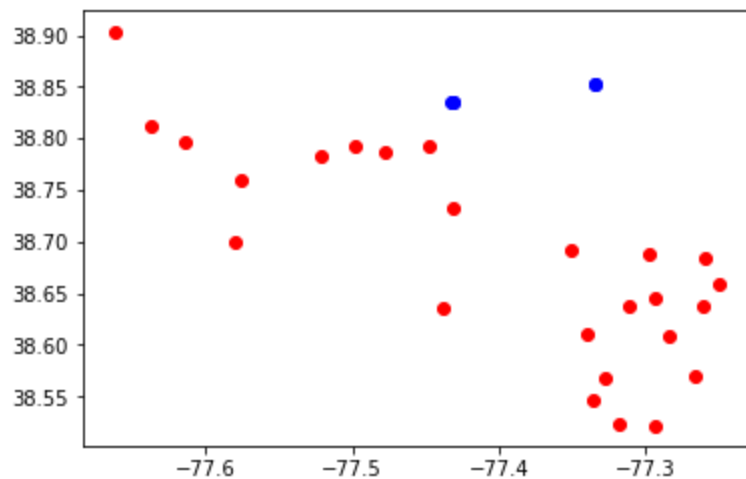
### 3.3 Clustering

I used the KMeans clustering algorithm to help me cluster the fast food restaurants into 24 different clusters. I had two purposes for doing this. The first reason was to set up the data into clusters to begin to clean it up better. The second reason

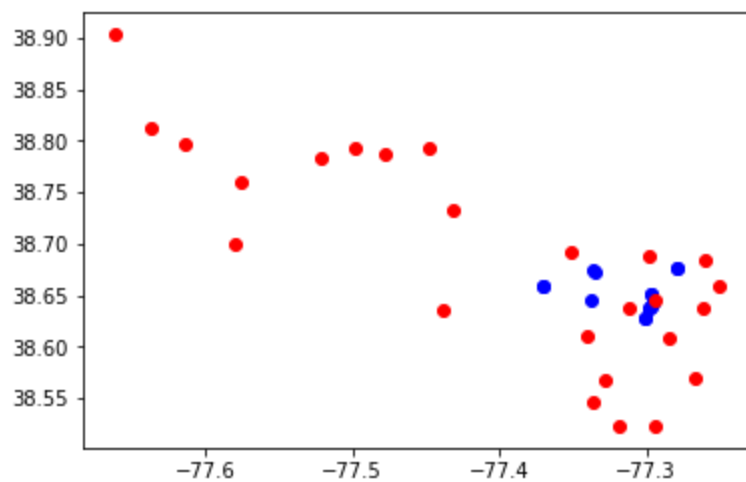
was to see where the KMeans clustering algorithm created the cluster centers for the data.

The next thing that happened was that I added the cluster labels to the data frame, and then split the data frame into 24 smaller data frames. After I have created the 24 smaller data frames, I then plot each of those dataframes on a scatter plot against the latitudes and longitude of the known neighborhoods of Prince William county.

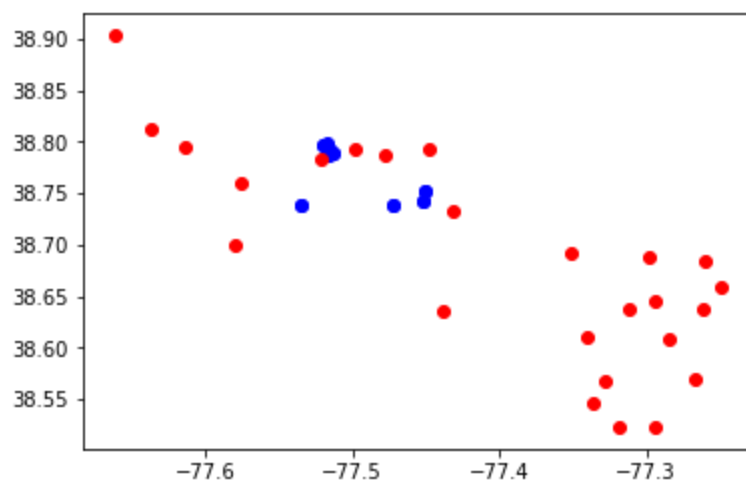
I analyzed each scatter plot and found out which clusters had data that I needed to use to make my conclusion. Below are the four clusters graphed on a scatter plot that had the data that I needed.



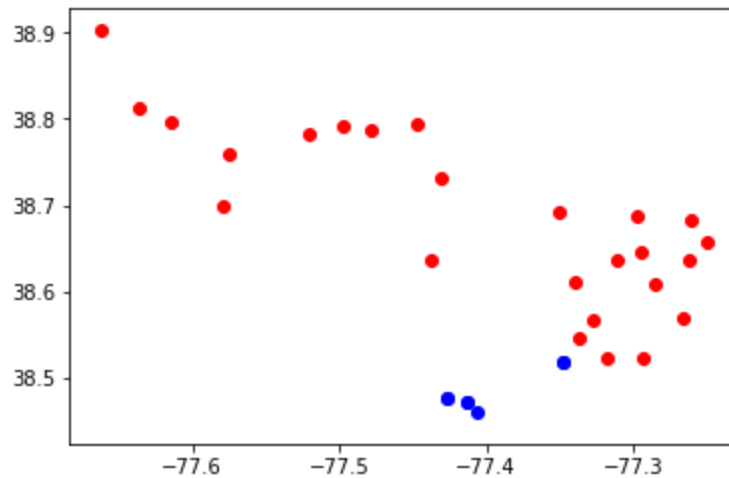
Cluster 4



Cluster 6



Cluster 12



As you can see that each of these clusters has blue dots representing the fast food restaurants are either mixed with the red dots representing the center of each neighborhood or very close to the red dots. The next thing I did was combine the four data frames of each of the clusters into one new dataframe.

### 3.4 Classification

Next was to clean up the data frame of any duplicated data and then reindex the data frame. After that I decided to examine each of the fast food restaurants and figure out which neighborhood it goes to. This took some time and I was able to figure out by looking at the latitude and longitude of each fast food restaurant and then compare them to see which neighborhoods were the closest before graphing them.

## 4.0 Analysis

After I counted which neighborhoods had fast food restaurants. Then I created a dataframe showing the neighborhoods and how many fast food restaurants each

neighborhood had. Afterwards I created a bar graph showing the results of the classification as shown below.



From the bar graph I was able to determine which neighborhoods are to be selected to answer my problem.

#### 4.1 Predictions

Using my own knowledge of Prince William County and the data that I analyzed, I have come to the conclusion of the top 5 neighborhoods. The top five neighborhoods are Haymarket, Gainesville, Nokesville, and Montclair.

#### 5.0 Conclusion

In conclusion, I was able to answer the problem that this project was based on.