Car_Price_Assignment

March 13, 2025

[278]:
```
#Import the libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[280]:
```
# Load Data
df = pd.read_csv('CarPrice_Assignment.csv')
df
```

[280]:

	car_ID	symboling	CarName	fueltype	aspiration	…
0	1	3	alfa-romero giulia	gas	std	…
1	2	3	alfa-romero stelvio	gas	std	…
2	3	1	alfa-romero Quadrifoglio	gas	std	…
3	4	2	audi 100 ls	gas	std	…
4	5	2	audi 100ls	gas	std	…
..	…	…	…	…	…	…
200	201	-1	volvo 145e (sw)	gas	std	…
201	202	-1	volvo 144ea	gas	turbo	…
202	203	-1	volvo 244dl	gas	std	…
203	204	-1	volvo 246	diesel	turbo	…
204	205	-1	volvo 264gl	gas	turbo	…

	doornumber	carbody	drivewheel	enginelocation	wheelbase	…	…
0	two	convertible	rwd	front	88.6	…	
1	two	convertible	rwd	front	88.6	…	
2	two	hatchback	rwd	front	94.5	…	
3	four	sedan	fwd	front	99.8	…	
4	four	sedan	4wd	front	99.4	…	
..	…	…	…	…	…	…	…
200	four	sedan	rwd	front	109.1	…	
201	four	sedan	rwd	front	109.1	…	
202	four	sedan	rwd	front	109.1	…	
203	four	sedan	rwd	front	109.1	…	
204	four	sedan	rwd	front	109.1	…	

	enginesize	fuelsystem	boreratio	stroke	compressionratio	horsepower	…
0	130	mpfi	3.47	2.68	9.0	111	

```
1           130        mpfi       3.47    2.68            9.0         111
2           152        mpfi       2.68    3.47            9.0         154
3           109        mpfi       3.19    3.40           10.0         102
4           136        mpfi       3.19    3.40            8.0         115
..          …          …          …       …               …           …
200         141        mpfi       3.78    3.15            9.5         114
201         141        mpfi       3.78    3.15            8.7         160
202         173        mpfi       3.58    2.87            8.8         134
203         145         idi       3.01    3.40           23.0         106
204         141        mpfi       3.78    3.15            9.5         114

     peakrpm citympg  highwaympg    price
0       5000      21          27  13495.0
1       5000      21          27  16500.0
2       5000      19          26  16500.0
3       5500      24          30  13950.0
4       5500      18          22  17450.0
..       …     …           …       …
200     5400      23          28  16845.0
201     5300      19          25  19045.0
202     5500      18          23  21485.0
203     4800      26          27  22470.0
204     5400      19          25  22625.0

[205 rows x 26 columns]
```

[282]:
```
#Inspecting the data frame.
print(df.info())
```

```
<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
None
```

[284]:
```
print(df.describe())
```

```
           car_ID   symboling   wheelbase   carlength    carwidth   carheight  \
count  205.000000  205.000000  205.000000  205.000000  205.000000  205.000000
mean   103.000000    0.834146   98.756585  174.049268   65.907805   53.724878
std     59.322565    1.245307    6.021776   12.337289    2.145204    2.443522
min      1.000000   -2.000000   86.600000  141.100000   60.300000   47.800000
25%     52.000000    0.000000   94.500000  166.300000   64.100000   52.000000
50%    103.000000    1.000000   97.000000  173.200000   65.500000   54.100000
75%    154.000000    2.000000  102.400000  183.100000   66.900000   55.500000
max    205.000000    3.000000  120.900000  208.100000   72.300000   59.800000

         curbweight  enginesize   boreratio      stroke  compressionratio  \
count   205.000000  205.000000  205.000000  205.000000        205.000000
mean   2555.565854  126.907317    3.329756    3.255415         10.142537
std     520.680204   41.642693    0.270844    0.313597          3.972040
min    1488.000000   61.000000    2.540000    2.070000          7.000000
25%    2145.000000   97.000000    3.150000    3.110000          8.600000
50%    2414.000000  120.000000    3.310000    3.290000          9.000000
75%    2935.000000  141.000000    3.580000    3.410000          9.400000
max    4066.000000  326.000000    3.940000    4.170000         23.000000

       horsepower      peakrpm     citympg  highwaympg         price
count  205.000000   205.000000  205.000000  205.000000    205.000000
mean   104.117073  5125.121951   25.219512   30.751220  13276.710571
std     39.544167   476.985643    6.542142    6.886443   7988.852332
min     48.000000  4150.000000   13.000000   16.000000   5118.000000
25%     70.000000  4800.000000   19.000000   25.000000   7788.000000
50%     95.000000  5200.000000   24.000000   30.000000  10295.000000
75%    116.000000  5500.000000   30.000000   34.000000  16503.000000
max    288.000000  6600.000000   49.000000   54.000000  45400.000000
```

[286]:
```
print(df.head())
```

```
   car_ID  symboling                   CarName fueltype aspiration doornumber  \
0       1          3        alfa-romero giulia      gas        std        two
1       2          3       alfa-romero stelvio      gas        std        two
2       3          1  alfa-romero Quadrifoglio      gas        std        two
3       4          2               audi 100 ls      gas        std       four
4       5          2                audi 100ls      gas        std       four

       carbody drivewheel enginelocation  wheelbase  …  enginesize  \
0  convertible        rwd          front       88.6  …         130
1  convertible        rwd          front       88.6  …         130
2    hatchback        rwd          front       94.5  …         152
3        sedan        fwd          front       99.8  …         109
4        sedan        4wd          front       99.4  …         136

   fuelsystem  boreratio  stroke compressionratio horsepower  peakrpm citympg  \
0        mpfi       3.47    2.68              9.0        111     5000       21
1        mpfi       3.47    2.68              9.0        111     5000       21
2        mpfi       2.68    3.47              9.0        154     5000       19
3        mpfi       3.19    3.40             10.0        102     5500       24
4        mpfi       3.19    3.40              8.0        115     5500       18

   highwaympg    price
0          27  13495.0
1          27  16500.0
2          26  16500.0
3          30  13950.0
4          22  17450.0

[5 rows x 26 columns]
```

[288]:
```
#Data Cleaning

def check_null_percentage(df):
    null_counts = df.isnull().sum()
    null_percentage = (null_counts / len(df)) * 100
    return pd.DataFrame({'Null Count': null_counts, 'Null Percentage':␣
 ↪null_percentage})
```

[290]:
```
null_df = check_null_percentage(df)
print(null_df)
```

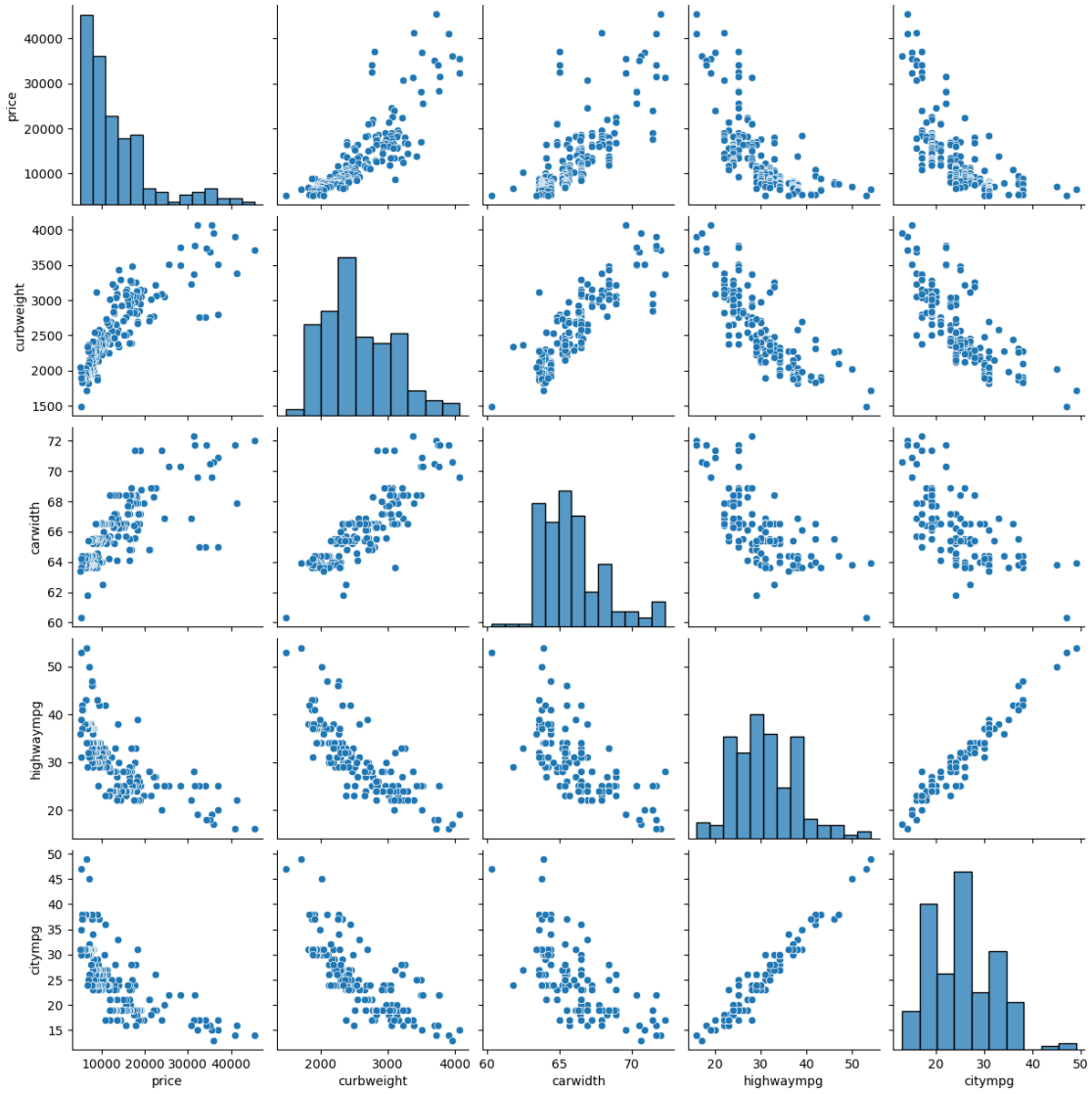
```
                Null Count  Null Percentage
car_ID                   0              0.0
symboling                0              0.0
CarName                  0              0.0
fueltype                 0              0.0
```

```
aspiration          0          0.0
doornumber          0          0.0
carbody             0          0.0
drivewheel          0          0.0
enginelocation      0          0.0
wheelbase           0          0.0
carlength           0          0.0
carwidth            0          0.0
carheight           0          0.0
curbweight          0          0.0
enginetype          0          0.0
cylindernumber      0          0.0
enginesize          0          0.0
fuelsystem          0          0.0
boreratio           0          0.0
stroke              0          0.0
compressionratio    0          0.0
horsepower          0          0.0
peakrpm             0          0.0
citympg             0          0.0
highwaympg          0          0.0
price               0          0.0
```

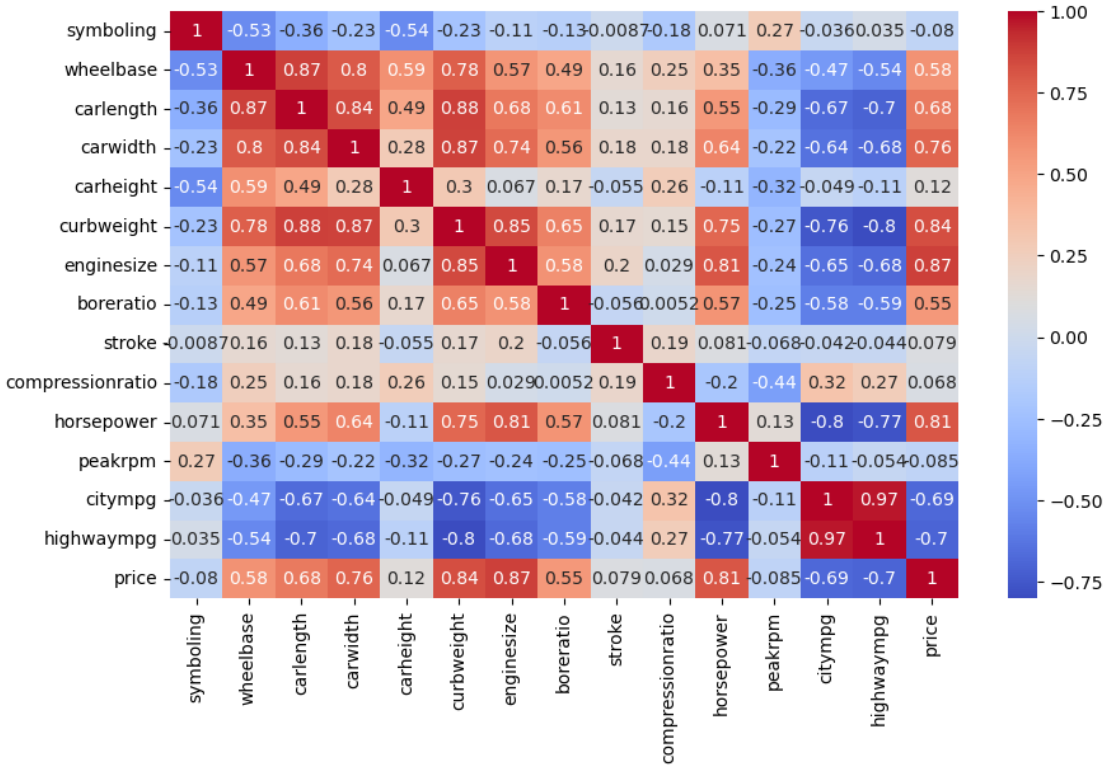
[292]:
```python
# Dropping unwanted columns
drop_columns = ['car_ID']  # Unnecessary column
df = df.drop(columns=drop_columns, axis=1)
```

[294]:
```python
#Sorting Column
df = df.sort_values(by='price', ascending=False)
```

[296]:
```python
# Data Visualization
sns.pairplot(df[['price', 'curbweight', 'carwidth', 'highwaympg', 'citympg']])
plt.show()
```



[297]:
```
plt.figure(figsize=(10, 6))
numeric_df = df.select_dtypes(include=[np.number])
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.show()
```



[298]:
```
#Data Preparation
# Fetching the all company names

car_names = df['CarName'].unique().tolist()
print(car_names)
```

```
['buick regal sport coupe (turbo)', 'bmw x5', 'buick century special', 'porsche
boxter', 'bmw x3', 'jaguar xk', 'jaguar xf', 'buick skylark', 'buick opel isuzu
deluxe', 'porsche cayenne', 'porcshce panamera', 'jaguar xj', 'buick skyhawk',
'bmw x4', 'buick century luxus (sw)', 'buick century', 'buick electra 225
custom', 'bmw z4', 'audi 4000', 'volvo 264gl', 'volvo 246', 'porsche macan',
'volvo 244dl', 'bmw x1', 'nissan kicks', 'volvo 144ea', 'volvo diesel', 'audi
5000', 'saab 99e', 'nissan clipper', 'mazda rx-7 gs', 'mazda glc', 'saab 99gle',
'peugeot 604sl', 'peugeot 504', 'audi 5000s (diesel)', 'audi 100ls', 'toyota
cressida', 'nissan teana', 'peugeot 505s turbo diesel', 'bmw 320i', 'volvo 145e
(sw)', 'toyota corolla liftback', 'volvo 245', 'mercury cougar', 'alfa-romero
stelvio', 'alfa-romero Quadrifoglio', 'toyota corona', 'toyouta tercel', 'toyota
starlet', 'mazda glc 4', 'audi fox', 'saab 99le', 'mitsubishi g4', 'mitsubishi
mirage g4', 'nissan fuga', 'audi 100 ls', 'volkswagen rabbit custom', 'nissan
otti', 'nissan dayz', 'alfa-romero giulia', 'volkswagen rabbit', 'peugeot 304',
'dodge coronet custom (sw)', 'honda civic', 'plymouth duster', 'mitsubishi
outlander', 'peugeot 504 (sw)', 'volkswagen dasher', 'subaru dl', 'vw dasher',
'toyota corolla', 'subaru r2', 'toyota mark ii', 'mazda 626', 'toyota tercel',
```

```
'isuzu D-Max', 'mazda glc deluxe', 'toyota celica gt', 'mazda glc custom',
'honda civic (auto)', 'honda accord', 'mazda rx-4', 'subaru tribeca', 'vw
rabbit', 'subaru baja', 'renault 5 gtl', 'nissan nv200', 'toyota corolla
tercel', 'volkswagen super beetle', 'toyota celica gt liftback', 'renault 12tl',
'subaru r1', 'honda civic 1300', 'nissan rogue', 'plymouth valiant', 'dodge dart
custom', 'isuzu D-Max V-Cross', 'honda prelude', 'toyota carina', 'dodge coronet
custom', 'volkswagen 411 (sw)', 'mazda glc custom l', 'toyota corona liftback',
'volkswagen type 3', 'mitsubishi pajero', 'nissan note', 'volkswagen model 111',
'volkswagen 1131 deluxe sedan', 'dodge d200', 'plymouth cricket', 'toyota
corolla 1600 (sw)', 'nissan juke', 'subaru brz', 'vokswagen rabbit', 'plymouth
fury gran sedan', 'dodge colt (sw)', 'nissan latio', 'subaru trezia', 'nissan
titan', 'nissan leaf', 'honda civic 1500 gl', 'honda accord lx', 'toyota corolla
1200', 'honda civic cvcc', 'subaru', 'nissan gt-r', 'mitsubishi montero',
'toyota corona hardtop', 'mazda rx2 coupe', 'isuzu MU-X', 'plymouth satellite
custom (sw)', 'dodge colt hardtop', 'chevrolet vega 2300', 'honda accord cvcc',
'dodge challenger se', 'chevrolet monte carlo', 'plymouth fury iii', 'dodge
monaco (sw)', 'mitsubishi lancer', 'maxda glc deluxe', 'dodge rampage', 'Nissan
versa', 'mitsubishi mirage', 'toyota corona mark ii', 'maxda rx3', 'chevrolet
impala']
```

[299]:
```
car_names = [name.split()[0] for name in car_names]
print(set(car_names))  # Print unique brand names
```

```
{'buick', 'toyouta', 'saab', 'volkswagen', 'volvo', 'mitsubishi', 'porcshce',
'Nissan', 'porsche', 'mercury', 'renault', 'alfa-romero', 'isuzu', 'mazda',
'jaguar', 'audi', 'subaru', 'maxda', 'vw', 'toyota', 'peugeot', 'dodge',
'plymouth', 'vokswagen', 'chevrolet', 'honda', 'nissan', 'bmw'}
```

[ ]:
```

```

[301]:
```
df['CarName'] = df['CarName'].replace({
    'toyouta': 'toyota',
    'porcshce': 'porsche',
    'vokswagen': 'volkswagen',
    'maxda': 'mazda',
    'alfa-romero':'alfa-romeo',
    'vw':'volkswagen',
})
```

[302]:
```
df
```

[302]:

	symboling	CarName	fueltype	aspiration	\
74	1	buick regal sport coupe (turbo)	gas	std	
16	0	bmw x5	gas	std	
73	0	buick century special	gas	std	
128	3	porsche boxter	gas	std	
17	0	bmw x3	gas	std	
..	…	…	…	…	

```
76           2                mitsubishi mirage      gas        std
150          1             toyota corona mark ii     gas        std
50           1                         maxda rx3     gas        std
18           2                  chevrolet impala     gas        std
138          2                           subaru      gas        std

    doornumber      carbody drivewheel enginelocation  wheelbase  carlength  \
74         two      hardtop        rwd          front      112.0      199.2
16         two        sedan        rwd          front      103.5      193.8
73        four        sedan        rwd          front      120.9      208.1
128        two  convertible        rwd           rear       89.5      168.9
17        four        sedan        rwd          front      110.0      197.0
..         …          …        …            …        …        …
76         two    hatchback        fwd          front       93.7      157.3
150        two    hatchback        fwd          front       95.7      158.7
50         two    hatchback        fwd          front       93.1      159.1
18         two    hatchback        fwd          front       88.4      141.1
138        two    hatchback        fwd          front       93.7      156.9

     …  enginesize  fuelsystem  boreratio stroke compressionratio  \
74   …         304        mpfi       3.80   3.35              8.0
16   …         209        mpfi       3.62   3.39              8.0
73   …         308        mpfi       3.80   3.35              8.0
128  …         194        mpfi       3.74   2.90              9.5
17   …         209        mpfi       3.62   3.39              8.0
..   …         …          …        …    …                 …
76   …          92        2bbl       2.97   3.23              9.4
150  …          92        2bbl       3.05   3.03              9.0
50   …          91        2bbl       3.03   3.15              9.0
18   …          61        2bbl       2.91   3.03              9.5
138  …          97        2bbl       3.62   2.36              9.0

     horsepower peakrpm  citympg  highwaympg    price
74          184    4500       14          16  45400.0
16          182    5400       16          22  41315.0
73          184    4500       14          16  40960.0
128         207    5900       17          25  37028.0
17          182    5400       15          20  36880.0
..          …     …      …         …       …
76           68    5500       37          41   5389.0
150          62    4800       35          39   5348.0
50           68    5000       30          31   5195.0
18           48    5100       47          53   5151.0
138          69    4900       31          36   5118.0

[205 rows x 25 columns]
```

[303]:
```
# Creating a new feature 'car_stability'
df['car_stability'] = df['wheelbase'] / df['carlength']
df
```

[303]:

```
     symboling                          CarName fueltype aspiration  \
74           1  buick regal sport coupe (turbo)      gas        std
16           0                         bmw x5      gas        std
73           0          buick century special      gas        std
128          3                 porsche boxter      gas        std
17           0                         bmw x3      gas        std
..         ...                            ...      ...        ...
76           2              mitsubishi mirage      gas        std
150          1          toyota corona mark ii      gas        std
50           1                      maxda rx3      gas        std
18           2              chevrolet impala      gas        std
138          2                        subaru      gas        std

    doornumber      carbody drivewheel enginelocation  wheelbase  carlength  \
74         two      hardtop        rwd          front      112.0      199.2
16         two        sedan        rwd          front      103.5      193.8
73        four        sedan        rwd          front      120.9      208.1
128        two  convertible        rwd           rear       89.5      168.9
17        four        sedan        rwd          front      110.0      197.0
..         ...          ...        ...            ...        ...        ...
76         two    hatchback        fwd          front       93.7      157.3
150        two    hatchback        fwd          front       95.7      158.7
50         two    hatchback        fwd          front       93.1      159.1
18         two    hatchback        fwd          front       88.4      141.1
138        two    hatchback        fwd          front       93.7      156.9

     ...  fuelsystem  boreratio  stroke compressionratio horsepower  peakrpm  \
74   ...        mpfi       3.80    3.35              8.0        184     4500
16   ...        mpfi       3.62    3.39              8.0        182     5400
73   ...        mpfi       3.80    3.35              8.0        184     4500
128  ...        mpfi       3.74    2.90              9.5        207     5900
17   ...        mpfi       3.62    3.39              8.0        182     5400
..   ...         ...        ...     ...              ...        ...      ...
76   ...        2bbl       2.97    3.23              9.4         68     5500
150  ...        2bbl       3.05    3.03              9.0         62     4800
50   ...        2bbl       3.03    3.15              9.0         68     5000
18   ...        2bbl       2.91    3.03              9.5         48     5100
138  ...        2bbl       3.62    2.36              9.0         69     4900

    citympg  highwaympg    price  car_stability
74       14          16  45400.0       0.562249
16       16          22  41315.0       0.534056
73       14          16  40960.0       0.580971
```

```
128      17          25  37028.0       0.529899
17       15          20  36880.0       0.558376
..      …           …       …              …
76       37          41   5389.0       0.595677
150      35          39   5348.0       0.603025
50       30          31   5195.0       0.585167
18       47          53   5151.0       0.626506
138      31          36   5118.0       0.597196

[205 rows x 26 columns]
```

[311]:
```
# Dropping highly correlated features
drop_features = ['carlength', 'carwidth', 'curbweight', 'wheelbase',␣
 ↪'highwaympg']
df = df.drop(columns=drop_features, axis=1)
```

[313]:
```
# Encoding categorical variables using dummy encoding
df = pd.get_dummies(df, drop_first=True)
```

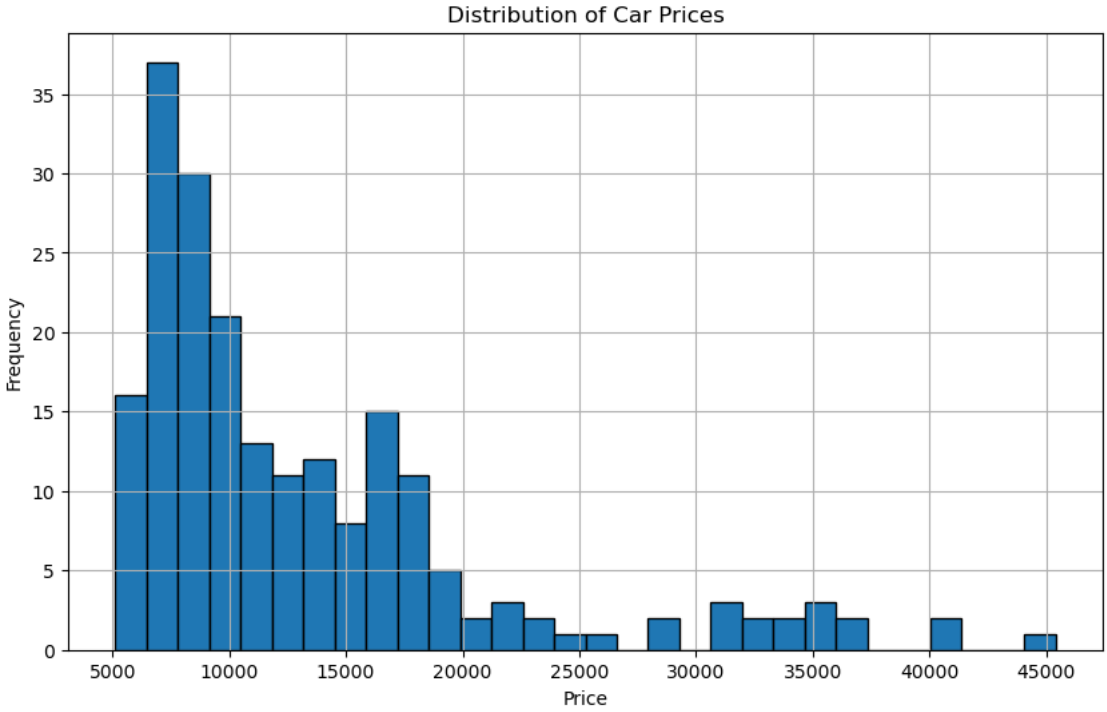
[318]:
```
df.columns = df.columns.str.replace(' ', '_')
df
```

[318]:
```
     symboling  carheight  enginesize  boreratio  stroke  compressionratio  \
74           1       55.4         304       3.80    3.35               8.0
16           0       53.7         209       3.62    3.39               8.0
73           0       56.7         308       3.80    3.35               8.0
128          3       51.6         194       3.74    2.90               9.5
17           0       56.3         209       3.62    3.39               8.0
..         …         …           …         …       …                  …
76           2       50.8          92       2.97    3.23               9.4
150          1       54.5          92       3.05    3.03               9.0
50           1       54.1          91       3.03    3.15               9.0
18           2       53.2          61       2.91    3.03               9.5
138          2       53.7          97       3.62    2.36               9.0

     horsepower  peakrpm  citympg    price  …  cylindernumber_three  \
74          184     4500       14  45400.0  …                 False
16          182     5400       16  41315.0  …                 False
73          184     4500       14  40960.0  …                 False
128         207     5900       17  37028.0  …                 False
17          182     5400       15  36880.0  …                 False
..          …       …       …       …     …                   …
76           68     5500       37   5389.0  …                 False
150          62     4800       35   5348.0  …                 False
50           68     5000       30   5195.0  …                 False
18           48     5100       47   5151.0  …                  True
138          69     4900       31   5118.0  …                 False
```

```
     cylindernumber_twelve  cylindernumber_two  fuelsystem_2bbl  \
74                   False               False            False
16                   False               False            False
73                   False               False            False
128                  False               False            False
17                   False               False            False
..                     …                   …              …
76                   False               False             True
150                  False               False             True
50                   False               False             True
18                   False               False             True
138                  False               False             True

     fuelsystem_4bbl  fuelsystem_idi  fuelsystem_mfi  fuelsystem_mpfi  \
74             False           False           False             True
16             False           False           False             True
73             False           False           False             True
128            False           False           False             True
17             False           False           False             True
..               …               …               …                …
76             False           False           False            False
150            False           False           False            False
50             False           False           False            False
18             False           False           False            False
138            False           False           False            False

     fuelsystem_spdi  fuelsystem_spfi
74             False            False
16             False            False
73             False            False
128            False            False
17             False            False
..               …                …
76             False            False
150            False            False
50             False            False
18             False            False
138            False            False

[205 rows x 186 columns]
```

[320]:
```
# Data Analysis - Histogram & Bar Charts
plt.figure(figsize=(10, 6))
df['price'].hist(bins=30, edgecolor='black')
plt.xlabel('Price')
plt.ylabel('Frequency')
```

```
plt.title('Distribution of Car Prices')
plt.show()
```



[324]:
```
# Generate Insights and save the report
df.describe().to_csv('data_analysis_report.csv')
```

[ ]:

[ ]: