Best Neighborhood to Open a Restaurant – Arlington, VA

Introduction:

Several entrepreneurs and restaurant connoisseurs are looking to open different types of restaurants in Arlington, VA and they want to know which neighborhoods are going to be the best for each type of restaurant. There are a variety of factors that can affect this decision, such as:

- A City's Population (its size, density, and ethnic/cultural/socioeconomic backgrounds)
- Surrounding Competition (the type of restaurants and total number of restaurants.
- Zoning Restrictions
- Property Values
- Local Taxes & Fees

So, to start our search, we will focus on the different types of restaurants that make up each neighborhood and classify them based on their similarities.

Data and Cleaning:

- Arlington Neighborhoods: 'https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Arlington_County,_Virginia'
- Geospatial Coordinates: Nominatim (from the Geopy library)
- Venues & Venue Information: Foursquare API

```
['Alcova Heights',
 'Arlington Forest',
 'Arlington Heights',
 'Arlington Ridge',
 "Arlington View / Johnson's Hill",
 'Ashton Heights',
 'Aurora Highlands',
 'Aurora Hills',
 'Ballston',
 'Barcroft',
 'Bellevue Forest',
 'Bluemont',
 'Bon Air',
 'Boulevard Manor',
 'Brandon Village']
```

	Neighborhood	Latitude	Longitude
0	Alcova Heights	38.8646	-77.0972
1	Arlington Forest	38.8689	-77.1131
2	Aurora Highlands	38.8528	-77.0684
3	Aurora Hills	38.8515	-77.0641
4	Ballston	38.882	-77.1115
5	Barcroft	38.8559	-77.1039
6	Bellevue Forest	38.9143	-77.1136
7	Bluemont	38.8747	-77.133
8	Bon Air	38.8732	-77.1266
9	Brandon Village	38.8757	-77.1158
10	Buckingham	38.8734	-77.1066
11	Carlin Springs	38.8772	-77.1118
12	Cherrydale	38.8971	-77.1083
13	Claremont	38.8432	-77.1047
14	Clarendon	38.8871	-77.0952
15	Columbia Forest	38.854	-77.1103

Beautiful Soup – Web Scrapping

Nominatim + Pandas

Foursquare API

- Nominatim is not perfect, so for simplicity, any neighborhoods with no coordinates or coordinates that do not make geological sense are filtered out (go from 73 to 50 neighborhoods).
- Call Foursquare API and find 924 venues in these 50 neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alcova Heights	38.864557	-77.097201	Redbox	38.868374	-77.097198	Video Store
1	Alcova Heights	38.864557	-77.097201	Burger King	38.860737	-77.094868	Fast Food Restaurant
2	Alcova Heights	38.864557	-77.097201	7-Eleven	38.868449	-77.097067	Convenience Store
3	Alcova Heights	38.864557	-77.097201	El Ranchero Migueleno	38.860710	-77.095183	Mexican Restaurant
4	Alcova Heights	38.864557	-77.097201	Alcova Heights	38.861586	-77.101470	Basketball Court

Restaurants

- Filter venues to focus only on restaurants.
- Filter neighborhoods further by only focusing on neighborhoods with 5 or more restauarants
- Restructure data to see Top 5 most common types of restaurants (based on percent total of a neighborhood's restaurants)

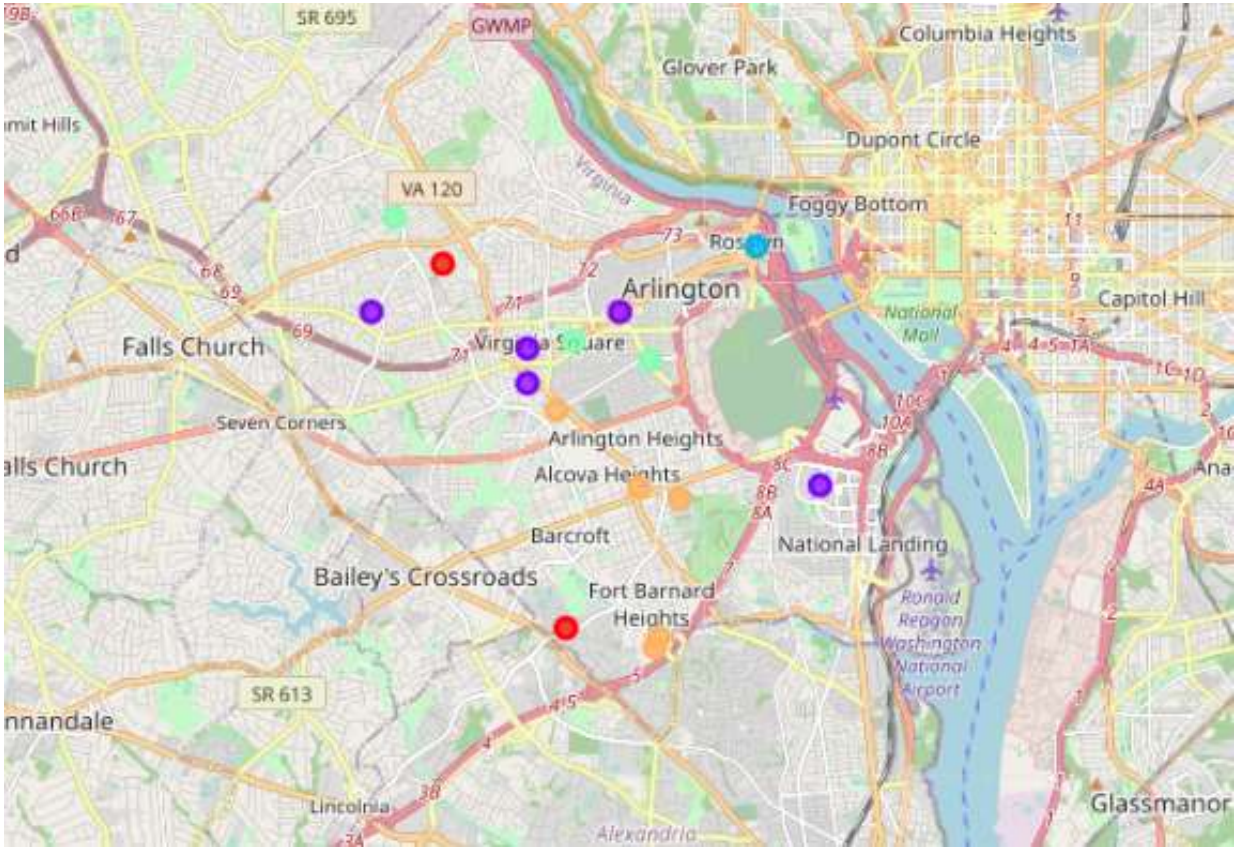
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ballston	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Restaurant	Indian Restaurant
1	Buckingham	Latin American Restaurant	Mexican Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant
2	Carlin Springs	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Mediterranean Restaurant	New American Restaurant
3	Claremont	Fast Food Restaurant	American Restaurant	Chinese Restaurant	Latin American Restaurant	Vietnamese Restaurant
4	Clarendon	American Restaurant	Vietnamese Restaurant	Persian Restaurant	Eastern European Restaurant	French Restaurant
5	Columbia Heights	Thai Restaurant	American Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Fast Food Restaurant
6	Garden City	Indian Restaurant	Mexican Restaurant	Thai Restaurant	Szechuan Restaurant	Sushi Restaurant
7	High View Park	Fast Food Restaurant	Afghan Restaurant	American Restaurant	Indian Restaurant	Italian Restaurant
8	Lyon Park	Korean Restaurant	Chinese Restaurant	South American Restaurant	Indian Restaurant	Mediterranean Restaurant
9	Pentagon City	Vietnamese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Portuguese Restaurant
10	Randolph Square	American Restaurant	Mexican Restaurant	Ramen Restaurant	Italian Restaurant	Indian Restaurant

Clustering

- Use K-Means Clustering Algorithm
- Since we are trying to find groups that have not been explicitly labeled, the k-means clustering algorithm is a good choice.
- Group neighborhoods into one of 5 clusters.

<u>Key</u>

- Red – Cluster 0
- Purple – Cluster 1
- Blue – Cluster 2
- Green – Cluster 3
- Gold – Cluster 4



Cluster 0 & Cluster 1

Cluster 0

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
13	Claremont	Fast Food Restaurant	American Restaurant	Chinese Restaurant	Latin American Restaurant	Vietnamese Restaurant
30	High View Park	Fast Food Restaurant	Afghan Restaurant	American Restaurant	Indian Restaurant	Italian Restaurant

Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Ballston	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	Restaurant	Indian Restaurant
11	Carlin Springs	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Mediterranean Restaurant	New American Restaurant
14	Clarendon	American Restaurant	Vietnamese Restaurant	Persian Restaurant	Eastern European Restaurant	French Restaurant
37	Pentagon City	Vietnamese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Portuguese Restaurant
47	Westover	Thai Restaurant	American Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Italian Restaurant

Cluster 2 & Cluster 3

Cluster 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
42	Rosslyn	Mediterranean Restaurant	Portuguese Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Mexican Restaurant

Cluster 2 only contains one neighborhood, Rosslyn, so it must have a unique list of most common venues. This could mean it is a niche neighborhood and hard to get a foot hold in.

Cluster 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
28	Garden City	Indian Restaurant	Mexican Restaurant	Thai Restaurant	Szechuan Restaurant	Sushi Restaurant
32	Lyon Park	Korean Restaurant	Chinese Restaurant	South American Restaurant	Indian Restaurant	Mediterranean Restaurant
45	Virginia Square	Afghan Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Peruvian Restaurant	Fast Food Restaurant

Cluster 3 contains some middle eastern and southeast Asian restaurants.

Cluster 4

Cluster 4

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Buckingham	Latin American Restaurant	Mexican Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant
16	Columbia Heights	Thai Restaurant	American Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Fast Food Restaurant
40	Randolph Square	American Restaurant	Mexican Restaurant	Ramen Restaurant	Italian Restaurant	Indian Restaurant
43	Shirlington	American Restaurant	Mexican Restaurant	Ramen Restaurant	Italian Restaurant	Indian Restaurant
46	Westmont	Mexican Restaurant	Thai Restaurant	Fast Food Restaurant	American Restaurant	Ethiopian Restaurant

Conclusion & Next Steps

This clustering analysis gives our stakeholders an idea of feature restaurant types prominent in each neighborhood in Arlington, VA.

- Ideally, the results will aid in the neighborhood selection process for opening a new restaurant.
- It will help in finding the right balance between a market oversaturated in one type of restaurant and a competitive one.
- Finally, it will help reduce risk and improve ROI.

Moving forward to improve this analysis, it would be a good idea to include the remaining neighborhoods in Arlington that were left out due to bad or no location data. It would also be interesting to take financial, demographic, and popularity data into consideration as it would offer a more pinpointed and final neighborhood selection.