**Location to establish a Food truck in Kolkata**

Using K-means clustering to find the optimal neighborhood in Kolkata for opening a Food truck



1. **Introduction**
   1. **Background**

Kolkata, the city of Joy, is known for its exquisite culture that resonates from its love for art & literature. But when it comes to their cuisine, the versatility of it would leave one spoilt for choices. While the foodie can relish in the joys of it, being a [restaurateur](https://www.google.com/search?sxsrf=ALeKk01PR-NLf7T_xF38BDVw_F5xYJaNOQ:1590898902139&q=a+restaurateur&spell=1&sa=X&ved=2ahUKEwi41r2foN3pAhUAzTgGHdQHD5wQkeECKAB6BQgREJEB) in Kolkata is no mean task. The competition is immense considering the different palettes served in different parts of the city, not to mention the stiff local competition that exists.

For any new food business to setup, it would take immense research to find a suitable spot in the city to be profitable. Despite being a red ocean, it cannot be ignored that the ocean is big enough still for a brand to fit in, thanks to the food crazy people of the city.

This project helps to do an elementary analysis on the optimal neighborhood under KMC, where a food truck might start.

**Business Problem**

A client wants to set up a food truck in Kolkata. The business problem is to analyze the neighborhoods and cluster them such that a head start can be made in terms of shortlisting neighborhoods to park the food truck in.

* 1. **Interests**

A market survey spanning multiple areas and days would involve cost and time. This analysis can help the client getting a headstart into zeroing on the clusters where one can possibly look to start/expand the business.

Neighborhood data combined with restaurants details fetched from Foursquare API would help to cluster optimal neighborhoods. K-means clustering algorithm would be used for modeling the data.

1. **Data acquistion and Cleaning**
   1. Data Sources

Ward wise details were scrapped from Wikipedia for Kolkata Municipal Corporation. It included youth population below 6 years, literacy rate for the ward and the geographical co-ordinates for the same.

#### URL: <https://en.wikipedia.org/wiki/Category:Municipal_wards_of_Kolkata>

Foursquare API (explore function) was used to fetch the data for the ward for schools, tutorials in the ward.

The final dataset would look similar to this

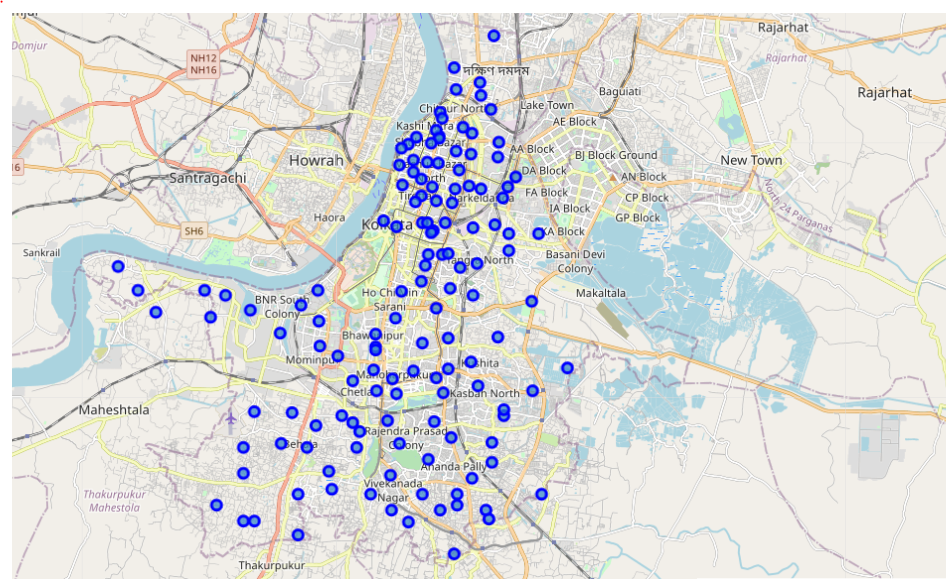
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ward No. | Area | Latitiude | Longitude | Pop. Below 6 years | Literacy rate | No. of Restaurants/Food outlets |
| 1 | Cossipore | 22.617889 | 88.370556 | 7.59% | 86.41% | XYZ |

* 1. Data Cleaning

One of the dependencies of the data scrapped was limitation of data fetched using Foursquare API. Many wards did not have any restaurant data and foursquare API restricts API calls to 500 for a free sandbox account. Therefore, only those wards were taken into consideration, where in restaurant details were available.

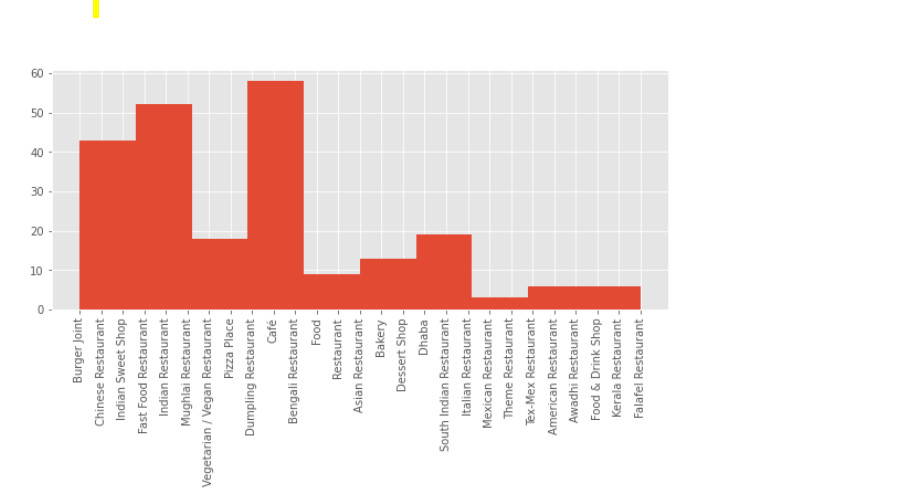
The literacy rate was converted to decimal for exploratory data analysis

Map of wards under KMC:



1. Exploratory Data analysis

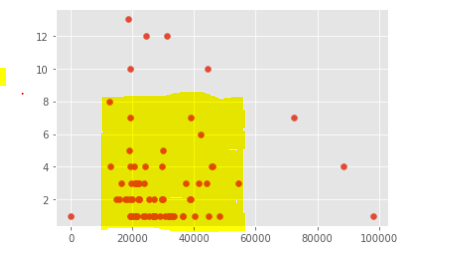
Distribution of different types of restaurants in the selected wards



The factors the have been taken into account while creating the model are:

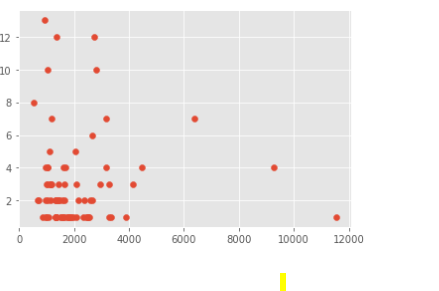
|  |  |
| --- | --- |
| Factor | Hypothesis |
| Total Population | More the Population, more chance of sales, better suited to open a Food truck |
| Literacy rate | As a proxy for Spending power i.e. more Spending power(literacy rate), more chance of sales, better to open a restaurant |
| No. of restaurants | More restaurants in the ward, more competition, less favourable to open |
| Pop. Below 6 years | More pop. below 6 years, more youth in future, more fast food sales growth in future |

**Scatter plot of total population vs No. of restaurants**



There seems to be little co-relation b/w Total populations vs No. of restaurants in the area with co-relation co- efficient of 0.23.

**Scatter plot of Population below 6 years vs No. of restaurants**

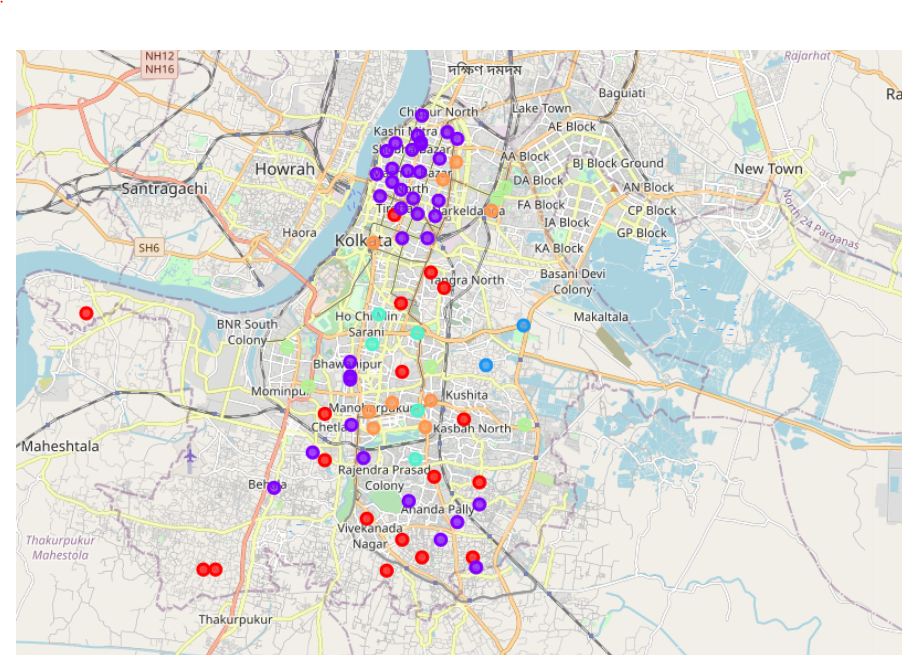


There seems to a linear relationship b/w population below 6 years vs. no. of restuarants for lower populations. Co-relation co-efficient is 0.09.

1. Predictive Modelling & Cluster analysis

K-means clustering with K=6 was used to cluster the neighborhoods

Here is map of the K-means modeled KMC wards



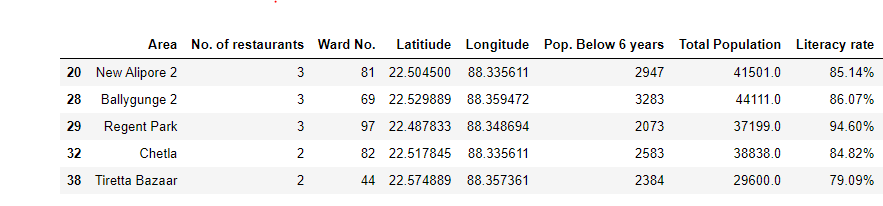
**Cluster 1**

Total Population – **MID**

Literacy rate – **MID [relatively to other sectors]**

No. of restaurants - **LOW**

Pop. Below 6 years – **LOW**



This is one of the optimal neighborhoods to open a food truck in as competition is low, average population size and literacy rate, which would mean decent household income.

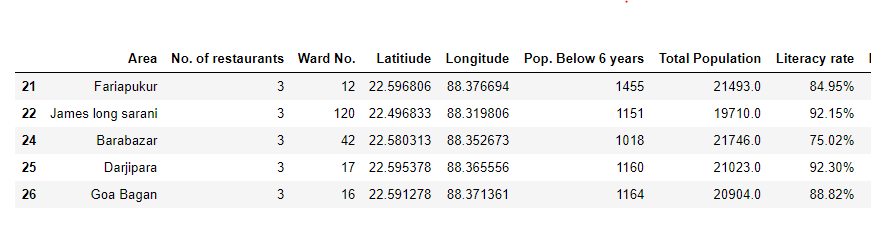
**Cluster 2**

Total Population – **LOW**

Literacy rate – **MID [relatively to other sectors]**

No. of restaurants - **LOW**

Pop. Below 6 years – **LOW**



This is not one of the clusters where one can comprehend opening a food truck owing to low population numbers and moderate spending power

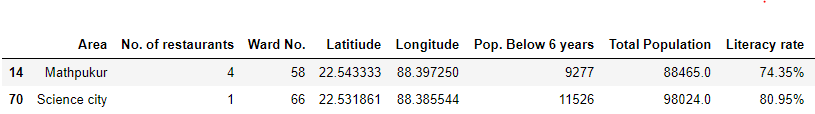
**Cluster 3**

Total Population – **HIGH**

Literacy rate – **LOW [relatively to other sectors]**

No. of restaurants - **LOW**

Pop. Below 6 years – **HIGH**



One of the more optimal neighborhoods with less competition and high population, although literacy rate i.e. spending power can be major deterrent.

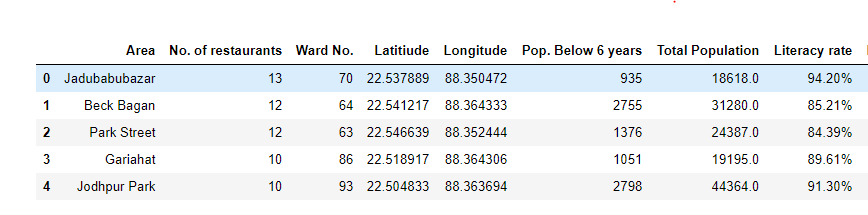
**Cluster 4**

Total Population – **LOW**

Literacy rate – **HIGH [relatively to other sectors]**

No. of restaurants – **HIGH**

Pop. Below 6 years – **LOW**



Extremely unfavorable neighborhood for opening a food truck with low population and high competition.

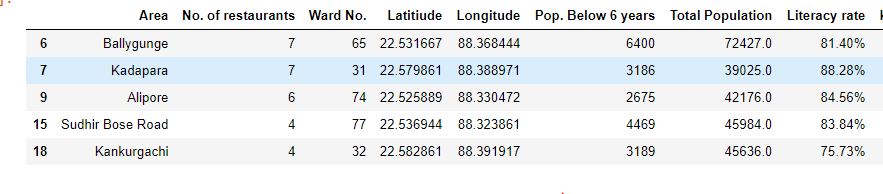
**Cluster 5**

Total Population – **HIGH**

Literacy rate – **MID [relatively to other sectors]**

No. of restaurants - **MID**

Pop. Below 6 years –  **HIGH**



High spending power and population are attractive factors for this cluster but competition also moderate. Therefore, it can be considered as one of the optimal clusters.

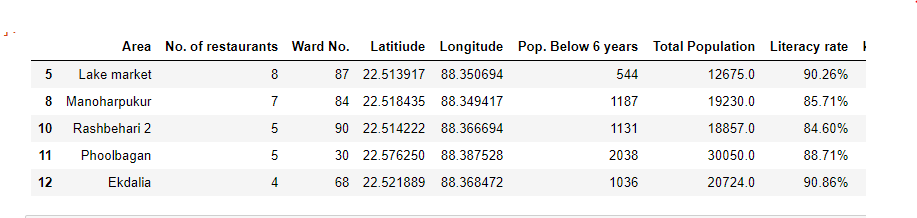
**Cluster 6**

Total Population – **LOW**

Literacy rate – **MID [relatively to other sectors]**

No. of restaurants - **HIGH**

Pop. Below 6 years – **MID**

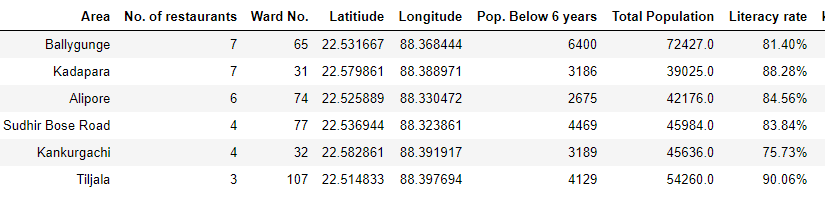


Not a favorable cluster with low population, high competition and moderate spending power.

1. Conclusions

Cluster 1, 3 and 5 are the favorable clusters for opening a food truck. On close analysis, it could be seen that cluster 5 could be the ideal neighborhoods for opening a food truck with high population, moderate competition and spending power.

Neighborhoods ideal for opening a food truck



1. Future Directions
   1. The model has lot of short comings in terms of lack of restaurant data for a lot wards, right predictor variables with literacy rate put as a proxy etc
   2. Further data addition would lead to refinement of the clusters.
   3. Analysis of K elbow Visualizer and squared error can be done for choosing the best K. The inclusion of such has been kept out of scope of the project.
2. References

#### <https://en.wikipedia.org/wiki/Category:Municipal_wards_of_Kolkata>