CAR MARKET ANALYSIS - U.S.

FOR I.M. VEHICLES

By MX Data

AGENDA

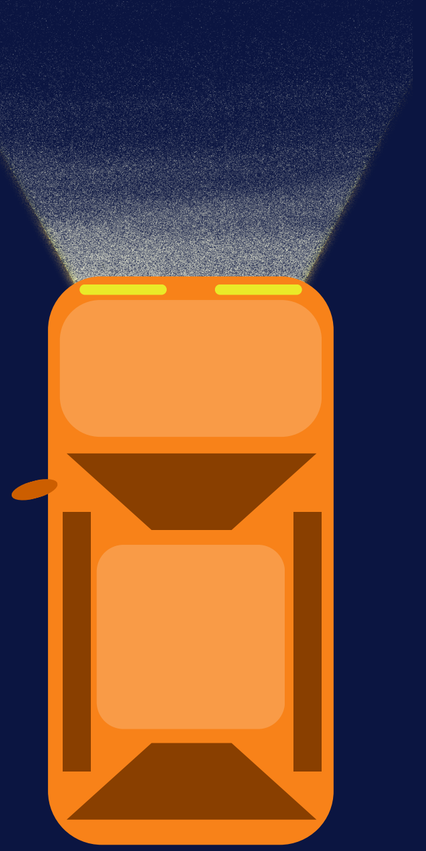
1. BUSINESS PROBLEM
2. DATA OBSERVATIONS
3. DATA CLEANING
4. REGRESSION
5. RECOMMENDATIONS
6. LIMITATIONS

BUSINESS PROBLEM

I.M. Vehicles --> U.S. market & produce cars
I.M. Vehicles focus: affordability & safety

Looking to know:

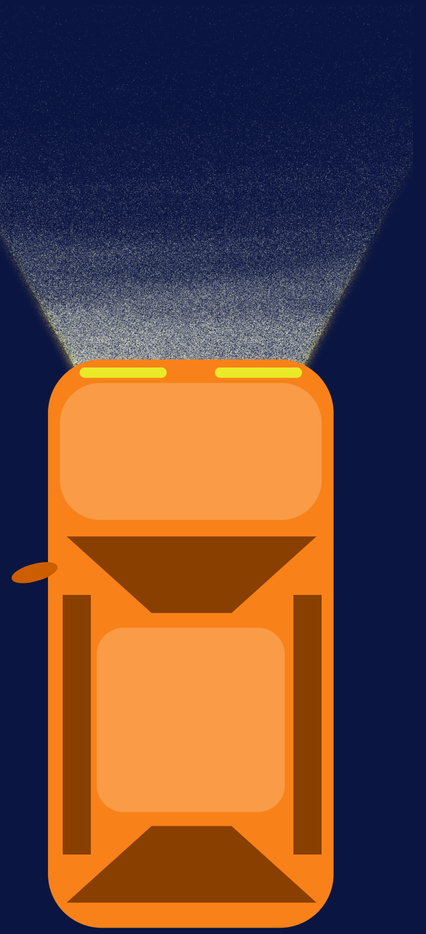
- The factors contributing to car prices in the U.S.
- Which of these factors contribute the most to price
- What choices should be made/avoided for these factors that will fit their focus



BUSINESS PROBLEM

Dataset contains

- Info about cars in the U.S.
- Based on market surveys



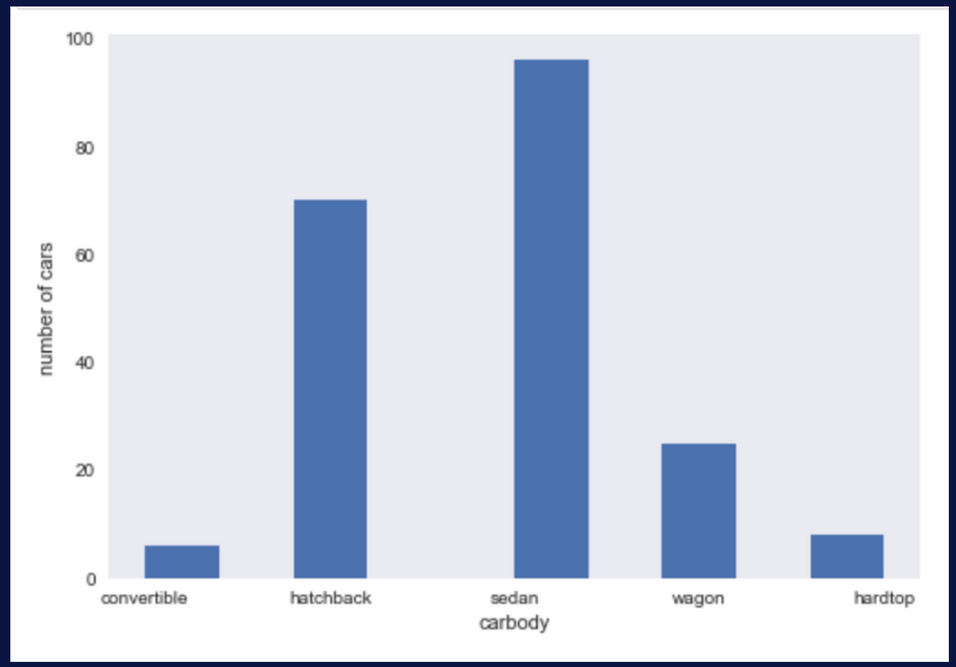
DATA OBSERVATIONS

Dataset contained

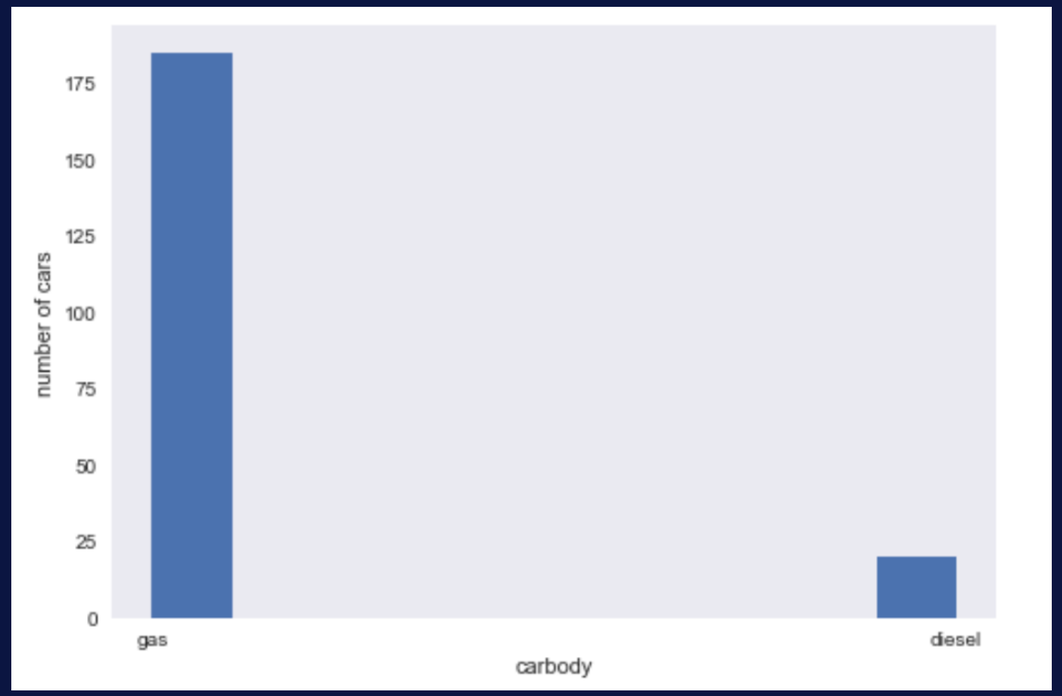
- 26 variables
- E.g. fuel type, aspiration, number of doors, car body type, car width, engine type, horsepower

MOST FREQUENT CAR

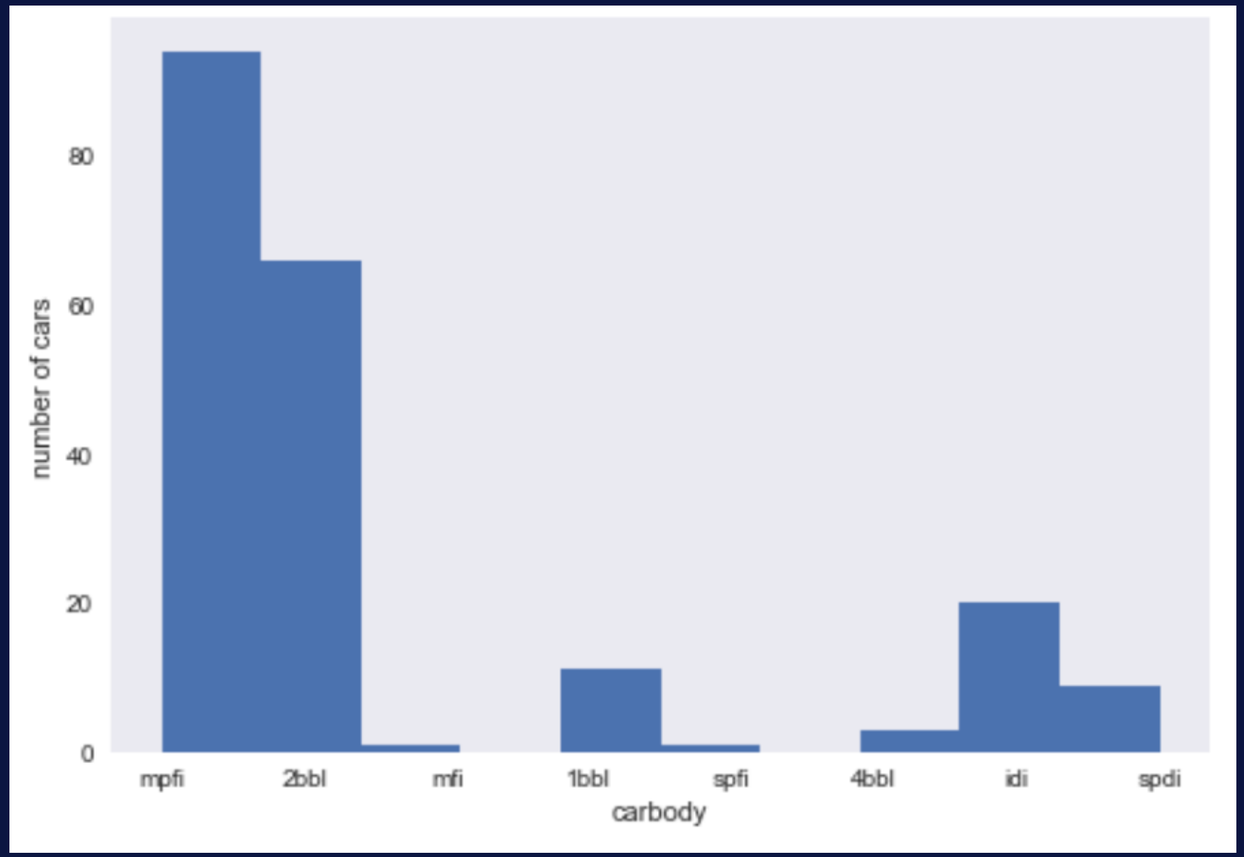
MOST FREQUENT CAR BODY



MOST FREQUENT FUEL TYPE



MOST FREQUENT FUEL SYSTEM



DATA TYPES

Categorical data (objects)

Numerical data (floats & ints)

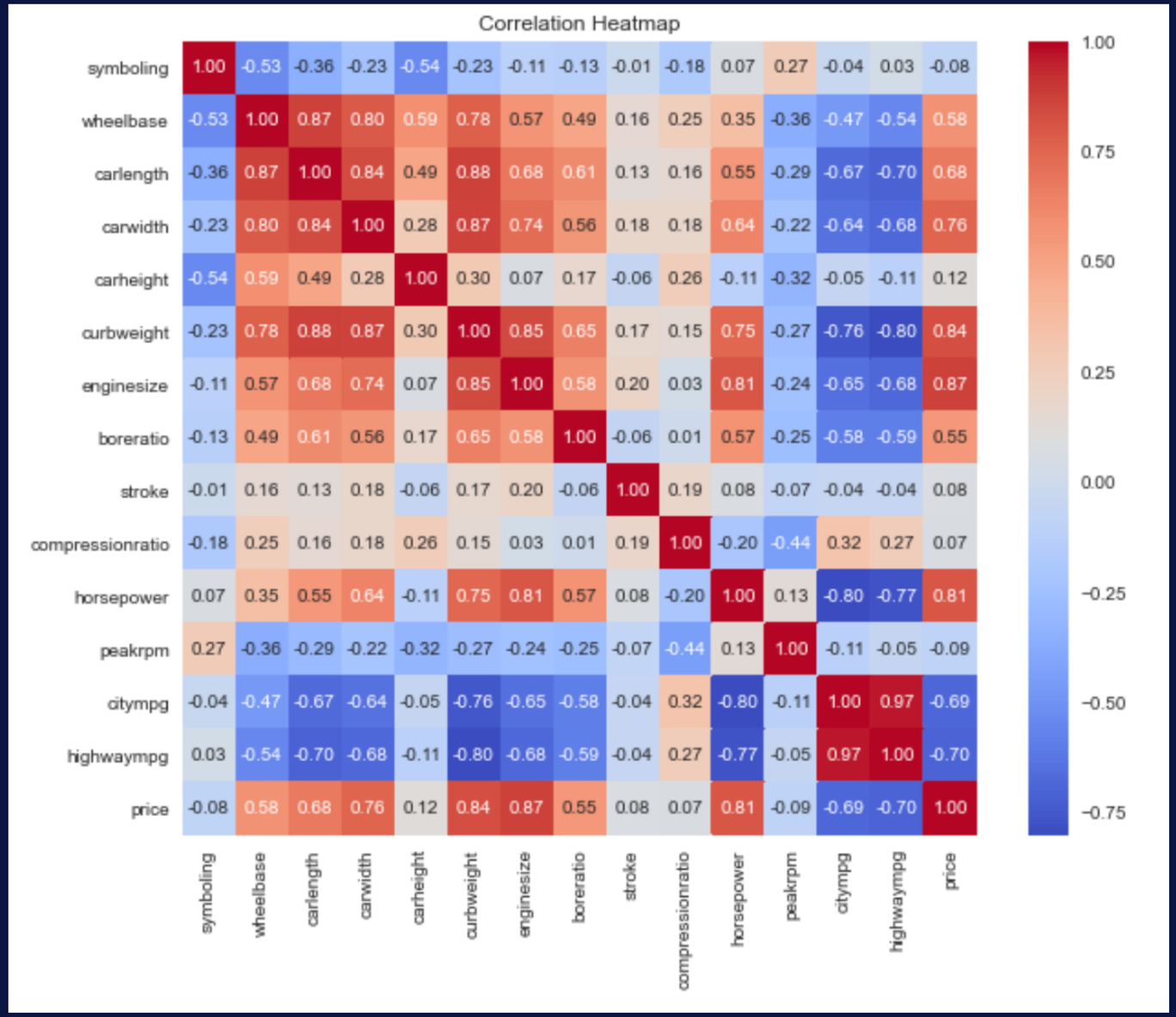
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   car_ID             205 non-null    int64
 1   symboling          205 non-null    int64
 2   CarName            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)
memory usage: 41.8+ KB
```

DATA CLEANING

- Car_ID column dropped
- Label Encoding to fix categorical data
- Outliers analysis
- Correlation between variables

DATA CLEANING

- Correlation between variables
  - carlength & wheelbase
  - curbweight and carwidth
  - highwaympg and citympg



REGRESSION MODELLING

1. BASE MODEL
2. MODEL 2 - USING TOP CORRELATED VARIABLES
3. MODEL 3- SCALING, REDUCING MULTICOLINARITY
4. MODEL 4 (FINAL AND CHOSEN)

FINAL MODEL

Variables added/removed

- Horsepower
- Engine size
- Engine location
- Car width

Reduced multicolinearity

Out[61]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.846
Model:	OLS	**Adj. R-squared:**	0.838
Method:	Least Squares	**F-statistic:**	106.5
Date:	Sun, 08 Oct 2023	**Prob (F-statistic):**	1.47e-23
Time:	12:04:31	**Log-Likelihood:**	-580.68
No. Observations:	62	**AIC:**	1169.
Df Residuals:	58	**BIC:**	1178.
Df Model:	3		
Covariance Type:	nonrobust		

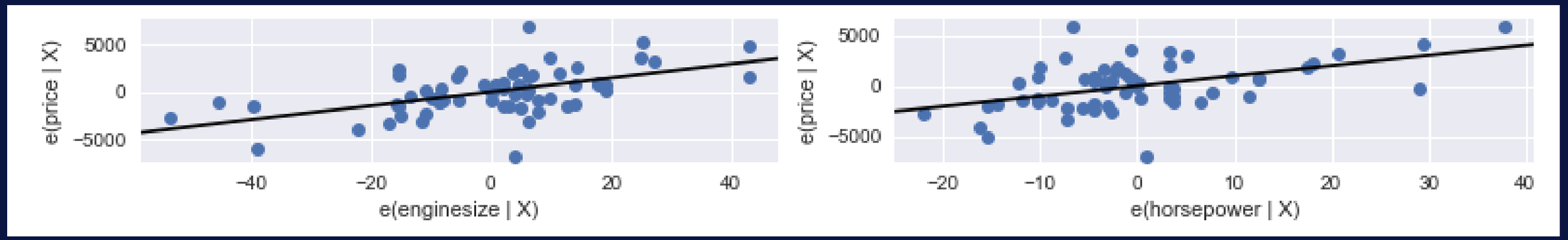
	coef	std err	t	P>|t|	[0.025	0.975]
0	2419.9851	627.782	3.855	0.000	1163.343	3676.627
1	2285.3774	752.083	3.039	0.004	779.919	3790.836
2	2577.8277	650.461	3.963	0.000	1275.789	3879.867
4	-1.066e+05	3049.865	-34.945	0.000	-1.13e+05	-1e+05

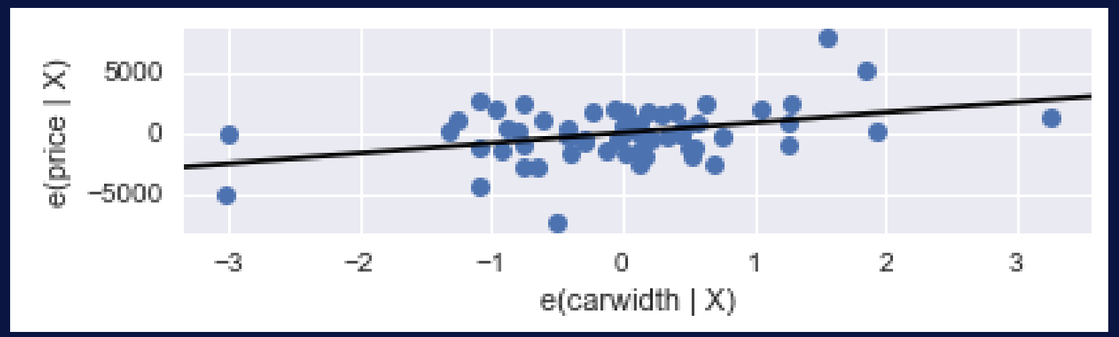
Omnibus:	3.707	**Durbin-Watson:**	1.981
Prob(Omnibus):	0.157	**Jarque-Bera (JB):**	3.538
Skew:	0.576	**Prob(JB):**	0.170
Kurtosis:	2.794	**Cond. No.**	13.1

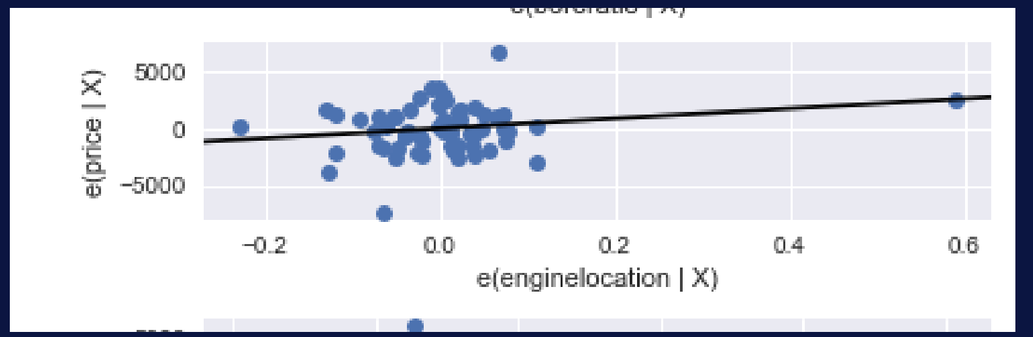
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

MODEL 4

Regression line and plots

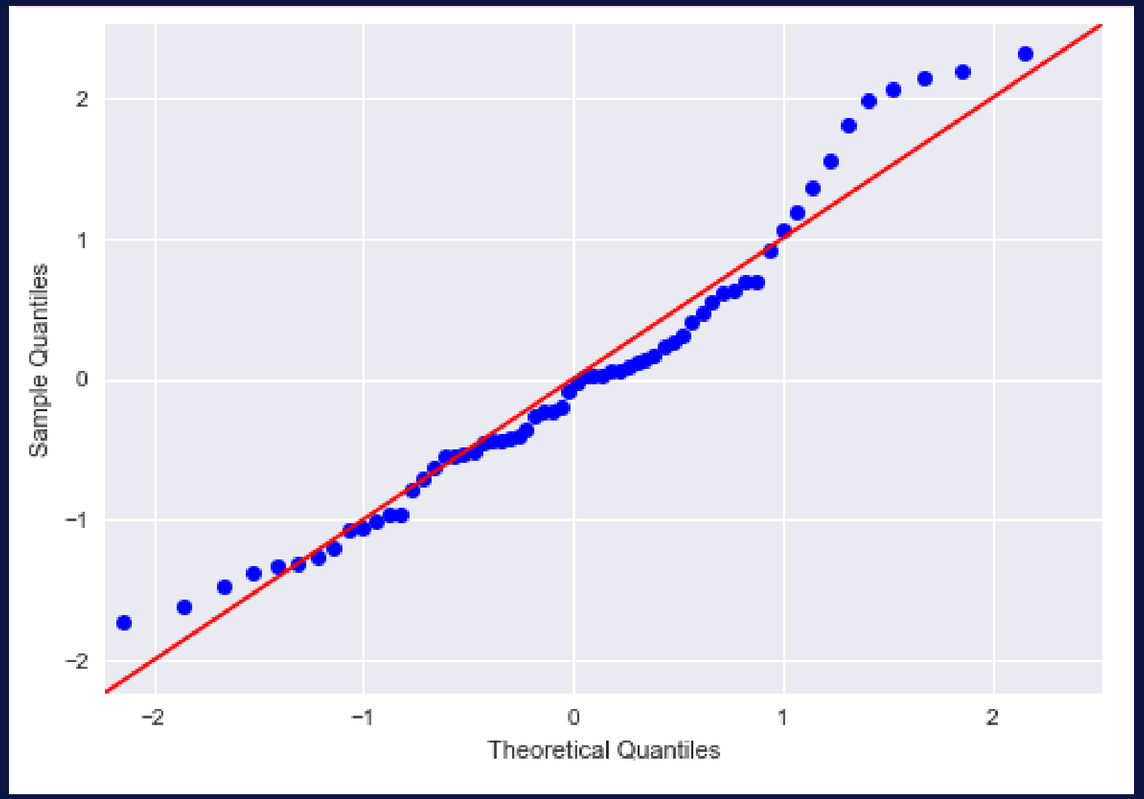


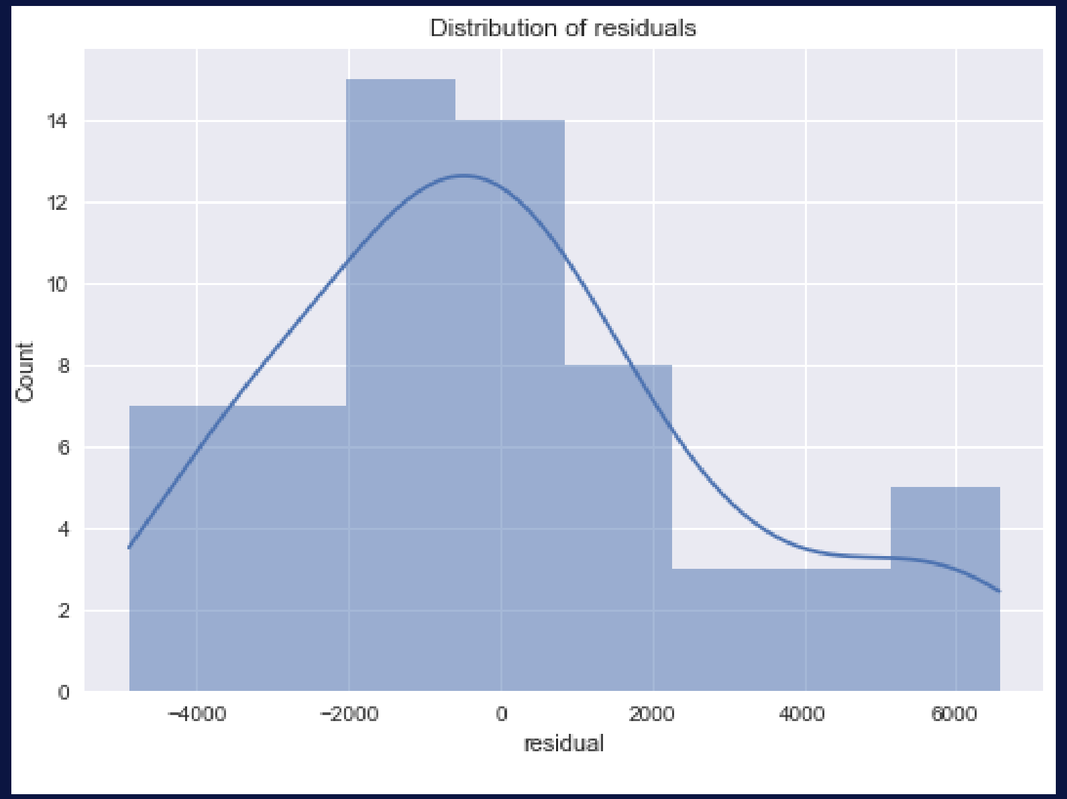




MODEL 4

Residuals





RECOMMENDATIONS

Consider the variables

- Horsepower
- Engine Size
- Engine Location
- Car Width

RECOMMENDATIONS

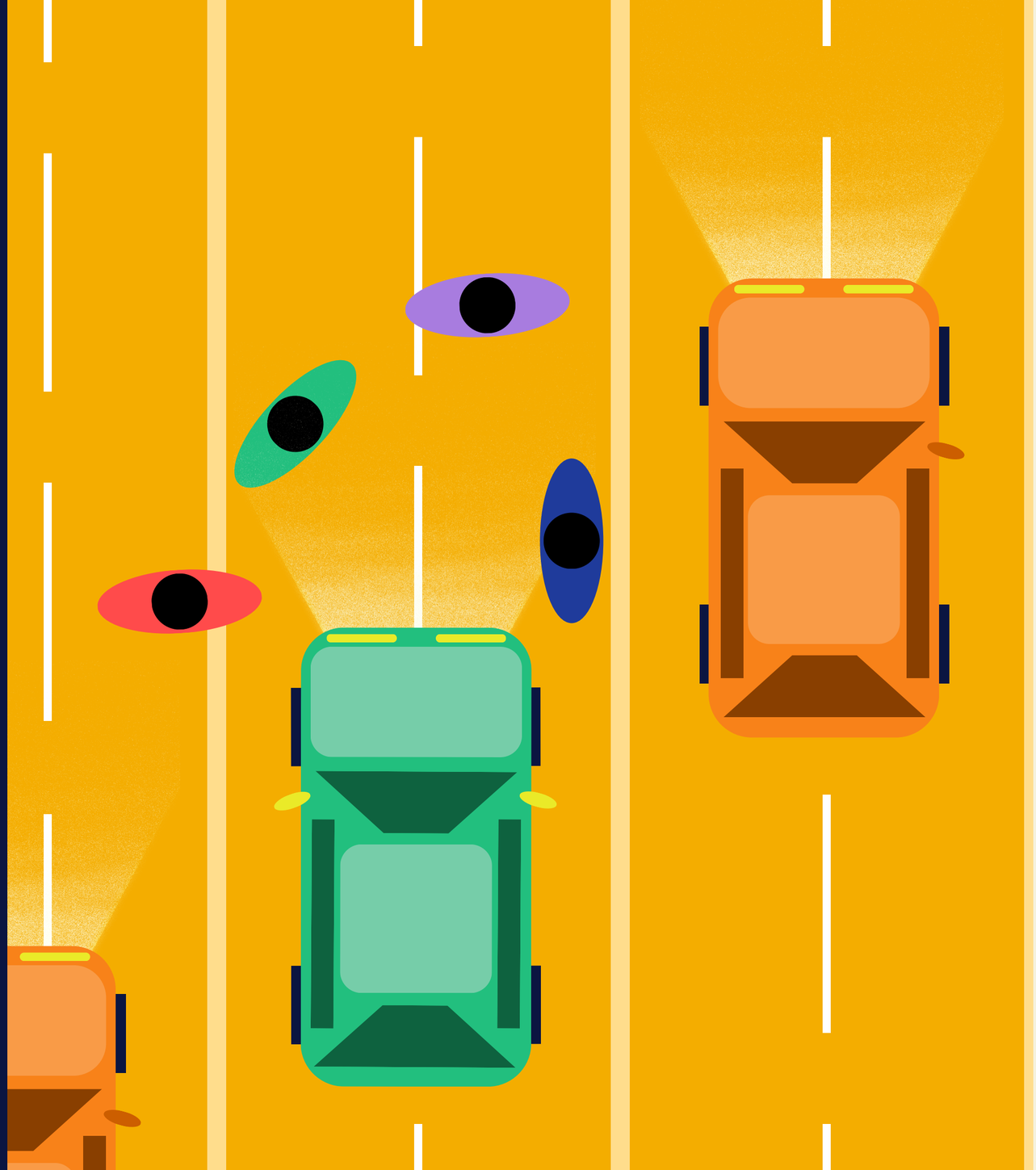
- Horsepower: mean was 104hp (77kW)
  - Safety of vehicle
  - Fuel economy
- Engine Size: average was 1.26L.
  - Bigger engine uses more fuel (not so efficient)
- Engine Location: front vs rear
  - Front
- Car Width: average was around 165cm
  - consider target market
  - practicality may be more imporant

LIMITATIONS

Dataset size

Continuation of the models

More recent data could be used



THANKYOU

kaggle.com/datasets/hellbuoy/car-price-prediction/datat