

Description of the Data to be Used in the Project

There is a various amount of infrastructure to be considered in Mumbai.

For this project, the following data is required.

- Mumbai pin code (scraped from web source)- Source is Mumbai7.com, contains the list pin codes and postal office names.
- For the kinds of infrastructure in each neighborhood, Foursquare API will be used. The venues can be easily filtered therefore, and it will be easy to access the various places.
- Geospace data is needed lastly, to get the latitudes and longitudes based on the postal offices in Mumbai.

This project will make use of skills in data science such as web scraping from Mumbai7.com, working with location data from the Foursquare API, data cleaning, wrangling and using Folium maps to visualize the data. K-means clustering will also be used to cluster up the good infrastructure locations.