
RECOMMENDATION SYSTEM TO START A RESTAURANT BUSINESS IN NORTHERN VIRGINIA

FEBRUARY 10

IBM Applied Data Science Capstone Project
Authored by: Reeva Bhatkal



Contents

I.	INTRODUCTION	3
	Background Discussion	3
	Problem Description	4
II.	DATA	5
III.	METHODOLOGY.....	6
	Business Understanding.....	6
	Exploratory Data Analysis	6
	Problem Approach and Using K-Means Clustering	7
IV.	RESULT.....	10
	Cluster 0 (Most Common: American Restaurant & Seafood Restaurant)	12
	Cluster 1 (Most Common: Depends on neighborhood).....	13
	Cluster 2 (Most Common: Greek & Chinese Restaurant)	15
	Cluster 3 (Most Common: Fast Food & American Restaurant).....	16
	Cluster 4 (Most Common: Thai Restaurant)	18
	Cluster 5 (Most Common: Vietnamese Restaurant).....	19
	Cluster 6 (Most Common: Chinese & Vietnamese Restaurant)	19
V.	DISCUSSION	21
VI.	CONCLUSION	22

I. INTRODUCTION

Background Discussion

“Northern Virginia” (NoVA) is an area within the Commonwealth of Virginia and is part of the Washington DC metropolitan area. It is located in the northeastern part of the Commonwealth of Virginia, along the borders of Maryland and Washington, D.C. It provides recreational opportunities for a wide range of interests and lifestyles. From sports teams in all major leagues to historic landmarks, national and state parks, rivers, ample retail options, and cultural offerings, the region boasts a quality of life that attracts residents and visitors.

Northern Virginia is considered one of the country’s most international regions and has the highest number of ethnic groups and nationalities represented in Virginia. Close proximity to the nation’s capital provides all the advantages of accessibility to key federal agencies, with a competitive edge of highly skilled talent at cost-competitive wages. Northern Virginia’s unique blend creates an ideal environment for all types, from recent graduates eager to live in a diverse urban environment to families looking for comfortable housing solutions in the nearby suburbs. It is one of the fastest growing, diverse communities in the United States.

In recent years, restaurants are slowly becoming a driving force in Virginia’s economy. They provide jobs and build careers for thousands of people and play a vital role in local communities throughout the state. There are over 15,000 eating and drinking establishments in Virginia, with annual sales over \$18 billion. In the state, there are over 378,000 restaurant and food service jobs – 9% of total employment in Virginia. Nationwide, restaurant industry reported sales of \$863 billion in 2019 (up from \$590 billion in 2010). The industry is on the move, having added 3 million jobs since 2009 and expected to add 1.6 million new jobs by 2029.

Considering Northern Virginia’s diversity and ethnicity it is evident that starting a restaurant business would earn you more money comparatively than most of the other businesses. Although, with more profitable business there comes the most competition. This article can serve as one of the guides to start a restaurant business based on a

particular cuisine like Indian, Italian, American, etc., by providing a specific location. The number of restaurants in a specific location categorized based on cuisine and population distribution based on ethnicity and culture are some of the features considered for analysis.

Problem Description

Opening a restaurant is an exciting and challenging business proposition and choosing the right location can help drive success. Northern Virginia has achieved exceptionally strong local economy and high quality of life through its focus on economic development efforts of the community. These efforts have resulted in the counties being the largest and strongest economy in the Washington Metropolitan Area.

Restaurants are not only a valuable piece of our economy but was identified as a top initiative where there were opportunities to optimize the usage of cross agency resource and service to improve the experience for the restaurateur. In a dynamic market like Northern Virginia, prospective business owners should expect excellence in operations and customer services.

It is evident that to survive in such competitive market it is very important to strategically plan and various factors need to be studied to decide on location such as:

- Population and demographics
- Type of Restaurant (Dine-in, Carryout, Drive-through)
- Who are the competitors in that location?
- Your Target Market
- Restaurant Concept (Franchise, new-concept restaurant)
- Location

And the list goes on...

Success Criteria & Target Audience:

The objective is to locate and recommend stakeholders, which neighborhood of Northern Virginia will be the best choice to start a restaurant business based on cuisine. The success criteria of the project will be a good recommendation of neighborhood

choice based on lack of such restaurants considering cuisine as a factor in that location and a suitable target audience who has the purchasing power.

II. DATA

To build a recommendation model, following datasets and information are considered for analysis:

- There are 95 counties and 39 independent cities in Virginia. To segment the districts and explore them, we will essentially need a dataset that contains the 95 counties and the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is:
https://data.opendatasoft.com/explore/embed/dataset/geonames-postal-code@public-us/table/?sort=place_name&q=US&refine.admin_name1=Virginia
- Data provided by the United States Census Bureau from [https://www.virginia-demographics.com/counties by population](https://www.virginia-demographics.com/counties_by_population) is utilized to get more insights about Northern Virginia city and counties. The data provided knowledge about the population density, immigrants' country of origin and many more.
- We will be using Foursquare API to get information about available restaurants for a given city and county in Northern Virginia. The API also provided information about restaurant styles based on cuisine.

III. METHODOLOGY

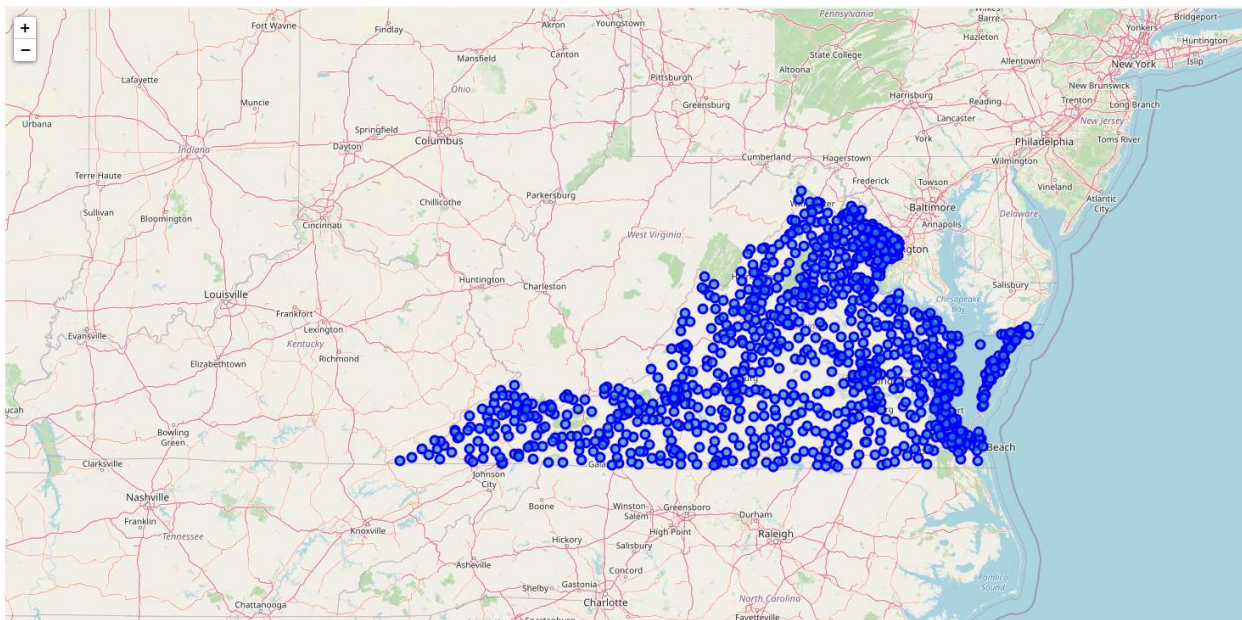
Business Understanding

Our main goal is to get optimum location for new restaurant business in Northern Virginia based on cuisine.

Exploratory Data Analysis

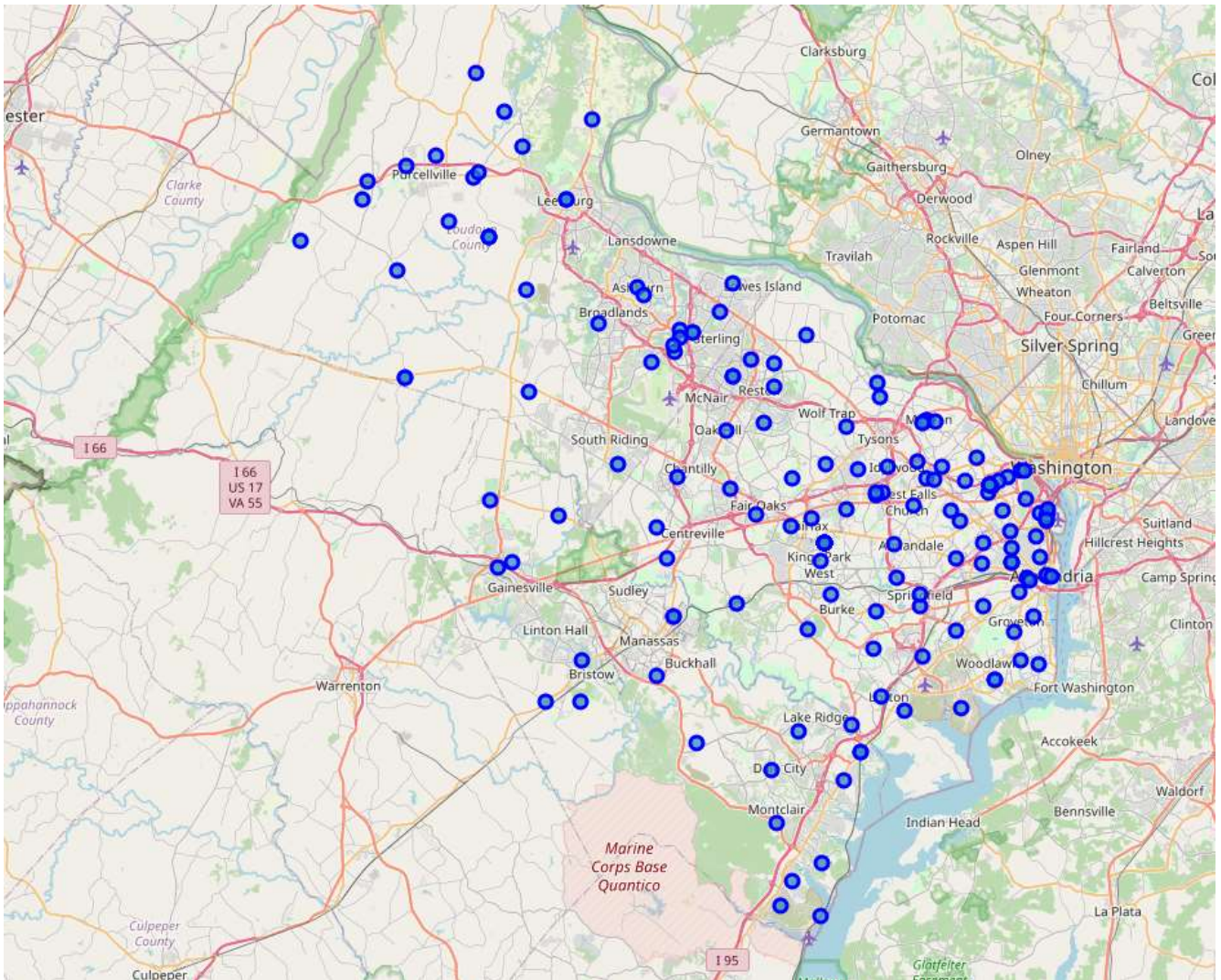
The Commonwealth of Virginia is divided into 95 counties, along with 38 independent cities that are considered county-equivalents. For the purpose of this project we will focus mainly on Northern Virginia. Northern Virginia is mainly divided into 4 counties including 5 independent cities. Geopy and folium libraries are used to create a map to visualize neighborhoods of Virginia.

Map of Virginia with all counties and independent cities:



All the blue markers on the map above are for neighborhoods. It is evident that the city is more congested at the north and widespread in the south.

Map of counties and cities in Northern Virginia:



Problem Approach and Using K-Means Clustering

The process starts with scrapping the web page containing information about Virginia city's present in each county using pandas library. During the process Pandas data frame is constructed with rows having city and corresponding county information. The neighborhoods in the resulting data frame must be plotted on a map using folium library. To do so, latitude and longitude value for each neighborhood is determined using geopy library. After getting latitude and longitude values, it is merged with the original data frame containing neighborhood and borough information appropriately. Finally, each neighborhood is plotted on folium map.

	Postal Code	City	County	latitude	longitude
0	23358	Hacksneck	Accomack	37.6393	-75.8650
1	24522	Appomattox	Appomattox	37.3530	-78.8224
2	22205	Arlington	Arlington	38.8836	-77.1395
3	22214	Arlington	Arlington	38.8688	-77.0739
4	22952	Lyndhurst	Augusta	37.9744	-78.9361

Next, Foursquare API is used to explore the neighborhoods and segment them. First, data frame with all the venues information provided by foursquare for the given latitude and longitude values. Foursquare API returns 2919 results with 276 unique categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arlington	38.8836	-77.1395	The Italian Store	38.885179	-77.140250	Italian Restaurant
1	Arlington	38.8836	-77.1395	Toby's Homemade Ice Cream & Coffee	38.885283	-77.140743	Ice Cream Shop
2	Arlington	38.8836	-77.1395	Westover Market Beer Garden	38.885637	-77.141413	Beer Garden
3	Arlington	38.8836	-77.1395	Lost Dog Cafe	38.885656	-77.141698	Sandwich Place
4	Arlington	38.8836	-77.1395	Stray Cat Café	38.885387	-77.141564	American Restaurant

Now, only restaurants are extracted from venue category list. The resulting data frame has 276 unique categories available in Northern Virginia.

Next, one hot encoding is performed on the resulting data frame for each neighborhood. The results have 703 unique restaurants in Northern Virginia with 50 different style of cuisines.

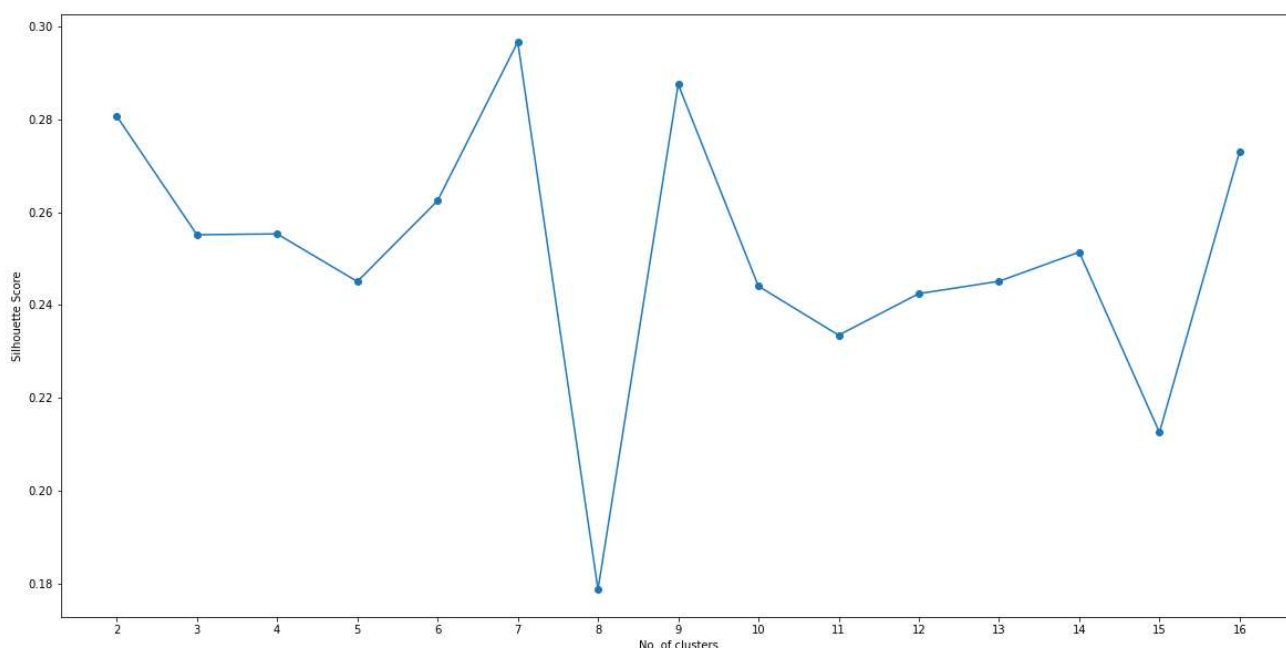
Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Asian Restaurant	Belgian Restaurant	Cajun / Creole Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	...	South Indian Restaurant	Southern / Soul Food Restaurant	Spanish Restaurant	Sushi Restaurant	Tapas Restaurant	Tex-Mex Restaurant	Thai Restaurant	Turkish Restaurant	Vegetarian / Vegan Restaurant
0 Arlington	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
4 Arlington	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
7 Arlington	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0
8 Arlington	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
26 Arlington	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0

5 rows x 51 columns

Rows are grouped by neighborhood to determine the frequency of occurrence of each restaurant. A new data frame is created with each row assigned for neighborhood and its corresponding top ten common restaurants based on cuisine.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alexandria	American Restaurant	Mexican Restaurant	Italian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Restaurant	New American Restaurant	Thai Restaurant
1	Annandale	American Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant	German Restaurant	French Restaurant
2	Arlington	American Restaurant		Mexican Restaurant	Mediterranean Restaurant	Indian Restaurant	Restaurant	Sushi Restaurant	Chinese Restaurant	Italian Restaurant	Thai Restaurant
3	Ashburn	Chinese Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant	German Restaurant	French Restaurant
4	Bristow	Mexican Restaurant	Vietnamese Restaurant	Thai Restaurant	American Restaurant	Sushi Restaurant	Chinese Restaurant	Ethiopian Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant

Finally, k-means clustering is performed on the data frame to check the pattern for each neighborhood and get the information about the top ten common restaurants for each neighborhood. Before fitting the data frame best value of k for k-means clustering is determined by based on silhouette_score from sklearn.metrics. It is observed from the graph below that, k = 8 would yield more better results for the computation using k-means clustering.



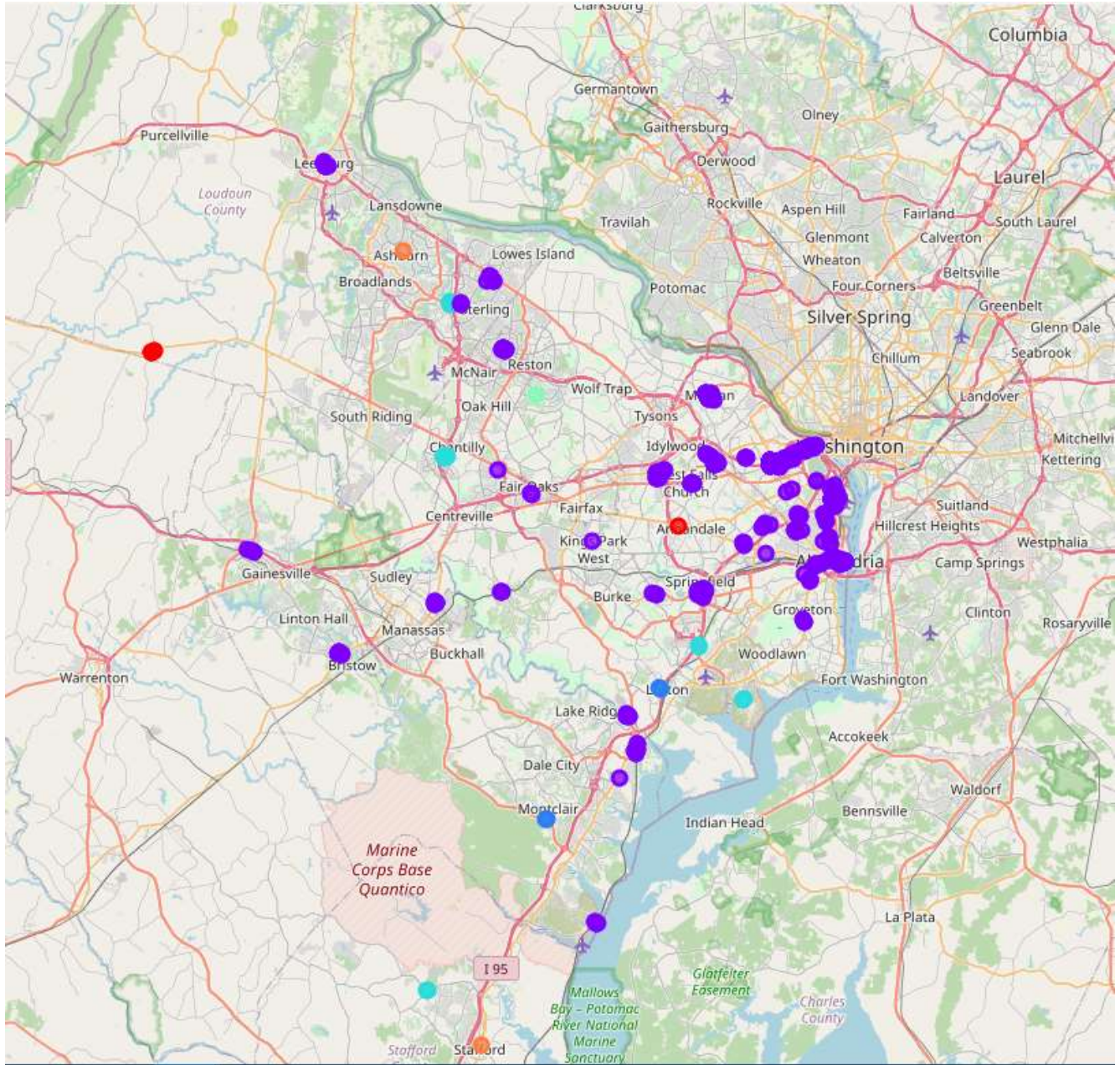
IV. RESULT

Neighborhood K-Means clustering based on mean occurrence of venue category:

All 7 clusters follow unique pattern for top ten common restaurants for a particular neighborhood. The detail shows the number of neighborhoods assigned to each cluster. Cluster 1 has the highest neighborhoods of 669 and cluster 4 and 5 has the least with 1.

```
: 1      669
   3      19
   0       7
   2       4
   6       2
   5       1
   4       1
   Name: Cluster Labels, dtype: int64
```

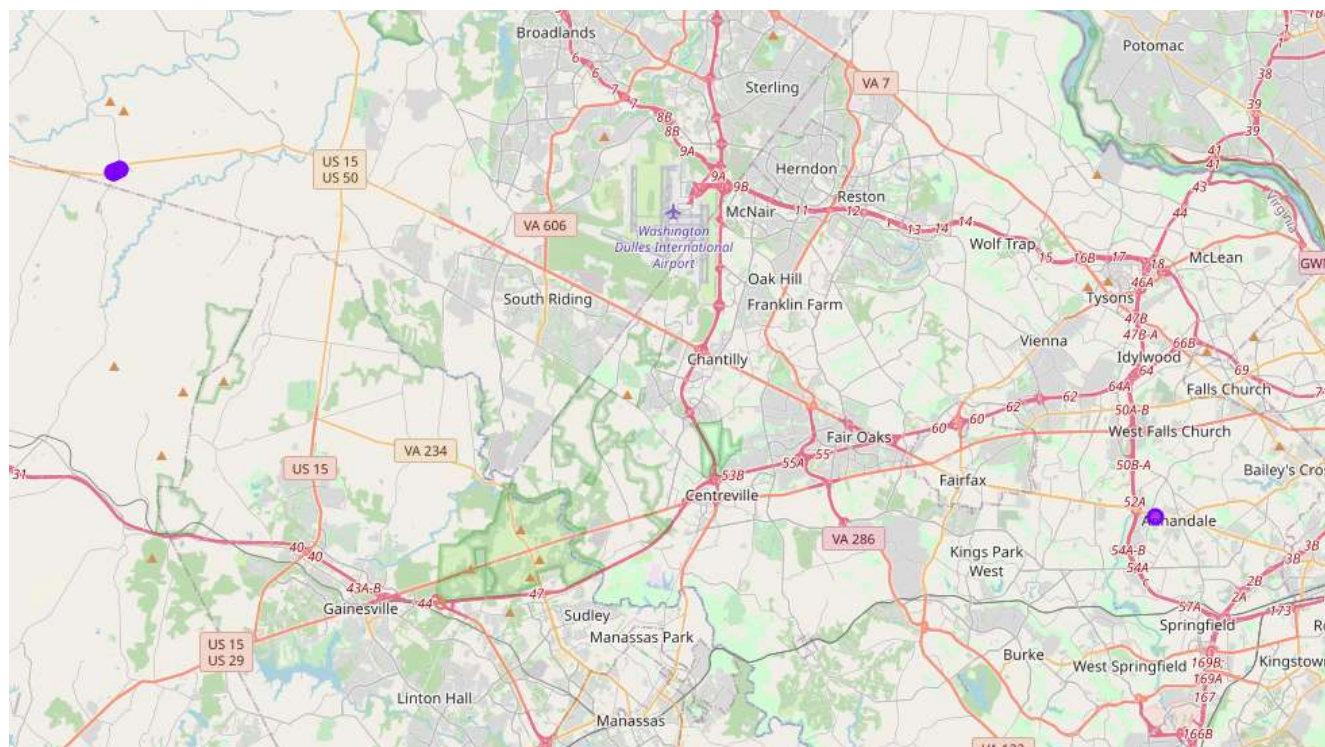
The clustered map is shown below:



The resulting data frames for each cluster are shown below.

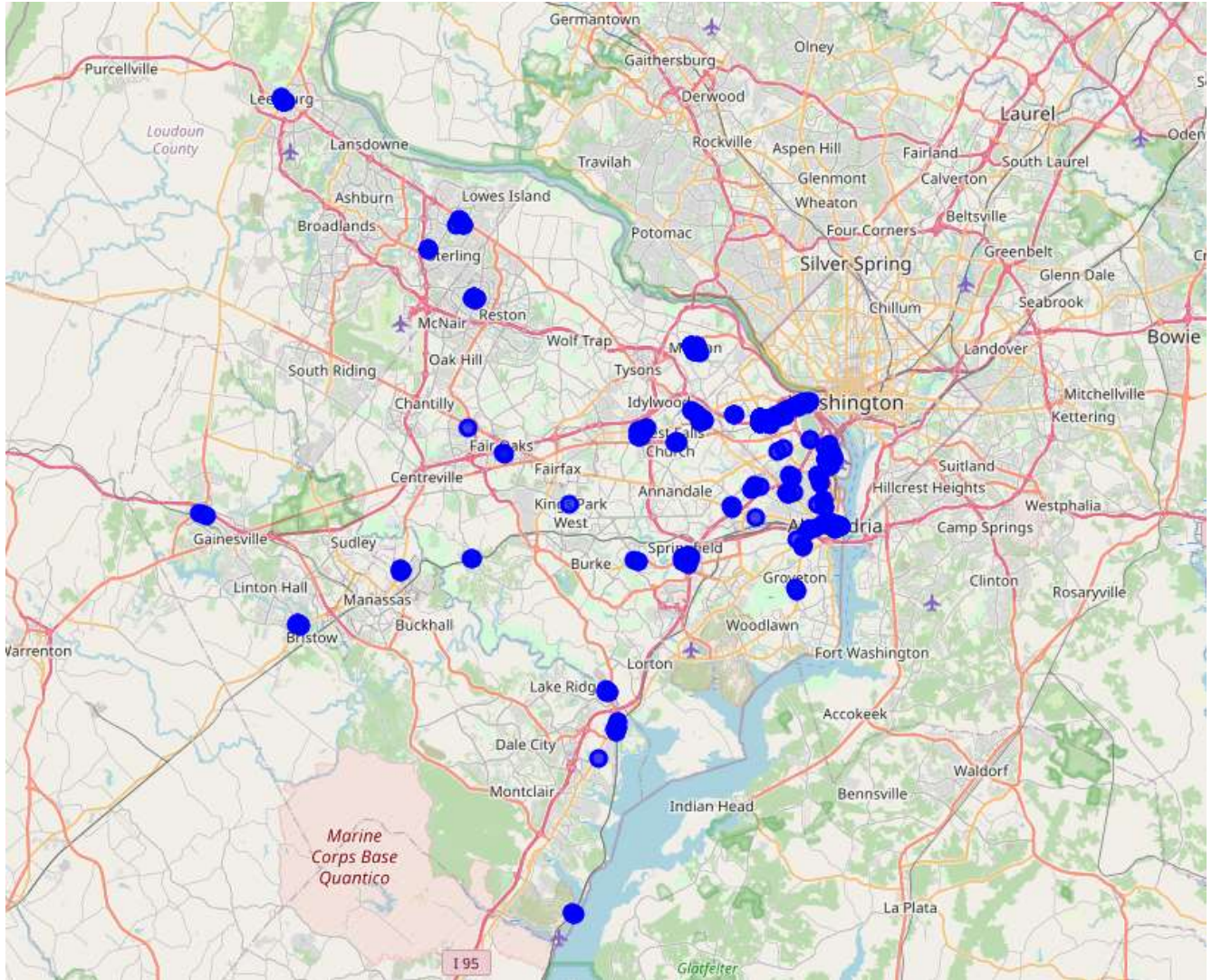
Cluster 0 (Most Common: American Restaurant & Seafood Restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
79	Middleburg	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant
83	Middleburg	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant
86	Middleburg	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant
326	Annandale	American Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant	Italian Restaurant
994	Middleburg	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant
998	Middleburg	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant
1001	Middleburg	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant



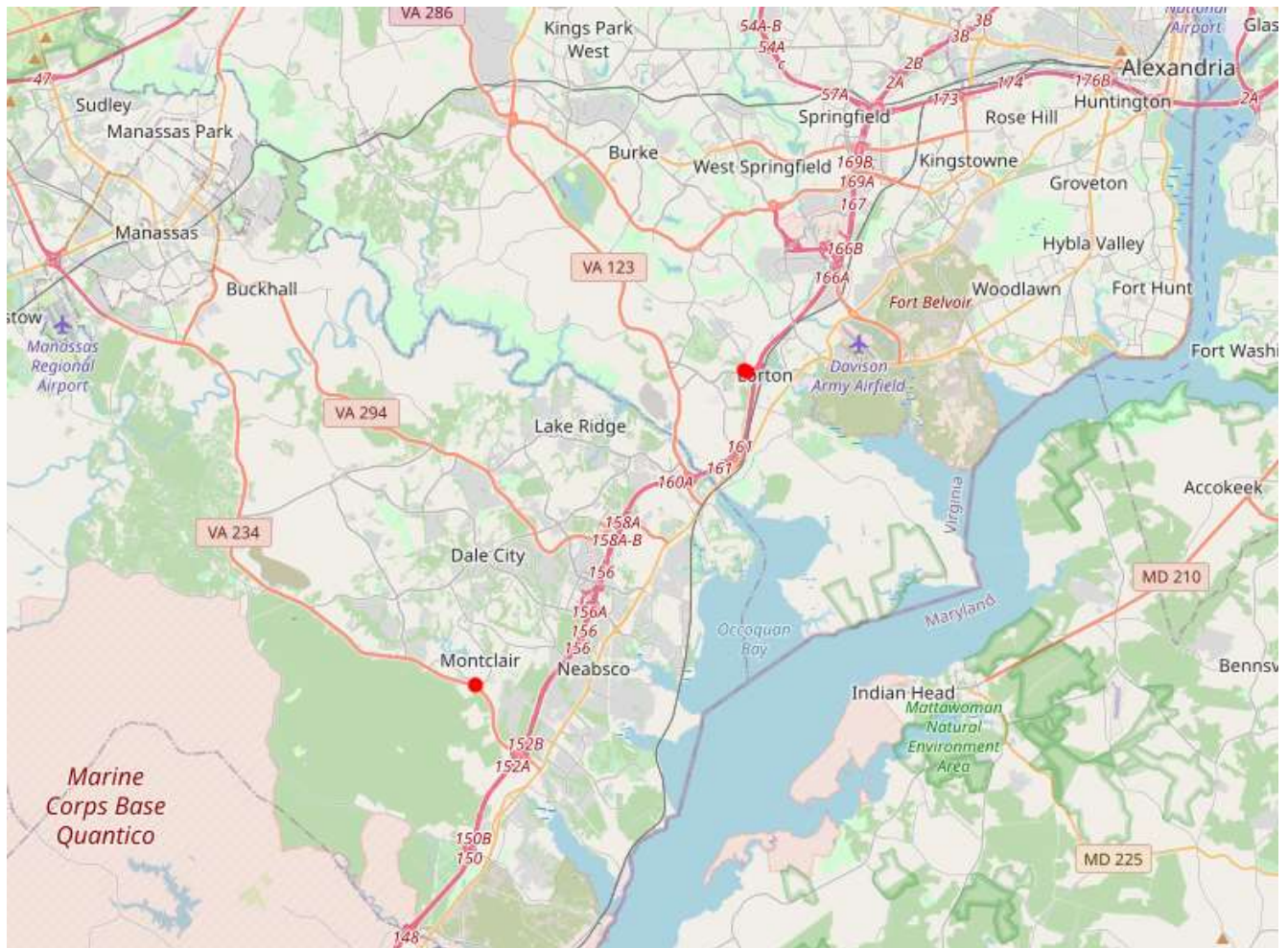
Cluster 1 (Most Common: Depends on neighborhood)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arlington	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Indian Restaurant	Restaurant
4	Arlington	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Indian Restaurant	Restaurant
7	Arlington	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Indian Restaurant	Restaurant
8	Arlington	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Indian Restaurant	Restaurant
26	Arlington	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Indian Restaurant	Restaurant
...
2904	Alexandria	American Restaurant	Mexican Restaurant	Italian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant
2905	Alexandria	American Restaurant	Mexican Restaurant	Italian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant
2906	Alexandria	American Restaurant	Mexican Restaurant	Italian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant
2907	Alexandria	American Restaurant	Mexican Restaurant	Italian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant
2918	Alexandria	American Restaurant	Mexican Restaurant	Italian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant



Cluster 2 (Most Common: Greek & Chinese Restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1124	Dumfries	Greek Restaurant	Chinese Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
1125	Dumfries	Greek Restaurant	Chinese Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
1841	Lorton	Greek Restaurant	Filipino Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
1843	Lorton	Greek Restaurant	Filipino Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant



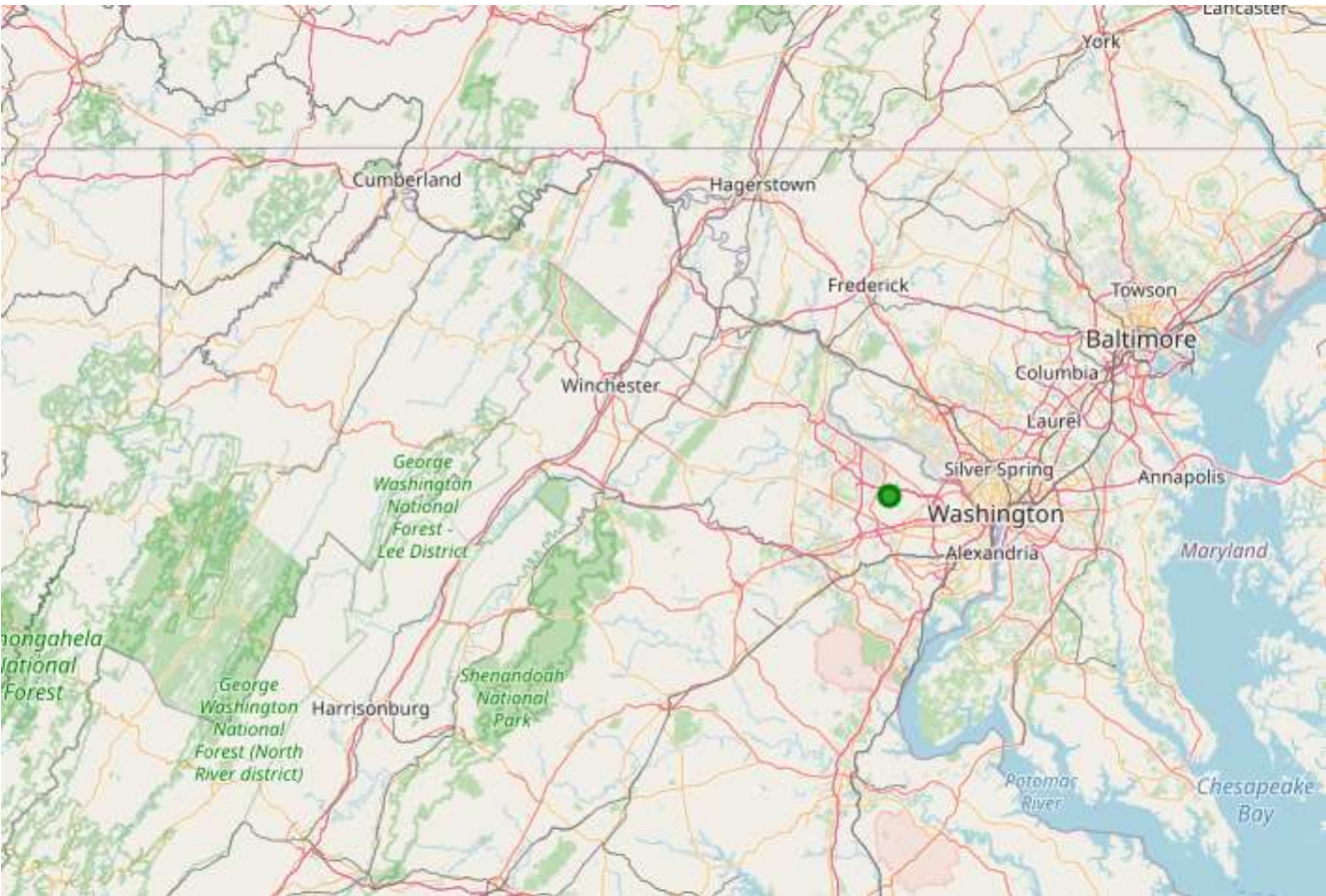
Cluster 3 (Most Common: Fast Food & American Restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
526	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
528	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
529	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
533	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
538	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
539	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
543	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
545	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
548	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
549	Chantilly	Fast Food Restaurant	American Restaurant	Japanese Restaurant	Mexican Restaurant	Ethiopian Restaurant
609	Newington	American Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
612	Newington	American Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
802	Dulles	Kebab Restaurant	Asian Restaurant	Fast Food Restaurant	Ethiopian Restaurant	Japanese Restaurant
809	Dulles	Kebab Restaurant	Asian Restaurant	Fast Food Restaurant	Ethiopian Restaurant	Japanese Restaurant
815	Dulles	Kebab Restaurant	Asian Restaurant	Fast Food Restaurant	Ethiopian Restaurant	Japanese Restaurant
1084	Fort Belvoir	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
1089	Fort Belvoir	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
2890	Garrisonville	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant
2891	Garrisonville	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant



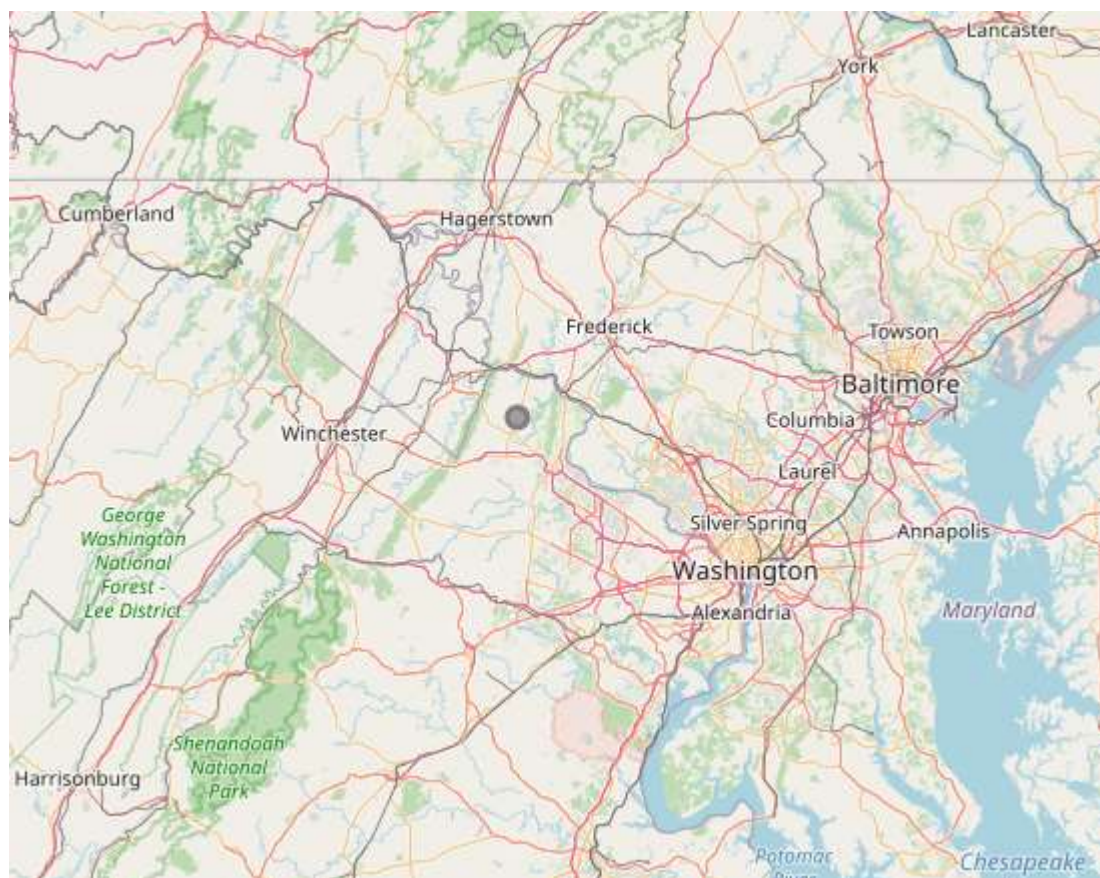
Cluster 4 (Most Common: Thai Restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2138	Reston	Thai Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Italian Restaurant	Indian Restaurant



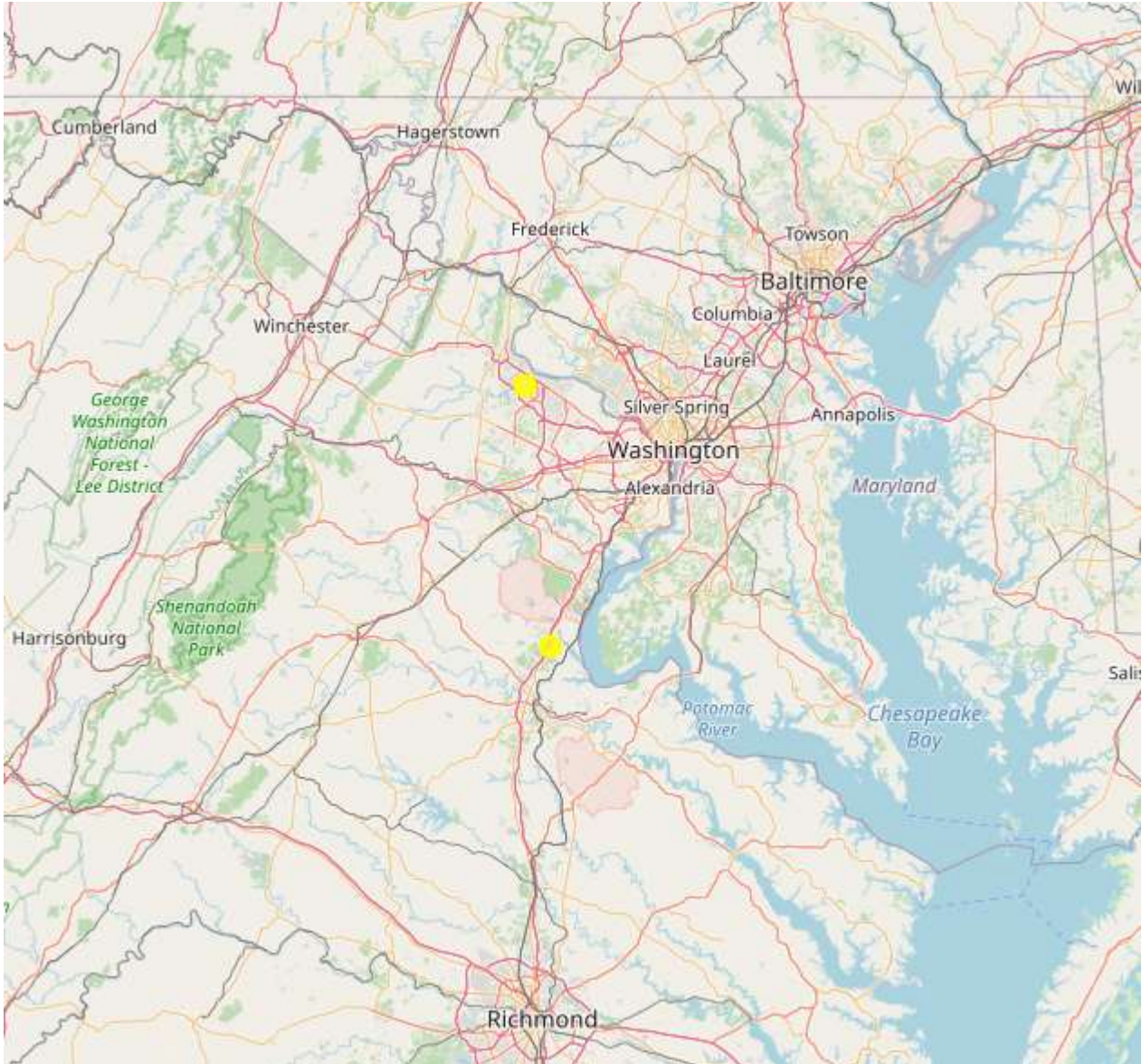
Cluster 5 (Most Common: Vietnamese Restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1888	Lovettsville	Restaurant	Vietnamese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant



Cluster 6 (Most Common: Chinese & Vietnamese Restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
791	Ashburn	Chinese Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant	Italian Restaurant
1433	Stafford	Chinese Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Japanese Restaurant	Italian Restaurant



V. DISCUSSION

Based on the observations noted from the map in the Results section, most of the restaurants are concentrated in the central area of Fairfax county, with the highest number in cluster 1 and moderate number in cluster 0 and 3.

The results can be approached in two ways:

- 1) On one hand, cluster 4 and 5 has very low number to no restaurants in the neighborhoods. This represents a great opportunity and high potential areas to open new restaurant as there is very little to no competition from existing restaurants. Meanwhile, restaurants in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of restaurants.
- 2) From another perspective, the results also show that the oversupply of restaurants mostly happened in the northern area of the city, with the suburb area still have very few restaurants. Therefore, this project recommends property developers to capitalize on these findings to open new restaurants in neighborhoods in cluster 4 and 5 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new restaurants in neighborhoods in cluster 4 and 5 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 1 which already have high concentration of restaurants and suffering from intense competition.

VI. CONCLUSION

There is always room for improvement and hence the above solution I have provided can also be improved for best results depending upon the data we have.

This analysis is performed on limited data. If good amount of data is available there is scope to come up with better results. If there are lot of restaurants probably there is lot of demand. Virginia has so many restaurants, yet certain neighborhood or counties do not have a specific cuisine restaurant available.

As per the neighborhood or restaurant type mentioned like Greek Restaurant further analysis can be performed. A venue with the lowest risk and competition can be identified. In the future, it is a good idea to collect more information about revenue, population, rent payments, crime level, future projects, etc. It will help select a good location inside each county.