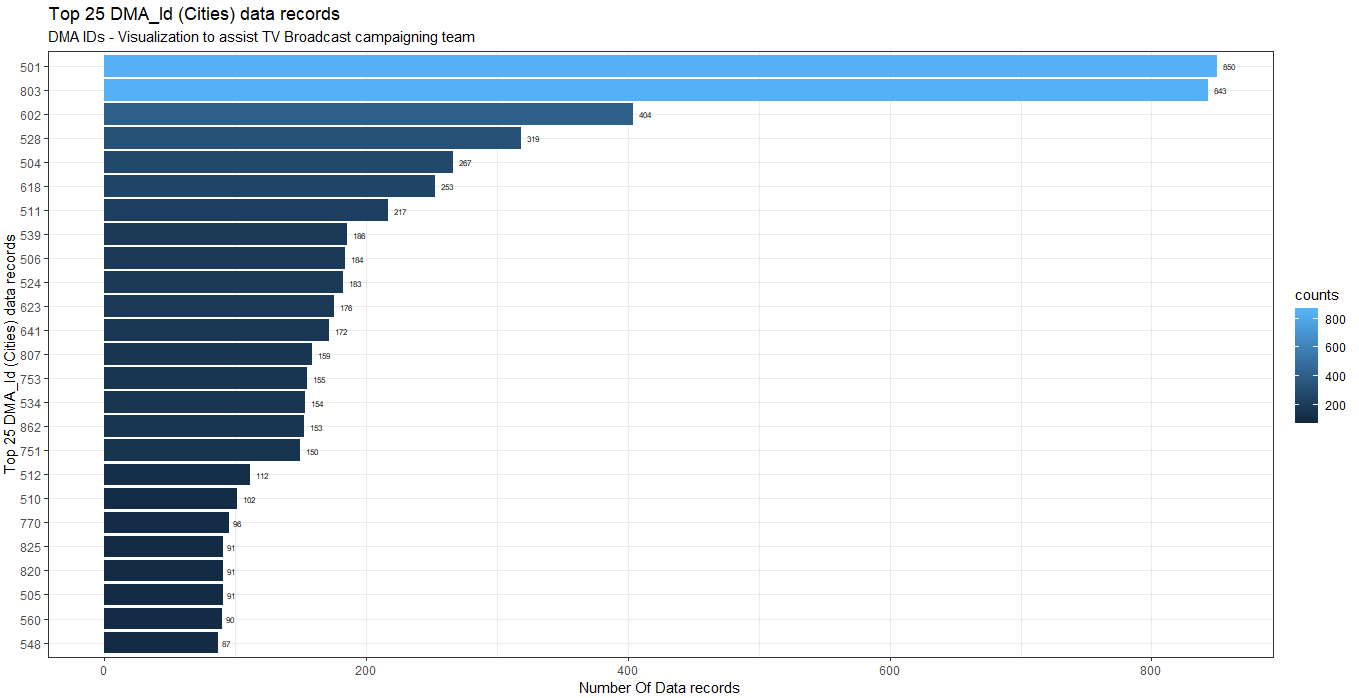
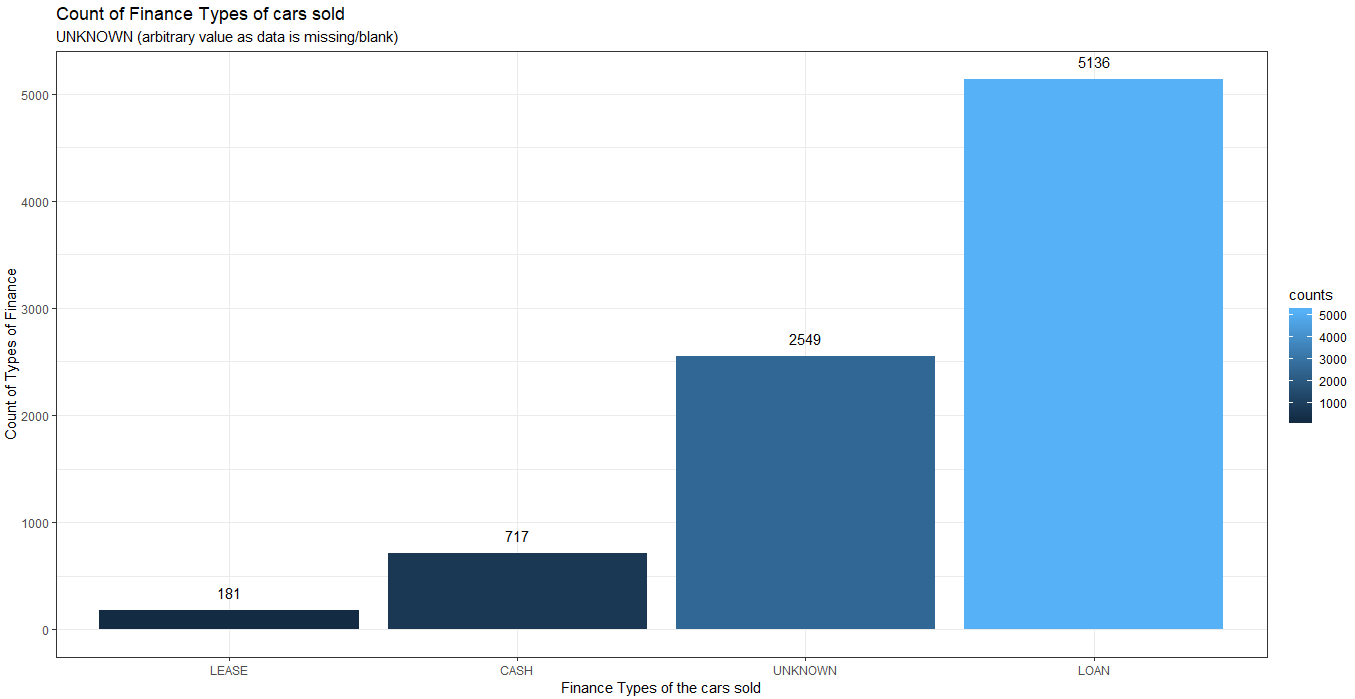
Visualizations and Story Telling of data.

1. Visualization focused to help better TV campaigning.



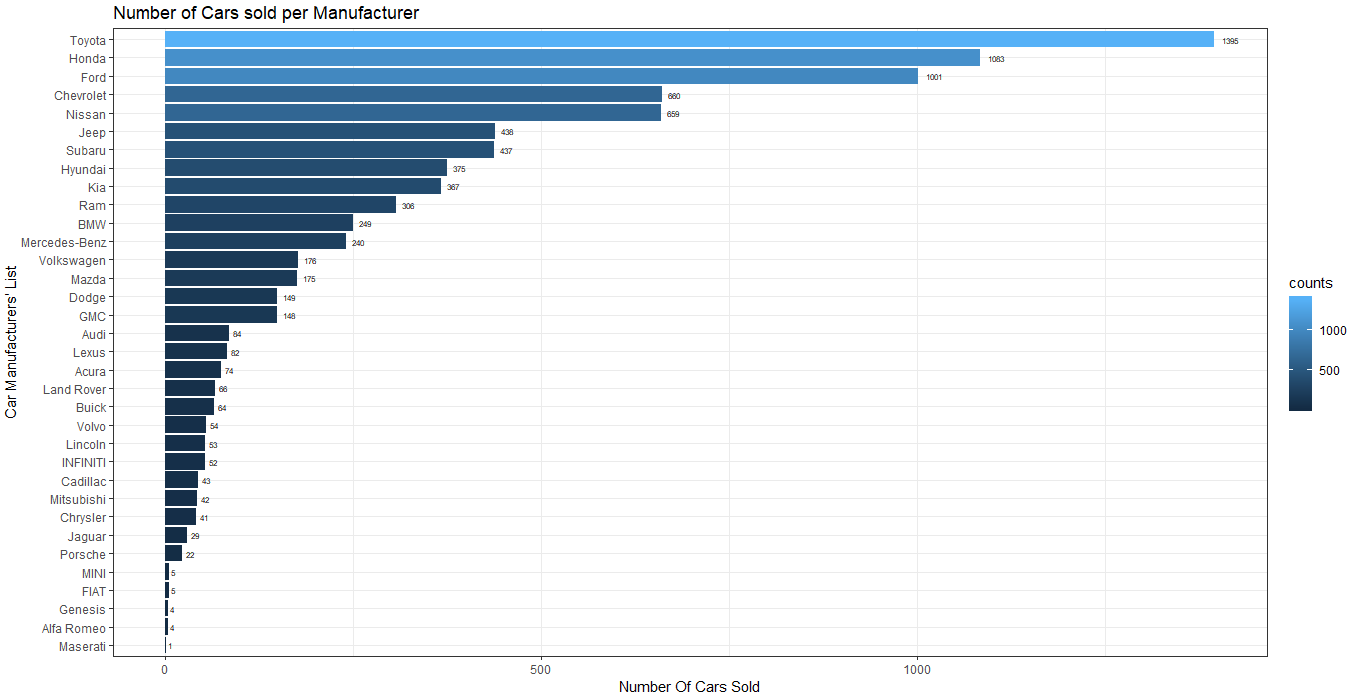
501 – New York City and 803 – Los Angeles have 850 and 843 car sales respectively. Therefore, campaigning of TrueCar platforms in DMA cities where there is high access to cable TV would benefit us and our connected dealers. In these visualizations, I have displayed only Top 25 DMA cities where car sales are highest.

1. Finance Types

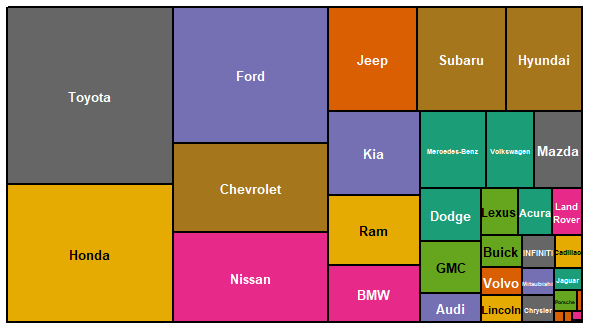


“Loan” finance type is the highest count - 5136 and second is the arbitrary values “Unknown” with count – 2549. TrueCar does have partners in financial institutions, thus if dealers can update finance types well, it would be helpful.

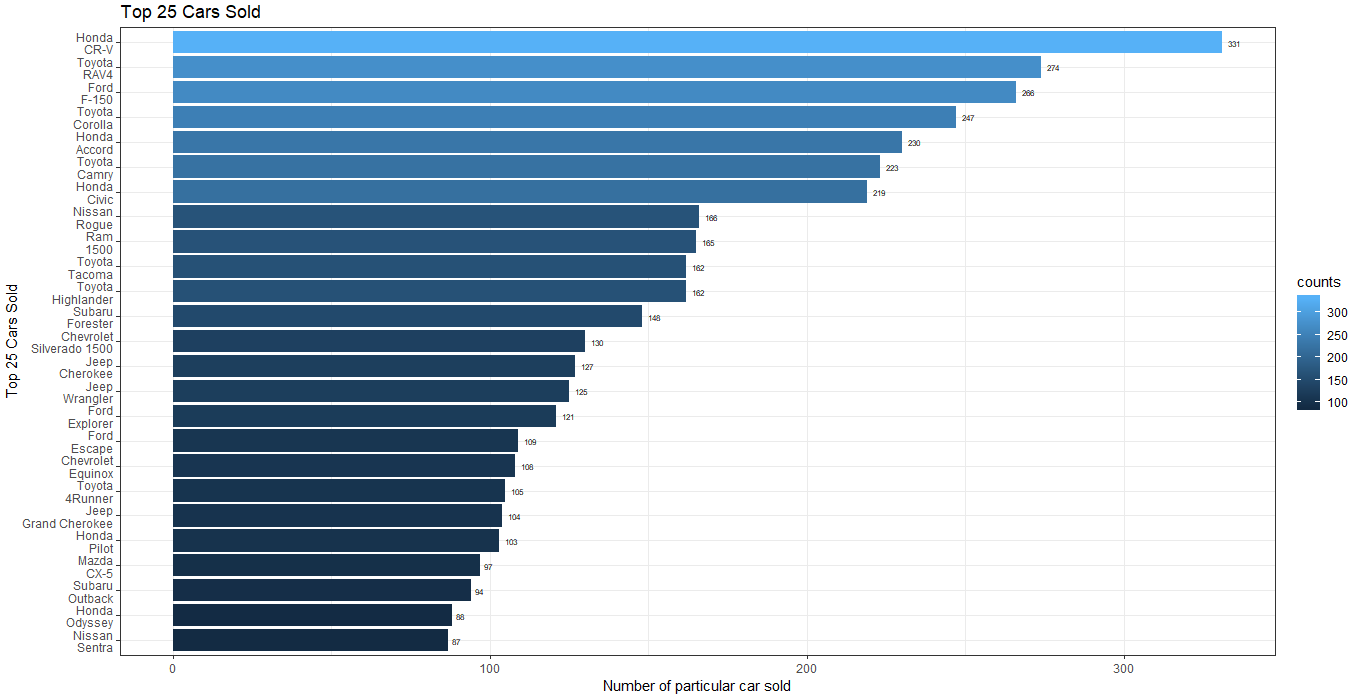
1. Number of Cars sold per manufacturer.



Toyota and Honda cars are sold the most with values 1395 cars and 1083 cars respectively.

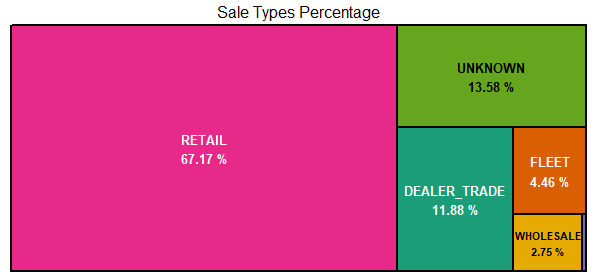


1. Top 25 cars sold



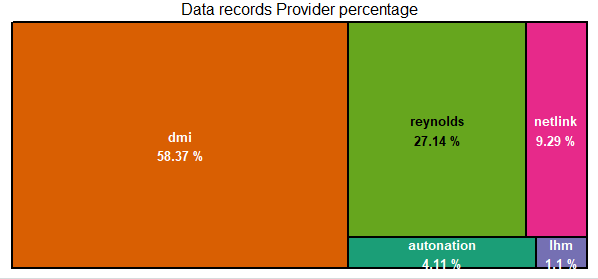
Honda CRV is the most sold car with 331 sold.

1. Dealer Sale Types Percentage



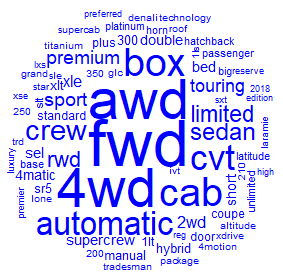
“Retail” Sale Type is the highest with 67.17% of all sales and 0.16% is “House Deal” sale.

1. Data records’ providers’ percentage



“dmi” is the highest data provider of the lot.

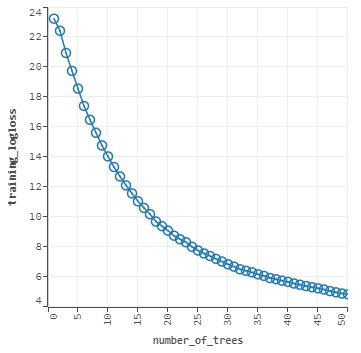
1. Word Cloud of Style\_name description



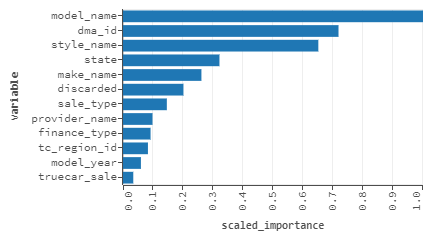
4wd, fwd and awd cars are equally sold. “CVT” and “automatic transmission cars are sold more than manual or stick shifts. Sedan and Touring car model types are sold more.

Machine Learning models:

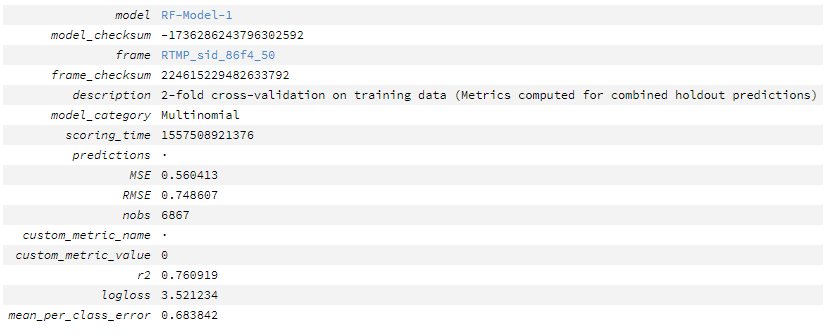
1. In the first model, I wanted to investigate how “curve\_rating” column is affected by other variables. Thus, I chose only categorical values columns for analysis. I have used Random Forest algorithm to better understand the data points with trees = 50 and depth of 20 along with cross validation of 2 folds.
   1. Training Logloss:



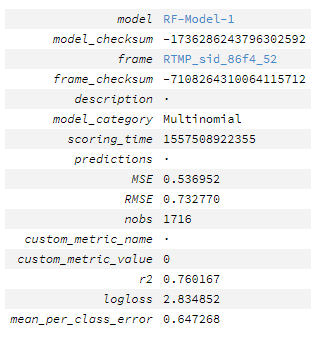
* 1. Important variables affecting “curve\_rating” variable with co-efficient values (scaled importance).

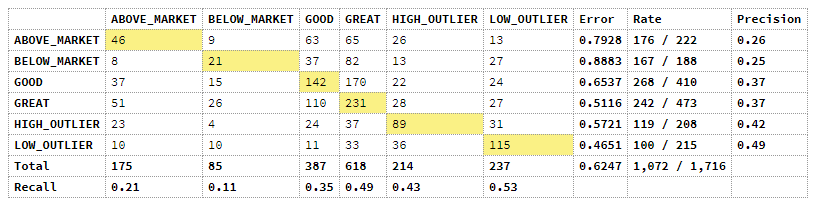


* 1. Model Information and Statistics values of this regression analysis.

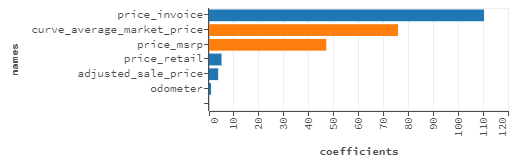


* 1. Prediction of the model and confusion Matrix.



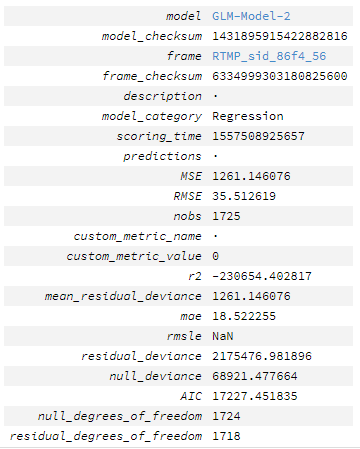


1. In the second model, I wanted to investigate how “price\_ratio” column is affected by other variables. Thus, I chose only numerical values columns for analysis. I have used Generalized Linear Regression algorithm to better understand the data points along with cross validation of 2 folds.
   1. Co-efficient and how positively/negatively “price\_ratio” variable is affected.

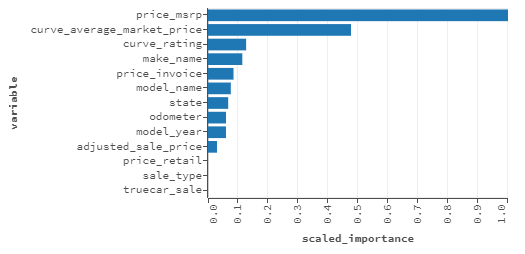


“price\_invoice”, “price\_retail”, “adjusted\_sale\_price” and “odometer” are positively affecting “price\_ratio”. “curve\_average\_market\_price” and “price\_msrp” are negatively affecting “price\_ratio”. We know that priceRatio = Adjusted Sale Price/priceMSRP. Thus, price\_msrp is inversely related to price\_ratio and thus, negatively affects price\_ratio value. The interesting part is that “curve\_average\_market\_price” i.e. the value at which similar car is sold elsewhere is also negatively affecting “price\_ratio” value. This can be deduced that “adjusted\_sale\_price” i.e TrueCar set value is affected. In reality, the customer always buys goods from different platforms where he/she gets the goods at a cheaper price. And most often or always, dealer sells cars at a price lesser than MSRP value. Thus, fluctuating curve\_average\_market\_price value affects TrueCar’s adjusted\_sale\_price value and thereby, negatively impacts “price\_ratio” value.

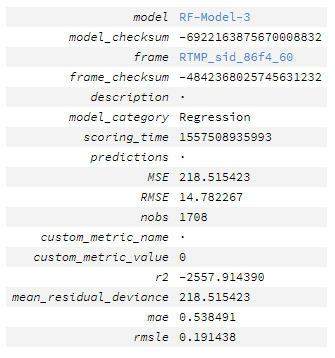
* 1. Prediction results



1. In the second model, I wanted to investigate how “price\_ratio” column is affected by all other variables. I have used Random Forest algorithm to better understand the data points with trees = 50 and depth of 20 along with cross validation of 2 folds.
   1. Important variables affecting “price\_ratio” variable with co-efficient values (scaled importance).



* 1. Prediction results



What I miss doing in this data?

Principal component analysis, unsupervised learning algorithm, could have been done. Because grouping of cars’ manufacturers and models could have provided us an idea of how the market works and clusters might have given an idea of grouping cars on lines of models.

TrueCar Website:

The website is well built and keeping UX in mind, I found the website easy in navigation. I liked that a user upfront can choose between new and used car types with brands mentioned. After selection of new type, I found out that with selection of new exterior colors, the car showed up as a favicon changes accordingly which is a niche feature.

Updates:

1. While selecting the options (Checkbox) for customization of new car, the easy shifting of checkbox selection can be done. Currently, we must deselect the check box and select the new.
2. Feedback button in the used car section can be enhanced like make it a floating button.

Other than these two updates, I did not get an idea of feature enhancement in the website. A new feature creation like “Auto chatbot” can be created in the used car section which can provide a humanized touch to the website.