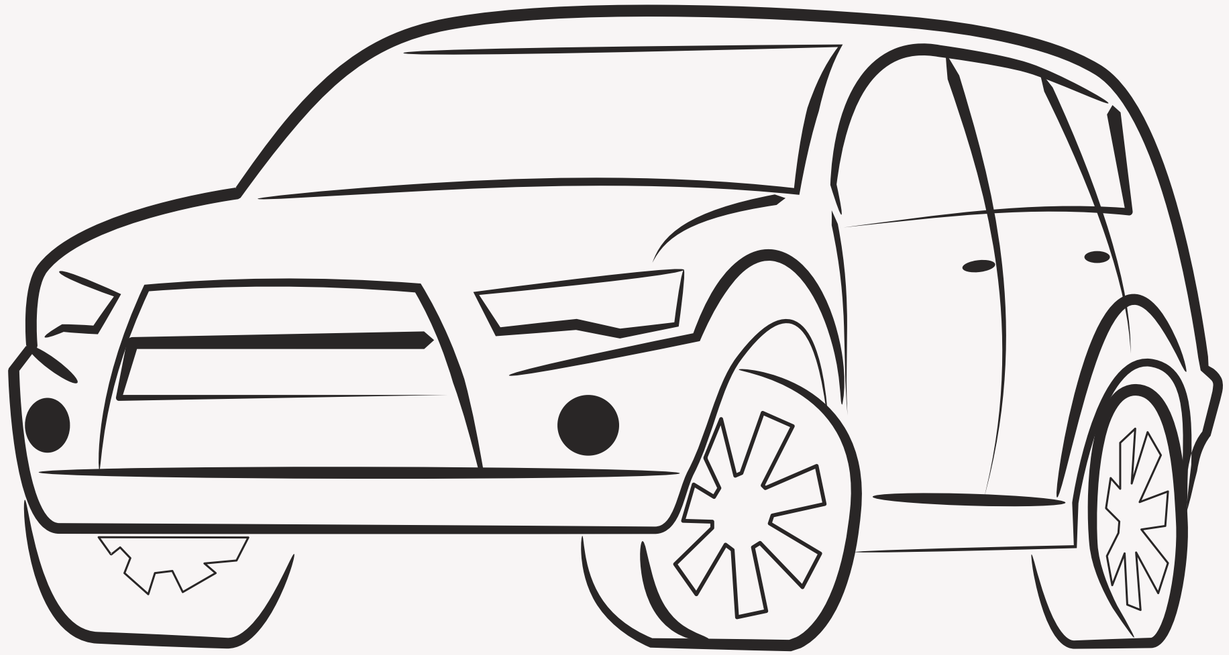
ML-based Car Price Prediction

Dismantling Data | Building Model



Agenda

- Project Overview
- Demo - How it will look like?
- Development Process
- Quick Data Insights
- Model Choice & Outcomes
- How model works
- Next Steps
- Q&A

Our Objective

"To build a model to recommend unit price($000) for cars by using past contracts data in year 2160"

Constraints

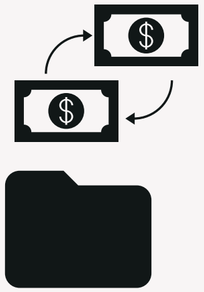
- Predict retail price based on retail data as input to a model which includes past contracts.
- None of the current industry information is relevant as we are dealing in 2160 year

Part 2

Demo

*Disclaimer
This demo depicts how the system works. It needs lot of work to have good look and feel,

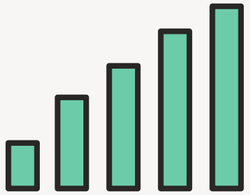
Model Development Process



Input data from past contracts

- Data validation
- Cleaning

- Data Transformation
- Make data model ready

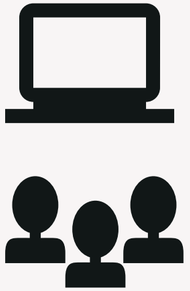


- Hypothesis Testing
- Exploratory data analysis

Disclaimer*

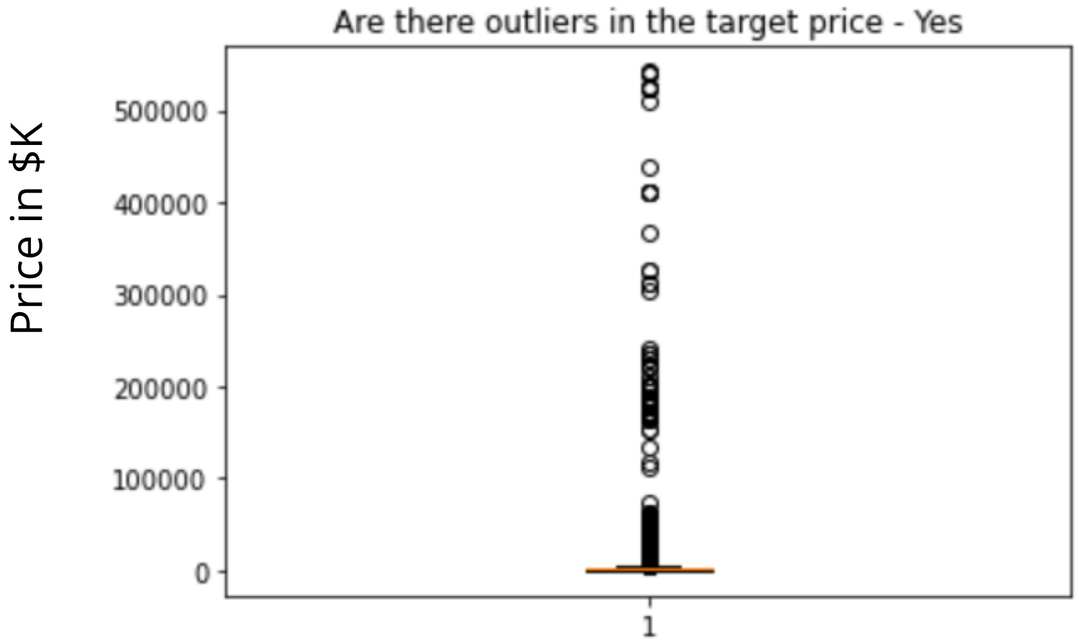
- Adjust the knowledge discovery during experiments. This is used in this use case.

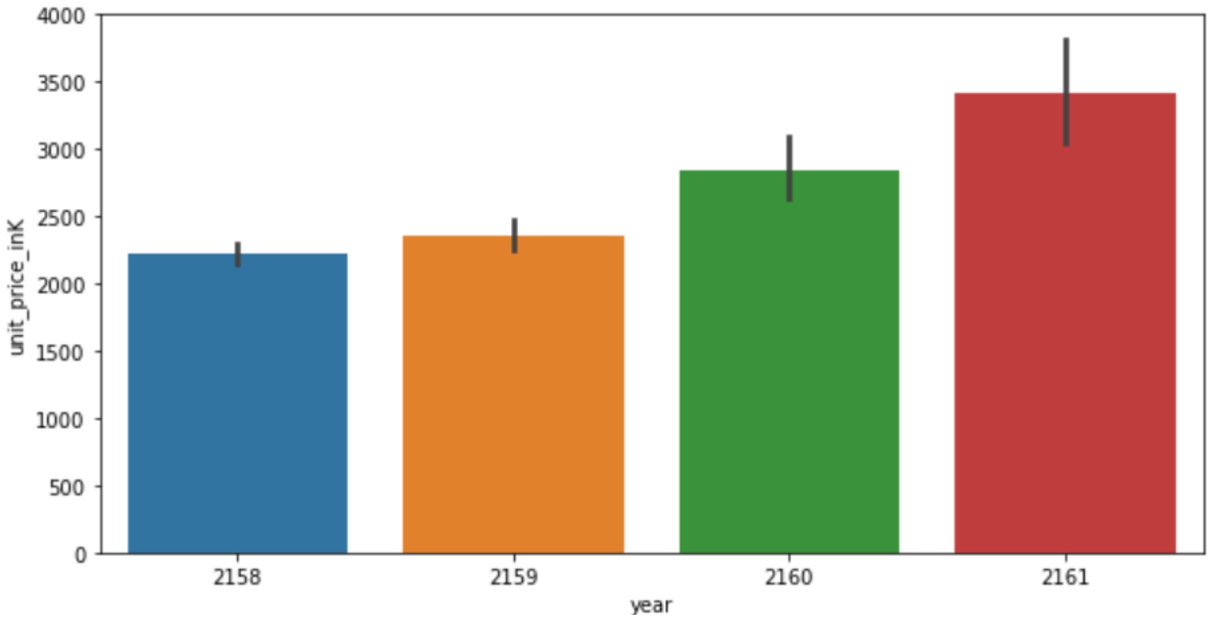
- Model development
- Experiments



- Model Deployed
- Serves the business needs

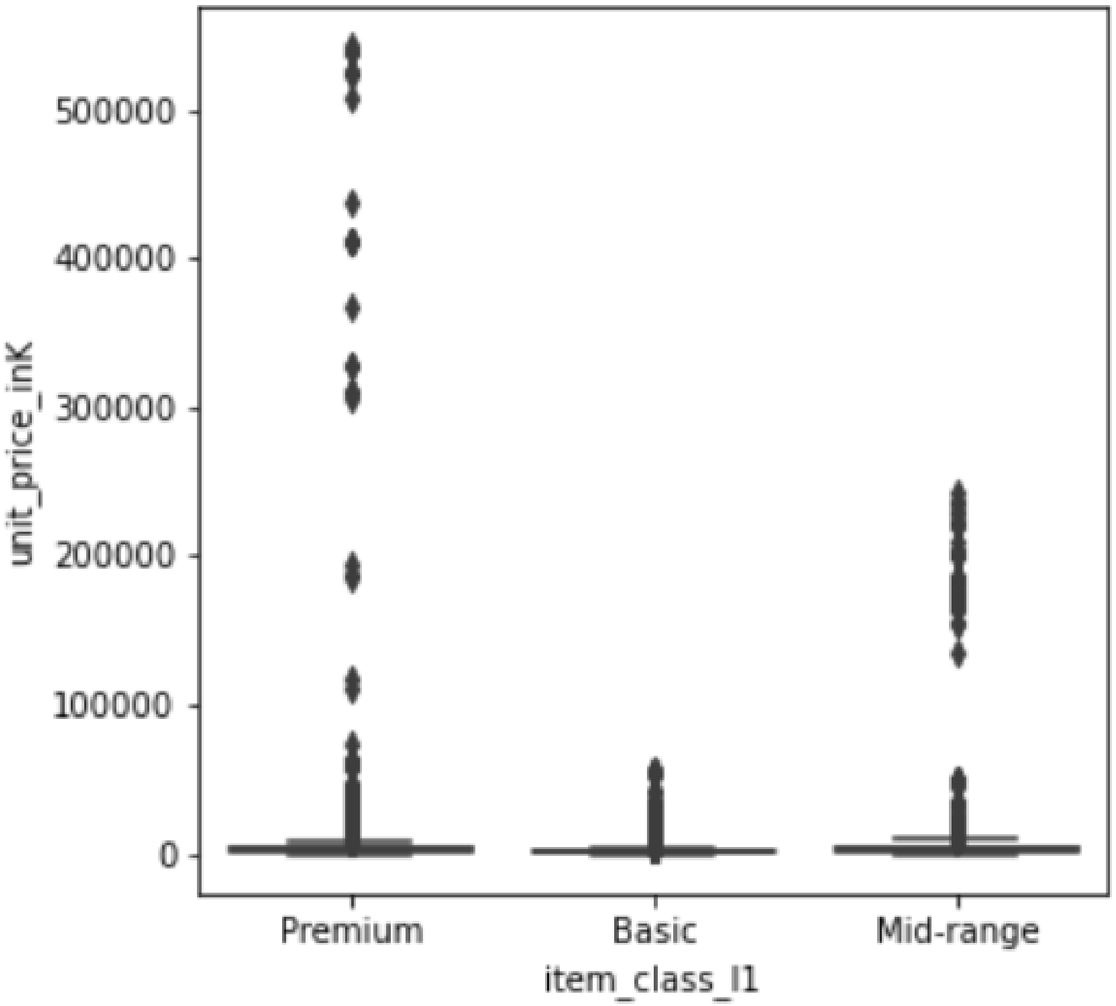
Data Insights 1



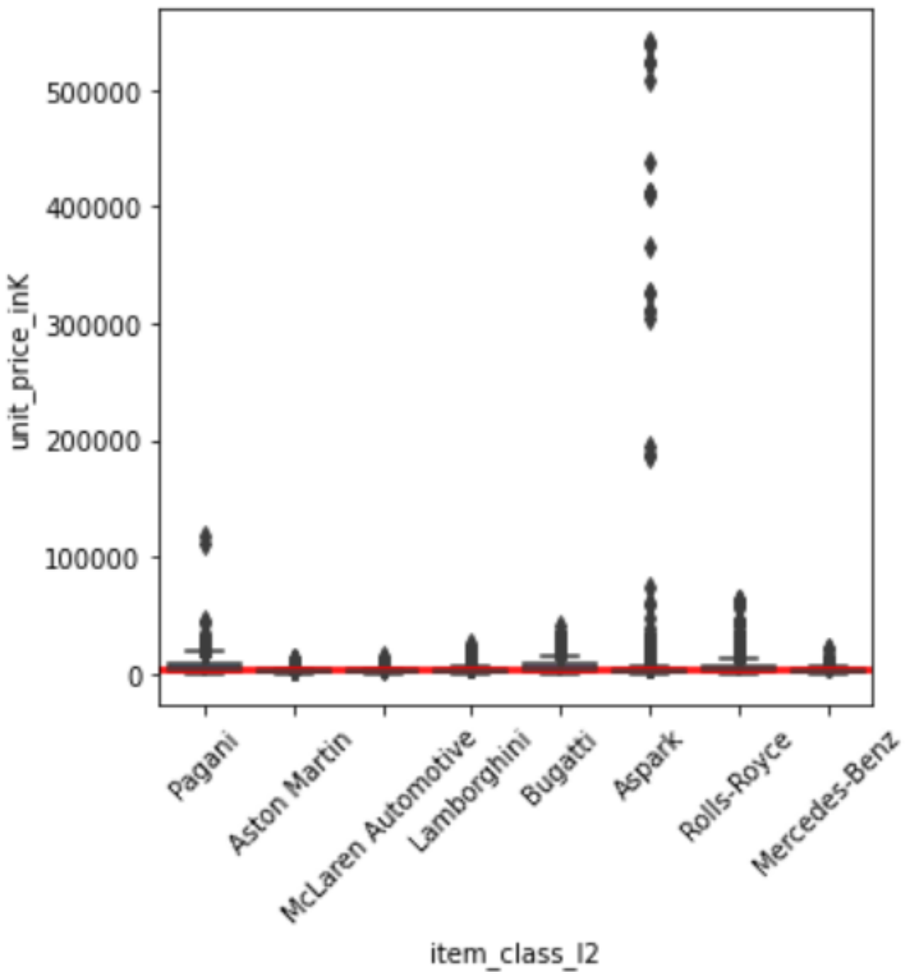


- Outliers in unit_price could be due to higher-end cars.
- Mean value of unit price is seeing an upside trend

Data Insights 2

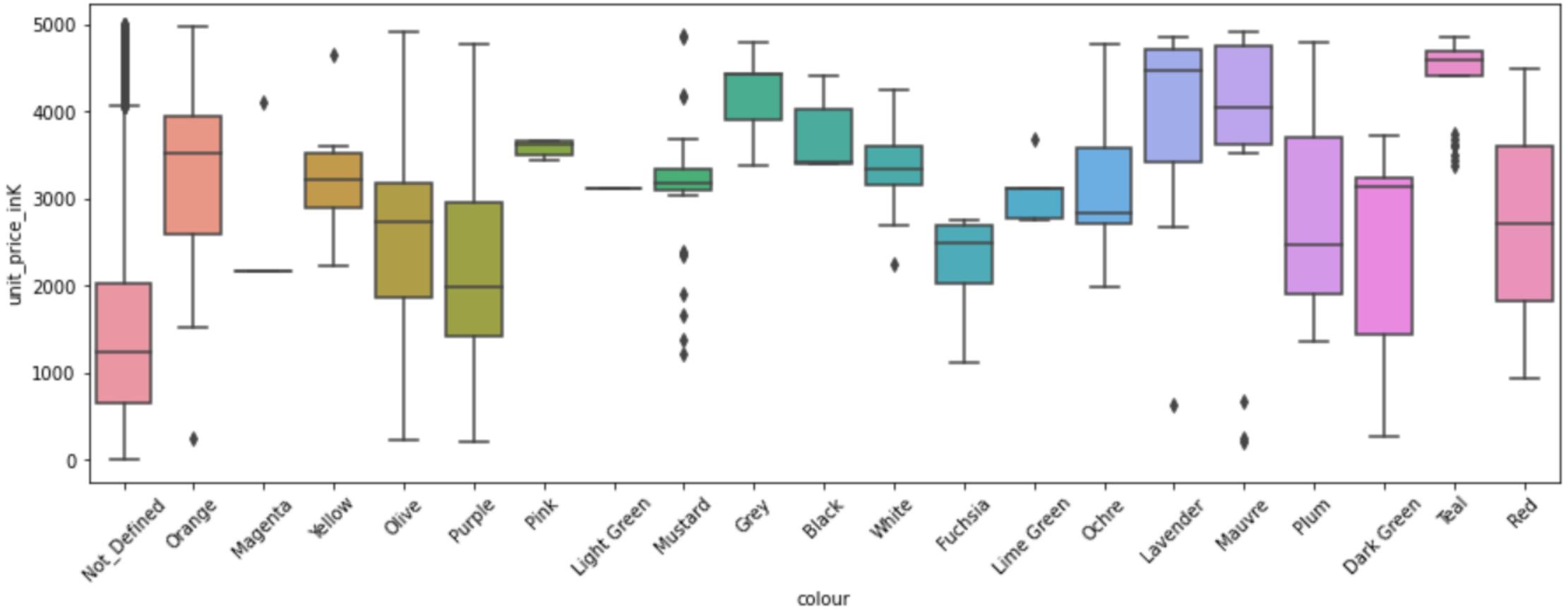


This is the preimum range class of brand type Vs price



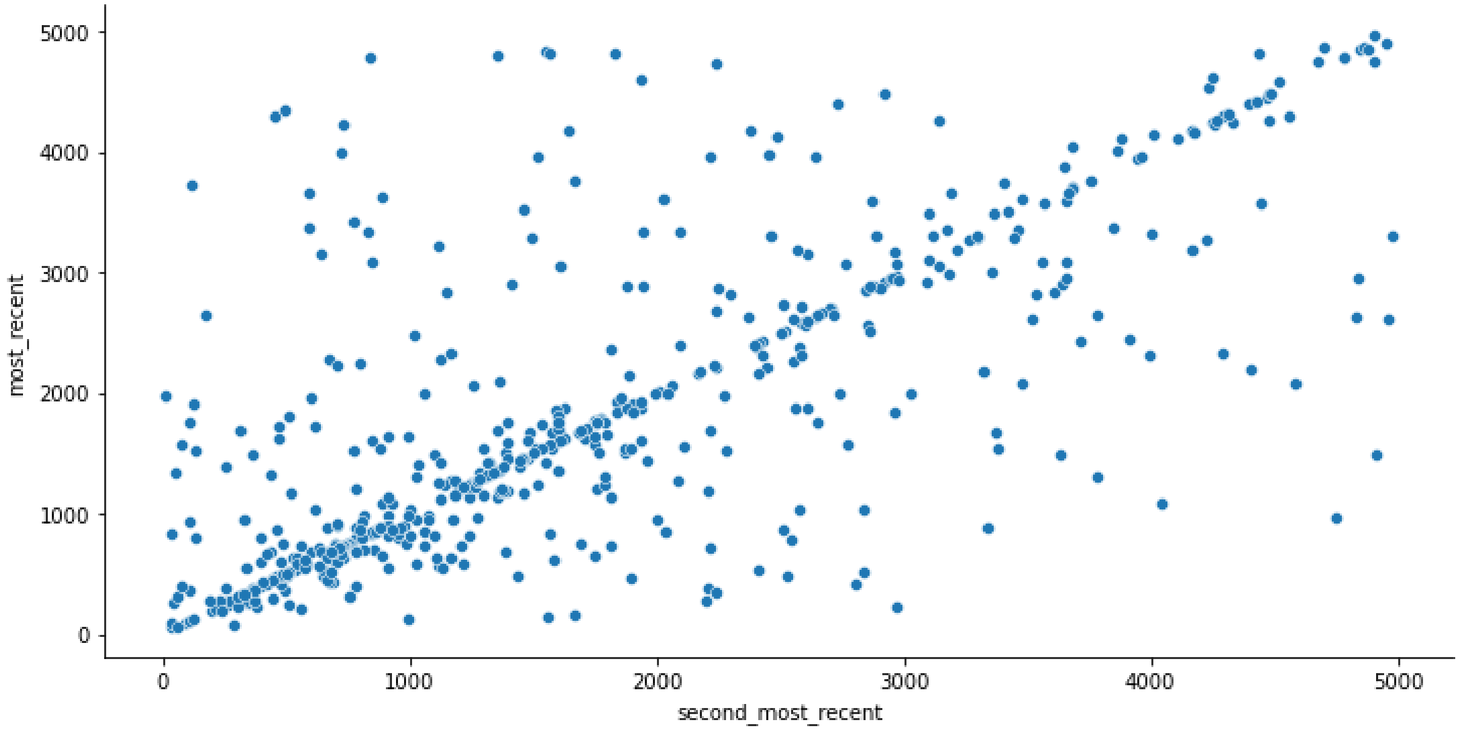
- Understanding more about the car classes

Data Insights 3



- Filter based on lower car segment
- if color is defined, it tends to get better price

Data Insights 4



- Filter based on lower car segment
- Recent price has linear corelation with the last paid price

Model Outcome

Model Used

- Random Forest

Outcome - Validation Data

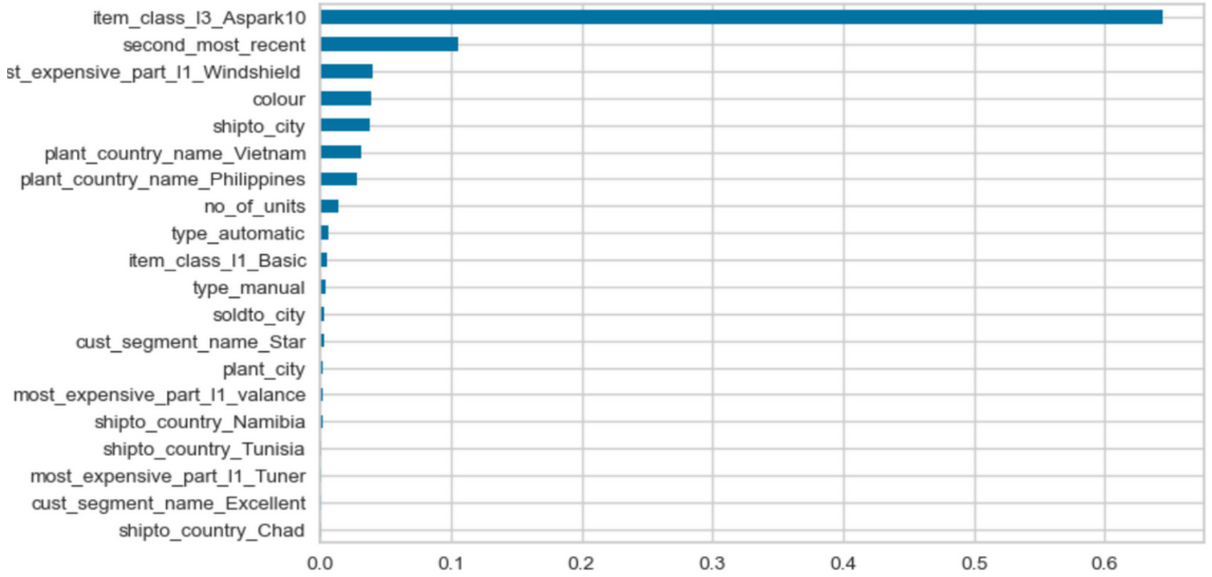
- 95% of the time, the model is able to get the price close to the actual price
- Model price brought more revenue as predicted revenues are higher than actual revenue

Test Model only on retail data

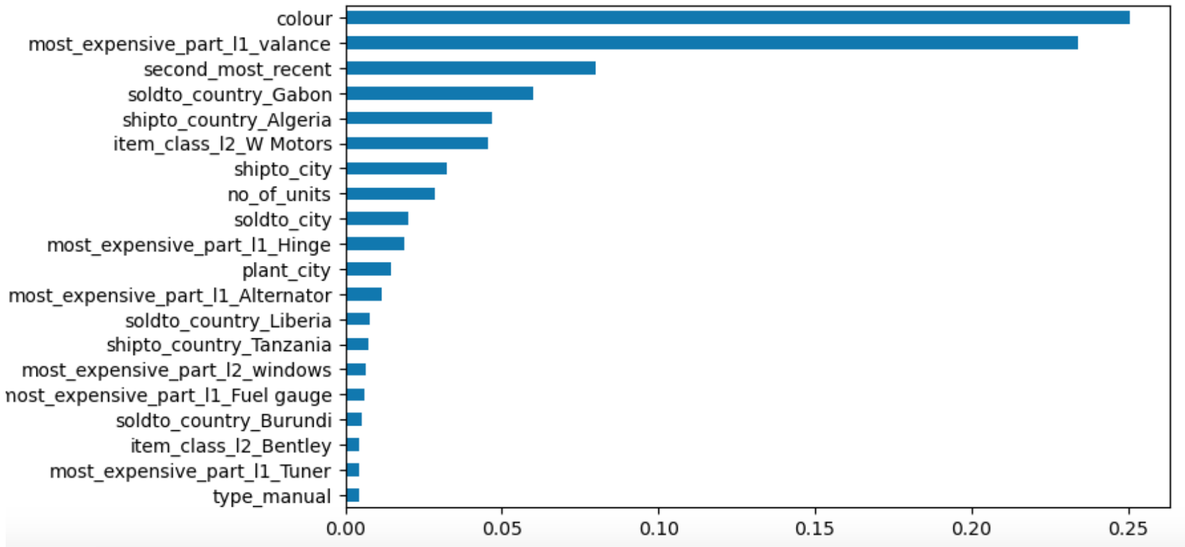
- 65% times model predicts the price closer to the actual price
- Predicted price are in the range of +/- 322 of actual price

Experiment No.	Model	R2_Score	RMSE	Decision
1	Linear Regression	0.14	266454	Model does not fit, we need to have complex models as we have higher dimensions
2	Random Forest (default)	0.72	4867	Model has high RMSE value, removed outlier class features from the model
3	Random Forest (Tuned)	0.95	1445	Current Model – It is overfitting

Model Decision Interpretation



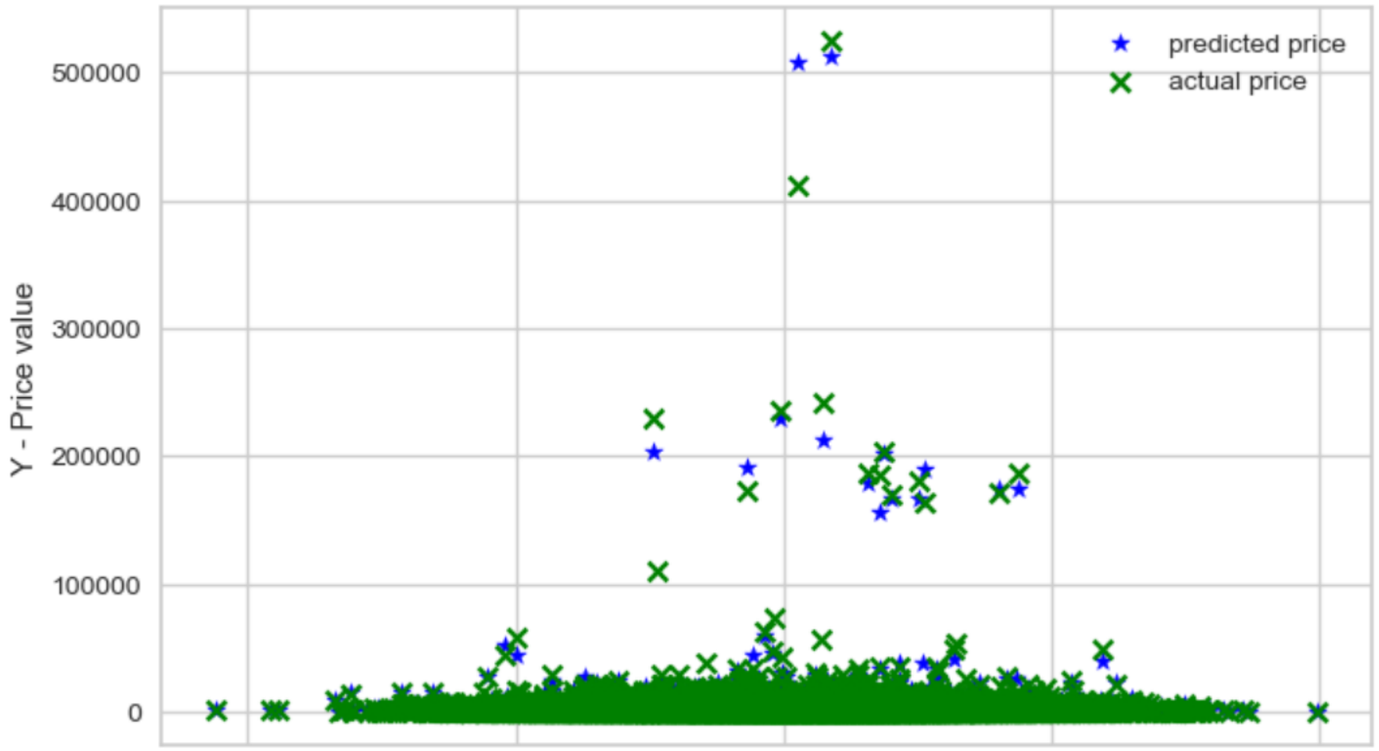
Model with complete dataset

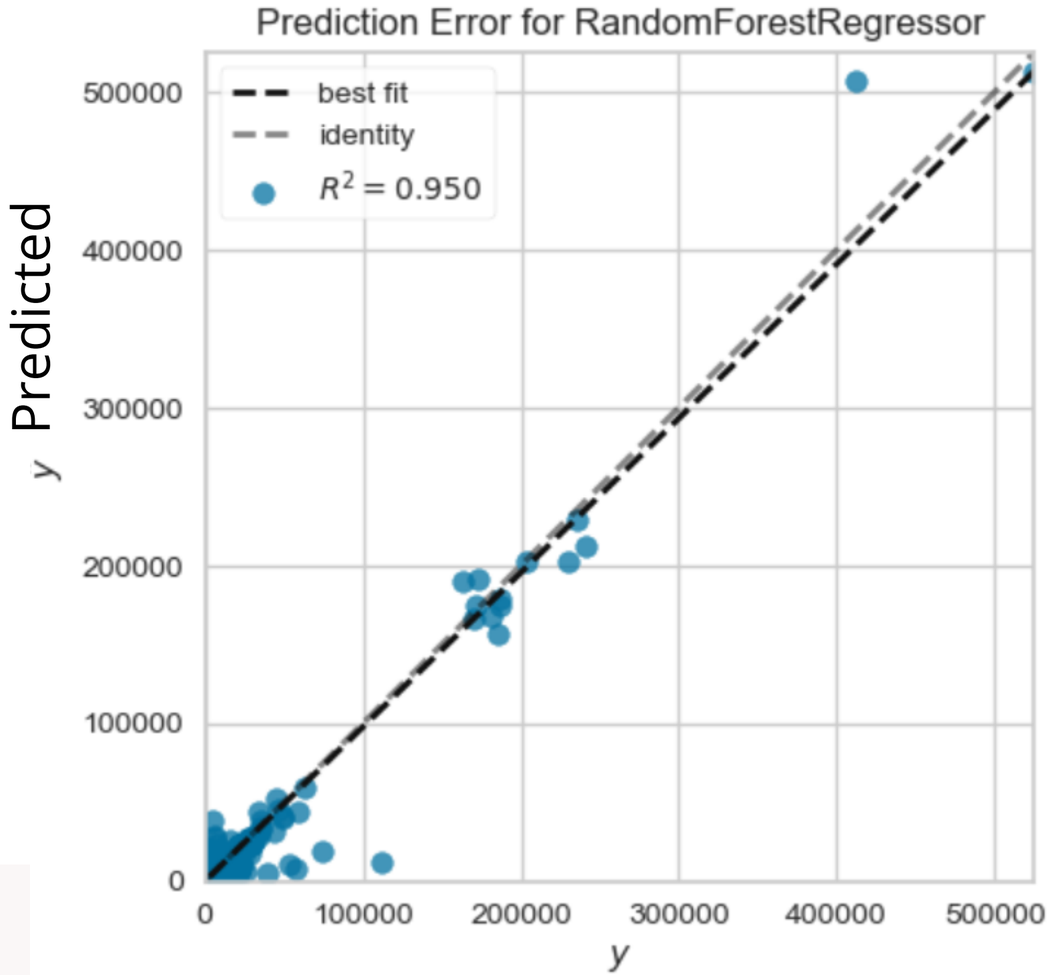


Model with removing outlier of car class

*Feature weights in taking decisions

Results Interpretation

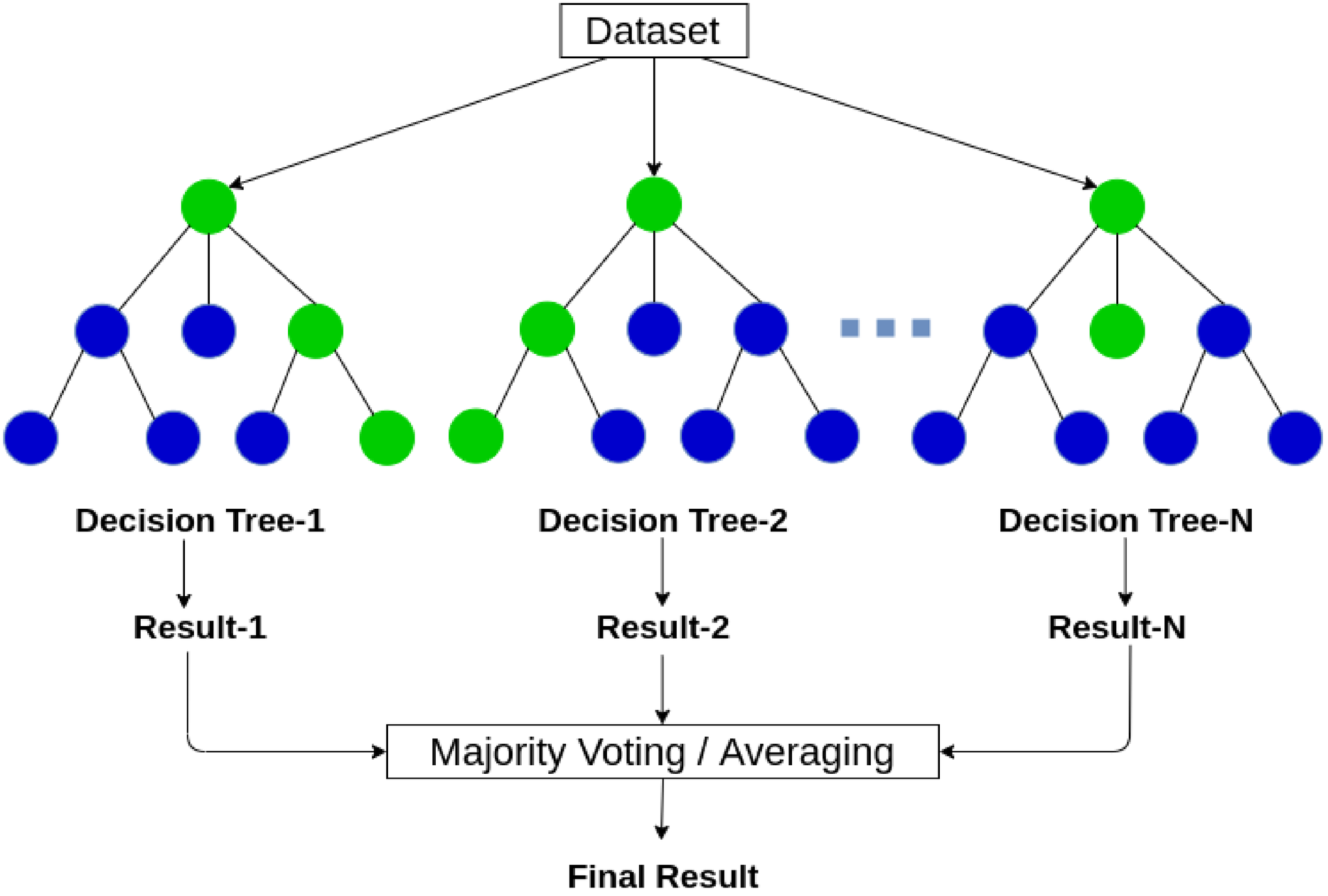




Actual

How Random Forest works?

"Everyone has different way to predict"



Img source : www.analyticsvidya.com

Next Steps

To Make it Better

REAL TIME TESTING

Model shall be deployed and tested in live supervised enviroment

EXPERIMENT MORE

Improve the model with better experiments such as incorporating feedbacks from SMEs

BUILD TRUST WITH USERS

Build more local explanations of the model

Thank You.