The Battle of the Neighborhoods

IBM Applied Data Science Capstone

I. Lahlou

May 2020

Table of contents

1. Introduction: Business Problem

An entrepreneur is looking for an ideal location to target for a new Ice Cream Shop in an area within Fairfax County, Virginia.

He wants to use a quantitative approach to identify the 3 best locations within the target search area. He will then pick his preferred location from these 3 identified locations.

His selection criteria include:

- Low Ice Cream Shops per capita
- Located in a commercial area
- High foot traffic

2. Data

This problem can be solved using population data, location data and other datasets and applying some algorithms to best address the entrepreneur's selection criteria.
Data to be used in this analysis:

- The most recent population data will be extracted from Fairfax County, Virginia Open Geospatial Data web site http://data-fairfaxcountygis.opendata.arcgis.com/datasets/314bfe4019754952a715be3a33384d9d_0.csv
    - The population data contains the average population per household along with each household geographic location.
    - Data will first need to be reviewed and cleaned.

- The Fairfax County boundary will also be extracted from Fairfax County, Virginia Open Geospatial Data web site https://opendata.arcgis.com/datasets/58cf8abd870e47aeb1be8911983d2d44_15.geojson
- To identify commercial areas within the Target area, Land Use data of Fairfax county will be extracted from ArcGIS Open Data web site: https://opendata.arcgis.com/datasets/64abd9d8f80146a4924727dc42dfba86_0.geojson

Once the data is collected, processed and explored, an in-depth analysis will be performed and conclusions will be drawn. The results will then be communicated to the entrepreneur for feedback.

3. Methodology

- The search area targeted by the entrepreneur includes just a small section within the county. A polygon representing this area will therefore be created to filter only the population within the target search area.
- The target area polygon will then be divided into a grid of equal sized clusters representing sub-neighborhoods.
- The population count sum of each sub-neighborhood will then be computed and assigned to the cluster's centroid location.
- Venue data will be extracted from the Foursquare location data, reviewed and structured.
- Land use data will then be filtered and mapped to display only commercial areas within the target search area.
- A selection will be made by process of elimination. We will eliminate clusters without commercial areas, then clusters with high Ice Cream shops per capita and finally we will select the best 3 locations out of the remaining clusters based on general knowledge of the foot traffic in those clusters.

4. Analysis

Acquiring and cleaning county population data:

The first step in the analysis is to extract the most recent population data in Fairfax County. The population data contains the average population per household along with each household geographic location.

Data is first reviewed and cleaned. The table below shows the raw data:

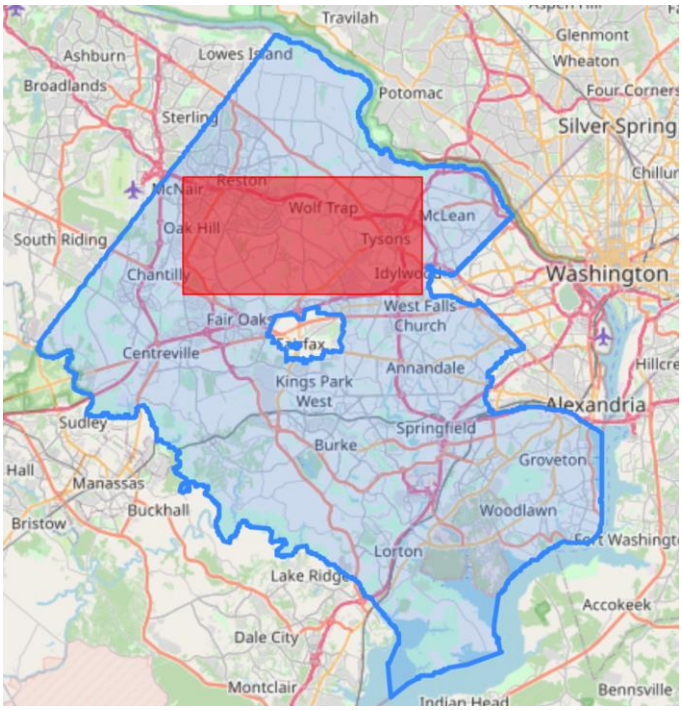
	X	Y	OBJECTID	PIN	PARCE_ID	CURRE_POPUL	LOW_ESTIM_POPUL	HIGH_ESTIM_POPUL	VALID_FROM	VALID_TO
0	-77.241452	38.897249	1	0393 28 0084	314692	3.0423	2.8774	3.2093	NaN	2019-01-01T05:00:00.000Z
1	-77.241194	38.897127	2	0393 28 0085	285535	3.0423	2.8774	3.2093	NaN	2019-01-01T05:00:00.000Z
2	-77.240933	38.897012	3	0393 28 0086	356281	3.0423	2.8774	3.2093	NaN	2019-01-01T05:00:00.000Z
3	-77.240667	38.896900	4	0393 28 0087	228776	3.0423	2.8774	3.2093	NaN	2019-01-01T05:00:00.000Z
4	-77.240376	38.896784	5	0393 28 0088	333761	3.0423	2.8774	3.2093	NaN	2019-01-01T05:00:00.000Z

The table below shows the data after cleaning:

	ID	Latitude	Longitude	Pop
0	1	38.897249	-77.241452	3.0423
1	2	38.897127	-77.241194	3.0423
2	3	38.897012	-77.240933	3.0423
3	4	38.896900	-77.240667	3.0423
4	5	38.896784	-77.240376	3.0423

Mapping to visualize Fairfax County and the search area:

Blue: County boundary
Red: Search area

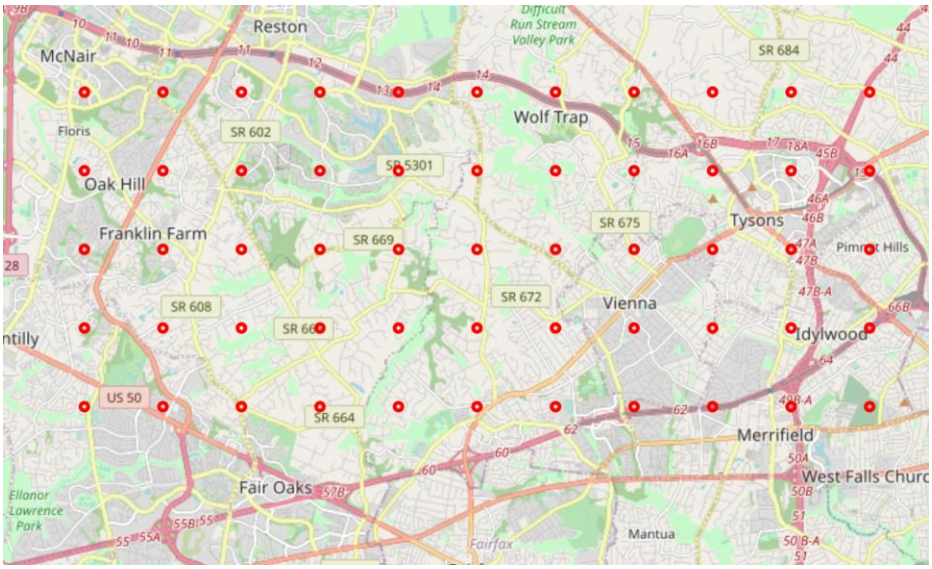


Dividing the search area into a grid of an equal sized clusters and computing population per cluster:

The target area polygon is then divided into a grid of equal sized clusters representing sub-neighborhoods. The spacing between the 55 center locations is 2.32 km.

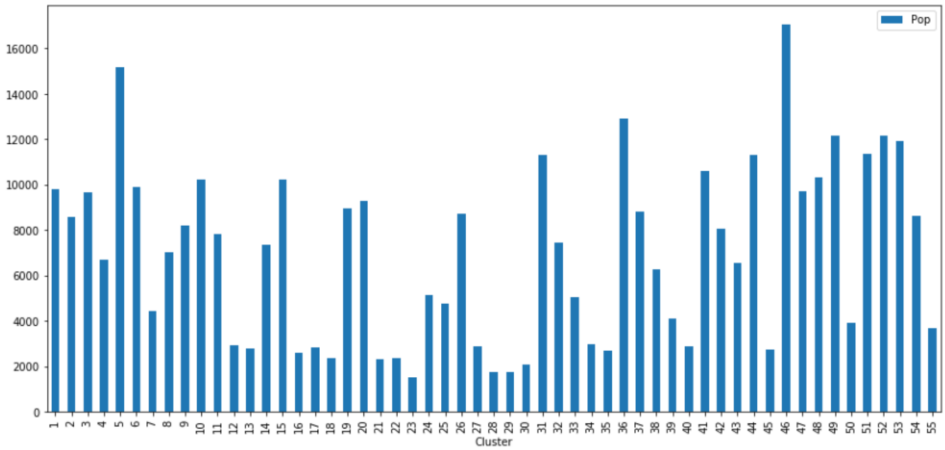
The population count sum of each sub-neighborhood is then computed and assigned to the cluster's centroid location.

Red: Cluster Center Locations



Population per cluster:

	Cluster	Pop	Lat	Long
0	1	9818.0	38.880000	-77.41
1	2	8572.0	38.896222	-77.41
2	3	9656.0	38.912440	-77.41
3	4	6697.0	38.928655	-77.41
4	5	15151.0	38.944866	-77.41



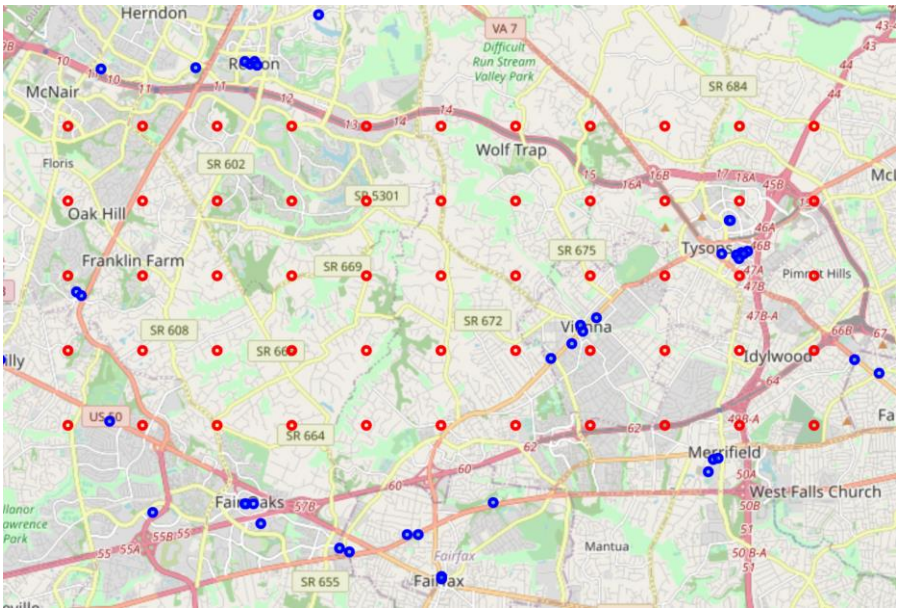
Venue data:

Venue data is extracted from the Foursquare location data, reviewed and structured.

There are 67 ice cream shops within 15 kms from the center of the search area.

	name	categories	lat	lng
0	Pitango Gelato	Ice Cream Shop	38.958326	-77.359211
1	Ben & Jerry's	Ice Cream Shop	38.858604	-77.355992
2	Freddy's Frozen Custard & Steakburgers	Ice Cream Shop	38.863287	-77.291277
3	Sugar Mama's Ice Cream	Ice Cream Shop	38.852587	-77.331517
4	Ben & Jerry's	Ice Cream Shop	38.900370	-77.266271

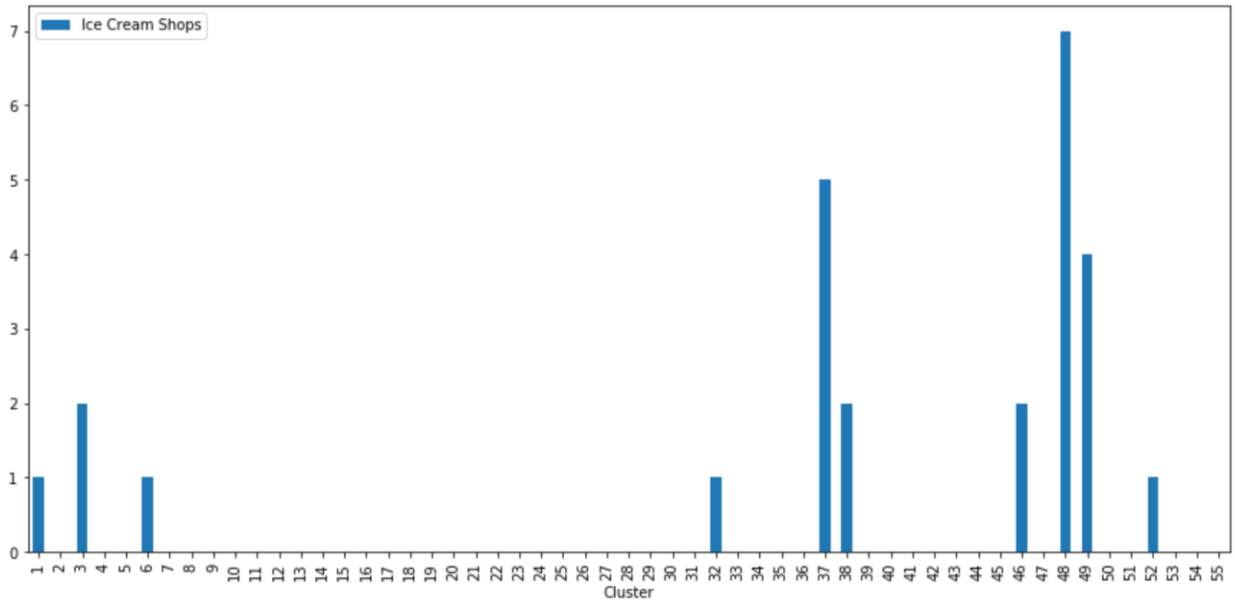
Red: Cluster Center Locations
Blue: Ice Cream Shops



The Ice Cream Shops within 1275 meters of each centroid are then identified and their counts aggregated.

	Cluster	Pop	Lat	Long	name	distance_m
0	1	9818.0	38.880000	-77.410000	sweetFrog Premium Frozen Yogurt	1016.975766
1	3	9656.0	38.912440	-77.410000	Baskin-Robbins	431.920098
2	3	9656.0	38.912440	-77.410000	Zinga	579.491727
3	6	9914.0	38.880000	-77.389159	sweetFrog Premium Frozen Yogurt	797.168420
4	32	7441.0	38.896222	-77.284955	Molly's Yogurt	888.638102

Ice Cream Shops per Cluster:



Importing County Land-Use data:

Land use data in Fairfax County is imported, cleaned and mapped to display only commercial areas within the target search area.

CATEG	OBJECTID
Agricultural	1
Commercial	230
High-density Residential	96
Industrial, light and heavy	33
Institutional	198
Low-density Residential	2121
Medium-density Residential	1710
Open land, not forested or developed	1363
Public	17
Recreation	136
Utilities	54

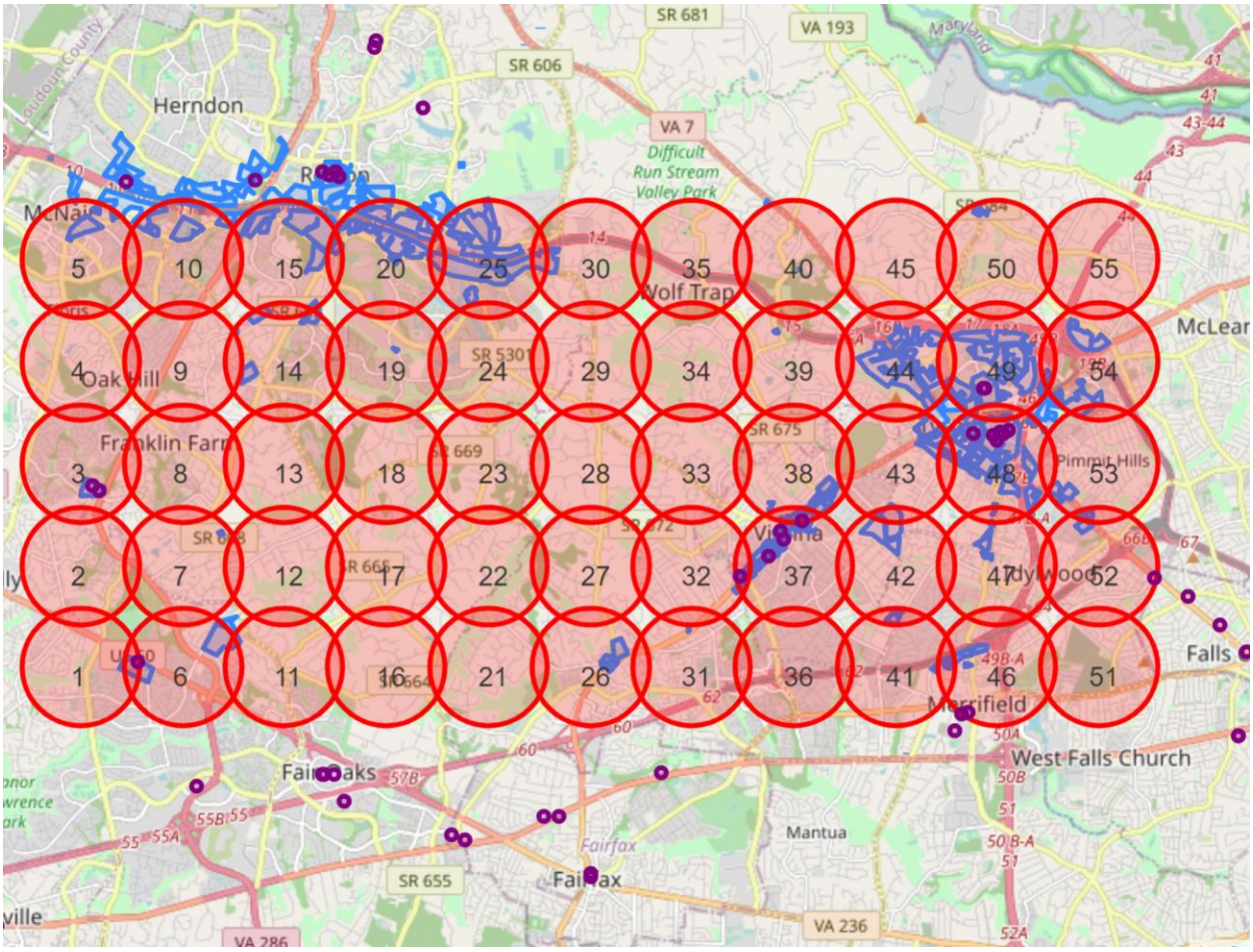
	OBJECTID	ACRES	VALID_FROM	VALID_TO	CATEG	Shape__Are	Shape__Len	ID	geometry
0	2358	4.80972	1546214400000	1546318800000	Commercial	209511.304688	2036.967903	0	POLYGON ((-77.26911 38.89861, -77.26931 38.898...
1	2361	0.57077	1546214400000	1546318800000	Commercial	24862.960938	630.826742	0	POLYGON ((-77.38097 38.90129, -77.38145 38.901...
2	2363	7.56844	1546214400000	1546318800000	Commercial	329681.382813	2939.873316	0	POLYGON ((-77.22452 38.90056, -77.22451 38.900...
3	2364	10.77443	1546214400000	1546318800000	Commercial	469334.382813	3157.926594	0	POLYGON ((-77.24276 38.89787, -77.24276 38.897...
4	2365	4.69036	1546214400000	1546318800000	Commercial	204311.859375	2772.353873	0	POLYGON ((-77.26632 38.90053, -77.26557 38.901...

5. Results and Discussion

The selection is made by process of elimination.

First, clusters without commercial areas are eliminated. Then, clusters with high Ice Cream shops per capita are removed. Finally, we will select the best 3 locations out of the remaining clusters based on general knowledge of the foot traffic in those clusters.

Blue = Commercial Areas
Purple = Ice Cream Shops
Red = Clusters 1-55



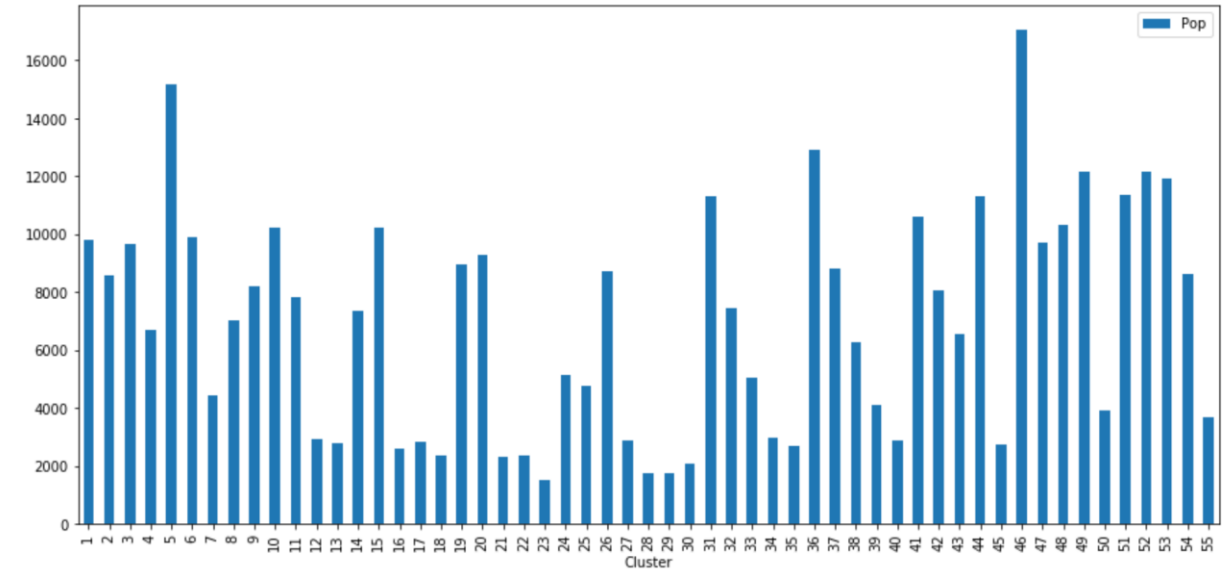
As per map above, the commercial areas available within the search area are limited to 32 of the 55 clusters:
1,3,5,6,7,9,10,11,14,15,19,20,25,26,30,31,32,37,38,39,41,42,43,44,46,47,48,49,50,52,53 and 54.
We limit our selection to these 32 clusters since we cannot open an Ice Cream Shop in residential areas.

The total population per cluster as well as Ice Cream shops per capita are illustrated in the Charts below.

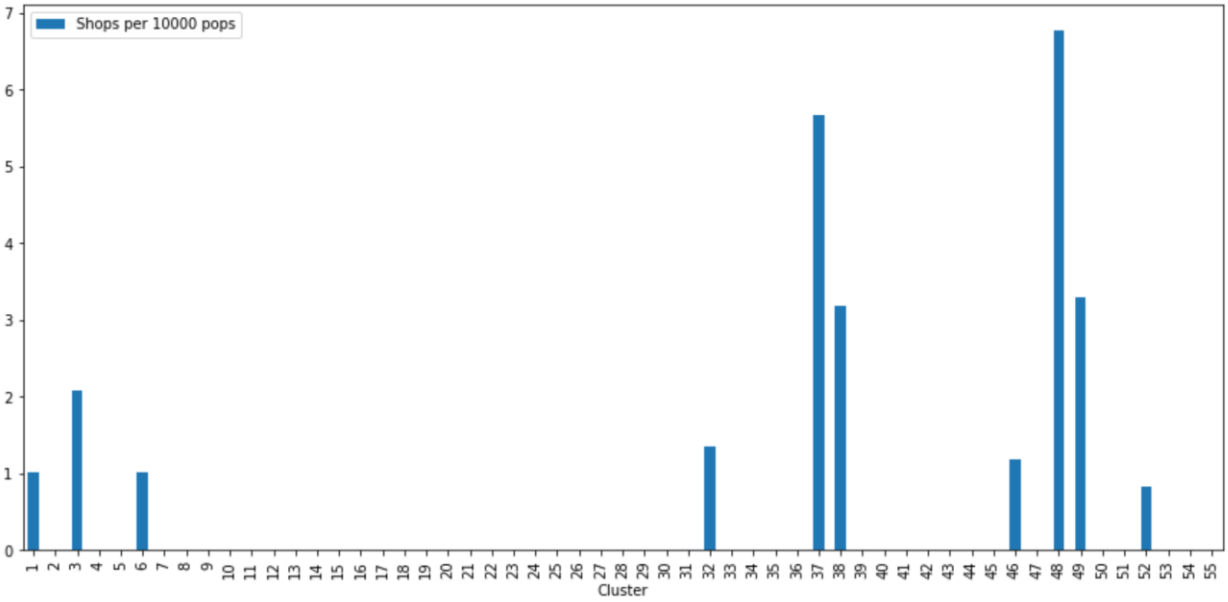
Based on the data, we further eliminate 10 Clusters 1,3,6,32,37,38,46,48,49 and 52 since they have the highest Ice cream shops per capita.

Out of the remaining 22 clusters which include commercial areas and their Ice Cream shops per capita is low, the following clusters stand out as best possible locations:
5,10,14,19,20,26,31,41,44,53 and 54.

Population per cluster:



Shops per 10,000 pops:



6. Conclusion

Further analysis of the remaining 11 clusters based on general Knowledge of foot traffic and adjacent clusters population reveals that the top 3 best locations for an Ice Cream Shop are clusters: **5, 44 and 47**

Our methodology and final selection will be communicated to the entrepreneur for feedback. Additional study can be conducted as needed based on the feedback received.