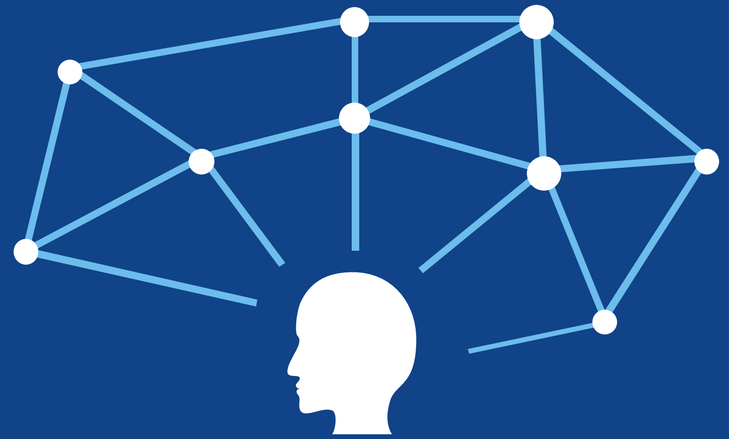
# Predict Price of the Used Cars

Amjad Alotaibi

# Problem Statement

**This model will help predict the price of used cars and help buyers know the real value of cars, so prices cannot be manipulated.**

# Data Source



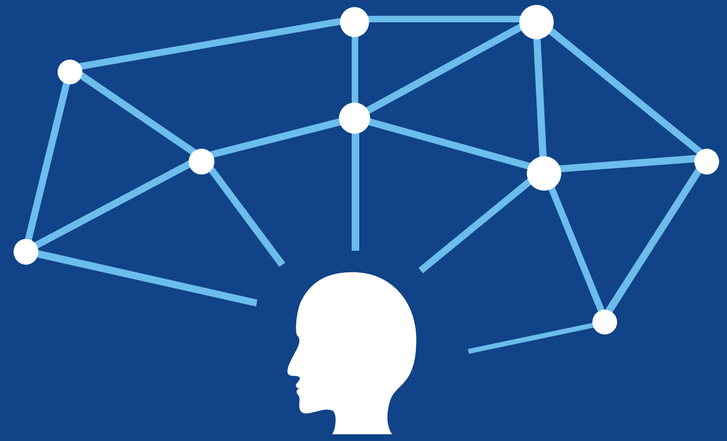
“

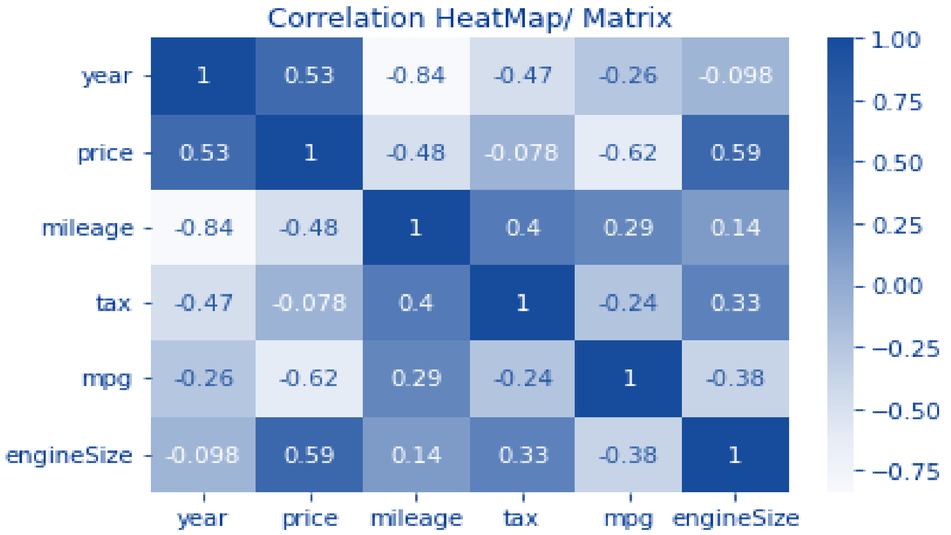
I used the dataset from Kaggle.com. It contains information about used Audi cars with many different features such as model, year, price, engine size, etc. This dataset contains more than 10669 records.

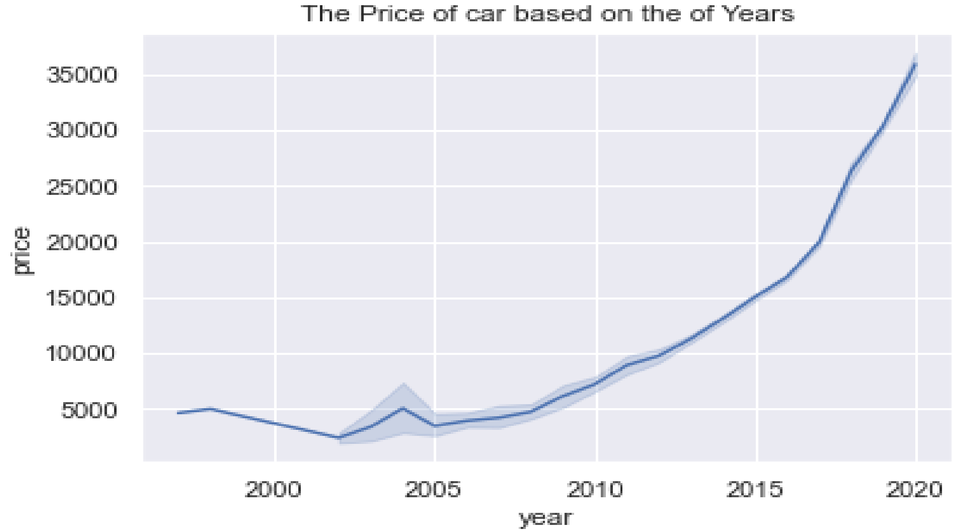
# Data Preparing

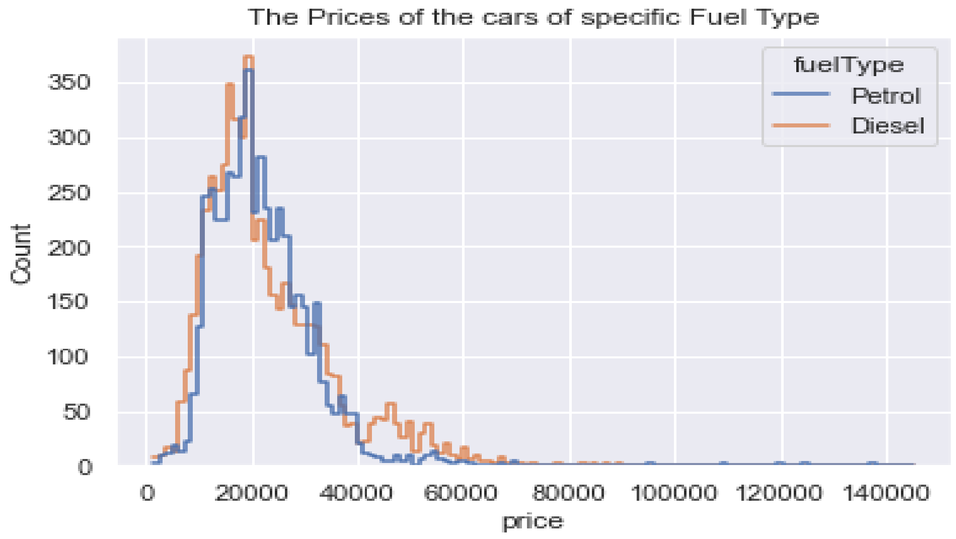
- Missing values
- Outliers
- Label Encoding Technique

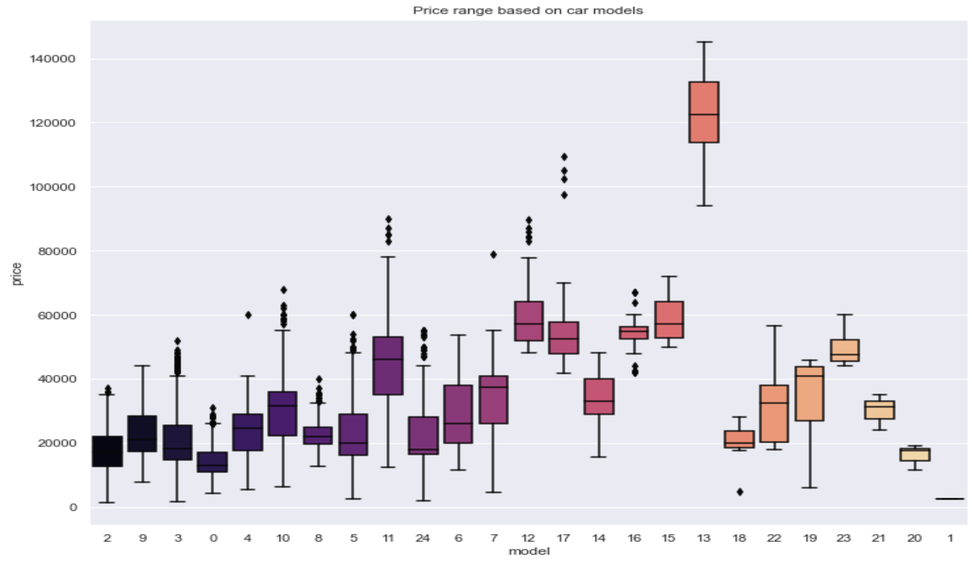
# Exploratory Data Analysis




Correlation HeatMap/ Matrix
year
price
mileage
tax
mpg
engineSize
1
0.53
-0.84
-0.47
-0.26
-0.098
0.53
1
-0.48
-0.078
-0.62
0.59
-0.84
-0.48
1
0.4
0.29
0.14
-0.47
-0.078
0.4
1
-0.24
0.33
-0.26
-0.62
0.29
-0.24
1
-0.38
-0.098
0.59
0.14
0.33
-0.38
1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


The Price of car based on the of Years
price
year
35000
30000
25000
20000
15000
10000
5000
2000
2005
2010
2015
2020


The Prices of the cars of specific Fuel Type
fuelType
Petrol
Diesel
Count
price
350
300
250
200
150
100
50
0
0
20000
40000
60000
80000
100000
120000
140000

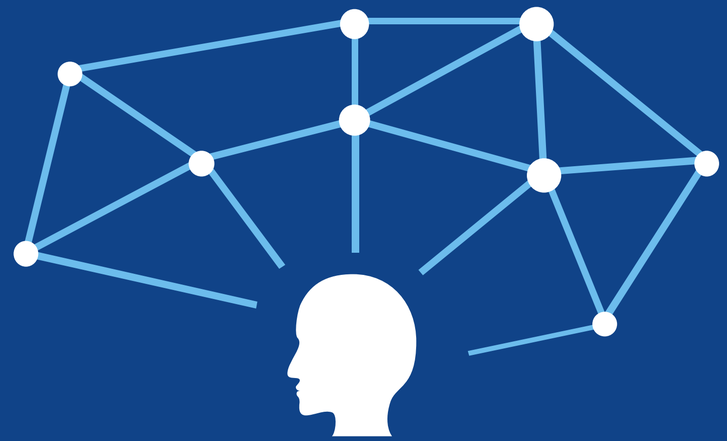

Price range based on car models
price
model
140000
120000
100000
80000
60000
40000
20000
0
2
9
3
0
4
10
8
5
11
24
6
7
12
17
14
16
15
13
18
22
19
23
21
20
1

“

# Train and Test Data

The data was split into two different parts for model creation: 65% Train and 35% Test.

# Models



# The accuracy of models

| | Linear Regression | Random Forest | CatBoost(r2) |
|---|---|---|---|
| **Accuracy** | **80.23%** | **95.56%** | **96%** |

# Conclusion

**As it is clear to us, we can say that CatBoost is the most reliable model because it has the highest accuracy rate.**

# Thank You