**Location to establish a Tutorial/Hobby center In Kolkata**

Using K-means clustering to find the optimal neighborhood in Kolkata for opening a Tutorial/Hobby center for Kids



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

India has been touted as the fastest growing nation for quite some time now. A significant reason is its demography, where youth population (ranging between 15-24 years) constitutes ~34% of its huge ~135 crore population. With increasing number of young crowd and rising household income, the demand for quality education has increased manifolds.

Indian education industry is estimated to be b/w 20000-25000 crores with roughly 367 universities & 18000 colleges. In addition to the mainstream sector, there exists a support eco system for this industry, which constitutes of the Tutorial/hobby centers which focuses on a student’s auxiliary skills to become employable. This would include extra-curricular centers, Foreign language classes, personality development centers, course tutors etc.

Nowadays, children are pushed to these centers at a young age so that, cometh the moment, they are expected to excel at the rat race. This project revolves around these centers, geographically focused in Kolkata (coz that’s my home town) and analyzes the optimal neighborhood for opening such centers.

* 1. **Business Problem**

A client seeks to open a tutorial/hobby center in Kolkata, India. Which neighborhood would be ideal to open for such a center?

Our reasoning is based on Literacy rate of the neighborhood, young population below 6 years and the competition in the area. We will be using Foursquare API along with census data to analyze the results.

* 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 possible look to start/expand the business.

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.

Foursquare API 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 tutorials |
| 1 | Cossipore | 22.617889 | 88.370556 | 7.59% | 86.41% | XYZ |

* 1. Data Cleaning

1. Exploratory Data analysis
2. Predictive Modelling
3. Cluster Analysis
4. Conclusions
5. Future Directions
6. References
   1. <http://www.investinindia.com/industry/education/education-industry>