**Capstone Data Science Project Proposal: Motor Carrier Performance Analysis**

Line 1: Authors Name/s per 1st Affiliation

Line 2: Author’s Name/s per 1st Affiliation

Line 3 (of Affiliation): Dept. name of organization

Line 4: name of organization, acronyms acceptable

Line 5: City, Country

Line 6: e-mail address if desired

***Abstract- To be written after completion of the project***

1. Introduction

There is no question that the United States motor carrier industry is the nation’s largest freight transport sector, and its subsectors are numerous and varied. According to [2], it is responsible for transporting approximately $10.5 trillion worth of goods annually, accounting for over four percent of total gross domestic product (GDP). Cantor et al. [1] argue that the trucking business provides most hauling services in the country and is almost “the service provider in the last mile” solution to the customer. Miller and Saldanha [3] and Miller et al. [4] have stated that truck drivers, particularly their driving performance, influence how customers perceive motor carriers and play an integral part in their success. Consequently, this project report aims to get meaningful conclusions about the company operations by performing analytics on three datasets, including the company where I currently work. Although motor carrier performance analysis has been of interest to investigators, a systematical exploration into the issue has been hampered by the lack of available data; this current study combines the three datasets to get meaningful information.

In the ten years I have been working in the trucking industry, I was fortunate enough to gather experience in almost every department of a motor carrier company. Each firm mainly consists of three departments: accounting/billing, safety, and dispatch. Similarly, using datasets from the factoring company and the international fuel tax agreement report, this paper intends to get predictive analytics and explore the following essential questions: (1) which broker the company has the most loads with and information like in each quarter which was top brokers by numbers and revenue. Then see if there is a pattern that would give what should be expected in the quarters to come. (2) get information on who pays the loads in the shortest time and which broker takes more time to pay the invoice. Based on which broker/customers are top clienteles I want to check what is expected if the company decides to remove the factoring company and start billing invoices directly. (3) the fundamental analysis from the last dataset is which are the top ten states with the most percentage of the miles driven and how the fuel ratio was. I plan to use Orange to apply data mining techniques on the dataset and Bayesian optimization in Python as this will give an idea about the brokers with the most revenue or one with the most loads.

1. Data

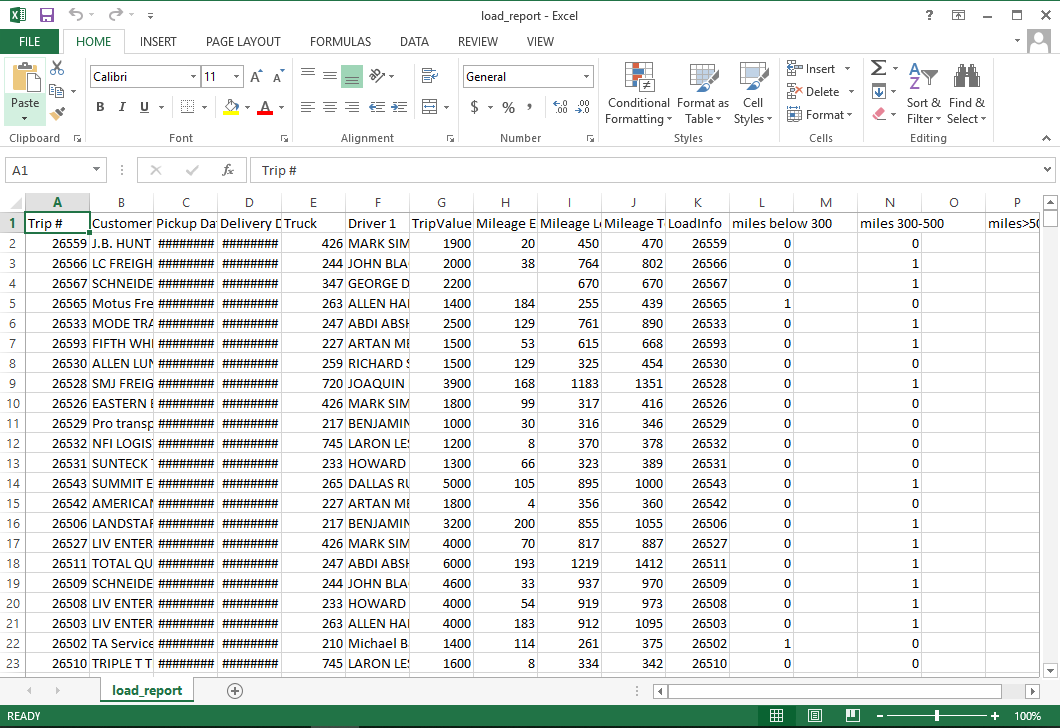
**Dataset (1)**

Orange software was utilized in this dataset to project comparisons between amounts of pays between brokers within 45 days. It shows how the productivities of the brokers rank per load to be trucked. The data shows that the broker with the most wages is productive; hence, the system offers more jobs.

**Analysis**

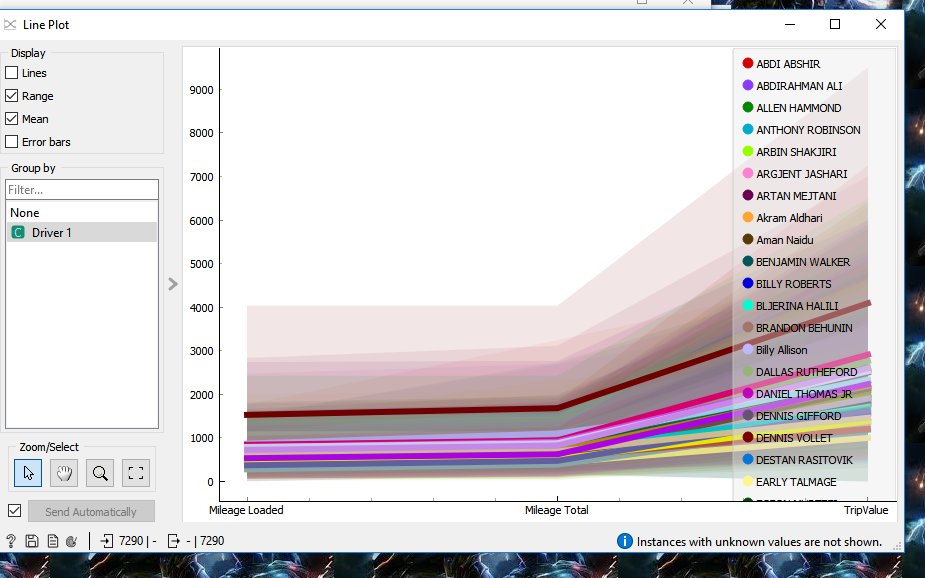
The analysis was done using tables and graphs to represent the data. The table contains information about the brokers and the amount paid to them. The amount represents the advance given for a load to be trucked. Below is a sample part of the data used in the study.

Figure 1: Load Report



A graph was plotted to show the amount that was paid in days for each broker. The data was plotted using Orange after being normalized and filled, as highlighted in figure 2 below. The broker with the highest amount of pays within 45 days was Argent Jashari. Despite the presence of systematic biases, a pattern was sighted on this data, and it showed that with an increase in the total mileage, there was an increase in the trip value for each trip. The trend, however, had outliers who maintained a less steep curve within the same duration. In this dataset, I also evaluated the fuel ratio for the quarters in each quarter from 2020 to Q3 of 2021. It gives insight into which states took up the most fuel, with low mileage being covered in those states. The ratio was given by the proportion of miles given to gallons of fuel consumed in that quarter in the given state.

Figure 2: A Graph Showing Amount Paid In Days for Each Broker.



In this sample dataset for Q1 2020, the fuel ratio was calculated, and then the dataset was arranged in order from largest to smallest in terms of fuel ratio. In that quarter, Detroit was the state with the highest fuel ratio. Accordingly, this inference that more fuel was used in that state than the miles covered. The first five states for this quarter were DE, FL, GA, IA, and ID. Below is a table of states and fuel ratios for Q1 2020 and only covers the top five states with the highest fuel ratio. The same was done for other quarters.

Table 1: States and Fuel Ratio for Q1 2020

|  |  |
| --- | --- |
| DE | 23.786 |
| FL | 13.723 |
| GA | 11.O46 |
| IA | 9.344 |
| ID | 9.152 |

Table 2: Fuel Ratio for Q2 2020 (Top Five States)

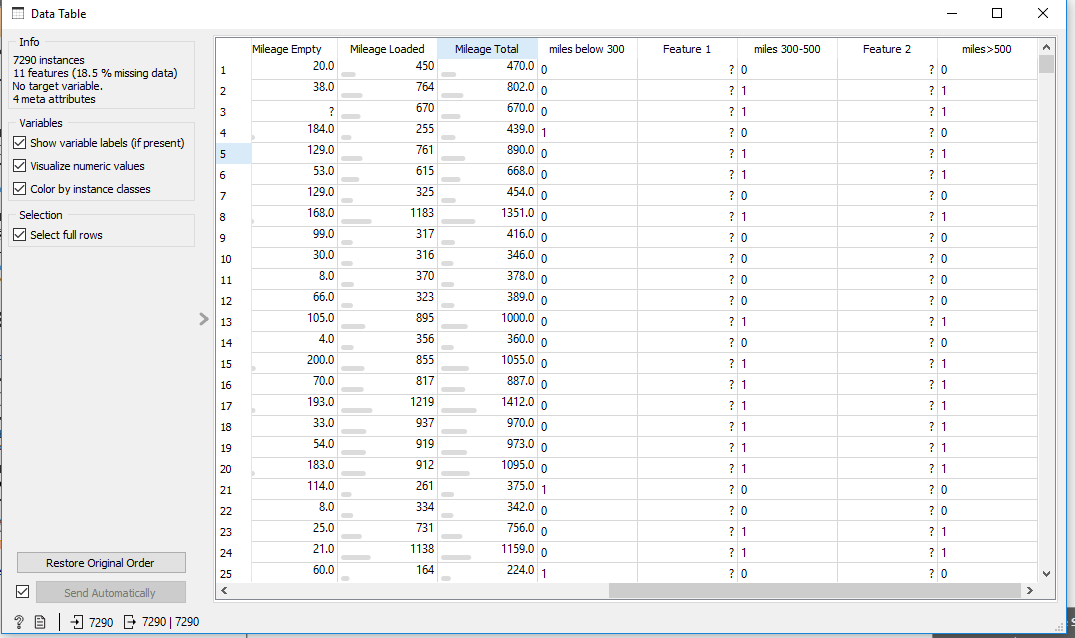
|  |  |
| --- | --- |
| AL | 63.876 |
| AZ | **63.245** |
| AR | **41.004** |
| CA | **31.176** |
| CO | **29.562** |

For Q2 2020, AL had the highest fuel ratio, which does not show any pattern with the previous quarter. The analysis, however, gave a pattern where Detroit appeared more times on the top five than any other state. Accordingly, this brings to the conclusion that much fuel is used in that state. Based on that data, trucks should reduce fueling in this state to improve efficiency.

**DATASET (3)**

In this dataset, I analyzed mileage covered by different drivers. It was split into three parts: below 300 miles, 300 to 500 miles, and above 500 miles. The following Orange data table shows a sample of the data that was used in this analysis. The last three rows in the data table represent the drivers that traveled within the labeled distance.

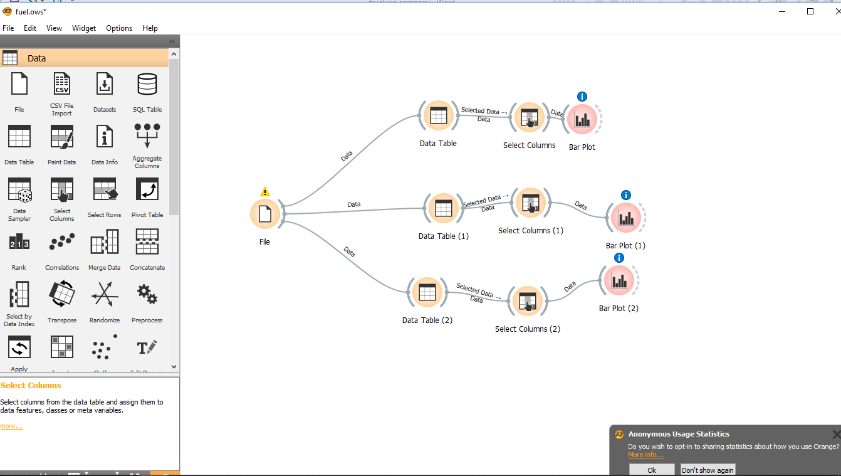
Figure 3: Dataset 3 Sample Data



**Analysis**

The dataset was supposed to give a hint on what distance most of the drivers prefer. I analyzed this by creating box graphs on Orange, which I used to conclude the length of the trip majority of the drivers prefers. I achieved this by plotting line graphs of a subset of 200 items for each driver on each length of the trip. Below is an Orange workflow for the given data.

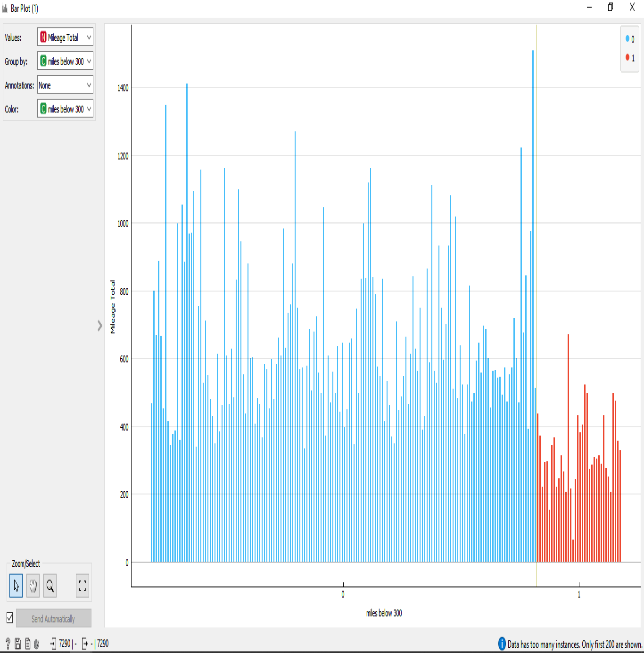
Figure 4: An Orange Workflow for the Data



The described graphs are shown in figures five to eight below, and columns filter the data. The red bars represent an accurate value, while the blue bars represent the false value. From the graphs, the following conclusions were drawn:

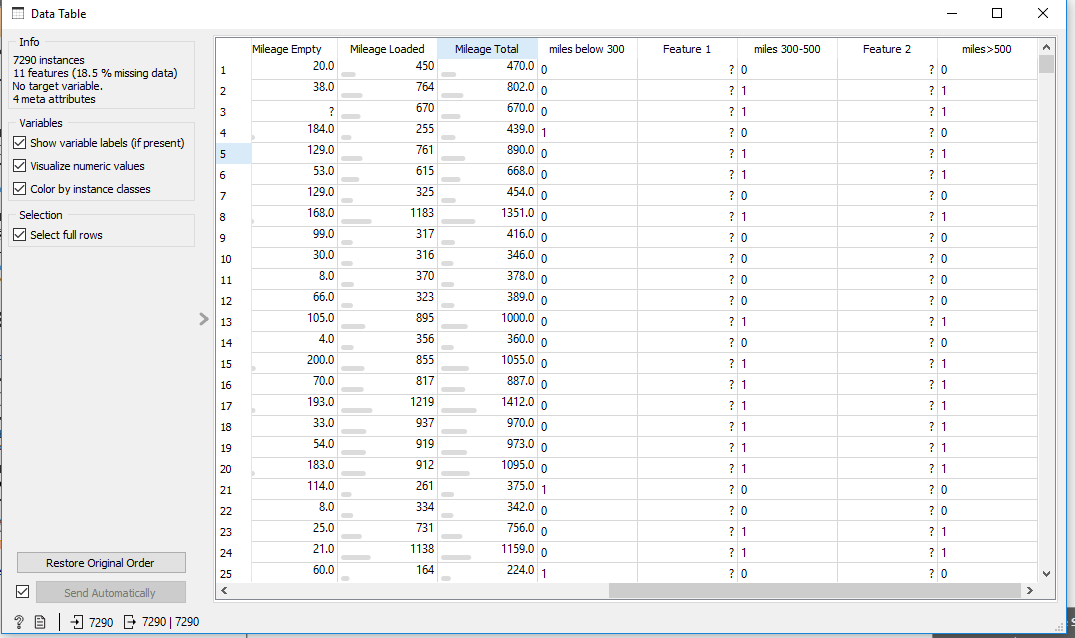
1. **Trips below 300 miles.**

For this length of trip, a more significant number of drivers preferred to take a trip above 300 miles. It, therefore, makes the company take into consideration taking businesses with longer trips. As a result, increasing uptake of the appointments among the drivers and increasing efficiency.The plotting was done by giving jobs on miles below 300 miles a value of one and zero contrariwise. The plots were then done, and concluded that contracts above 300 miles had a higher density. Below are graphs showing the results.

Figure 5: Trips below 300 Miles

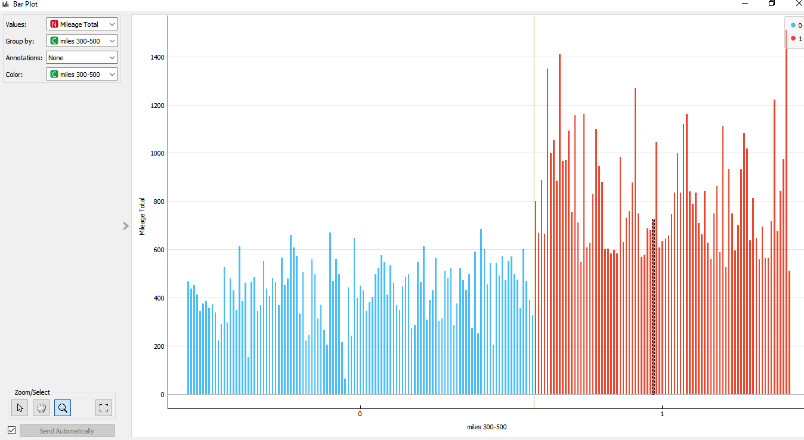
The blue lines are the number of drivers who went on a trip with a distance of more than 300 miles. The data table below contains rows with either one or zero to represent whether the driver went below or above 300 miles. From that data, it is arguably true that most trips should range between 300 miles to ensure efficiency.

Figure 6: Mileage Total



1. **300 To 500 Miles**

The dataset contains the number of drivers who took a trip between 300 and 500 miles. It was derived from the purchase summary database by checking the length of trips where drivers were loaded. For trips with a mileage between 300 and 500 miles, a value of one was given and zero contrariwise. On this dataset, there is an almost equal distribution of fewer drivers taking more trips with a mileage outside of this range. Figure seven below shows a graph containing a plot of the density of drivers who took a trip with a mileage outside the range of 300 to 500 miles and those that took a trip outside this range.

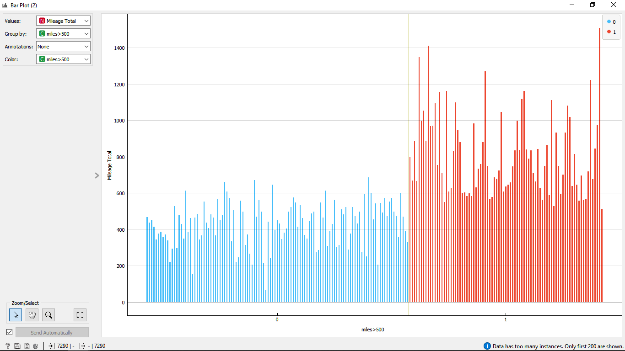
Figure 7:

The range has a lesser density. As a result, having less uptake leaves the dataset of trips taking more than 500 miles. An equal distribution of trips in this range can be obtained based on the driver’s preference for tours.

**Over 500 miles**

Most of the drivers took trips in this region. A higher frequency was observed in this region compared to the other two regions. It, therefore, leads to higher uptake of jobs in this region. The data shows that taking jobs in this region will generate higher revenue. However, the other two regions still produce a substantial amount of revenue. The graph for this region is shown below. It was plotted on data generated from a formula that returns a one if the trips were more than 500 miles and a zero for trips below 500 miles.

Figure 8:



1. Discussion
2. Conclusion and Recommendations

REFERENCES

[1] D. E. Cantor, T. M. Corsi, C. M. Grimm, and P. Singh, “Technology, firm size, and safety: Theory and empirical evidence from the US motor-carrier industry,” *Transportation Journal*, vol. 55, no. 2, pp. 149–167, 2016.

[2] J. P. Saldanha, J. W. Miller, C. Shane Hunt, and J. E. Mello, “Linking formal controls to motor carrier performance: Curvilinear and interaction effects,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 64, pp. 28–47, 2014.

[3] J. W. Miller and J. P. Saldanha, “A new look at the longitudinal relationship between Motor Carrier Financial Performance and Safety,” *Journal of Business Logistics*, vol. 37, no. 3, pp. 284–306, 2016.

[4] J. W. Miller, J. P. Saldanha, M. Rungtusanatham, and M. Knemeyer, “How does driver turnover affect motor carrier safety performance and what can managers do about it?,” *Journal of Business Logistics*, vol. 38, no. 3, pp. 197–216, 2017.