

# Data Acquisition

For any “data science project” data is of paramount importance. For this study, we needed data about neighborhoods in each of these metro cities. The data published by the government on postal codes for all India would serve us well for this study.

We will specifically download the CSV provided under <https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude>.

The data is separated into two csv:

1. Being the Mumbai data which shows us the Pin Code, Place to Visit and Location

	Place	Pincode	Location
0	Antop Hill	400037	Mumbai
1	B P T Colony	400037	Mumbai
2	B.P.Lane	400003	Mumbai
3	BEST Staff Quarters	400012	Mumbai
4	C G S Colony	400037	Mumbai

2. Being the Bangalore data which also shows us the Pin Code, Place to Visit and Location

	Place	Pincode	Location
0	A F Station Yelahanka	560063	Bangalore
1	Agram	560007	Bangalore
2	Air Force Hospital	560007	Bangalore
3	Amruthahalli	560092	Bangalore
4	Anandnagar	560024	Bangalore

In this study, we will download the CSV, read it into a pandas Dataframe and curate it to remove the data related to all other cities, towns, and places which are not Mumbai or Bangalore, since we are only interested in comparing these two biggest metro cities in India.

# Data Cleaning

After acquiring the two data, the grouping is done based on the Pin Code for each place such that no repeated data occurs for the same area having the same Pin Code. And we also, with the help of geocoders, find the longitude and latitude of the places to visit in the cities, aka with the help of **Google Maps API geocoding** to come at the final datasets.

	Pincode	Location	Place	Latitude	Longitude
0	400001	Mumbai	Bazargate , Elephanta Caves, M.P.T. , Stock Ex...	18.939031	72.837345
1	400002	Mumbai	Kalbadevi , Ramwadi , S. C. Court , Thakurdwar	18.947905	72.826741
2	400003	Mumbai	B.P.Lane , Mandvi , Masjid , Null Bazar	18.952760	72.834560
3	400004	Mumbai	Ambewadi , Charni Road , Chaupati , Girgaon ,...	18.954380	72.821556
4	400005	Mumbai	Asvini , Colaba Bazar , Colaba , Holiday Camp ...	18.915923	72.817565

*Final Mumbai Data*

	Pincode	Location	Place	Latitude	Longitude
0	560001	Bangalore	Bangalore Bazaar , CMM Court Complex , Dr. Amb...	12.979185	77.606623
1	560002	Bangalore	Bangalore Corporation Building , Bangalore City	12.964070	77.577647
2	560003	Bangalore	Malleswaram , Palace Guttahalli , Swimming Poo...	13.003656	77.569745
3	560004	Bangalore	Basavanagudi , Mavalli , Pampamahakavi Road , ...	12.945664	77.575075
4	560005	Bangalore	Fraser Town	12.998115	77.620842

*Final Bangalore Data*

The new dataset is used to generate the venues for each place using the **Foursquare API**.