Location Selection for Coffeehouse

# Problem Description and Audience:

Two women are interested in opening a coffeehouse in the Washington, D.C. area. They want to know which location(s) would be the best options for where to lease a storefront. This report will be formatted as a presentation to them about potential storefront locations.

# Data Sets and Analysis Plan

I plan to access Washington D.C. venue data for coffeehouses via the Foursquare API. Specifically, I will create a data set that includes neighborhoods and traffic flow in the DC area which I will retrieve from DC’s [open data repository](https://dc.gov/page/open-data). I will then call the Foursquare API to retrieve the locations of venues similar to the ones the clients want to build in the neighborhoods the clients are interested in. Next, I will append the details of those venues to the data set.

Once the dataset is complete, I will then use kmeans clustering to look at the venues in each neighborhood by the amount of users checking in at those areas.

This will result in being able to make a recommendation for which location the clients should choose for their new coffeehouse.

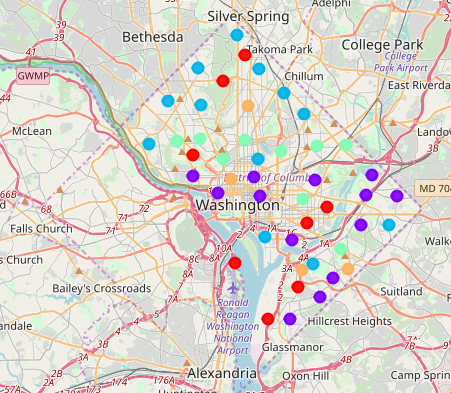
# Methodology

In order to get a final data set that includes the location of neighborhoods, coffee shops within the neighborhoods, and traffic patterns in the neighborhood. The analysis process is as follows:

* Read in a shape file of DC neighborhood clusters
* Access the centroid of each neighborhood cluster and merge it to the original shape file
* Clean this merged data set
* Make a GET request to the Foursquare API for the locations of coffee shops near the centroids of each neighborhood cluster
* Read in a shape file for DC traffic patterns
* Merge the neighborhood, Foursquare, and traffic sets together and clean the result
* Apply k-means clustering to look at similarities and differences between clusters

# Results

The K-means clustering method split the neighborhoods into five clusters based on how many coffee shops were already in the area and the amount of traffic in each area. The cluster with both the most traffic and the most coffee shops was cluster 4 (orange).



# Discussion

Because the traffic in the southeast corner of the city is most likely due to interstate travel between DC, Maryland, and Virginia, the client’s best options are to choose a storefront location near Dupont Circle or in the Petworth area of DC. These two areas have the most traffic and a high number of coffee shops. Foursquare provides access to a count of user check-ins at each location, and this data could greatly enhance the analysis by showing the client what types of coffee shops are popular in certain areas. It could additionally help with determining which locations within this neighborhood would be best for a shop.

# Conclusion

K-means clustering provided a list of neighborhoods in which it would be most beneficial for my client to open a new coffee shop. By accessing data about venues in the neighborhood clusters and traffic patterns, we were able to narrow down the best neighborhood options for the shop. Additional data could be used to enhance the clustering model and further limit the radius of the best location for the client, if necessary.