**Data Science Capstone Project:**

**K-Means Clustering to predict optimal Neighborhood for Construction of Residential Estate**

By

Karan Bhowmick

INTRODUCTION:

Mumbai is the Financial Capital of India. With a population of 20.4 million, with India’s ever-increasing life rate overpopulation has become a cause for concern. In a world pestered by overpopulation, availability of accommodation becomes the prime concern. To capitalize on the ever-increasing demand of Residential Areas. Real Estate has been dubbed the one of the best investment options. This Project tackles the issue of shortlisting a neighborhood with plentiful amenities, so as to receive the best prices on Residential Buildings/Estates. The location of a Residential Estate, is the deciding factor on Profitability and Demand of ownership. Thus, it is of utmost importance that the neighborhood chosen can augment returns.

Data Science and Machine Learning are two avenues that are paving the way for corporations worldwide. Predictive Analytics is taking over all departments in the industry. From Medicine to Finance, data is being used to gain insights into market trends. It is the need of the hour to make data-driven decisions to model a business.

Using Data Analytics and Machine Learning for Predictive Analytics, we can help streamline the decision-making process.

DATA ENGINEERING:

Data Requirements:

* Neighborhood Name
* Latitude Longitude of the Location
* Facilities around the neighborhood
* Parks, Cafes, Shops etc.

Data Procurement:

We can use BeautifulSoup Library in Python to extract the Neighborhood names from a trusted source. I used Wikipedia to scrape data on the neighborhoods in the Mumbai region of Maharashtra.

To scrape data regarding points of interest in these neighborhoods. Latitude, Longitude, Name and Category of these POIs can be extracted using Foursquare API.