

Using machine learning to find location for opening a new restaurant in Michigan.

1. INTRODUCTION AND BACKGROUND OF THE PROBLEM

Having a successful restaurant depends on multiple factors. Many restaurants fail because they didn't put enough consideration into the key factors. One of the key factors for a successful restaurant is choosing the location. The neighborhood contributes multiple factors to the success or failure of a restaurant. In this project machine learning will be used to select a potential spot for opening a restaurant in one of the Michigan counties.

The following three characteristics of the neighborhood will be used in choosing optimum location.

1.1 The presence of other businesses that attract people to your restaurants.

This includes sports stadium, major medical facility and theaters. For example, moviegoers often stop by to eat in restaurants before or after the show, therefore the presence of such attraction is a plus.

1.2 The population size.

The neighborhood should have enough people to support your restaurant. There should be enough people who live in or pass through the area regularly to keep you busy.

1.3 Income of the neighborhood

People should have extra money to spend. Neighborhood with higher income are expected to spend more money to keep your business running.

1.4 Scope of the project: There are multiple factors that will affect of the locations, however, in this project we will consider the above three factors in identifying the most optimum location to open a restaurant.

1.5 Target Audience

Entrepreneurs who are coming to Michigan and are planning to open a restaurant in one of the neighborhoods.

2. DATA

2.1 Description, Source and Use.

The Counties (Neighborhood), population size of the neighborhood and income per capital will be downloaded from **Wikipedia**¹. Search for the location will be at neighborhood level. That is neighborhoods for opening a restaurant will be identified. In doing so, both income and the population data will be used to identify the best neighborhoods for opening restaurant.

(83, 3)

	Neighborhood	Income	Population
0	Oakland	56138	1202362
1	Leelanau	32194	21708
2	Livingston	31609	180967
3	Washtenaw	31316	344791
4	Charlevoix	28403	25949

- Coordinates (Latitude and longitude) of the neighborhoods will be collected using **Geocoder** package. Those coordinates will be used to identify venues in each neighborhood.

	Neighborhood	Income	Population	Latitude	Longitude
1	Leelanau	32194	21708	44.938491	-85.811823
2	Livingston	31609	180967	41.968310	-86.542530
3	Washtenaw	31316	344791	42.253229	-83.838777
4	Charlevoix	28403	25949	45.316530	-85.260720
5	Midland	28363	83629	43.617080	-84.245060

- Venues from **Foursquare** will be extracted for each neighborhood. This data will be used to identify neighborhoods the presence and the number of businesses in the neighborhood. This will be used to identify a neighborhood with potential for new restaurant.

	Neighborhood	Income	Population	Latitude	Longitude	Number of restaurants
2	Livingston	31609	180967	41.96831	-86.54253	1.0
4	Charlevoix	28403	25949	45.31653	-85.26072	14.0
5	Midland	28363	83629	43.61708	-84.24506	24.0
9	Grand Traverse	27091	86986	42.69565	-82.92171	4.0
11	Eaton	25963	107759	41.91606	-83.39790	23.0

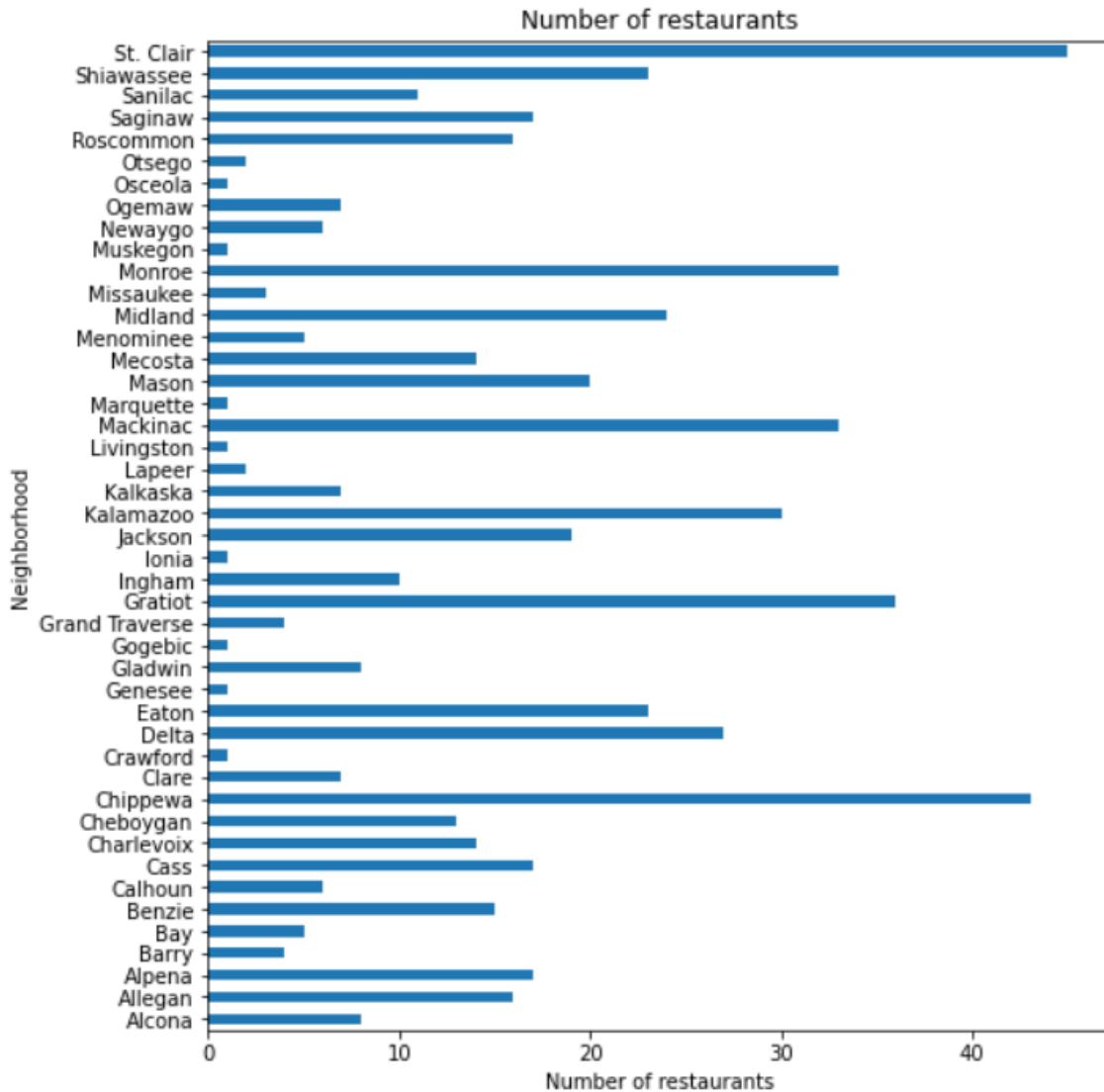
2.2 Data source: Wikipedia.

https://en.wikipedia.org/wiki/List_of_Michigan_locations_by_per_capita_income

3. METHOD

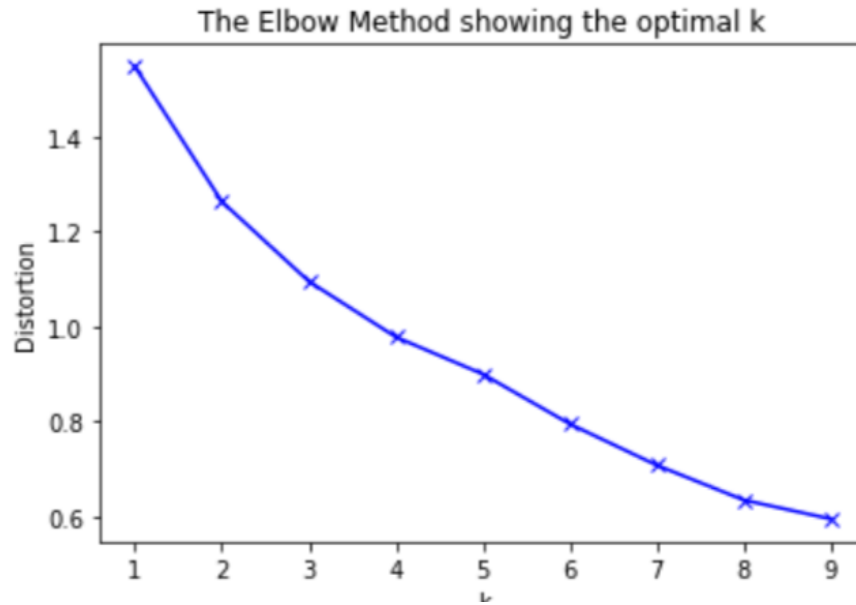
3.1 Data exclusion: In this project, neighborhoods with population between 10,000 and 700,000 will be considered. The reason is to exclude neighborhoods with few customers (low population) and highly competitive areas (high population). Similarly, neighborhoods that do not have restaurants and related venues will be excluded.

3.2 Number Of Restaurants Counted: Number of business that will have impact in opening new restaurant were counted. These are business are other restaurants, cafés, bars, theaters and shops. Neighborhoods with at least one venues that fulfill the above condition are counted in the number of restaurants. Total of 45 neighborhoods fulfilled the condition.



3.3 Machine learning: K-mean clustering is used to cluster similar neighborhoods in Michigan based on the population, number of restaurants and income per capital. The advantage of using K-mean is its algorithm is able to identify the non-linear structures and is best suited for real life complex data set.

K-elbow method was used to identify the optimum k value in this case was $k = 4$.

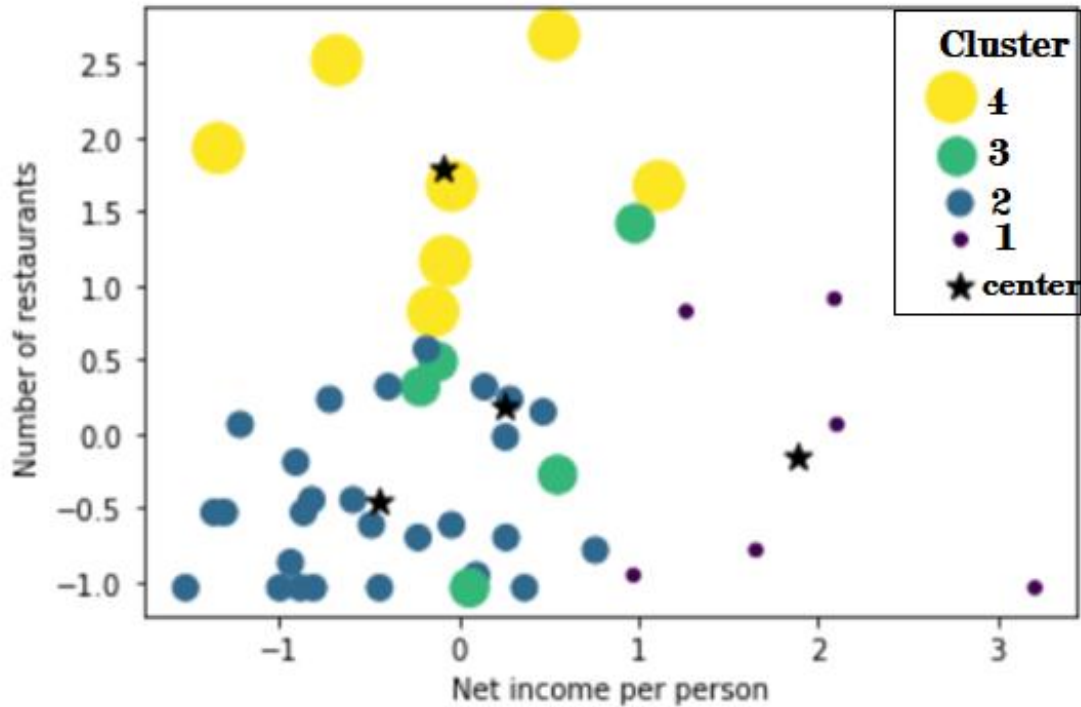


Optimum value of k-mean was calculated as selected as 4 from Elbow method and scatter plots of the normalized data (number of restaurants and income) and clusters was color and size coded.

4. RESULTS

4.1 Number of restaurants and income

Clusters are color coded cluster 1 [purple], 2 [blue], 3 [green] to cluster 4 [yellow] and size increase with cluster number. Black star marker represents the centroids.

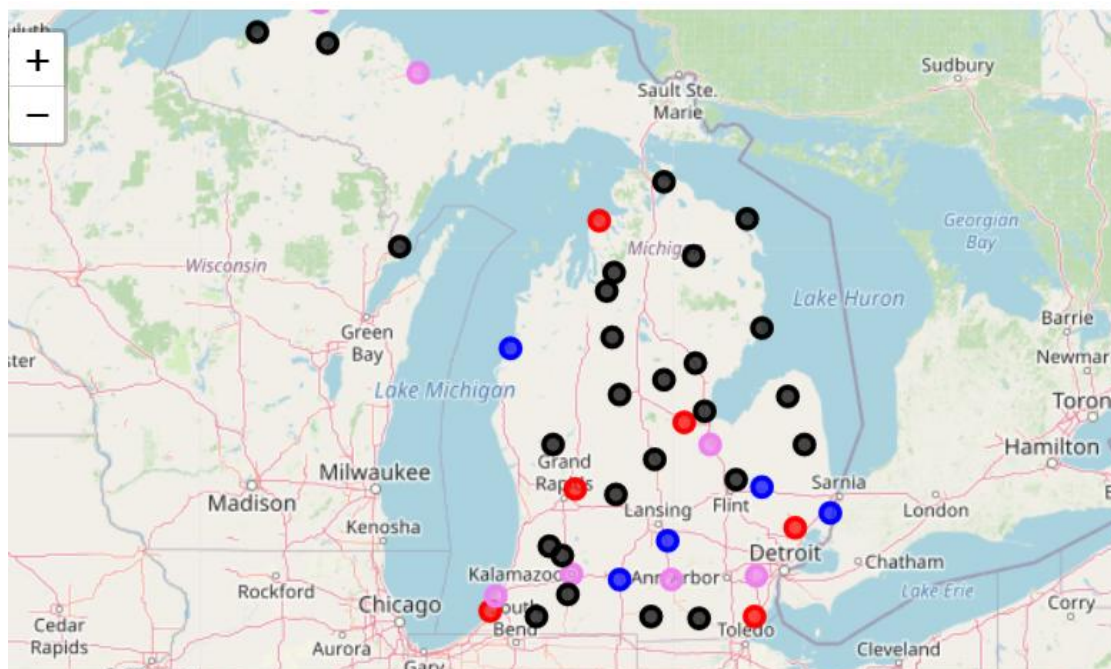


From the result it could be noted Cluster 1 has high income population but fewer number of restaurants. While Cluster 4 has the highest number of restaurants and average income.

Majority of the neighborhood is in Cluster 2 and they have low income and fewer restaurants. Neighborhoods in Cluster 3 have relatively higher number of income and more restaurants as compared to Cluster 2.

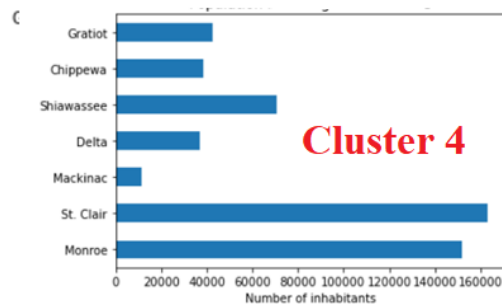
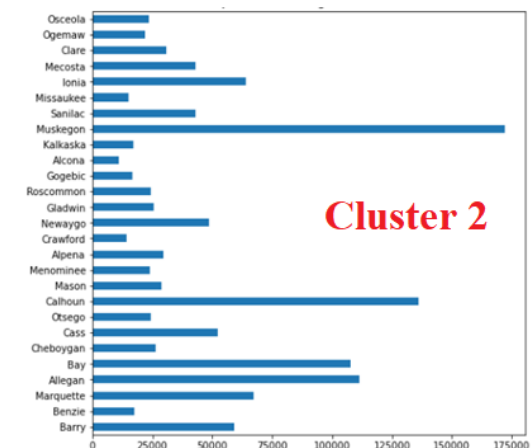
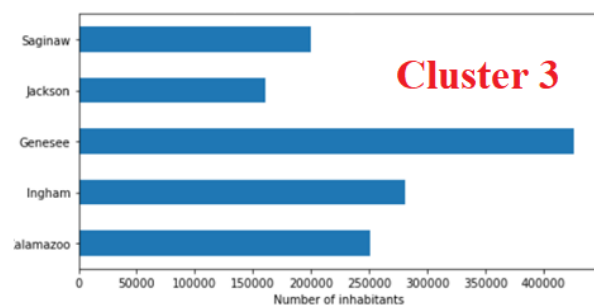
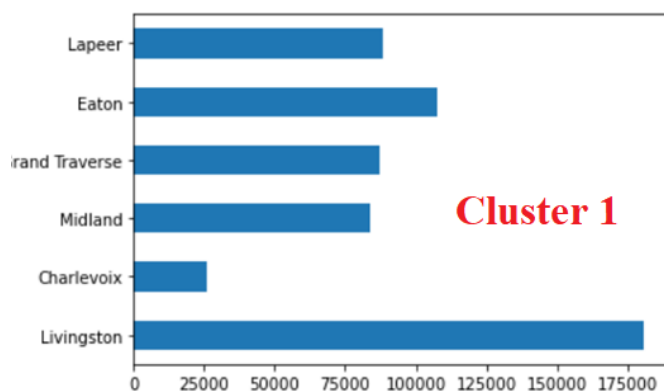
4.2 Number of restaurants and population

Michigan neighborhoods with four color coded clusters superimposed in the map.



Clusters 1,2 and 4 have a neighborhood with a maximum population of around 170,000.

Cluster 3 however has a neighborhood with highest population that is around 400,000 population.



5. DISCUSSION:

Neighborhoods in Cluster 1 could be considered to have the optimum location to open a restaurant. This is because it has population with highest income. Though, the number of restaurants are few it has a population as much as the other two clusters. This could be considered as less competitive location. As it can be seen in the population of cluster 3, **Livingstone** neighborhood has the highest number of population. Therefore it could be considered the as the first option.

As the second option because of its Cluster 3 there has highest population with mid income. Neighborhoods in Cluster 4 could be considered as the third option due high number of restaurants but low income of the population makes it risky.

Neighborhoods in Cluster 2 are the least to be considered for opening a new restaurant due to low number of restaurants coupled with low income population. This cluster also comprises the majority of the neighborhood.

6. Conclusion:

From the results it can be concluded that majority of the neighborhoods in Michigan might not be an optimum location to open a restaurant. Therefore, a careful consideration is needed when selecting a neighborhood to open a new restaurant. From this analysis we found Livingstone to be the most optimum neighborhood to open a restaurant.