**Impact of Intermodal Equipment Status on Intermodal Equipment Rental Fees (aka Per Diem)**

Robert A. Bresley

Bellevue University

DSC 530: Data Exploration and Analysis

Shankar Parajulee

March 5, 2022

**Impact of Intermodal Equipment Status on Intermodal Equipment Rental Fees (aka Per Diem)**

Intermodal containers (aka equipment) generally incur a rental fee for its use, which is an overhead cost for transportation companies which do not use their own equipment. These fees escalate in cost with longer rentals and can severely impact profit margins for transportation companies if equipment is not managed carefully. The primary intent of this data analysis is to determine if equipment that is currently placed at a customer location and waiting to be unloaded has a higher risk of incurring large rental fees than equipment in other statuses (empty, released loaded, etc.)

**Outcome of EDA**

Choosing variables from this data set was difficult, as the majority of them are categorical variables and analytical methods used to determine most impactful variables were unable to be used effectively. Furthermore, since this is a data set, I have personally produced from work data, there were multiple phases to cleaning this data set. First, I had to clean the data to use it professionally, which included many, many iterations of logic and calculation changes in order to improve the accuracy and validity of these aggregated results which are, mostly, aggregated from data entered by hand with few data validations present. Next, in order to use the data for this project I had to anonymize it to protect private information. Then I created calculated fields specific to the questions this project is intended to answer. Finally, I decided how to analyze records with missing data values. For the most part, I removed records which were missing values for the variables key to answering the primary question, like Days on Customer Clock.

While analyzing this data, it was important to keep in mind the context surrounding each record; it represents a transportation move in an industry that is fraught with exceptional circumstances. Keeping this fact in mind helps to better analyze the data and understand not only the outliers, but also what they might indicate and help understand if they are important or not.

**What do you feel was missed during the analysis?**

I feel that I missed an opportunity to discover variables in this data set that might contribute to higher per diem costs than the obvious variables I’m checking. Essentially, I feel that I might have aggregated this data too much and limited the data set to the point where analysis shows that these variables have an impact predominantly because they are the only variables being analyzed.

**Were there any variables you felt could have helped in the analysis?**

I feel that variables recording how many containers were at a customer location at a specific time would have been helpful in determining how important that location was in our overall per diem costs. I also feel that variables recording how long a piece of equipment stayed in each equipment status would have been extremely helpful. I’m finding that the issue with this data set is that it is a snapshot in time, and not a true history; while the equipment may have been on a customer clock for a long time, if it recently started moving again that high customer clock would be attributed to the released loaded status instead of the unloading status that actually took so long.

**Were there any assumptions made you felt were incorrect?**

I believe the largest assumption I made was that my snapshot data set would help me analyze historical data in regards to equipment status. While this data set still tells me which customers tend to have high per diem costs, and I can see which locations have a high number of days, I’m unable to directly analyze which equipment statuses are most responsible for our per diem costs. I also assumed that I could use categorical variables for much of my data analysis and this is not true, however this assumption could be partially corrected by transforming that categorical data into numerical representation.

**What challenges did you face, what did you not fully understand?**

I think the biggest challenge I faced was fitting the requirements to the questions I was trying to ask and fitting both of these to the data set I had available. At times it felt like pounding a square peg into a round hole. While I started off changing and updating and modifying my data set, due to the data cleaning and validation work involved with each modification I had to finally pick a point and freeze the data set as a final version. This prompted me to adjust slightly the question I was wanting to answer and took the entire project into a slightly different direction. What I didn’t understand until the end was that I probably needed to better filter and format my data for the tools I have available to me instead of focusing on the question. This would allow the tools to determine what the best questions were to ask. I finally realized that this is an important aspect of the difference between analysis and machine learning; analysis asks questions of data while machine learning uses data to identify important questions to ask.

**Outcome**

The final outcome of this analysis is that it is highly likely that a container that is currently placed at a customer location and is waiting to be unloaded represents a higher risk of high per diem charges than a container in a different container status, so these containers should be a focus when managing equipment. Based on this analysis and my learnings, I would also recommend developing a data set which specifically records the history of how much time each container spent in each status. This more targeted, but also more complete, set of data would offer a better data set to answer this specific question. I would also recommend creating a data set with a broader range of variables in order to use a myriad of Python and R tools to determine which variables offer the best predictive capabilities in regards to overall per diem costs.