

Data Source and Cleaning

The data for this Project is mainly acquired from the following link

<https://simplemaps.com/data/in-cities>

It provides a list of 213 prominent cities in India. Each row includes a city's latitude, longitude, state and other variables of interest such as population. This is a subset of all 478,635 places in India that you'll find in our World Cities Database. They have released this data subset for free under the MIT license. So because of this, we are free to use it for personal or commercial applications. And it can easily be downloaded in .csv, .json and .xlsx format directly from the website. A brief sample of the data is attached below.

city	lat	lng	country	iso2	admin	capital	population	population_proper
Mumbai	18.987807	72.836447	India	IN	Mahārāshtra	admin	18978000	12691836
Delhi	28.651952	77.231495	India	IN	Delhi	admin	15926000	7633213
Kolkata	22.562627	88.363044	India	IN	West Bengal	admin	14787000	4631392
Chennai	13.084622	80.248357	India	IN	Tamil Nādu	admin	7163000	4328063
Bengalūru	12.977063	77.587106	India	IN	Karnātake	admin	6787000	5104047
Hyderabad	17.384052	78.456355	India	IN	Andhra Pradesh	admin	6376000	3597816
Ahmadābād	23.025793	72.587265	India	IN	Gujarāt	minor	5375000	3719710
Hāora	22.576882	88.318566	India	IN	West Bengal		4841638	1027672
Pune	18.513271	73.849852	India	IN	Mahārāshtra		4672000	2935744
Sūrat	21.195944	72.830232	India	IN	Gujarāt		3842000	2894504
Mardānpur	26.430066	80.267176	India	IN	Uttar Pradesh		3162000	2823249
Rāmpura	26.884682	75.789336	India	IN	Rājasthān		2917000	2711758
Lucknow	26.839281	80.923133	India	IN	Uttar Pradesh	admin	2695000	2472011
Nāra	21.203096	79.089284	India	IN	Mahārāshtra		2454000	2228018
Patna	25.615379	85.101027	India	IN	Bihār	admin	2158000	1599920
Indore	22.717736	75.85859	India	IN	Madhya Pradesh		2026000	1837041
Vadodara	22.299405	73.208119	India	IN	Gujarāt		1756000	1409476
Bhopal	23.254688	77.402892	India	IN	Madhya Pradesh	admin	1727000	1599914
Coimbatore	11.005547	76.966122	India	IN	Tamil Nādu		1696000	959823
Ludhiāna	30.912042	75.853789	India	IN	Punjab		1649000	1545368

Since the data is mostly in a proper data frame format, there is not much cleaning to do on this data. I have cleared some of the unwanted columns from the table such as Country as it is obvious that we are analysis the country India, iso2 which is another short form of the country, capital and population proper.

We will then use the lat, log data of different cities and use it to the Foursquare API and get details about the venues which are surrounding the cities with in 10 km. Then those data which we acquire from the Foursquare will be merged with this data and we will make a detailed Exploratory Data Analysis to address our problem.