

The Battle of the Neighborhoods - Chennai

by,

Lingesh R

2. Data

In this project we will utilize three datasets in order to solve our problem. The first dataset is a list of municipalities. This data set was scraped from Wikipedia using BeautifulSoup. The next data set is one contain the latitude and longitude for each municipality. This dataset was obtaining using Geocoder package. The file containing the list of venues was obtained using the Foursquare API. We selected all venues not just restaurants in order to see where restaurants were more popular. These are the top 5 rows of these data sets.

Neighborhood data:

Neighbourhood	
0	Red Hills
1	Royapuram
2	Korukkupet
3	Vyasarpadi
4	Tondiarpet

Neighborhood with latitude and longitude coordinates:

	Neighbourhood	Latitude	Longitude
0	Red Hills	13.19543	80.18431
1	Royapuram	13.11394	80.29420
2	Korukkupet	13.11680	80.27726
3	Vyasarpadi	13.11778	80.25168
4	Tondiarpet	13.12923	80.28955

Combined with Foursquare API data:

	Municipalities	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Red Hills	13.19543	80.18431	Hotel Balaji Bavan	13.193716	80.185292	Indian Restaurant
1	Red Hills	13.19543	80.18431	Pvr Cinemas Skls Galaxy Mall	13.191586	80.186100	Movie Theater
2	Red Hills	13.19543	80.18431	Skls Galaxy Mall	13.191500	80.186038	Shopping Mall
3	Red Hills	13.19543	80.18431	Lakshmi Theatre	13.189894	80.188758	Multiplex
4	Red Hills	13.19543	80.18431	Universal Gym	13.190710	80.177946	Gym

Some of the features extracted from the data were, the latitude and longitude of each municipality, and from the foursquare API we extracted the venues, the venue category, and its geo-coordinates. All of this data can be used to help us in our analysis and problem solving, which will be discussed more thoroughly in the methodology and analysis sections.