The Battle of Neighborhoods - Week – 1

Capstone Project by Mathan Kumar M

Introduction and Data Collection:

ABC Petrol Company wants to launch a Petrol pump in some locality in Chennai city. The company approaches me to suggest the best location for the new petrol pump so that the profitability can be increased.

So, for this problem, we must collect the geo-coordinates of the neighborhood localities in the city of Chennai and as well as the geo-coordinates of Offices, other petrol pumps and the roads (meeting junctions) where the motion of vehicles is quite more.

Data Section:

The following table has the details of the datasets used for the given problem.

S.No	File Name	Description
1.	localities_ coords	This file contains the name, latitude and longitude (geo-coordinates) of the localities in and around the city of Chennai.
2.	chief_centers	This comma separated valued file has four columns: a. Name of the which is considered to be one of the chief centers in Chennai city, having chief petrol consumers b. Latitude of the place c. Longitude of the place d. Type of place – It contains the label of the place of which category it belongs to (example: School, Office, Petrol Pump, Others)
3.	transport_ coords	This file contains the latitude, longitude and the traffic intensity (based on the motion of vehicles in the road) which ranges from 0-1 where: 0 – No traffic 1 – Full traffic

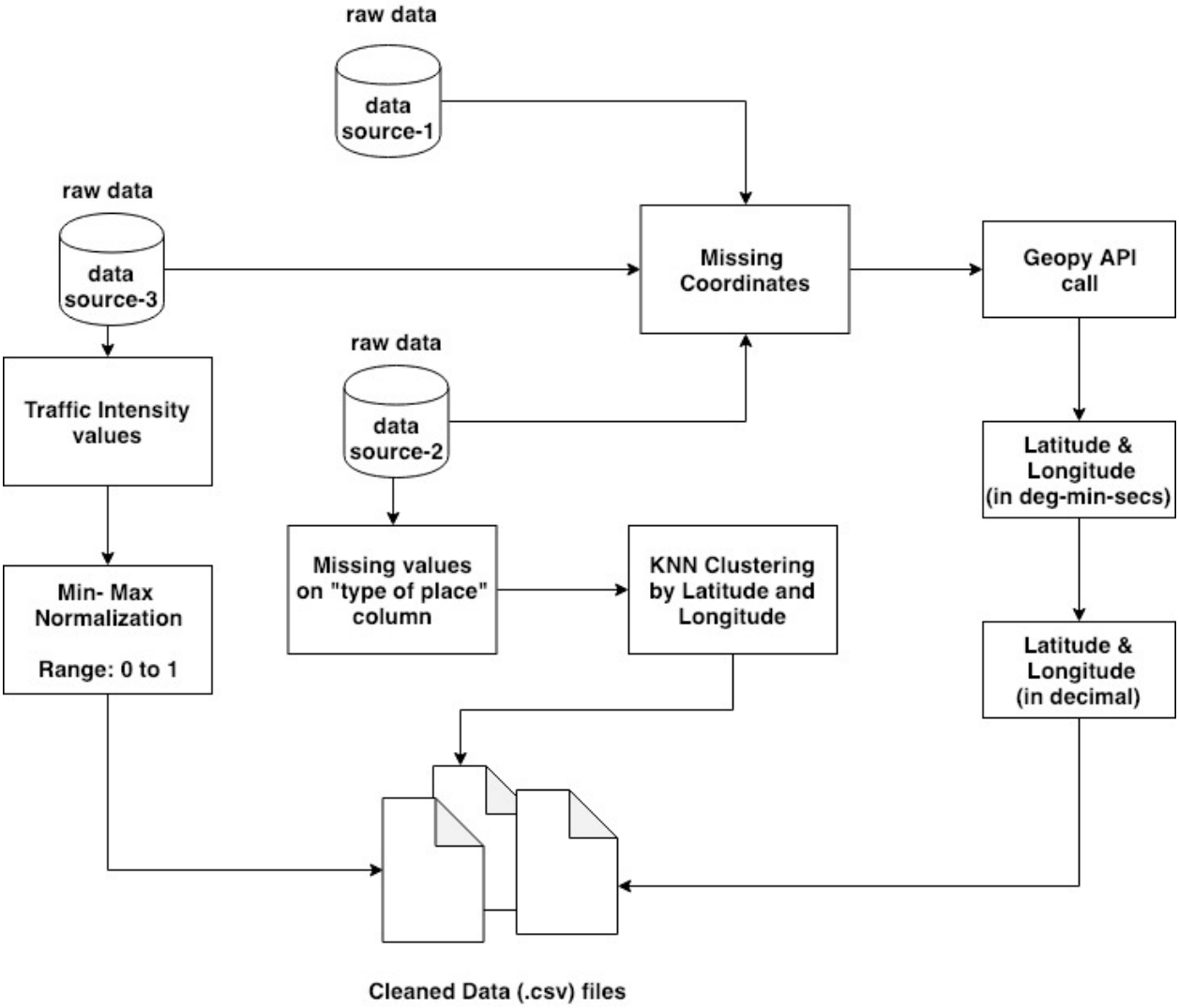
Data Sources:

1. https://github.com/mathan-madhav/coursera/blob/master/data_localities_coords.csv
2. https://github.com/mathan-madhav/coursera/blob/master/data_chief_centers.csv
3. https://github.com/mathan-madhav/coursera/blob/master/data_transport_coords.csv

Data Pre-processing:

This involves cleaning, normalizing and processing of missing data values to convert incomplete date to complete data.

The pre-processing step for the given problem is explained in the following diagram:



The final processed data is found in .csv format which can be imported as python data frames for processing and analysing.

Feature Extraction from Data:

The following features can be extracted from the data sources to solve the problem.

Feature Name	Weight Assignment	
Type of the place	Office	+1
	Schools	+0.5
	Petrol Pumps	-1
	Others (like Shopping Malls)	+0.5
Traffic Intensity	0.0 to 0.25	-1
	0.25 to 0.5	-0.5
	0.5 to 0.75	+0.5
	0.75 to 1.0	+1

Negative values indicate that the least likelihood of the locality to open the petrol pump whereas the positive values are more likelihood values for opening the petrol pumps in that particular locality.