

The Battle of Neighborhoods - Report

Capstone Project by Mathan Kumar M

1. Introduction and Business Problem:

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

1.1 Problem Background:

To suggest the best locality for ABC Petrol company, we need to gather the localities of Chennai city and analyze the places where target customers are high in numbers so that the company can gain more profit.

1.2 Problem Description:

In order to find the suitable place to launch a petrol pump we need to analyze the neighborhood localities in Chennai city and apply the following logics.

- The primary logic is to identify the roads connecting highways and localities with higher number of offices and buildings.
- The secondary logic is to find the locations of other petrol pumps and the locality with sparse number of petrol pumps will have higher opportunity to grab the attention of petrol consumers.

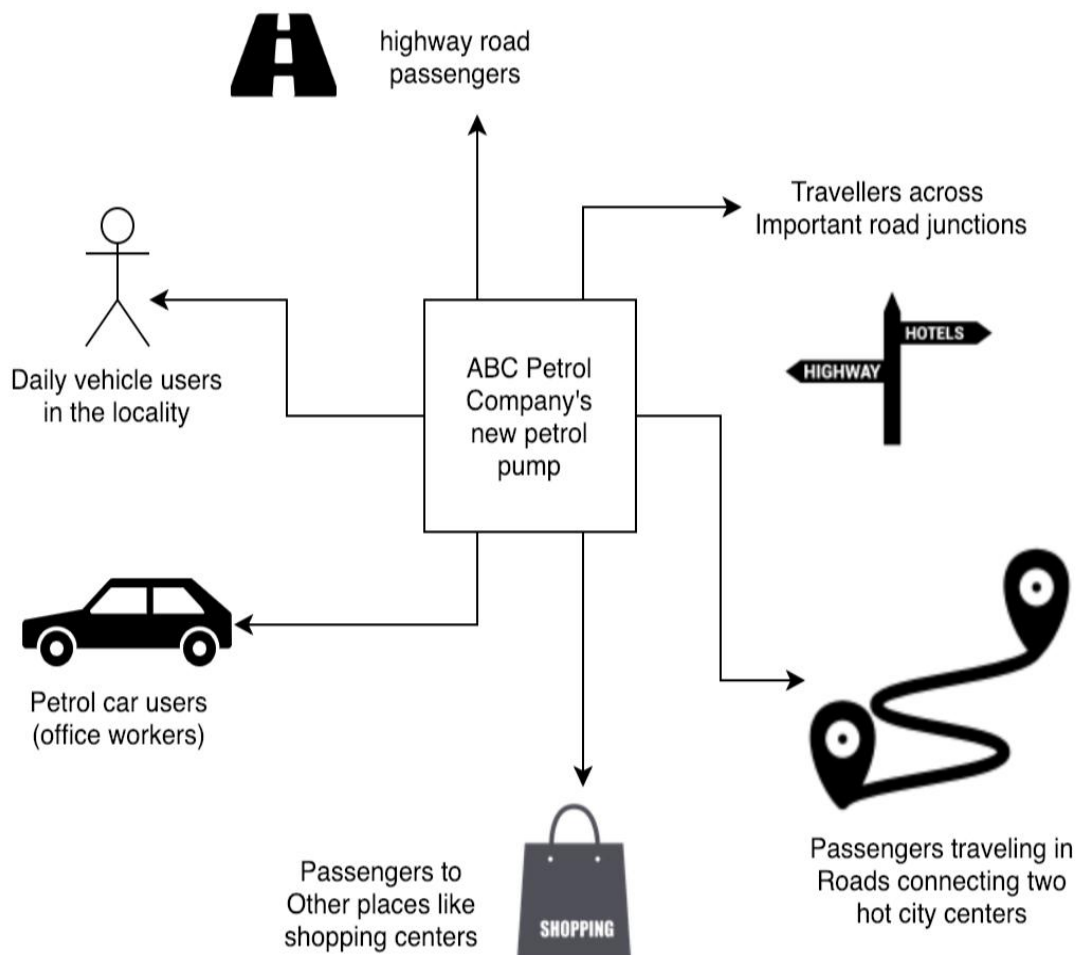
1.3 As a Business Strategic Problem:

From business point of view, the problem is to find a best location for the ABC Petrol Company to open its new petrol pump in Chennai city such that:

- the suppliers or distributors are nearby
- it must be known centre for the petrol pump services
- many of the people who work or live in the area are its business' target audience
- businesses in the area complement the petrol pump (for example shopping malls nearby)
- the costs of buying or leasing in the area are affordable and meet the needs of the business
- it must be a growing business hub with many opportunities in the near future.

1.4 Target Audience:

The Target audience are the daily vehicle users (cars, two-wheelers, and other vehicle users) who will depend on petrol fuel and petrol pumps for fuel source of their vehicles. This will help the ABC petrol company to gain profit.



1.5 Success Criteria:

Success criteria for this project would be a better recommendation for the ABC Company to open one of its branch in the suggested locality.

2. Data Section:

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.

2.1. Data description:

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:

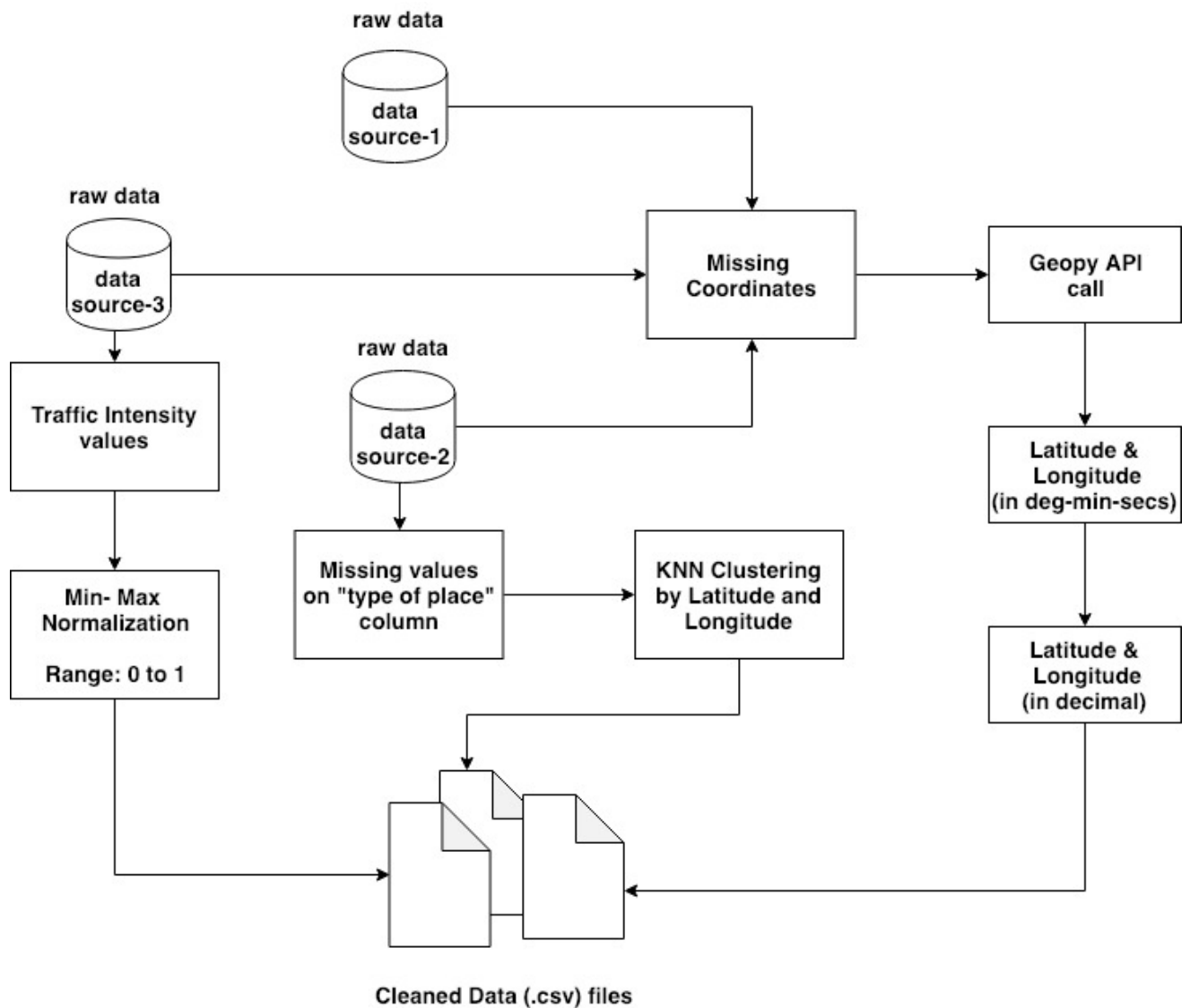
1. https://github.com/mathan-madhav/coursera/blob/master/data_localities_coords.csv
2. https://github.com/mathan-madhav/coursera/blob/master/data_chief_centers.csv
3. https://github.com/mathan-madhav/coursera/blob/master/data_transport_coords.csv

[These are the data sources after step 2.2 data pre-processing]

2.2. Data Pre-processing:

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

2.3. 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.

3. Methodology:

The methodology section contains two parallel pipelines from the cleaned data that we get from the Data cleaning section.

First pipeline:

From the cleaned data, the latitudes and longitudes (co-ordinates) of the localities are extracted. Using foursquare API call, the nearby locations are found and the details about the city centers (hot-spots) are identified.

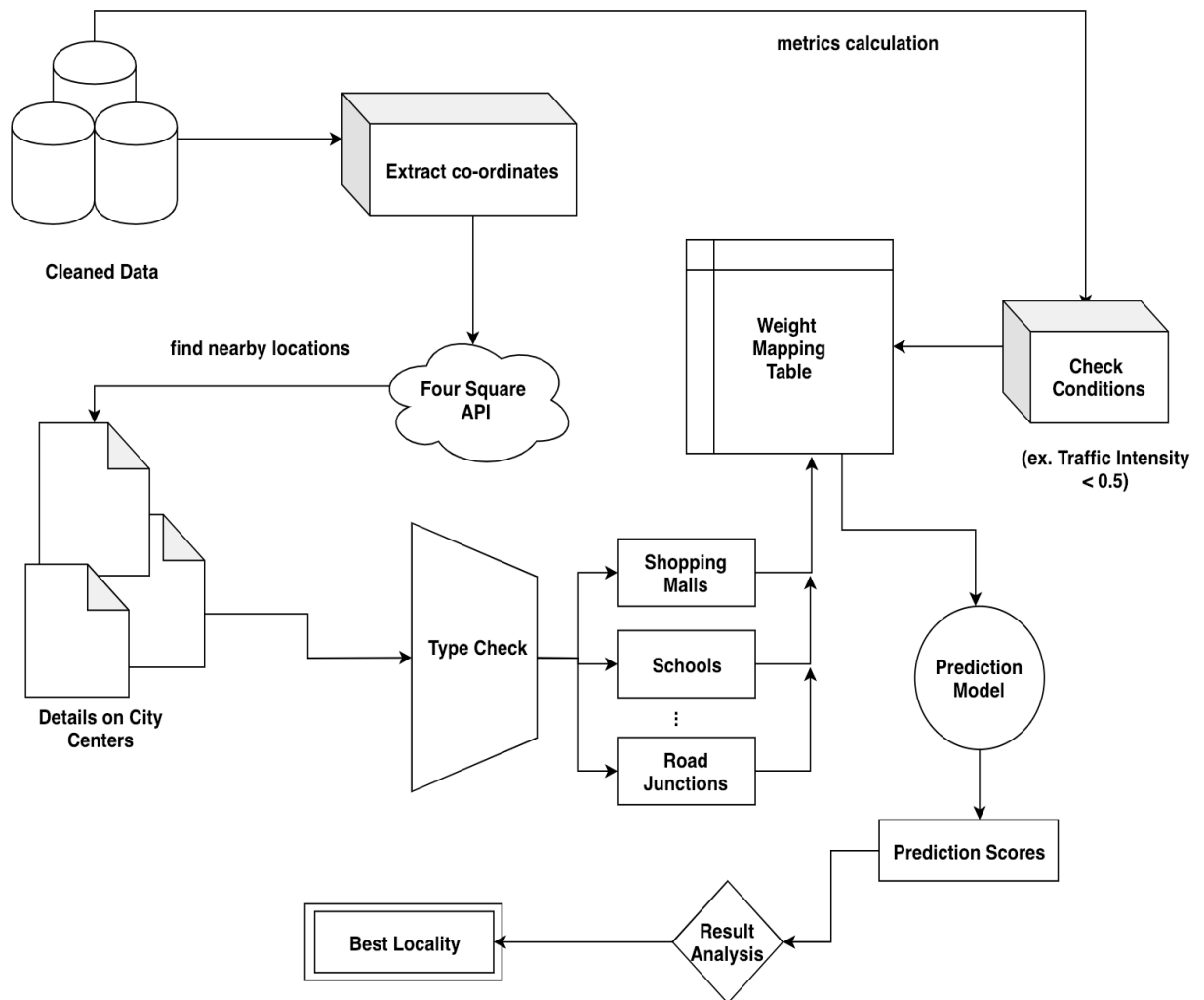
Using the identified city centers, the simple classifier (type-checker) checks for the type of city center. Few of the output labels are Shopping Malls, Schools, Road Junctions or others. For each type of the city center, a weight is associated using the **Weight Mapping table**.

The simple probabilistic prediction model predicts the features for every locality and emits the **prediction scores** (in range of 0-1)

Second pipeline:

This is a metrics calculation pipeline where features from the cleaned data like traffic intensity, locality neighborhood are extracted and rules are framed according to the weight assignment logic.

Correspondingly, the weight mapping table is updated with respect to the rules of weight assignment.



Result analysis is done by filtering the top-3 localities with highest prediction score and one among them can be treated as the best locality to launch a petrol pump by the ABC Petrol Company.

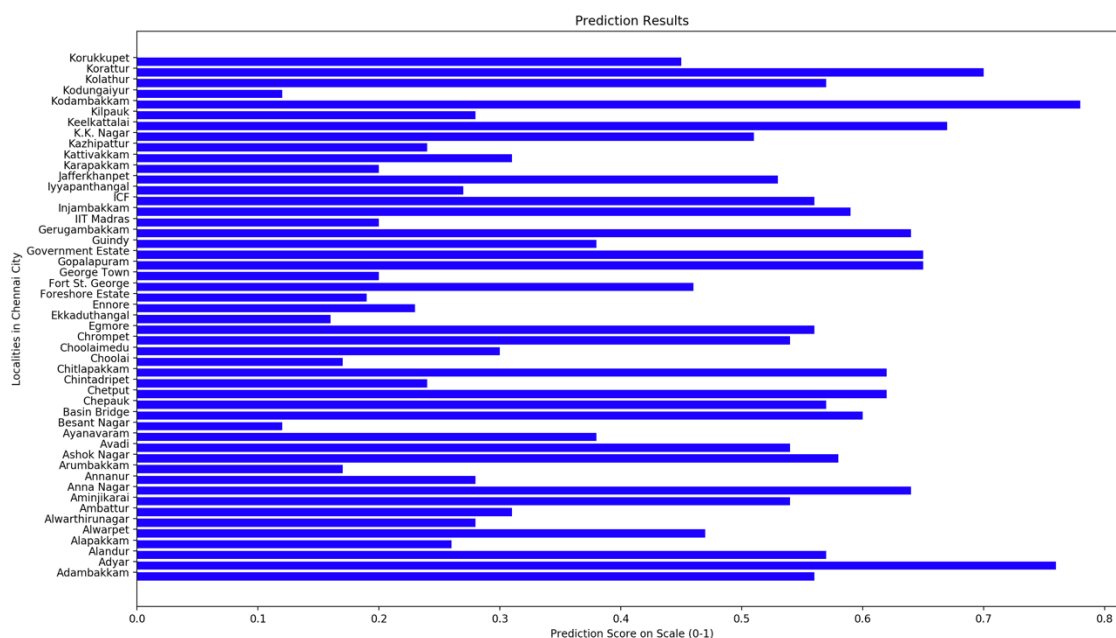
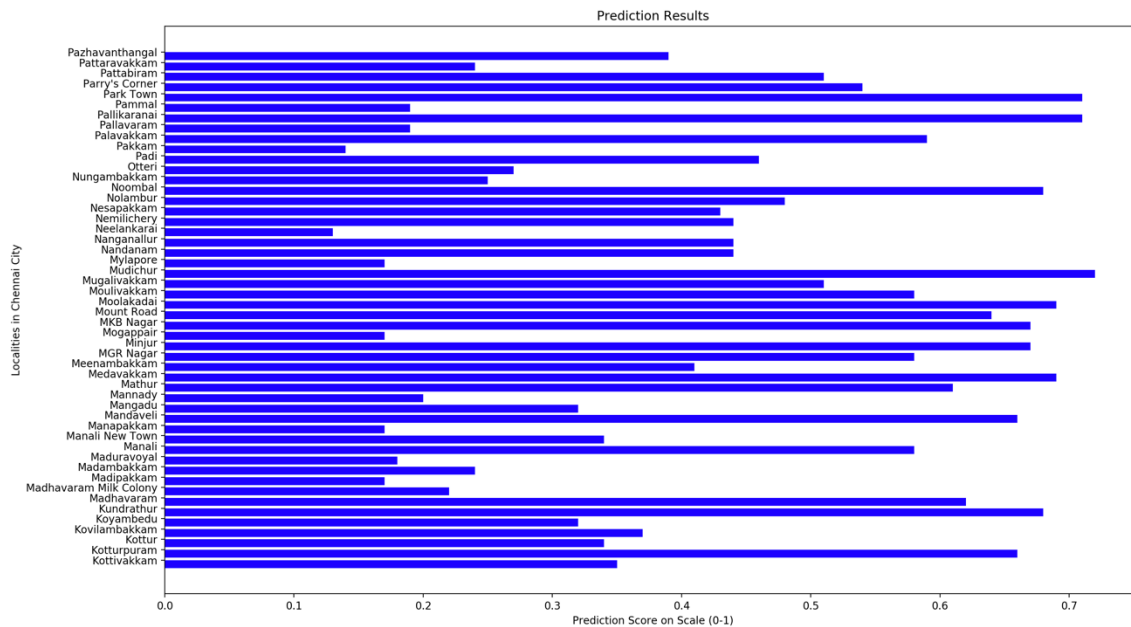
4. Discussion Section:

The following recommendations can be included in addition to the existing methodology:

- The details of the **distance between** different city centers can be taken into account so that the petrol pump can be opened in the midst of any 2 centers with higher number of customers
- Other secondary centers like **Car Mechanic Shops, Automobile Repair Shops** can be accounted based on the assumption that they may visit the petrol pump after repair.
- Thirdly, the **population density** can be used. Lesser the density lower the chance of opening a petrol pump because the customers will not prefer to move from dense to sparsely populated region to fill petrol.

5. Results Section:

Following are the prediction results (prediction scores – for each locality) obtained by implementing our prediction model following the proposed techniques in the methodology section:



The graph shows the plot between the locality and the prediction score on scale of 0 to 1. It can be noted that the prediction score varies from 0.13 (minimum prediction score) to as high as 0.8 (maximum prediction score).

6. Conclusion Section:

- Based on the prediction model and the methodology used, we come to a conclusion that opening a brand new petrol pump in the (1) **St.Thomas Mount Locality** (prediction score = 0.81) has the highest chance of making profit for the ABC company.
- It is based on the fact that the chosen locality is populated with a lots of petrol consumers and it comprises various popular city centers like Shopping Malls, Road Junctions and have high traffic intensity in and around the locality.
- Also, we can nominate the other equivalently eligible localities **(2) Kodambakkam and (3) Adyar** with same approximately prediction scores (0.78) and (0.76) respectively.
- The Business team of ABC company can analyse among the top 3 localities with high prediction scores and they can come up with a locality to launch their new petrol pump.