

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 1

Nama Mentor: Aditya Bariq Ikhsan

Nama:

- Evylia Yanuar Laily
- Taufik Syah Mauludin

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

- 1. Latar Belakang**
- 2. Explorasi Data dan Visualisasi**
- 3. Modelling**
- 4. Kesimpulan**

Latar Belakang

Latar Belakang Project

Sumber Data: <https://www.kaggle.com/datasets/hellbuoy/car-price-prediction>

Problem: **Regression**

Tujuan:

- Menentukan faktor-faktor yang mempengaruhi harga mobil

Explorasi Data dan Visualisasi

Business Understanding

- Perusahaan mobil Geely Auto Cina bercita-cita untuk memasuki pasar AS dengan mendirikan unit manufaktur dan memproduksi mobil secara lokal untuk memberikan persaingan di AS dan Eropa.
- Secara khusus, perusahaan ingin memahami faktor-faktor yang mempengaruhi harga mobil di pasar Amerika, karena mungkin sangat berbeda dari pasar Cina.



Data Cleansing

- Tidak terdapat missing value
- Tidak terdapat duplicated data

```
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   car_ID              205 non-null   int64  
1   symboling           205 non-null   int64  
2   CarName             205 non-null   object  
3   fueltype            205 non-null   object  
4   aspiration           205 non-null   object  
5   doornumber          205 non-null   object  
6   carbody             205 non-null   object  
7   drivewheel          205 non-null   object  
8   enginelocation      205 non-null   object  
9   wheelbase           205 non-null   float64 
10  carlength           205 non-null   float64 
11  carwidth            205 non-null   float64 
12  carheight           205 non-null   float64 
13  curbweight          205 non-null   int64  
14  enginetype          205 non-null   object  
15  cylindernumber      205 non-null   object  
16  enginesize          205 non-null   int64  
17  fuelsystem          205 non-null   object  
18  boreratio           205 non-null   float64 
19  stroke              205 non-null   float64 
20  compressionratio    205 non-null   float64 
21  horsepower          205 non-null   int64  
22  peakrpm             205 non-null   int64  
23  citympg             205 non-null   int64  
24  highwaympg         205 non-null   int64  
25  price               205 non-null   float64 
dtypes: float64(8), int64(8), object(10)
```


Data Cleansing

```
data unik dari CarName adalah :  
['alfa-romero giulia' 'alfa-romero stelvio' 'alfa-romero Quadrifoglio'  
'audi 100 ls' 'audi 100ls' 'audi fox' 'audi 5000' 'audi 4000'  
'audi 5000s (diesel)' 'bmw 320i' 'bmw x1' 'bmw x3' 'bmw z4' 'bmw x4'  
'bmw x5' 'chevrolet impala' 'chevrolet monte carlo' 'chevrolet vega 2300'  
'dodge rampage' 'dodge challenger se' 'dodge d200' 'dodge monaco (sw)'  
'dodge colt hardtop' 'dodge colt (sw)' 'dodge coronet custom'  
'dodge dart custom' 'dodge coronet custom (sw)' 'honda civic'  
'honda civic cvcc' 'honda accord cvcc' 'honda accord lx'  
'honda civic 1500 gl' 'honda accord' 'honda civic 1300' 'honda prelude'  
'honda civic (auto)' 'isuzu MU-X' 'isuzu D-Max' 'isuzu D-Max V-Cross'  
'jaguar xj' 'jaguar xf' 'jaguar xk' 'maxda rx3' 'maxda glc deluxe'  
'mazda rx2 coupe' 'mazda rx-4' 'mazda glc deluxe' 'mazda 626' 'mazda glc'  
'mazda rx-7 gs' 'mazda glc 4' 'mazda glc custom l' 'mazda glc custom'  
'buick electra 225 custom' 'buick century luxus (sw)' 'buick century'  
'buick skyhawk' 'buick opel isuzu deluxe' 'buick skylark'  
'buick century special' 'buick regal sport coupe (turbo)'  
'mercury cougar' 'mitsubishi mirage' 'mitsubishi lancer'  
'mitsubishi outlander' 'mitsubishi g4' 'mitsubishi mirage g4'  
'mitsubishi montero' 'mitsubishi pajero' 'Nissan versa' 'nissan gt-r'  
'nissan rogue' 'nissan latiao' 'nissan titan' 'nissan leaf' 'nissan juke'  
'nissan note' 'nissan clipper' 'nissan nv200' 'nissan dayz' 'nissan fuga'  
'nissan otti' 'nissan teana' 'nissan kicks' 'peugeot 504' 'peugeot 304'  
'peugeot 504 (sw)' 'peugeot 604sl' 'peugeot 505s turbo diesel'  
'plymouth fury iii' 'plymouth cricket' 'plymouth satellite custom (sw)'  
'plymouth fury gran sedan' 'plymouth valiant' 'plymouth duster'  
'porsche macan' 'porsche panamera' 'porsche cayenne' 'porsche boxer'  
'renault 12tl' 'renault 5 gtl' 'saab 99e' 'saab 99le' 'saab 99gle'  
'subaru' 'subaru dl' 'subaru brz' 'subaru baja' 'subaru rl' 'subaru r2'  
'subaru trezia' 'subaru tribeca' 'toyota corona mark ii' 'toyota corona'  
'toyota corolla 1200' 'toyota corona hardtop' 'toyota corolla 1600 (sw)'  
'toyota carina' 'toyota mark ii' 'toyota corolla'  
'toyota corolla liftback' 'toyota celica gt liftback'  
'toyota corolla tercel' 'toyota corona liftback' 'toyota starlet'  
'toyota tercel' 'toyota cressida' 'toyota celica gt' 'toyota tercel'  
'volkswagen rabbit' 'volkswagen 113i deluxe sedan' 'volkswagen model 111'  
'volkswagen type 3' 'volkswagen 411 (sw)' 'volkswagen super beetle'  
'volkswagen dasher' 'vw dasher' 'vw rabbit' 'volkswagen rabbit'  
'volkswagen rabbit custom' 'volvo 145e (sw)' 'volvo 144ea' 'volvo 244dl'  
'volvo 245' 'volvo 264gl' 'volvo diesel' 'volvo 246']
```

data unik dari Brand adalah :

```
['alfa-romero' 'audi' 'bmw' 'chevrolet' 'dodge' 'honda' 'isuzu' 'jaguar'  
'mazda' 'buick' 'mercury' 'mitsubishi' 'nissan' 'peugeot' 'plymouth'  
'porsche' 'renault' 'saab' 'subaru' 'toyota' 'volkswagen' 'volvo']
```

Data Cleansing

```
data unik dari symboling adalah :  
[ 3  1  2  0 -1 -2]
```

```
data unik dari symboling adalah :  
['Very_Risky' 'Neutral' 'Moderately_Risky' 'Safe' 'Moderately_Safe'  
'Very_Safe']
```

```
data unik dari fueltype adalah :  
['gas' 'diesel']
```

```
data unik dari aspiration adalah :  
['std' 'turbo']
```

```
data unik dari doornumber adalah :  
['two' 'four']
```

```
data unik dari enginelocation adalah :  
['front' 'rear']
```

```
data unik dari fueltype adalah :  
[1 0]
```

```
data unik dari aspiration adalah :  
[1 0]
```

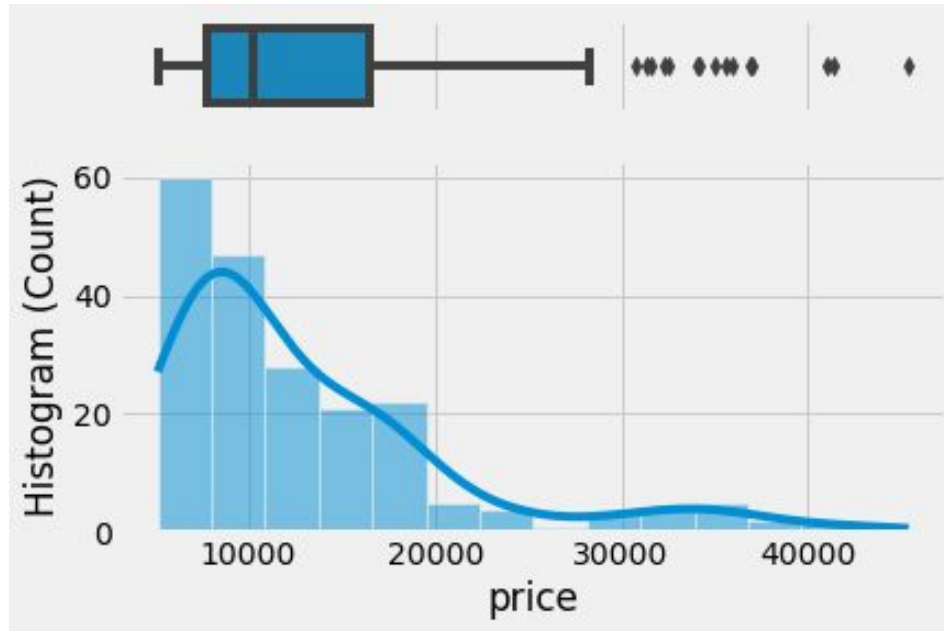
```
data unik dari doornumber adalah :  
[2 4]
```

```
data unik dari enginelocation adalah :  
[1 0]
```

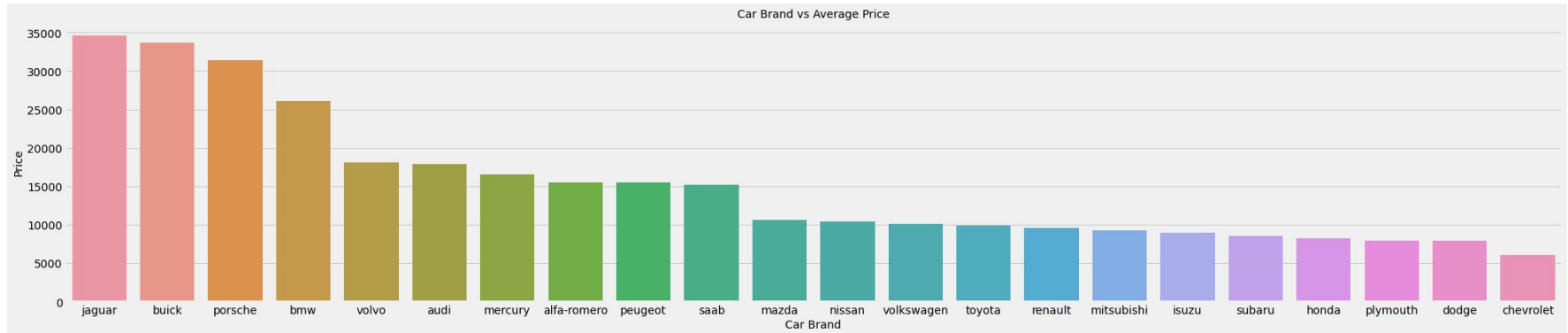
```
data unik dari cylindernumber adalah :  
['four' 'six' 'five' 'three' 'twelve' 'two' 'eight']
```

```
data unik dari cylindernumber adalah :  
[ 4  6  5  3 12  2  8]
```

Exploratory Data Analysis

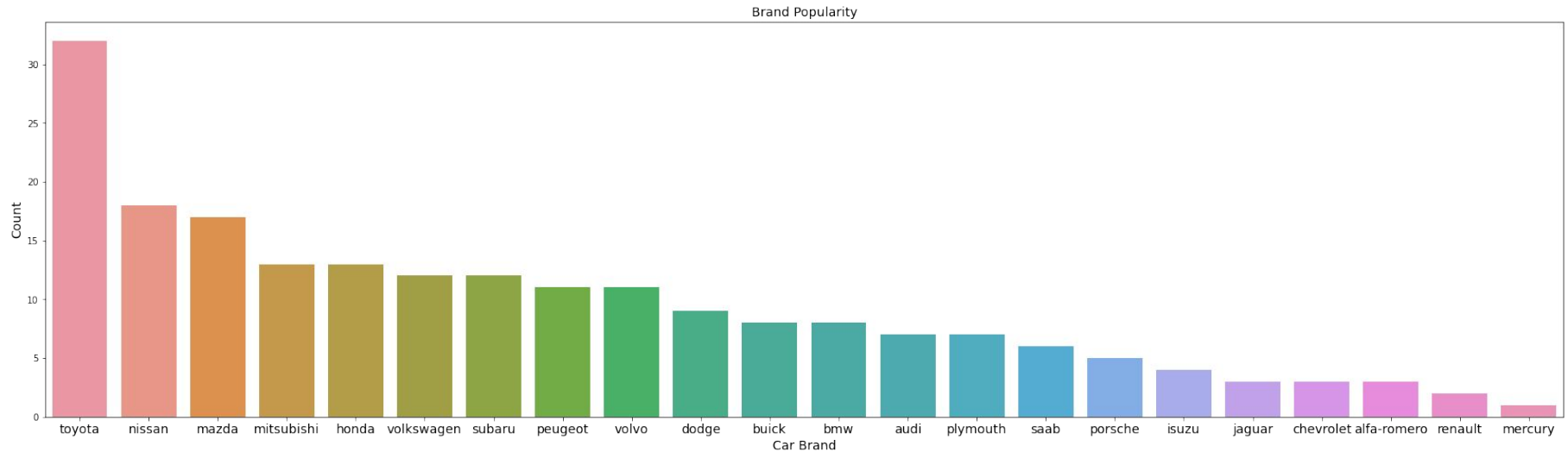


Exploratory Data Analysis



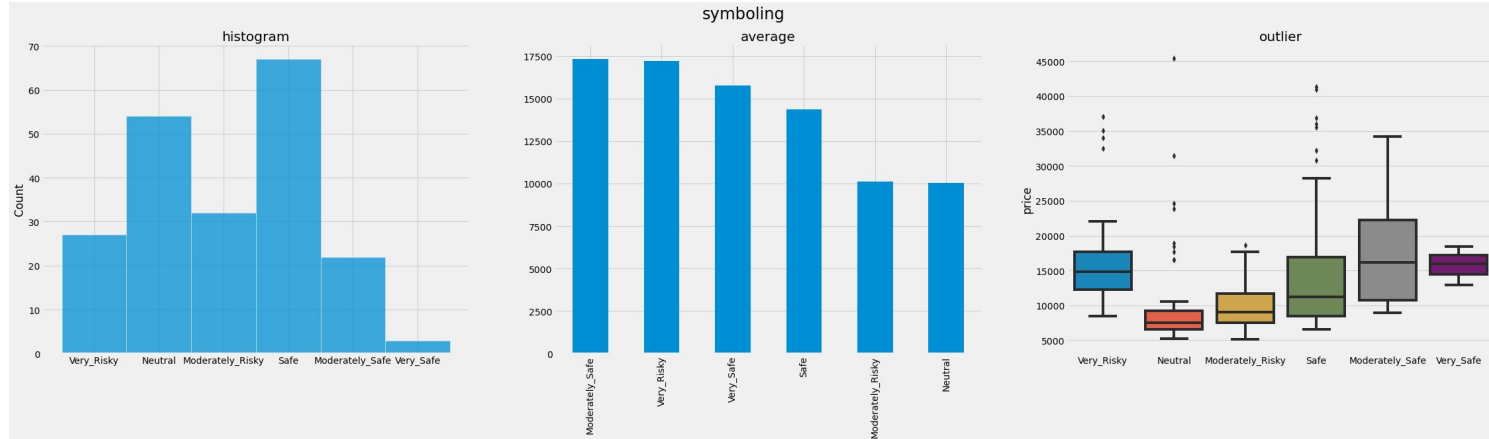
- Jaguar, Buick, dan Porsche memiliki harga rata-rata tertinggi.

Exploratory Data Analysis



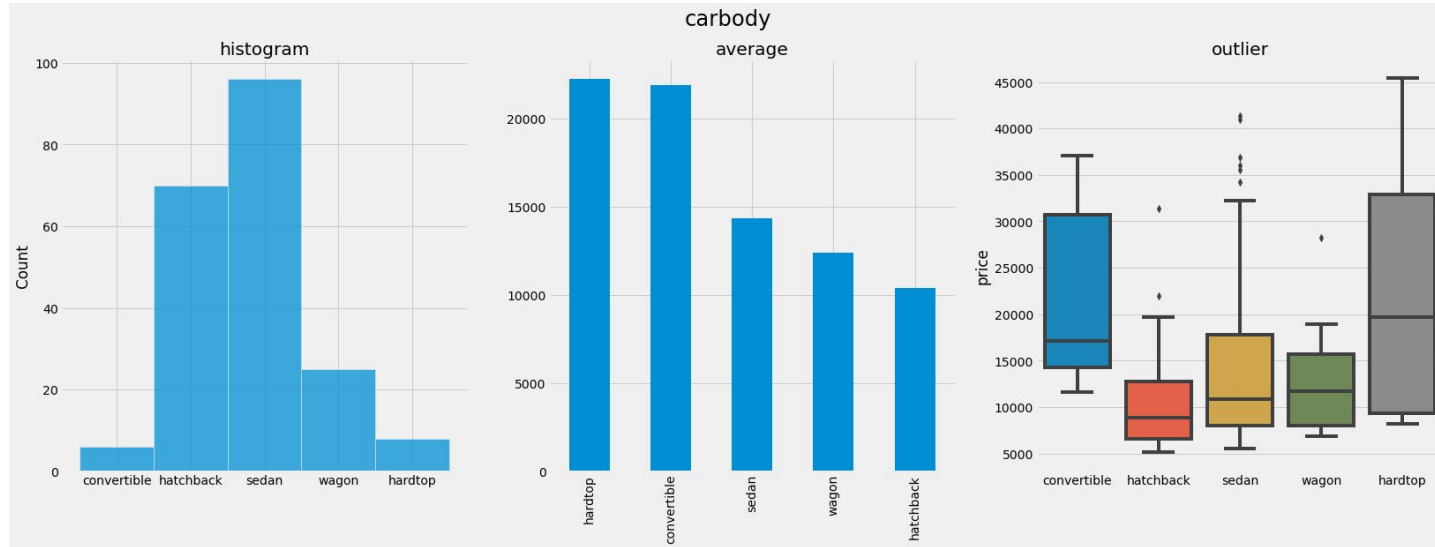
- Toyota merupakan mobil paling favorit.
- Mercury adalah Brand yang paling sedikit disukai.

Exploratory Data Analysis



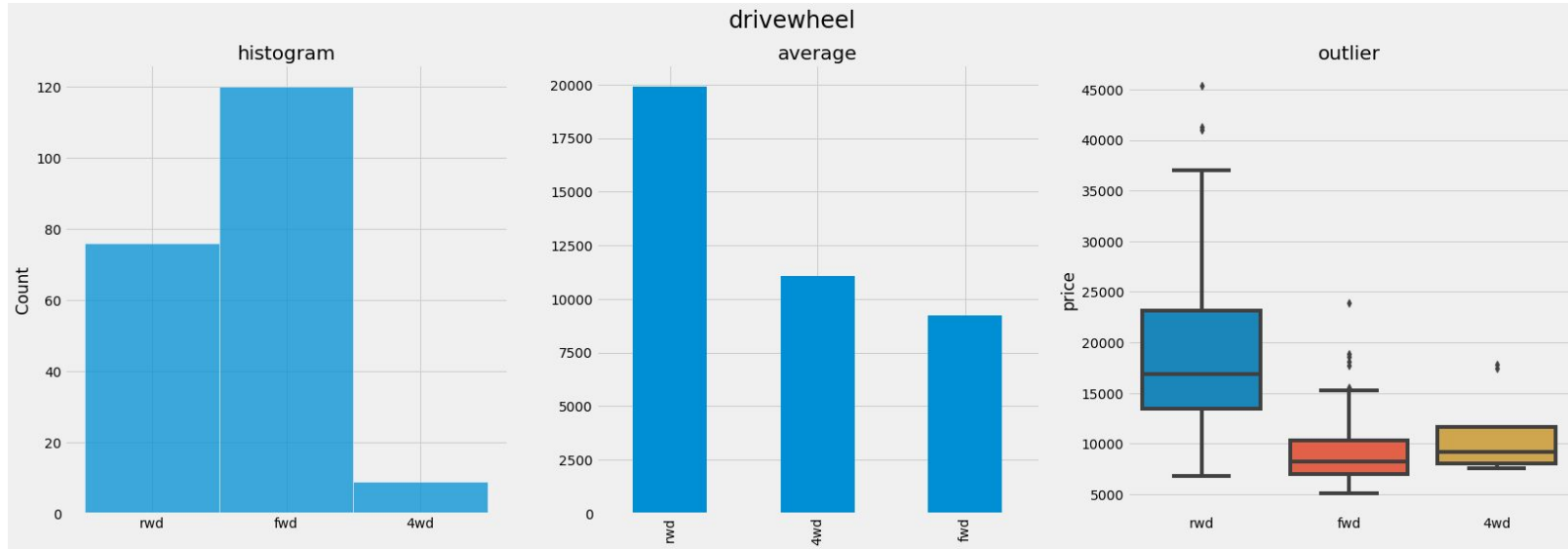
- Faktor symboling mobil yang paling banyak digunakan adalah tingkat safe.
- Tingkat Moderately_Safe dan Very_Risky memiliki harga tertinggi.
- Harga symboling Moderately_Safe lebih tinggi dibandingkan dengan yang lainnya.

Exploratory Data Analysis



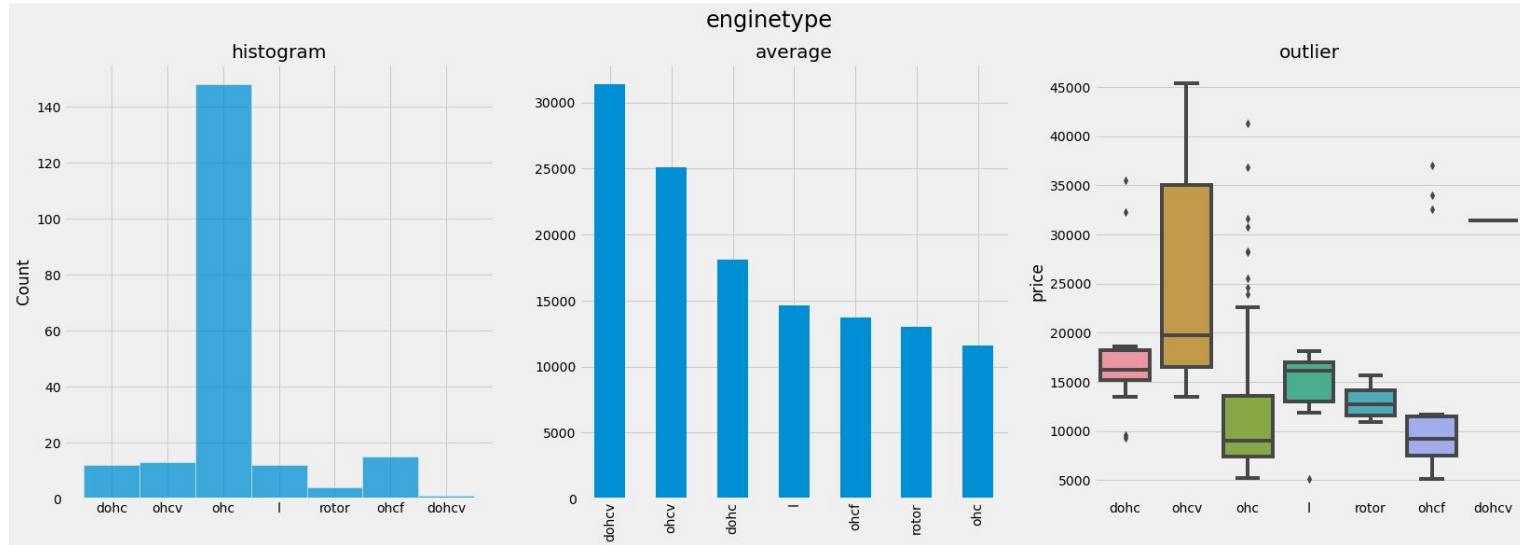
- Sebagian besar carbody mobil adalah sedan dan hatchback.
- Mobil dengan carbody hardtop dan convertible memiliki harga tertinggi.
- Semua jenis carbody relatif lebih murah dibandingkan dengan carbody hardtop.

Exploratory Data Analysis



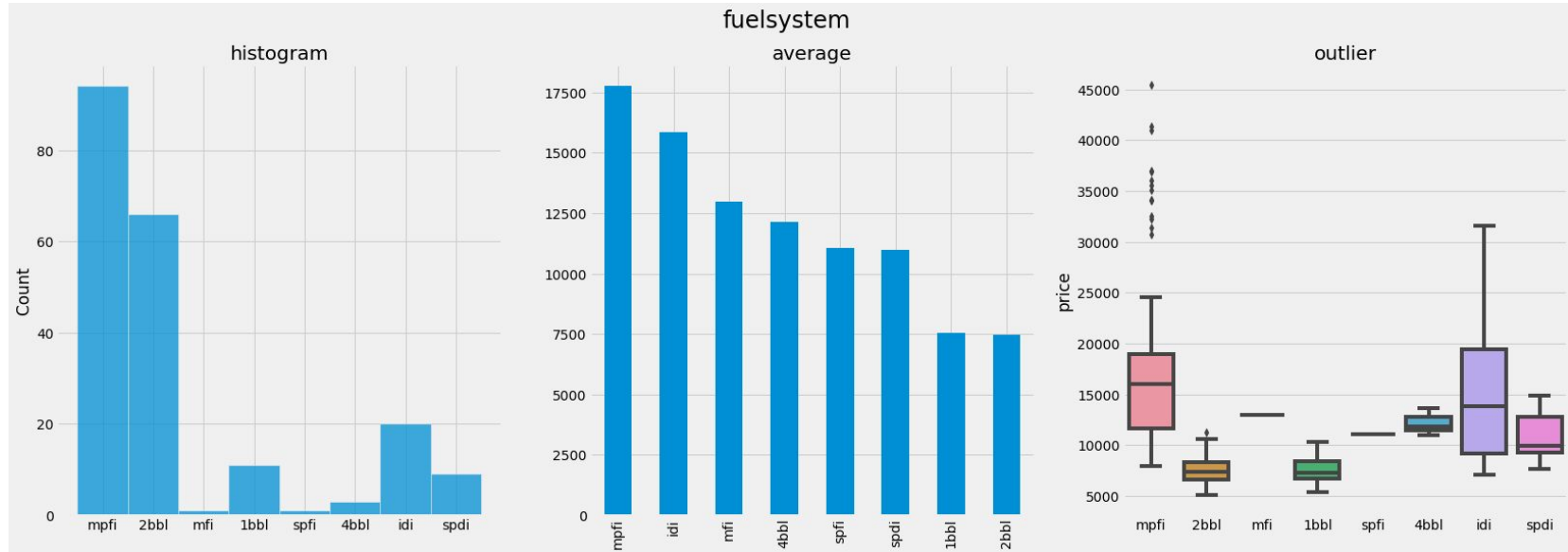
- Sebagian besar drivewheel mobil merupakan fwd dan rwd.
- Mobil dengan drivewheel rwd memiliki harga tertinggi.
- Semua jenis drivewheel relatif lebih murah dibandingkan dengan drivewheel rwd.

Exploratory Data Analysis



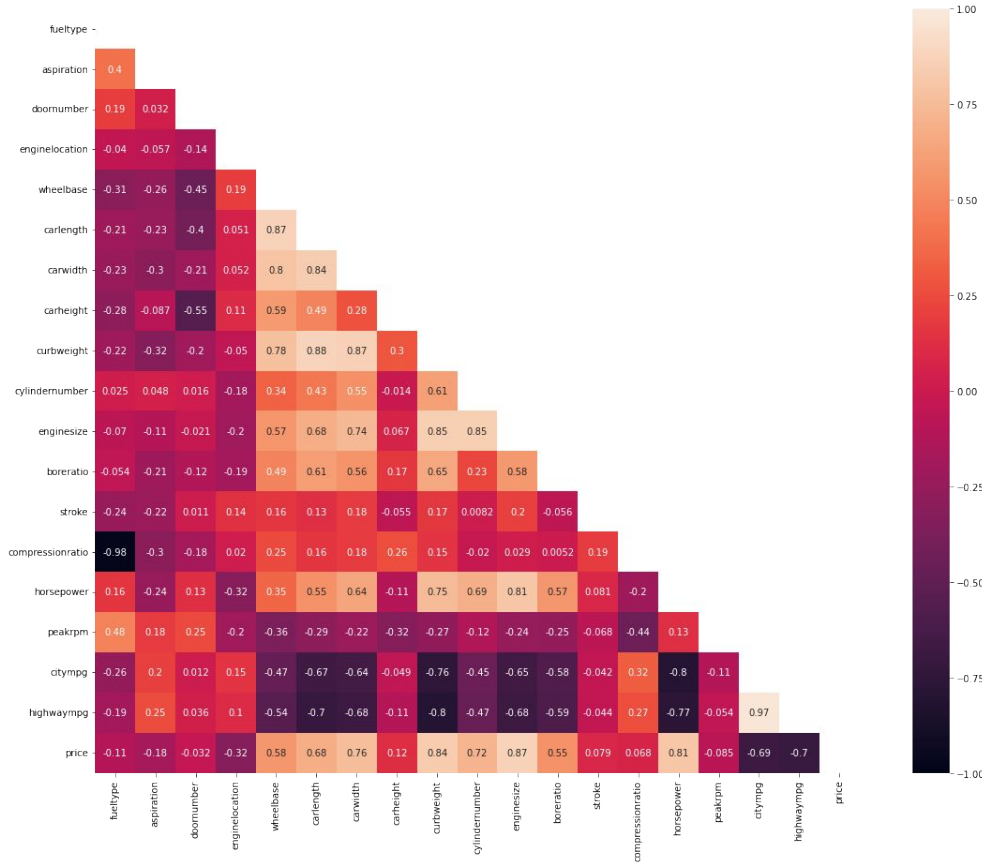
- Sebagian besar enginypemobil merupakan ohc.
- Mobil dengan enginetype dohc memiliki harga tertinggi.
- enginetype ohcv termasuk mobil dengan kisaran harga yang lebih tinggi.

Exploratory Data Analysis



- Sebagian besar fuelsystem mobil merupakan mpfi dan 2bbl.
- Mobil dengan fuelsystem mpfi memiliki harga tertinggi.
- Kisaran mobil dengan harga tinggi memiliki fuelsystem idi.

Exploratory Data Analysis



- wheelbase, carlength, dan boreratio berkorelasi kuat secara positif dengan price.
- carwidth, curbweight, cylindernumber, enginesize, dan horsepower berkorelasi sangat kuat secara positif dengan price.
- citympg dan highwaympg berkorelasi sangat kuat secara negative dengan price.

Modelling



Model Selection

Metode train test split (6:4)

Model yang digunakan:

- Linear Regression
- Ridge Regression
- ElasticNet Regression
- Lasso Regression
- Random Forest Regressor

Evaluasi metrik menggunakan MAE, RMSE, dan R^2

Model final yang digunakan adalah Model Random Forest Regressor tanpa Hyperparameter Tuning

Model Evaluation Metrics

Model	Linear Regression	Ridge Regression	ElasticNet Regression	Lasso Regression	Random Forest Regressor
MAE	1.761	1.699	2.160	1.720	1.400
RMSE	2.491	2.539	3.148	2.485	2.093
R^2	90.30%	89.92%	84.51%	90.34%	93.59%

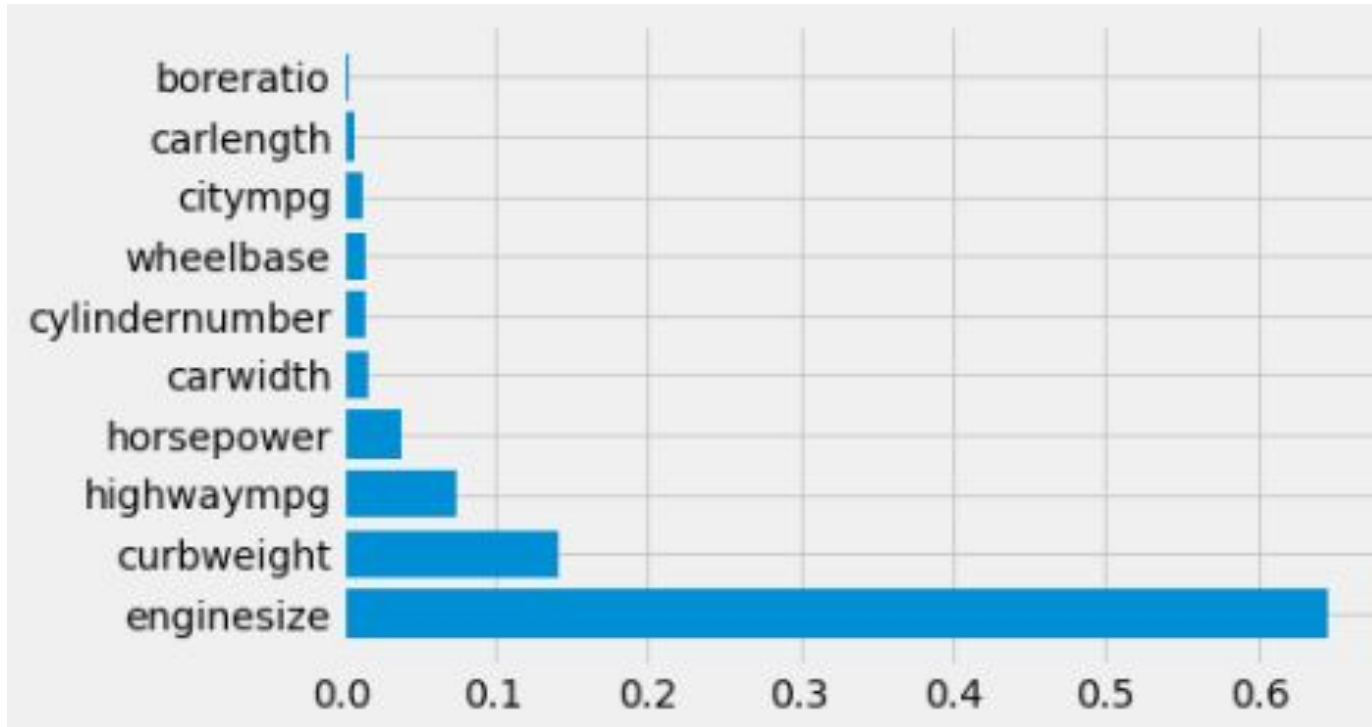
Model Evaluation Metrics

Model	Random Forest Regressor	
	tanpa Hyperparameter Tuning	dengan Hyperparameter Tuning
MAE	1.400	1.960
RMSE	2.093	2.687
R^2	93.59%	88.71%

Kolom

```
(['wheelbase', 'carlength', 'carwidth', 'curbweight', 'cylindernumber',  
  'enginesize', 'boreratio', 'horsepower', 'citympg', 'highwaympg',  
  'symboling_Moderately_Risky', 'symboling_Moderately_Safe',  
  'symboling_Neutral', 'symboling_Safe', 'symboling_Very_Risky',  
  'symboling_Very_Safe', 'carbody_convertible', 'carbody_hardtop',  
  'carbody_hatchback', 'carbody_sedan', 'carbody_wagon', 'drivewheel_4wd',  
  'drivewheel_fwd', 'drivewheel_rwd', 'enginetype_dohc',  
  'enginetype_dohcv', 'enginetype_l', 'enginetype_ohc', 'enginetype_ohcf',  
  'enginetype_ohcv', 'enginetype_rotor', 'fuelsystem_1bbl',  
  'fuelsystem_2bbl', 'fuelsystem_4bbl', 'fuelsystem_idi',  
  'fuelsystem_mfi', 'fuelsystem_mphi', 'fuelsystem_spdi',  
  'fuelsystem_spfi', 'Brand_alfa-romero', 'Brand_audi', 'Brand_bmw',  
  'Brand_buick', 'Brand_chevrolet', 'Brand_dodge', 'Brand_honda',  
  'Brand_isuzu', 'Brand_jaguar', 'Brand_mazda', 'Brand_mercury',  
  'Brand_mitsubishi', 'Brand_nissan', 'Brand_peugeot', 'Brand_plymouth',  
  'Brand_porsche', 'Brand_renault', 'Brand_saab', 'Brand_subaru',  
  'Brand_toyota', 'Brand_volkswagen', 'Brand_volvo'],  
 dtype='object')
```


Feature Importance



Conclusion

Conclusion

- Mobil Jaguar memiliki harga yang paling mahal.
- Mobil yang paling banyak disukai adalah Toyota.
- Manajemen perusahaan mobil Cina Geely Auto sebaiknya memperhatikan spesifikasi enginesize, curbweight, highwaympg, horsepower, dan carwidth untuk memproduksi mobil.

Terima kasih!

Ada pertanyaan?

zenius



Kampus
Merdeka
INDONESIA JAYA