

Machine-Learning Predictive Model

Tyler Busby, Elikem Tamaklo
and Fawaz Almoneef
Department of Industrial Engineering



Abstract

Ryder Inc., is a provider of heavy transportation equipment operating more than 110K vehicles in North America alone. At the end of a vehicle's leasing period Ryder sells these vehicles either as a retail or wholesale. Ryder would benefit from a model that will predict future selling prices (proceeds) of the vehicles based on their attributes and historical proceeds.

Introduction

Vehicle sales data for the past five years of sold was provided, including 50 unique variables for each record showing the characteristics of the vehicle being sold along with its proceeds see (Table1).

The goal was to build an effective model to predict proceeds based on key vehicle attributes, as shown in (Table1), where the characteristics were considered as independent parameters and proceeds dependent in the model.

Numerical	Categorical	
•Odometer Reading	•Vehicle ID	•Transmission Model
•Age in Selling Lot	•Purchase Type (New/Old)	•Engine Manufacturer
•Current Fair Market Value	•SAM Class Code	•Engine Model ID
•Projected Mileage	•(Tractor/trailer/Truck)	•Engine Horsepower
•Rear Axle Capacity	•Model Manufacturer	•Transmission Type
•Body Height	•Vehicle Condition	•Transmission Speed
•Body Width	•Vehicle Weight Class	•Refrigeration Unit Model
•Body Length	•Engine Brake Model	•Year-Month Sold
•Number of Axles	•Tire Size	•Model Year
•Initial Cost	•Suspension Type	•Accepted Date in Selling Lot
•Book Value	•Transmission Manufacturer	•SaleGr (Retail/Wholesale)

Table1 (Numerical & Categorical Variables)

Methods | Design | Analysis

- Visualizing the data showed that proceeds were significantly different based on vehicle type & vehicle condition (Figure1). Hence, we decided to divide the data into 3 different groups: tractor, trailer and truck and build a model for each
- Correlation and association analysis were performed on, both, the continuous and categorical variables, respectively, where variables with the highest correlation/association were used in the model and the rest were removed
 - Continuous variables:** Spearman correlational analysis was performed, result in (Table2), with "Current Fair Mkt Value" variable having the highest correlation (Figure2)
 - Categorical variables:** ANOVA testing was performed for the categorical variables and were one-hot-encoding for the Machine-Learning models
- After filtering the variables based on their significance, the models were trained and tested using 80% and 20% of the data, respectively, using R:
 - Models used were **Linear**, **Random Forest**, **KNearestNeighbor** and **xgbLinear** results in (Table3)

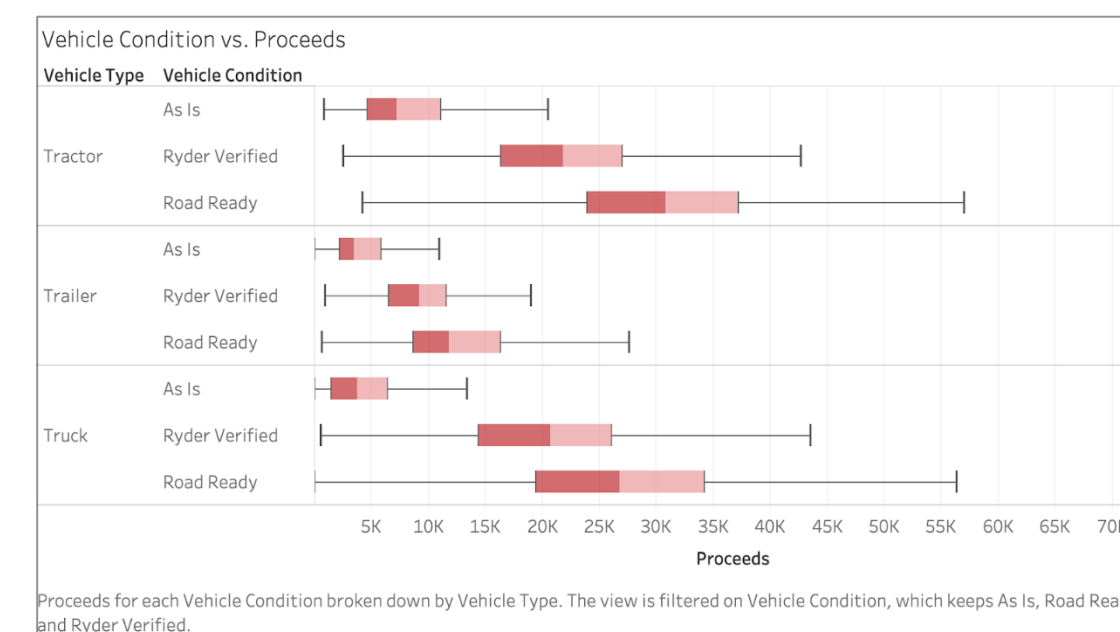


Figure1 (Vehicle Type vs Proceeds)

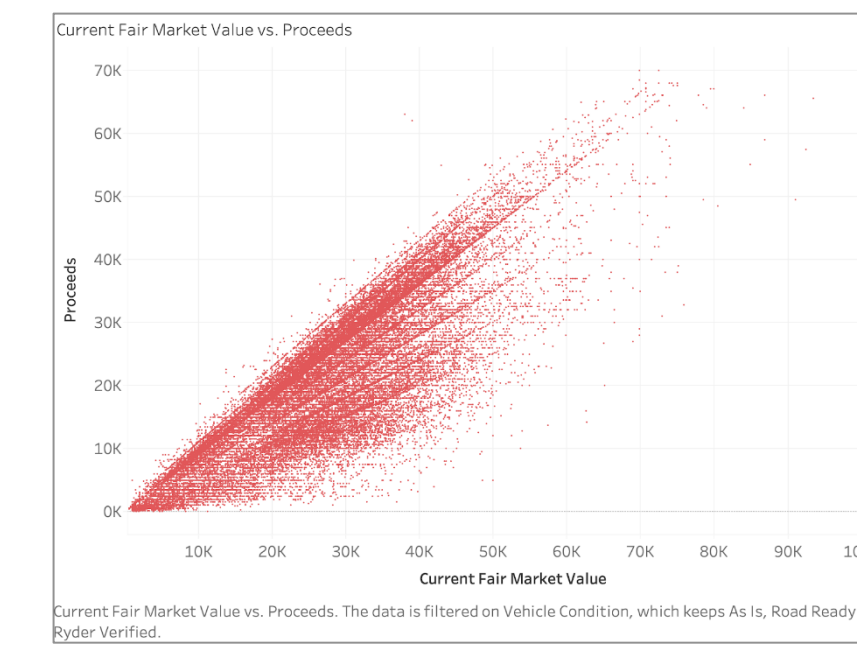


Figure2 (Fair Market Value vs Proceeds)

Results

Numerical Attribute	Spearman Coefficient		
	Tractor	Truck	Trailer
Vehicle Corrected Age	-0.30	-0.18	-0.57
Projected Mileage	-0.11	-0.18	-0.04
Odometer Reading	-0.11	-0.18	-0.04
Average Miles per Month	0.06	-0.09	0.06
Current Fair Market Value	0.80	0.69	0.81
Mile Year Amount	0.06	-0.09	0.06
Accumulated Depreciation	0.08	0.13	0.15
Engine Horsepower	0.27	0.14	
Suspension Capacity	0.29	0.14	-0.13
Axle Total Count	0.30	0.12	-0.18
Cost	0.31	0.31	0.25
Rear Axle Capacity	0.31	0.17	-0.05

Table2 (Spearman Coefficient Continuous Variables vs Proceeds)

	Tractor	Truck	Trailer
	R ²	R ²	R ²
Linear	0.80	0.74	0.85
KNN	0.88	0.84	0.91
RF	0.93	0.93	0.95
xgbLinear	0.94	0.93	0.95

Table3 (R² value for the different models)

Conclusion

In conclusion, the ANOVA and Spearman analysis helped us filter our variables down to 9, 6 and 7 variables for the models used for tractor, truck and trailer, respectively.

Using the filtered variables we were able to build our models and get R² values of **0.94**, **0.93** and **0.95** for tractor, truck and trailer, respectively using the most accurate Machine-Learning Model: **xgbLinear**.

As a result, this model will help Ryder Inc. better understand expected vehicle proceeds, with potential pricing implications, hence improving their liquidity and reducing associated liabilities.

Acknowledgments

It has been a pleasure working with the following people who provided a great help to us, thank you:

Dr. Jaime Buitrago
Lamis Amer
George Vargas from Ryder

