

Used Car Price Predictions on eBay

Team 6



Our Team



Chris
Chang

U98764222

cchang72@bu.edu

XX DRIVER LICENSE XX



Ting-Hsuan
Chen

U51060396

aliciac@bu.edu

XX DRIVER LICENSE XX



Haiyuan
Zhang

U68804194

haz155@bu.edu

XX DRIVER LICENSE XX



Mekhal
Raj R

U31160630

mekhal@bu.edu

XX DRIVER LICENSE XX



Yifang (Harry)
He

U94086175

yifanghe@bu.edu

XX DRIVER LICENSE XX



Mohammed
Al Qenae

U04709532

alqenae@bu.edu

XX DRIVER LICENSE XX

Table of Contents

01

Intro to Problem

Describing problem, the bigger picture, and dataset

02

EDA

Exploring features and correlation among variables

03

ML Models

Comparing ML models to see which one(s) performed best

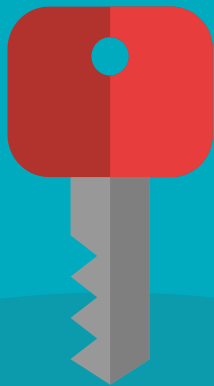
04

Conclusion

Summarizing results, challenges, and lessons we learned that can be put to practice

01

Intro to Problem



Objective / Bigger Picture

- To predict listing price of used cars on eBay based on predictors such as vehicle type, gearbox, fuel type, and kilometers on car.
- To provide insight/guidance for future buyers and sellers on the prices they should list or bid on eBay for used cars.



Our Dataset



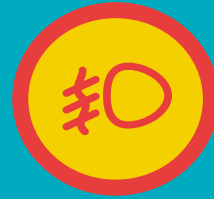
Data Source

<https://towardsdatascience.com/predicting-the-price-of-used-cars-891d13faf3fc>



Columns

19



Rows

50000



Variable Types

Character, Integer, Other

The features of our dataset

Price	Vehicle Type	Year of Registration	Gearbox	Power PS
Model	Kilometer	Fuel Type	Brand	Not Repaired Damage

Data Cleaning

1st



Dropped

Date Crawled, Seller,
Offer Type, abtest,
Month of Registration,
Date Created, Postal
Code, Last Seen, Name

2nd



Removed NAs,
zeroes, and blank
spaces

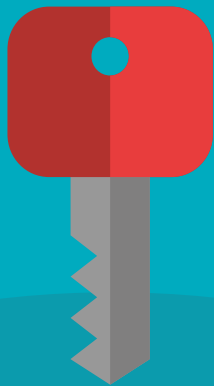
3rd



Filtered year of
registration and
price outliers

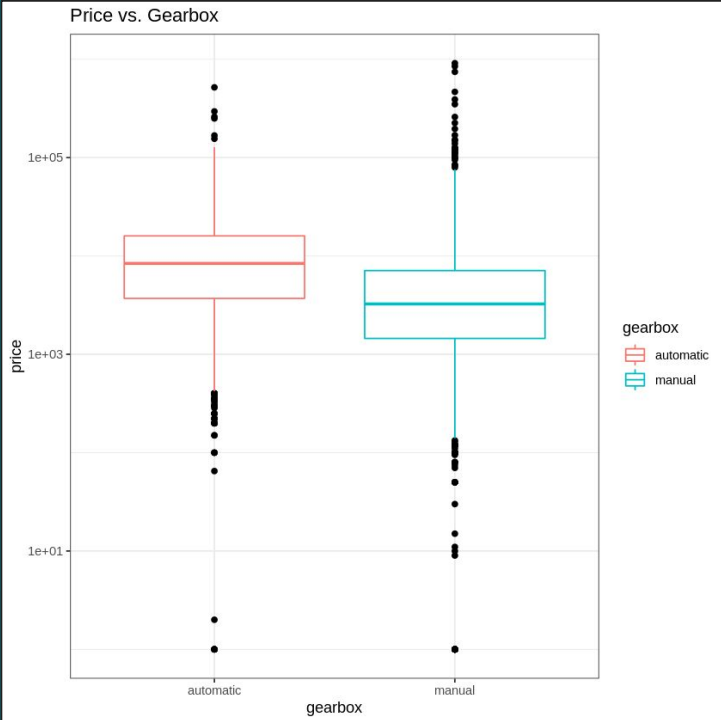
02

Exploratory Data Analysis

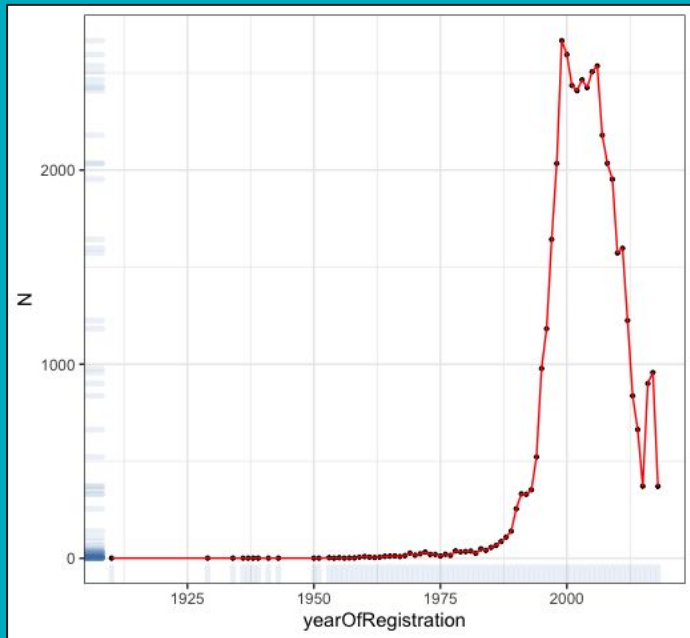


Gearbox

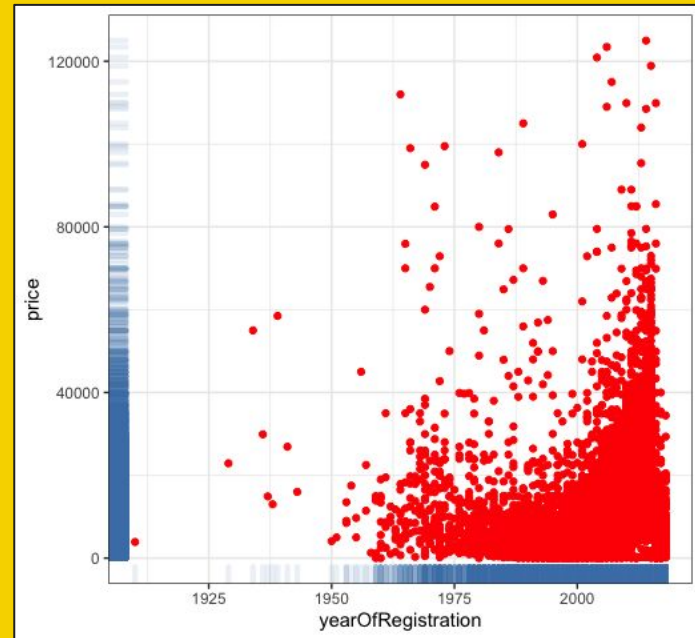
Price vs. Gearbox



Year of Registration



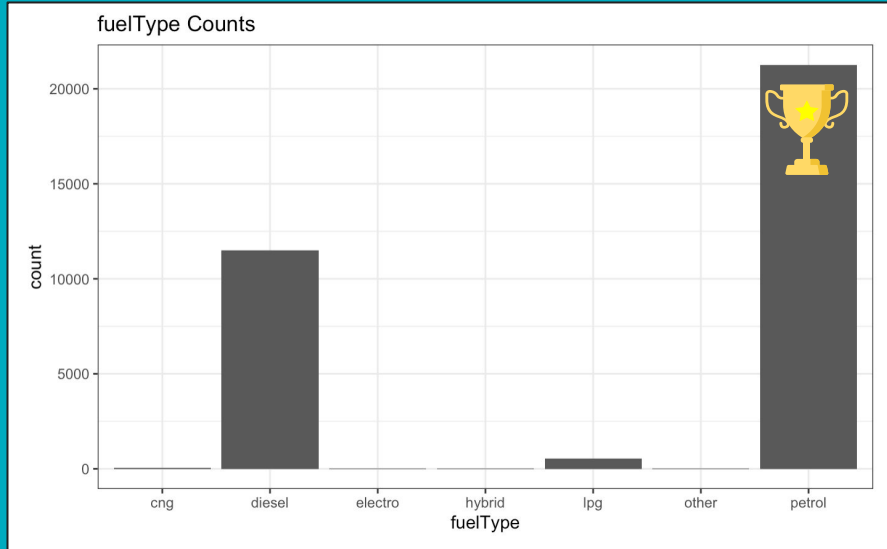
Counts of cars registered for each year



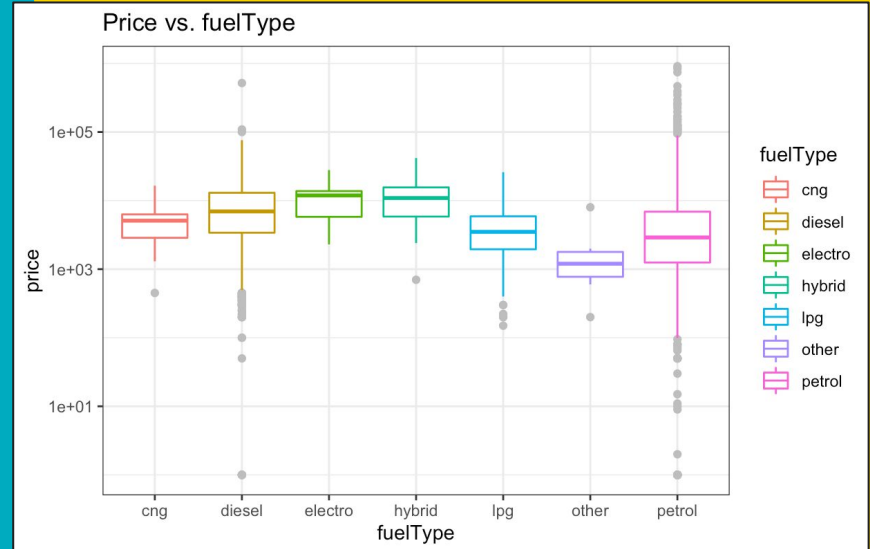
Relationship between price and registration year

Fuel Type

Count of Fuel Type

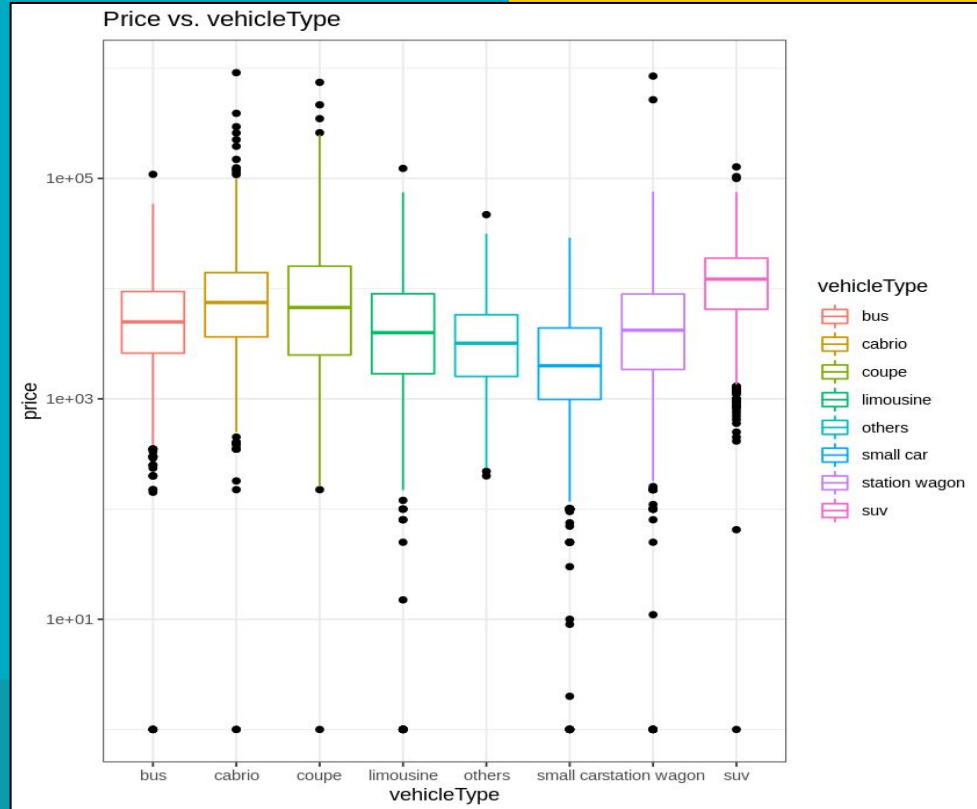


Price vs. Fuel Type



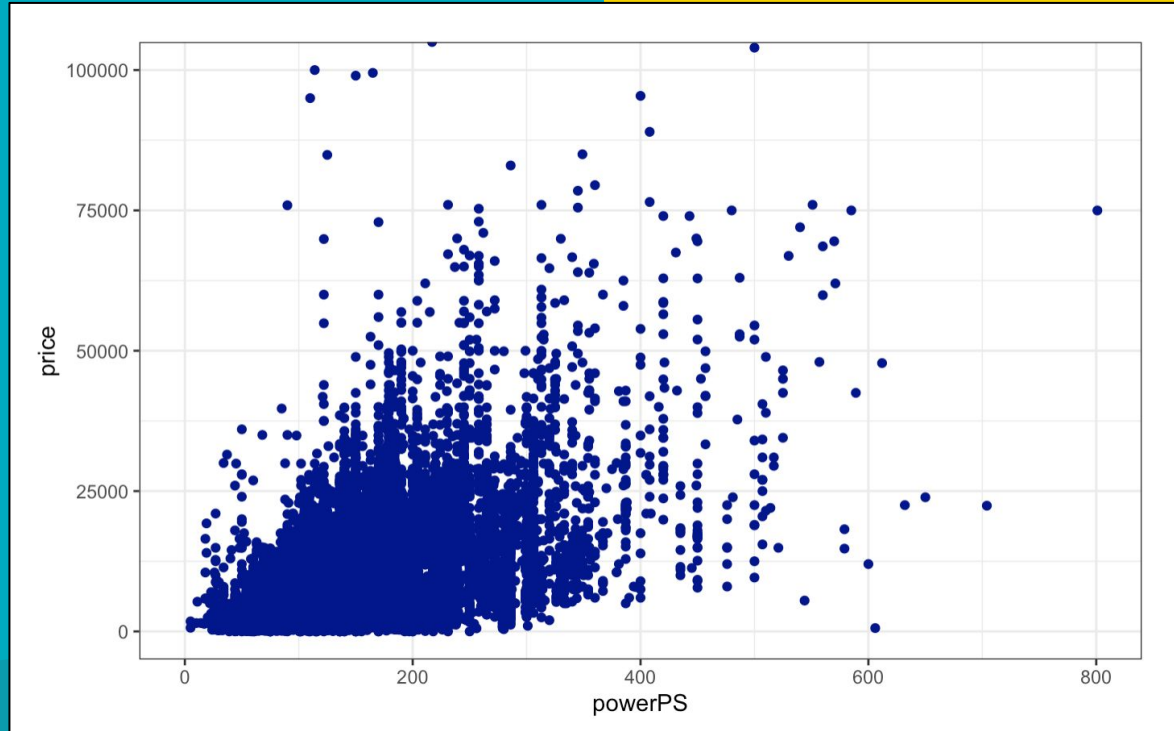
Vehicle Type

Vehicle Type vs. Price



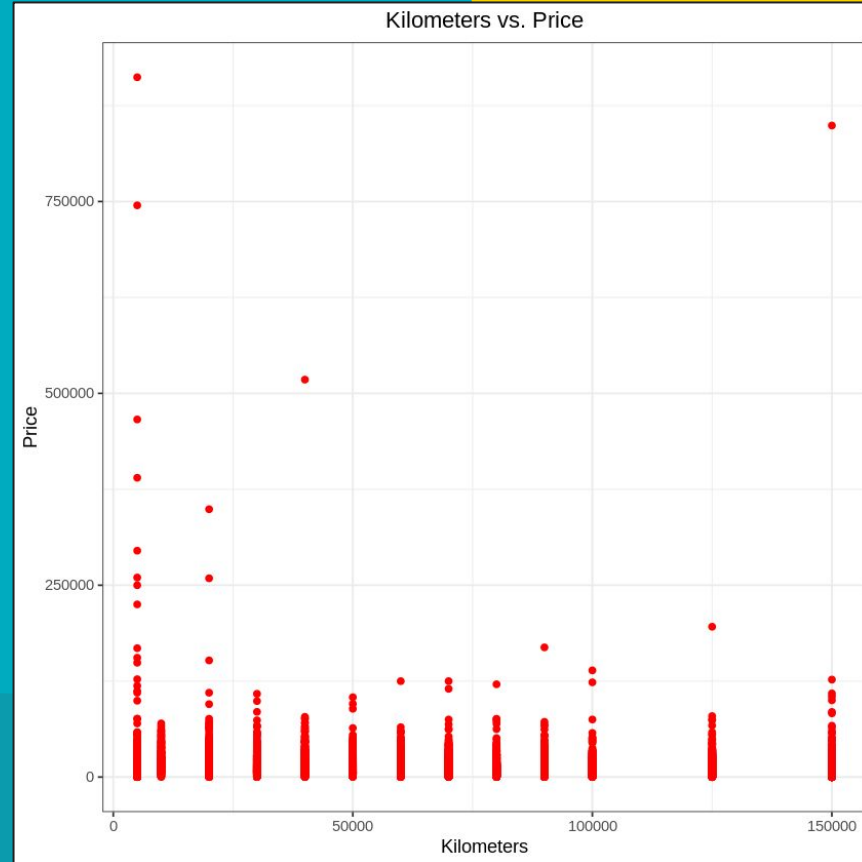
PowerPS

PowerPS vs. Price



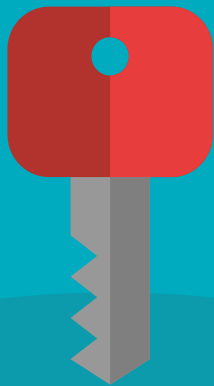
Kilometers

KM vs. Price



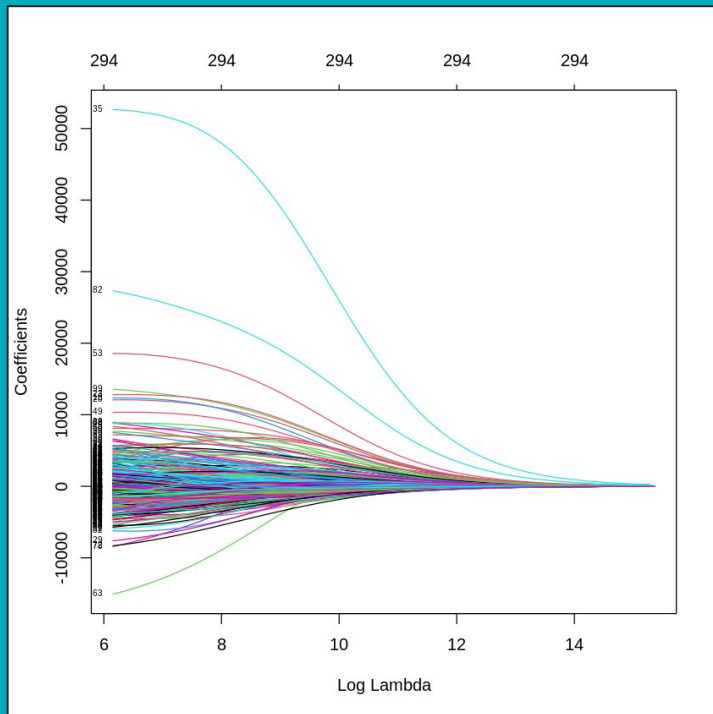
03

Machine Learning Models

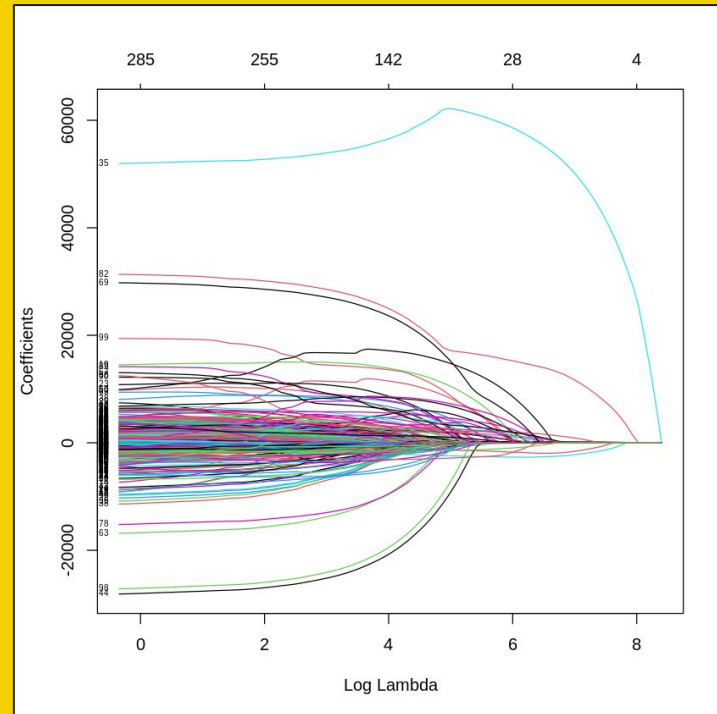


Ridge and Lasso

Test RMSE for OLS : 11938.486



Test RMSE for Ridge: 7155.59



Test RMSE for Lasso: 7195.74

The decision tree for predicting 'yearOffRegistration' starts with a root node 'yearOffRegistration < 2018' (7211, 100%). The tree splits into two main branches: 'yearOffRegistration < 2018' (left) and 'yearOffRegistration > 2018' (right). The left branch further splits into 'powerPS < 398' and 'powerPS >= 398'. The right branch splits into 'powerPS < 312' and 'powerPS >= 312'. The tree continues to split based on various features like 'brand_porsche', 'kilometer', 'yearOffRegistration', 'gearbox_manual', and 'fastType_disual', leading to terminal nodes with counts and percentages.

```

graph TD
    Root["yearOffRegistration < 2018  
7211  
100%"]
    Root --> L1L["powerPS < 398  
9102  
85%"]
    Root --> L1R["powerPS >= 312  
1781  
17%"]
    
    L1L --> L2L1["powerPS < 230  
4980  
82%"]
    L1L --> L2L2["gearbox_manual = 0  
1581  
2%"]
    
    L2L1 --> L3L1["yearOffRegistration < 2885  
3883  
24%"]
    L2L1 --> L3L2["powerPS < 122  
7247  
19%"]
    
    L3L1 --> L4L1["brand_porsche = 0  
3818  
55%"]
    L3L1 --> L4L2["yearOffRegistration >= 1880  
2739  
55%"]
    
    L4L1 --> L5L1["powerPS < 128  
5851  
54%"]
    L4L1 --> L5L2["powerPS < 148  
1581  
1%"]
    
    L5L1 --> L6L1["yearOffRegistration < 2012  
4032  
17%"]
    L5L1 --> L6L2["yearOffRegistration >= 2012  
2027  
37%"]
    
    L6L1 --> L7L1["1650  
28%"]
    L6L1 --> L7L2["3048  
15%"]
    L6L1 --> L7L3["5125  
8%"]
    L6L1 --> L7L4["3111  
3%"]
    L6L1 --> L7L5["5581  
5%"]
    L6L1 --> L7L6["8001  
5%"]
    L6L1 --> L7L7["1111  
4%"]
    L6L1 --> L7L8["1611  
3%"]
    L6L1 --> L7L9["2481  
5%"]
    L6L1 --> L7L10["4111  
1%"]
    L6L1 --> L7L11["4581  
2%"]
    L6L1 --> L7L12["5811  
4%"]
    L6L1 --> L7L13["8111  
2%"]
    L6L1 --> L7L14["1811  
2%"]
    L6L1 --> L7L15["2011  
1%"]
    L6L1 --> L7L16["2711  
1%"]
    L6L1 --> L7L17["3111  
3%"]
    L6L1 --> L7L18["3811  
5%"]
    L6L1 --> L7L19["4811  
5%"]
    L6L1 --> L7L20["5811  
5%"]
    L6L1 --> L7L21["6811  
5%"]
    L6L1 --> L7L22["7811  
5%"]
    L6L1 --> L7L23["8811  
5%"]
    L6L1 --> L7L24["9811  
5%"]
    
    L5L2 --> L8L1["yearOffRegistration < 2006  
1111  
7%"]
    L5L2 --> L8L2["yearOffRegistration >= 1889  
6018  
4%"]
    L5L2 --> L8L3["powerPS < 185  
1581  
1%"]
    L5L2 --> L8L4["kilometer <= 1581  
1311  
8%"]
    L5L2 --> L8L5["kilometer >= 2581  
2111  
1%"]
    
    L8L1 --> L9L1["1611  
3%"]
    L8L1 --> L9L2["2481  
5%"]
    L8L1 --> L9L3["4111  
1%"]
    L8L1 --> L9L4["4581  
2%"]
    L8L1 --> L9L5["5811  
4%"]
    L8L1 --> L9L6["8111  
2%"]
    L8L1 --> L9L7["1811  
2%"]
    L8L1 --> L9L8["2011  
1%"]
    L8L1 --> L9L9["2711  
1%"]
    L8L1 --> L9L10["3111  
3%"]
    L8L1 --> L9L11["3811  
5%"]
    L8L1 --> L9L12["4811  
5%"]
    L8L1 --> L9L13["5811  
5%"]
    L8L1 --> L9L14["6811  
5%"]
    L8L1 --> L9L15["7811  
5%"]
    L8L1 --> L9L16["8811  
5%"]
    L8L1 --> L9L17["9811  
5%"]
    
    L8L2 --> L10L1["1611  
3%"]
    L8L2 --> L10L2["2481  
5%"]
    L8L2 --> L10L3["4111  
1%"]
    L8L2 --> L10L4["4581  
2%"]
    L8L2 --> L10L5["5811  
4%"]
    L8L2 --> L10L6["8111  
2%"]
    L8L2 --> L10L7["1811  
2%"]
    L8L2 --> L10L8["2011  
1%"]
    L8L2 --> L10L9["2711  
1%"]
    L8L2 --> L10L10["3111  
3%"]
    L8L2 --> L10L11["3811  
5%"]
    L8L2 --> L10L12["4811  
5%"]
    L8L2 --> L10L13["5811  
5%"]
    L8L2 --> L10L14["6811  
5%"]
    L8L2 --> L10L15["7811  
5%"]
    L8L2 --> L10L16["8811  
5%"]
    L8L2 --> L10L17["9811  
5%"]
    
    L8L3 --> L11L1["1611  
3%"]
    L8L3 --> L11L2["2481  
5%"]
    L8L3 --> L11L3["4111  
1%"]
    L8L3 --> L11L4["4581  
2%"]
    L8L3 --> L11L5["5811  
4%"]
    L8L3 --> L11L6["8111  
2%"]
    L8L3 --> L11L7["1811  
2%"]
    L8L3 --> L11L8["2011  
1%"]
    L8L3 --> L11L9["2711  
1%"]
    L8L3 --> L11L10["3111  
3%"]
    L8L3 --> L11L11["3811  
5%"]
    L8L3 --> L11L12["4811  
5%"]
    L8L3 --> L11L13["5811  
5%"]
    L8L3 --> L11L14["6811  
5%"]
    L8L3 --> L11L15["7811  
5%"]
    L8L3 --> L11L16["8811  
5%"]
    L8L3 --> L11L17["9811  
5%"]
    
    L8L4 --> L12L1["1611  
3%"]
    L8L4 --> L12L2["2481  
5%"]
    L8L4 --> L12L3["4111  
1%"]
    L8L4 --> L12L4["4581  
2%"]
    L8L4 --> L12L5["5811  
4%"]
    L8L4 --> L12L6["8111  
2%"]
    L8L4 --> L12L7["1811  
2%"]
    L8L4 --> L12L8["2011  
1%"]
    L8L4 --> L12L9["2711  
1%"]
    L8L4 --> L12L10["3111  
3%"]
    L8L4 --> L12L11["3811  
5%"]
    L8L4 --> L12L12["4811  
5%"]
    L8L4 --> L12L13["5811  
5%"]
    L8L4 --> L12L14["6811  
5%"]
    L8L4 --> L12L15["7811  
5%"]
    L8L4 --> L12L16["8811  
5%"]
    L8L4 --> L12L17["9811  
5%"]
    
    L8L5 --> L13L1["1611  
3%"]
    L8L5 --> L13L2["2481  
5%"]
    L8L5 --> L13L3["4111  
1%"]
    L8L5 --> L13L4["4581  
2%"]
    L8L5 --> L13L5["5811  
4%"]
    L8L5 --> L13L6["8111  
2%"]
    L8L5 --> L13L7["1811  
2%"]
    L8L5 --> L13L8["2011  
1%"]
    L8L5 --> L13L9["2711  
1%"]
    L8L5 --> L13L10["3111  
3%"]
    L8L5 --> L13L11["3811  
5%"]
    L8L5 --> L13L12["4811  
5%"]
    L8L5 --> L13L13["5811  
5%"]
    L8L5 --> L13L14["6811  
5%"]
    L8L5 --> L13L15["7811  
5%"]
    L8L5 --> L13L16["8811  
5%"]
    L8L5 --> L13L17["9811  
5%"]
    
    L2L2 --> L14L1["kilometer <= 1581  
2581  
2%"]
    L2L2 --> L14L2["kilometer >= 2581  
1481  
2%"]
    
    L14L1 --> L15L1["1611  
3%"]
    L14L1 --> L15L2["2481  
5%"]
    L14L1 --> L15L3["4111  
1%"]
    L14L1 --> L15L4["4581  
2%"]
    L14L1 --> L15L5["5811  
4%"]
    L14L1 --> L15L6["8111  
2%"]
    L14L1 --> L15L7["1811  
2%"]
    L14L1 --> L15L8["2011  
1%"]
    L14L1 --> L15L9["2711  
1%"]
    L14L1 --> L15L10["3111  
3%"]
    L14L1 --> L15L11["3811  
5%"]
    L14L1 --> L15L12["4811  
5%"]
    L14L1 --> L15L13["5811  
5%"]
    L14L1 --> L15L14["6811  
5%"]
    L14L1 --> L15L15["7811  
5%"]
    L14L1 --> L1
```

Regression Tree

MSE Train

42,536,188

MSE Test

135,081,915

RMSE Train

6,521.977

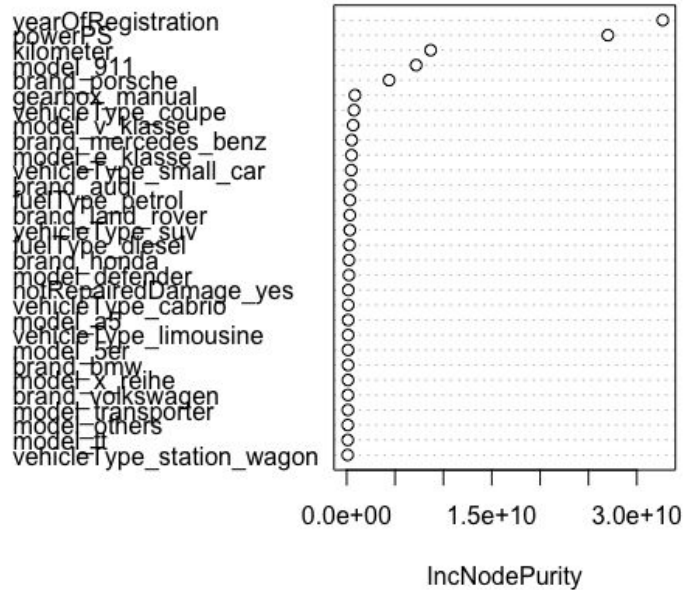
RMSE Test

11,622.47



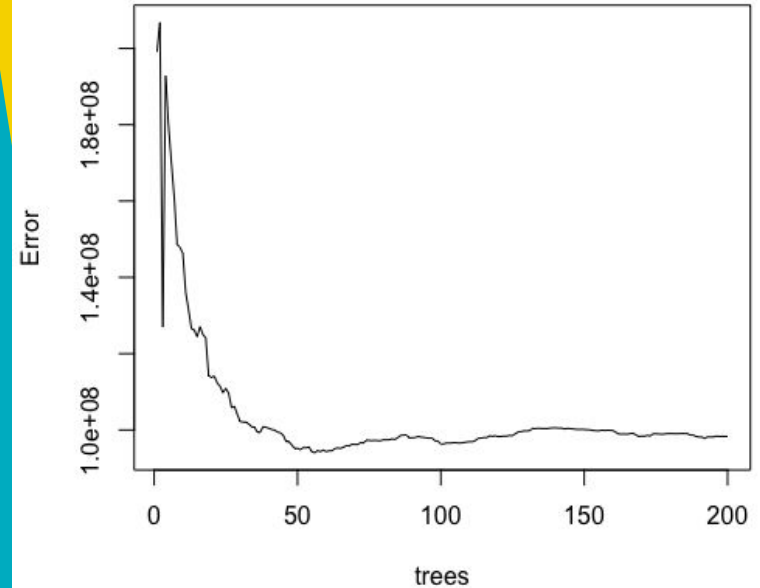
Random Forest

fit.rndfor



Predictor Importance

fit.rndfor



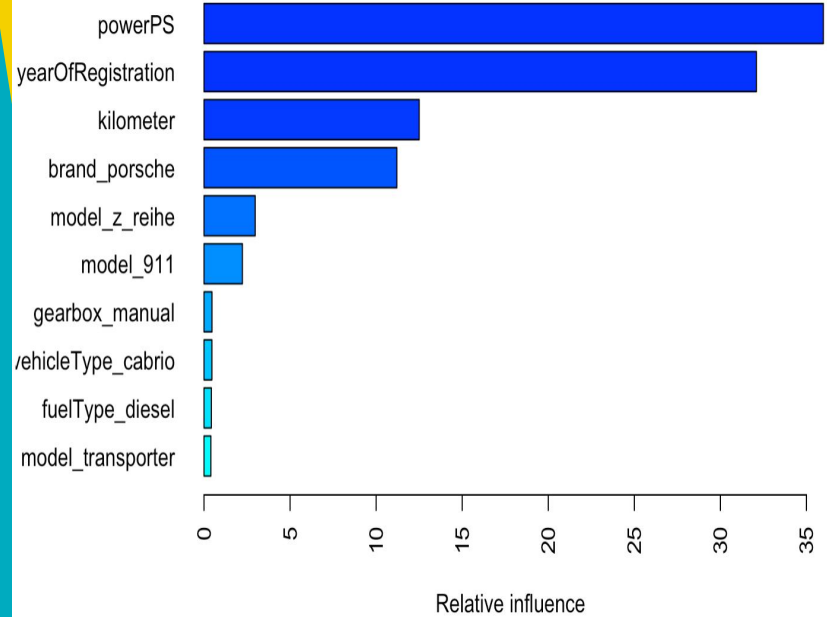
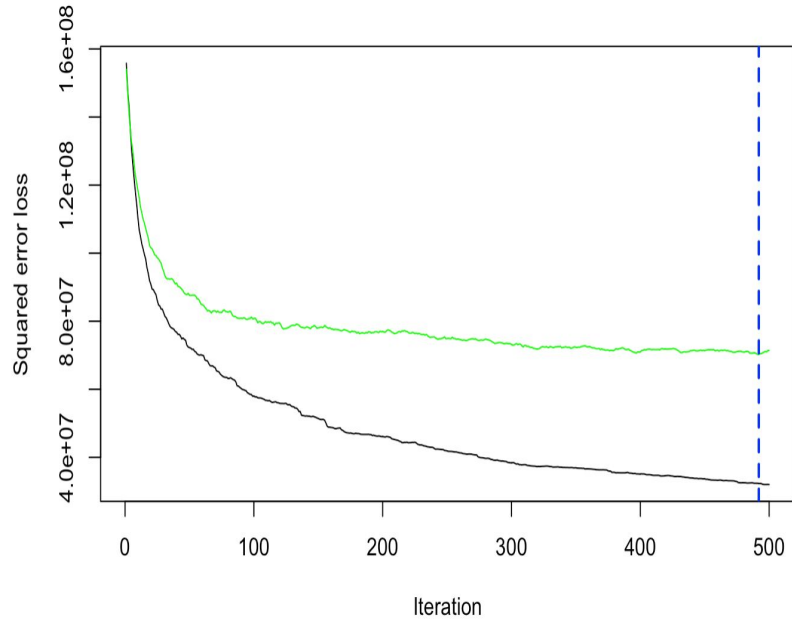
OOB Error

Random Forest

Train RMSE: 2091.875

Test RMSE: 4913.075

Boosting



Min Test RMSE:
8379.893

MSE Comparisons

OLS

RMSE Train : 9609.137

RMSE Test : 11938.49

Ridge

RMSE Train : 11442.35

RMSE Test : 7155.59

Lasso

RMSE Train : 11426.85

RMSE Test : 7195.74

Regression
Tree

RMSE Train : 6521.98

RMSE Test : 11622.47

Random
Forests

RMSE Train : 2091.875

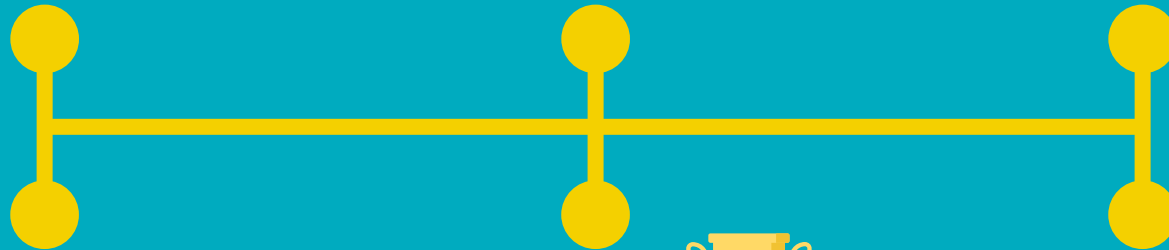
RMSE Test : 4913.075



GBM
Boosting

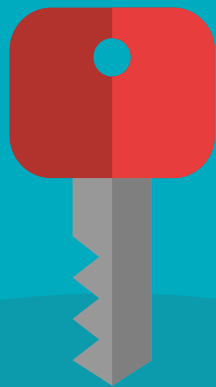
RMSE Train : 12395.44

RMSE Test : 8379.893



04

Conclusion





**Our biggest
challenge**

THANKS

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

