Identifying Long Delivery Times by LTL Classification

Seth Howells
seththowells@lewisu.edu
DATA-51000-002, FALL
Data Mining and Analytics
Lewis University

I. INTRODUCTION

The dataset under investigation is derived from an LTL (Less-Than-Truckload) logistics study from TRAN-SET (Transportation Consortium of South-Central States) [1]. The data is focused on the freight movement of all commodities between New Orleans to Oklahoma City, which is representative of port-to-hub logistics. LTL transportation mode is "a shipment whose weight and size characteristics allow it to be consolidated with other shipments in a single truckload, or to be shipped as is. [2]" Less-Than-Truckload shipments were designed to capitalize on the cost-savings of consolidating similar freight with similar lanes, i.e. Origin to Destination, as opposed to Full-Truckload (FTL) where few pallets are transported from Origin to Destination on a single, dedicated truck, which costs significantly more per pallet and diminishes eco-friendly modes.

In 2019, the LTL industry was more than $78 billion, which is up from $72 billion in 2017 [3]. As brick-and-mortar stores begin declining at an extraordinary rate due to the rise of consumer shipping preferences, the LTL industry stands to gain significant revenue boosts. However, the increase in revenue does not correlate to an increase of profits since every Purchase Order is dependent on the actual expense: constantly changing fuel surcharge rates, seasonality of commodities, duration of route (distance), and scalability (weight). Without proper freight consolidation and truck sharing measures, the billions increased into the market will do little to drive profits.

The purpose of the analysis is to identify new relations that specifically address freight consolidation and freight mobility through clustering analysis. This analysis addresses 3 main questions:

- What are the groups associated with the distance travelled and quantity of pallets in respect to the time it takes to deliver a Purchase Order?
- Does weight affect the time it takes to delivery an order?
- Is there an underlying relation between the weight of a shipment and the time it takes to deliver?

The future sections of this report describe the dataset, the methodology, results along with a discussion, and a conclusion. Section II contains a description of the dataset used for this analysis. The methodology for analysis is presented in section III. In section IV, I report and discuss the results. Finally, section V provides conclusions.

II. DATA DESCRIPTION

The variables under investigation are listed in the table below. Each variable is directly associated with LTL transportation. Distance, Weight, LTL Class, Origin, and Destination are components that affect the transportation cost. It is especially important to note that LTL Classifications are derived from the National Motor Freight Classification (NMFC). The LTL Classes are grouped by NMFC codes and by product description so that similar products are in the same class.

TABLE I. DATA ATTRIBUTES

Attribute	Type	Example Value	Description
ORIGIN	Nominal (string)	Location 38	Shipping point
DESTINATION	Nominal (string)	Location 23	Delivery point
DISTANCE	Numeric (real)	20.4524	Miles

Attribute	Type	Example Value	Description
QUANTITY	Numeric (integer)	11	Pallets/Parcels
WEIGHT	Numeric (integer)	709	Pounds (lbs.)
LTL CLASS	Numeric (integer)	60	NMFC classification
DELIVERY TIME	Numeric (real)	66.4707	Hours

III. METHODOLOGY

K-Means clustering analysis was performed on the data selected listed in SECTION II. A list of steps and flowchart below contains the steps in the analysis.

Steps 1-3 Attributes were selected from the dataset and preprocessed to normalize the data.
- Categorical data: divide by number of values
  - Treats categorical data like it is ordinal but varies between 0 and 1.
- Metric data:
  - Standardize the mean to equal 0
  - Standardize the variance to equal 1

Step 4 k-Means algorithm was performed. The Silhouette Scores found in TABLE II shows convergence occurring at 4 clusters. Therefore, the cluster analysis is comprised of 4 different clusters for investigation.

TABLE II. CHOOSING K CLUSTERS

k = number of clusters	Silhouette Scores
2	0.250
3	0.235
4	0.233

Step 5 Scatter Plot, Fig 1
- X-axis: Distance (Miles)
- Y-axis: Quantity (Pallets/Boxes)
- Apply color region by cluster
- Change data point size to Delivery Time (hours)

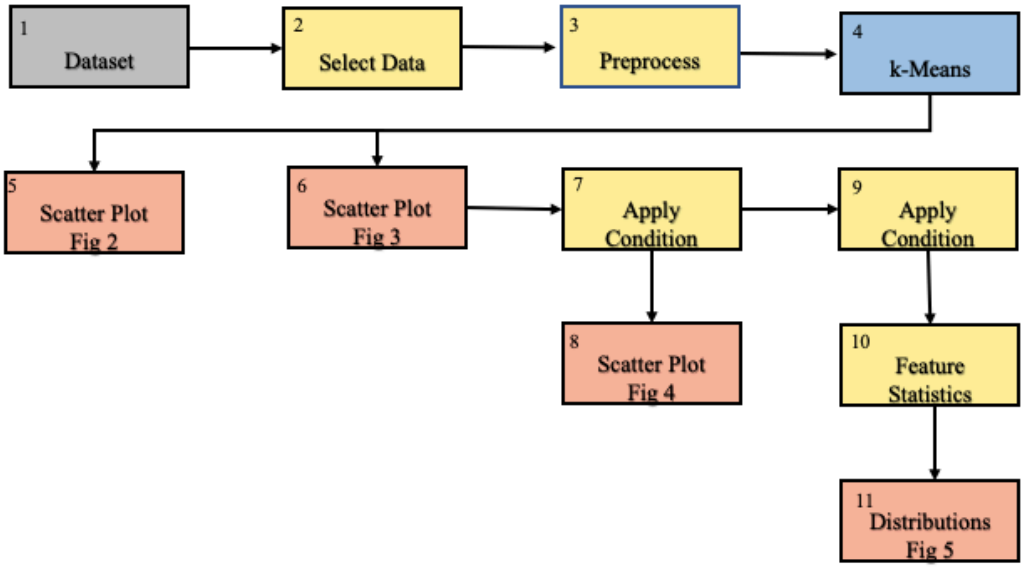
Step 6 Scatter Plot, Fig 2
- X-axis: Weight (lbs)
- Y-axis: Delivery Time (hours)
- Apply 5 regression lines: one for each cluster and one for average of clusters

Step 7 Apply condition: Cluster is C4 of Step 6

Step 8 Scatter Plot, Fig 3
- X-axis: Weight (lbs)
- Y-axis: Delivery Time (hours)
- Apply 5 regression lines: one for each cluster and one for average of clusters

Step 9 Apply condition: Cluster is C4 of Step 6 and LTL Class is between 0.913 and 0.914, which is the normalized value for LTL Class 85.

Steps 10-11 Select Delivery Time attribute for Distribution plot in Fig 4



IV. RESULTS AND DISCUSSION

LTL shipments optimize weight, quantity (or space), distance, and class in order to decrease the cost, which in return boost profit margin. Attributes WEIGHT, QUANTITY, DISTANCE, LTL CLASS were analyzed because the relations directly affect and LTL carrier's profitability. Upon further research, I found that LTL Class refers to a freight categorization system that groups similar commodities together. The groups of commodities can generally be determined by the product density.

The results of the analysis are visually explained in Fig 1, Fig 2, Fig 3, and Fig 4 and the following discusses the result of the figures.

Fig 1, Scatter Plot: Distance (Miles) vs Quantity (Pallets)

Scatter plot visualizes the four clusters based upon x-axis (Quantity), y-axis (Distance), and size of data points (Delivery Time). In efforts to normalize the data, the mean of the attributes was set to zero with the variance equal to one.

- Cluster 1 (blue) indicates that short distance and small quantities relates to LTL Class.
- Cluster 2 (red) indicates that as distance increases, the quantity decreases, and consequently that
- Cluster 3 (green) is clearly influenced by the quantity
- Cluster 4 (orange) indicates that mid-ranged mileage will consist of every type of LTL class because freight matching, otherwise known as consolidating, is easiest performed.

Fig 2, Scatter Plot with Regression: Weight (lbs.) vs Delivery Time (hours)

Delivery time, represented in hours, is critical to an LTL carrier's success because the Department of Transportation (DOT) regulates the hours of service (HOS) a driver is permitted to work per day [4]. In profitability terms, this means that increased delivery time (hours) equates to less deliveries made per day since the driver cannot go over the 10-hour HOS.

Fig 2 identifies clusters within a 2-dimensional concerning Weight (lbs) and Delivery time (hours). The color-coordinated regression lines further explain the effect of increased weight to delivery time. The overall correlation coefficient ($r = -0.14$) value indicates that there is not a significant relation. However, Cluster 3 (red) has a correlation coefficient of r = -0.35, which is the steepest correlation. Contradistinctively, Cluster 4 (orange) has a positive correlation, indicating that as weight increase, delivery time also increases.

Fig 3, Scatter Plot with Regression in Cluster 4: Weight (lbs.) vs Delivery Time (hours)

Fig 3 examines Cluster 4 in Fig 2 to determine underlying relationships between the positive correlation coefficient and its variables. Across all variables, LTL Class had the highest similarity within the cluster as represented by the color-coordinated data points and region. LTL Class between buckets 0.5 - 1.0 and 1.5 - 2.0 are in the path of the regression line. The scatter plot was standardized so that the mean represents zero while the variance is scaled to equal one. This indicates that the bulk of data points in Fig 3 has less than the average weight while more than average delivery time.

Fig 4, Frequency Distribution after Condition Applied: LTL Class 85 in Cluster 4

The underlying variable associated with the findings in Fig 3 are represented in a distribution graph Fig 4, where LTL Class equals 1 on the x-axis has the highest frequency. The highest frequency in Fig 4 represents LTL Class 85.

The significance of LTL Class of 85 in respect to the relationship between Weight and Delivery Time is that LTL Class 85 increases delivery time with increased weight. This contradistinction from the other clusters and LTL Classes provides a new perspective for further investigation into delivery times for LTL Class 85 commodities since increased delivery time reduces the amount of trips/stops a driver can make per day and less trips/stops per day equates to less revenue generated per day.

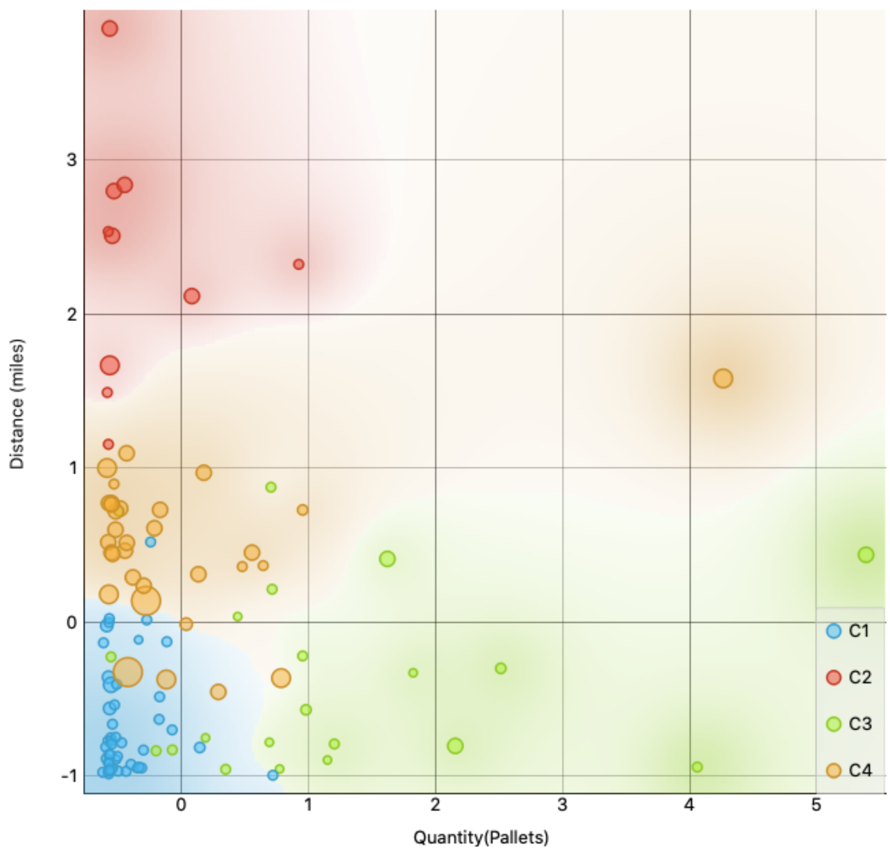


Fig. 1. Scatter Plot with Delivery Time (hours) size of points: Distance (Miles) vs Quantity (Pallets)

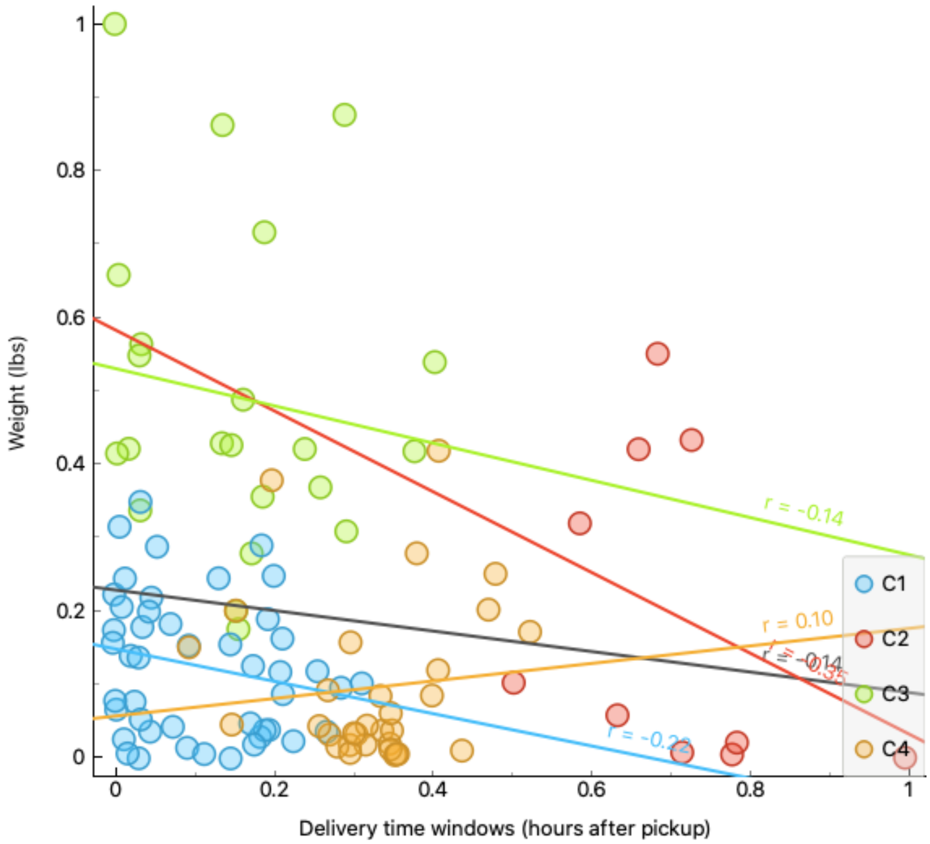


Fig. 2. Scatter Plot with Regression: Weight (lbs.) vs Delivery Time (hours)

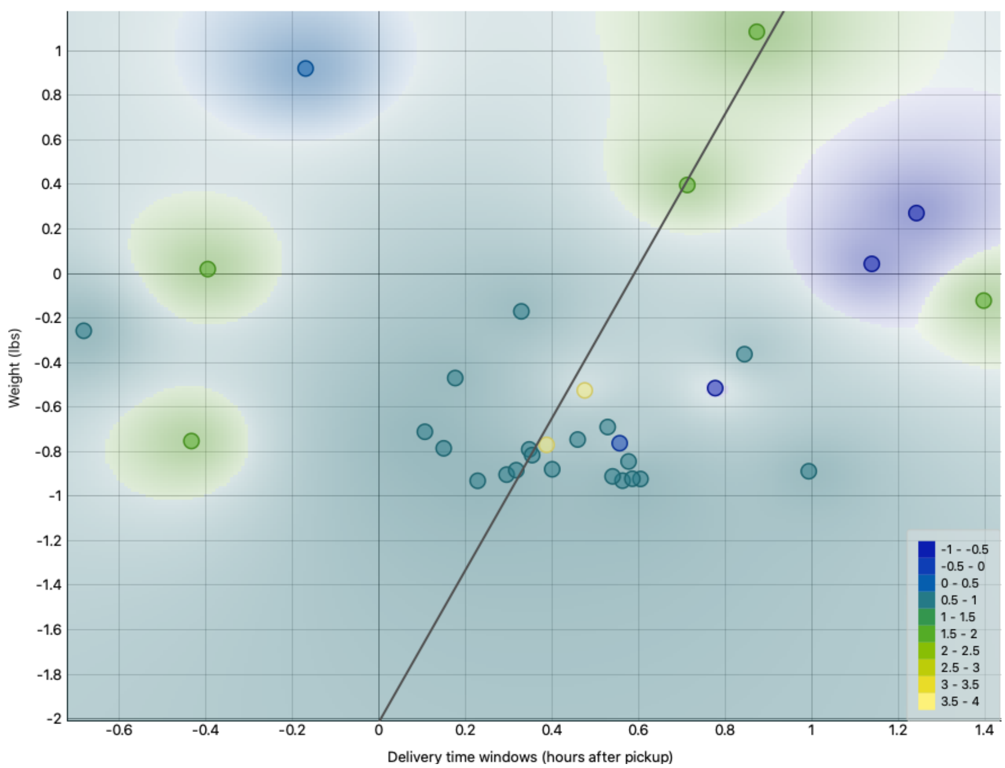


Fig. 3. Scatter Plot with Regression in Cluster 4: Weight (lbs.) vs Delivery Time (hours)

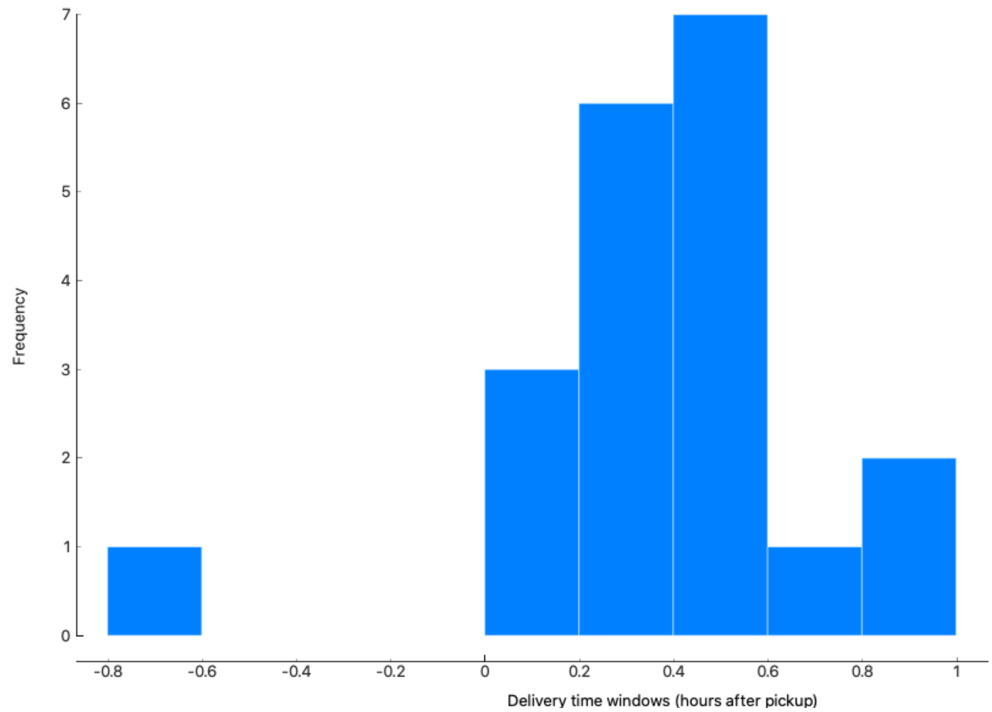


Fig. 4. Frequency Distribution after Condition Applied: LTL Class 85 in Cluster 4

V. CONCLUSIONS

The k-Means cluster analysis performed on the TRANSET dataset produced a new discovery that commodities within NMFC freight class 85 took longer to deliver despite averaging less weight per shipment. This was uncovered in Fig 2 where Cluster 4 (orange) had the only positive correlation coefficient out of all clusters. A positive correlation coefficient indicates that as weight increases, delivery time also increases. Fig 3 took a deeper look into the underlying variables influencing the positive regression line and found that LTL Class attribute had the most in common. Fig 4 confirmed previous findings with LTL Class 85 being the most frequent within Cluster 4. The distribution in Fig 4 clearly shows that almost all shipments in Cluster 4 had above average delivery time.

My recommendation upon this discovery is that further analysis should be monitored around shipments containing LTL Class 85 products because of the consistent lag in delivery time. As a result, any significantly above average time spent at delivery equates to missed time to deliver another shipment. In operational context, both planning and deployment, the time dimension is a critical component to an LTL carrier's success in freight management [5]. Mitigating the delivery time spent on LTL Class 85 can boost revenue since more stops can be made per day.

REFERENCES

[1] T. Liu and C. Zhao, "Impacts of freight consolidation and truck sharing on freight mobility." 2019.
[2] "Trucking Industry Glossary," Aug. 15, 2014. Accessed on: Sept. 25, 2020. [Online]. Available: https://www.dat.com/resources/glossary
[3] E. Mazareanu, "Size of the less-than-truckload market in the U.S. between 2017 and 2020." 10-Sep-2020.
[4] A. Scott, J. W. Miller, and A. T. Balthrop, "Did the electronic logging device mandate reduce accidents?" Jan-2019.
[5] S. Hernandez, G. Kalafatas, and S. Peeta, "A less-than-truckload carrier collaboration planning problem under dynamic capacities." 11-Jan-2011.