Warranty at Wabash

PURDUE UNIVERSITY | The Data Mine

Gannon Rice, Shree Bavana, Connor Sheriff, Victor Penedo Cutrim Inacio, Yunhu Kim, and Poomin Piyapanyawongse

WABASH™

INTRODUCTION

What does Wabash do?
Wabash National is a world-class manufacturer of advanced engineered solutions and services for transportation, logistics and distribution industries.

What is the problem?
Gap in understanding warranty costs/risks as it relates to configuration and newly developed technology.

What is our goal?
As undergraduate researchers, our priority working alongside Wabash is to utilize data to understand the impact and performance of product and configuration on warranty and creating models to help predict warranty costs for trucks and dry vans.

FALL

We generated EDAs on newly learned Wabash data to understand the concepts + purpose, along with beginning to clean data. Truck and Dry-Van Teams progressed from initial data analysis to data cleaning and configuration.

SPRING

After thorough cleaning of truck and dry van data and with assistance from out mentors, we were able to create functional models for warranty costs

RESEARCH METHODOLGY

- kaggle was our first stop for learning everything related to data analysis, learning about EDA and cleaning
- Our scrum master has been a great help in explaining important concepts that we used:
  - Exploratory Data Analysis
  - Linear Regression
  - Multiple Regression
  - Logistic Regression
  - Train Test Split
- Along with our TA, our mentors have guided us in understanding data analysis, forecasting and model building.

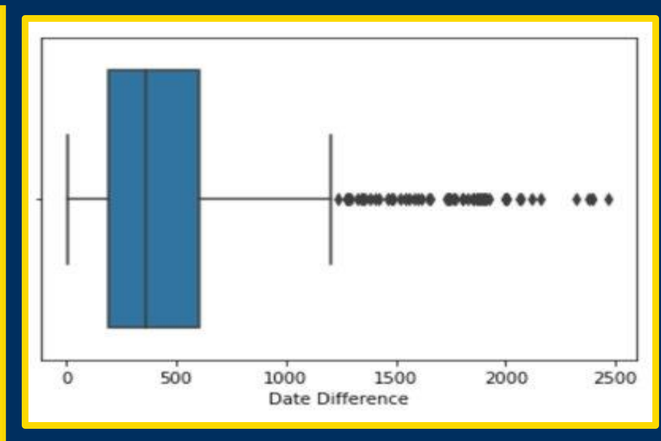
TRUCKS TEAM → DATA

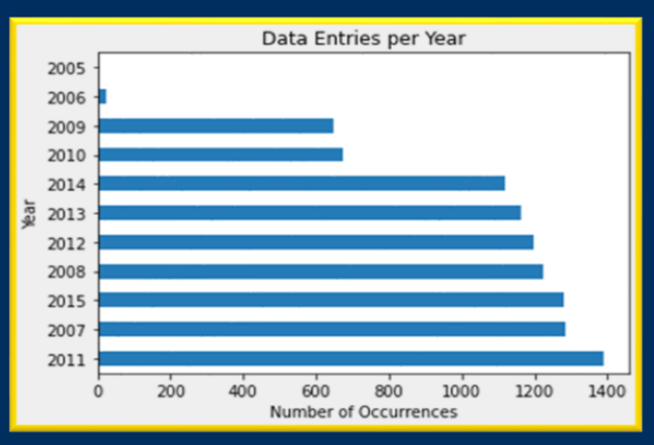
Data Used: Truck Schedule Files, Truck Config Files, Truck Warranty Sample File

Models: Linear Regression, Multiple Regression, Logistic Regression, etc.

Pre-processing:
- Cleaned up data through removing irrelevant columns and null values
- Fixed the index of data frames and merged columns for easier analysis
- Changed data types of variables to allow for model building

Results:
- Attempting to merge the Truck Config and Truck Warranty files through the Serial # column to identify the column with the greatest impact on the warranty.
- Merging files to increase number of variables we can correlate the Warranty Cost to.
- Tested and ran various models on the datasets to find any relationships between our variable of interest, the Serial #, and other variables from the Truck files.
- Model building to find relevant patterns between warranty and other variables, such as time and truck components, using linear regression, multiple regression, logistic regression, and autoregression models.





DATA → DRY VAN TEAM

Data Used: Dry Van Warranty.csv

Models: Multiple Linear Regression

Pre-processing:
- Performed the time series of DocumentDate and VINServiceDate columns.
- Examined correlations between input variables and Warranty cost
- Used Null Hypothesis significance testing to sort data based on correlation and statistical significance.
- Built a Multiple Linear Regression model and currently building a long short-term memory network(LSTM) as n alternative model.
- In the process of merging Configuration.csv and DryVanWarrnaty.csv files.
- Checking correlation and mergeability between RootCause and FailCode.

Results:
- Found uncorrelated relations between Input Variables and the Warranty Cost.
- No further examination on FailCode correlation due to lack of available part lists needed to further refine RootCause
- Will be implementing alternative model LSTM.

CONCLUSION

Conclusions:
The research project aims to understand the impact of product and configuration on warranty costs for transportation, logistics, and distribution industries. The team has cleaned and pre-processed data, including examining correlations and building multiple regression models, but results were uncorrelated. Therefore, the team is exploring alternative models and working on cleaning correlation data between FailCode and Serial #.

Future Goals:
To dive deeper into the models, continue to work with the data, refine the models, and get closer to finding correlation for warranty costs.

Acknowledgements:
Dr. Mark Ward, Maggie Ann Betz, David Glass, Devin Chatterji, Kali Lacy, John Dill, and the entire Wabash and Data Mine teams