Computational Complexity – MGMT590-051

Trucking Data Analysis Using Techniques Learnt in MGMT590-0051

Professor Name: Shai Vardi

Name: Amaan Ansari, Aditya Roy Choudhary

1.) **Dataset Overview**

The dataset consists of the travel routes for an India trucking company. It consists of different columns which contain information about journey on each route, such as origin location, destination location, longitude and Latitudes of origin and destination, type of vehicle etc. Our main goal is to implement what we have learned over the period of 8 weeks in context of a real-life project to provide insights and decision making for the client. We have tried to achieve it in multiple ways as you will read in Section 2.

```
['GpsProvider', 'BookingID', 'Market/Regular ', 'BookingID_Date',
 'vehicle_no', 'Origin_Location', 'Destination_Location', 'Org_lat_lon',
 'Des_lat_lon', 'Data_Ping_time', 'Planned_ETA', 'Current_Location',
 'DestinationLocation', 'actual_eta', 'Curr_lat', 'Curr_lon', 'ontime',
 'delay', 'OriginLocation_Code', 'DestinationLocation_Code',
 'trip_start_date', 'trip_end_date', 'TRANSPORTATION_DISTANCE_IN_KM',
 'vehicleType', 'Minimum_kms_to_be_covered_in_a_day', 'Driver_Name',
 'Driver_MobileNo', 'customerID', 'customerNameCode', 'supplierID',
 'supplierNameCode', 'Material Shipped', 'Curr_lat_lon'],
```

Figure 1. Columns in the Data Set

2.) **What were we trying to achieve?**

Our goal with this project was simple, to use the algorithms and techniques that we have learned to drive the business insights, provide automation, and assist Decision making.

➔ We started out with taking the raw data and cleaning it where we could have it in a useable format. We were able to derive some insights using exploratory data analysis with this.

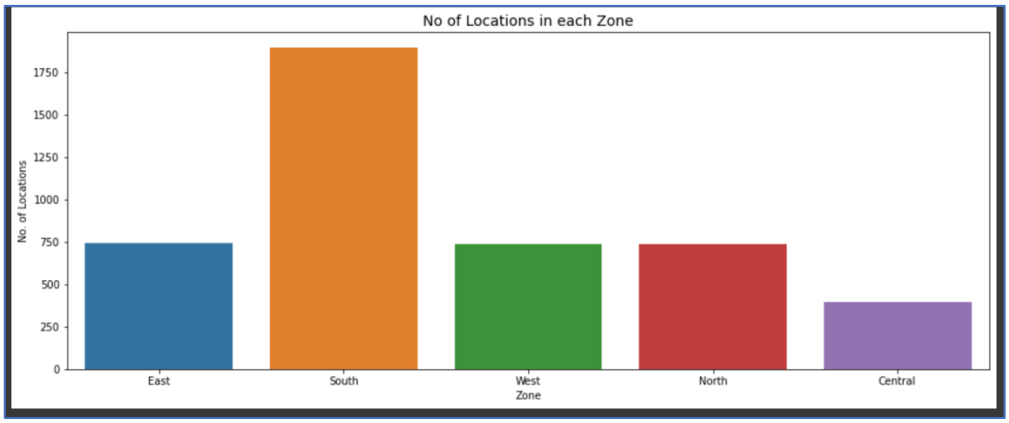


Figure 2. No of Origin/Destination - Zone Wise

➔ Post this we started clustering the data into different regions, this was done as India is a big nation and managing the complete fleet from a single position would cause problems for the company as India has different terrains and weathers across the

regions. So, what issues and services might occur/be required will vary vastly across these areas. We have divided the complete destination and origin data into 5 clusters based on geography of India. This would also help company in identifying the common issues and problems and will be able to tackle them in a more structured format. We have also superimposed it over the map of India so as for the easier for the company to visualize the distribution of its truck routes as well as newly formed clusters.

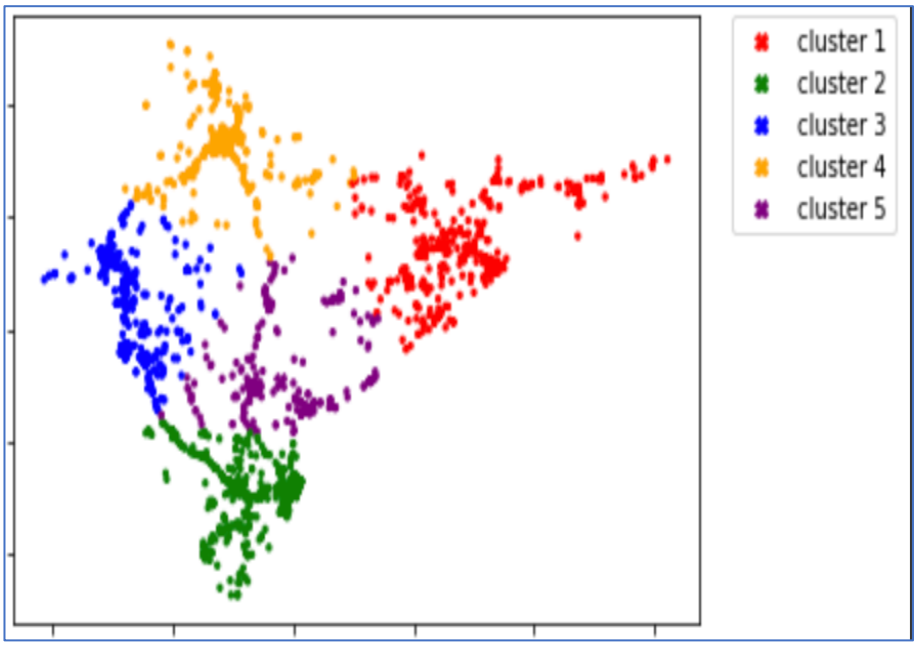


Figure 3. Final Clustering Results

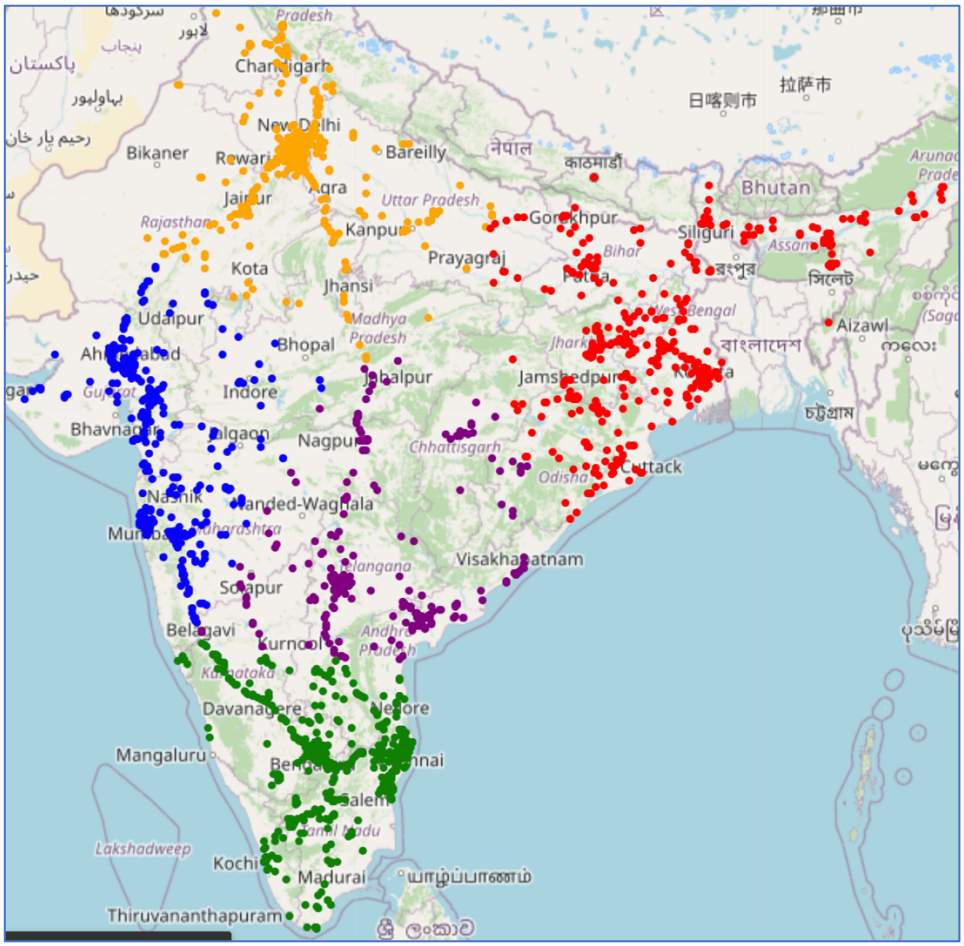


Figure 4. Superimposition of clusters on Indian Map

➔ The last step was great success in terms of breaking the company down into smaller divisions that could handle each location well but where would a newly added center go? To counter this confusion and eliminate the need for clustering again in the future. Which would change the clusters of previously assigned location as well (creating a lot of confusion to each location and hence a lot of volatility). We proceeded to label the Zones onto the existing dataset and perform an ML model to predict and classify any of the newly added location into its respective Zone. As soon as a new center's longitude and latitude are entered, the system will itself classify it as part of new cluster without disturbing any of the older centers.

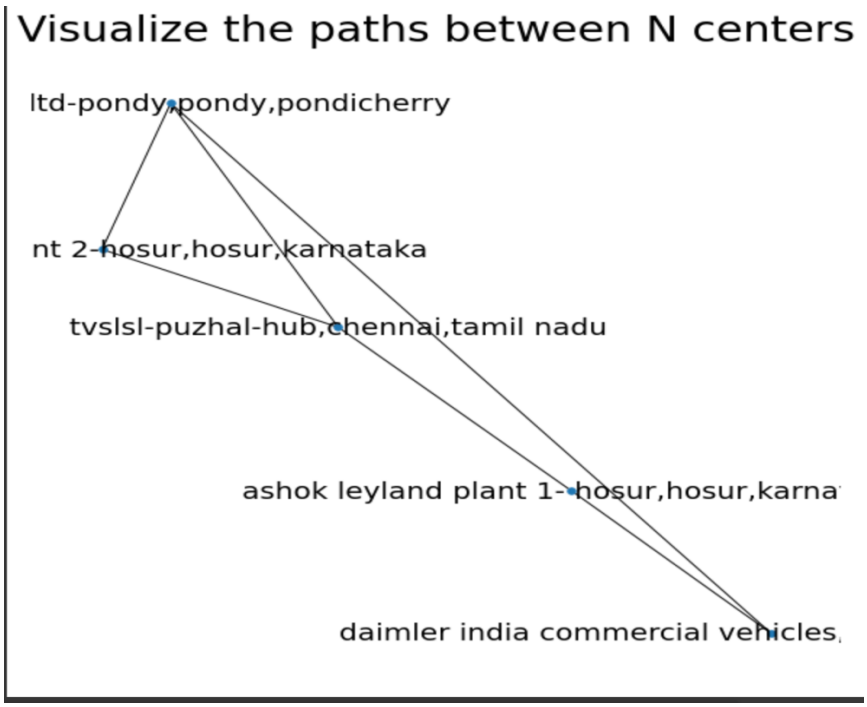
```
Enter the Latitude       30
Enter the Longitude      80
Coordinates:     [30. 80.]        Zones:  North
```

Figure 5. Classifying new location in Zones

➔ We have used Graph theory to an extensive limit to help the truck drivers identify the routes that they can take for a shorter trip from point A to point B when there's no direct trucking route between those two places. We have also used graph theory to visualize the trucking routes between any N number of stops that the company would want to see. This will help them in multiple ways, from keeping track of their trucks to identifying new routes in between the cities and how they can club the cities together based on average load/capacity of trucks. Such as instead of sending 2 small trucks to 2 cities nearby or who fall on the same route, they can send 1 medium sized truck to cater both pitstops.

```
Path taken from Source to Destination in the given graph
singaperumalkoil, kanchipuram, tamil nadu---> gopanapalli, krishnagiri, tamil nadu
---> mookandapalli, krishnagiri, tamil nadu
---> athipattu, tiruvallur, tamil nadu
```

Figure 6. Path to be Taken between two cities



➔ We have also had our spin on Classic Travelling Salesman problem and used it as Travelling Inspector Problem, In case an inspector wants to visit multiple cities what will be the shortest distance he would need to take for this.

```
[ ]  TSP_DP(arq, 1)

     16.204412264600315
```

Figure 7. Minimum Hamiltonian Cycle/TSP

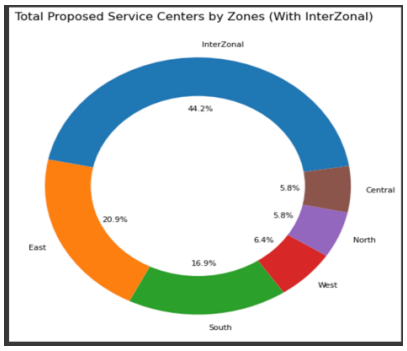
With this we could infer that the minimum distance an inspector would have to travel is about 16.2044 units

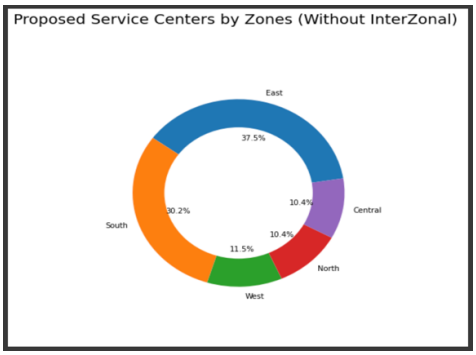
*(**Assumption**: Since our dataset did not have distances between the locations, we had to compute our distances based on the Euclidean Distance between the Coordinates of the different locations)*

➔ Finally, we have implemented a Linear Programming method to identify what will be the ideal number of Service Centers/Franchisee/Booking centers for our client that will help cover the region with minimum number of centers and minimum cost to reach each place.

```
Number of Centers required accross the country and by Zone
Nationwide:      136
East Zone:       36
South Zone:      29
West Zone:       11
North Zone:      10
Central Zone:    10
Inter Zonal:     76
```

Figure 8. No of Suggested Service Centres





3.) **What techniques did we use?**

➔ **K-Means Clustering: (Machine Learning)** We used K-Means clustering to identify the initial clusters for the company to divide the different destination/origin points based on the previous data to create 5 Zones for the company

➔ **Support Vector Machines: (Machine Learning)** We have used SVMs to map out the already zoned boundaries and classify any newly formed travel origin/destination into pre-existing clusters

➔ **Bellman Ford Algorithm: (Graph Theory)** We have used bellman form algorithm to suggest drivers the shortest trucking route that he can take between two cities that do not have any direct trucking routes.

➔ **Sub-Graph Creation: (Graph Theory)** We have created on demand weighted subgraphs to help company understand how the current trucking route network between X number of cities of their choice looks like.

➔ **Dynamic Programming:** We used dynamic programming to Implement Travelling salesman problem using Dynamic Programming. As it is computationally better than regular brute force algorithm when the number nodes is greater than 10.

➔ **Minimum Vertex Cover: (Linear Programming)** We used linear programming to identify the locations for proposing new Service/Booking Centers using minimum cover vortex.

4.) **What more could have been done?**

➔ **Cost and Profit analysis:** We did not have data for the trips and difference in operation cost for Mini and large trucks else we could have done a cost benefit analysis of mini vs large truck

➔ **Truck Breakdown Data:** This would have helped us fine tune the Vertex/Service center cover