Mielage Prediction Project

In [2]:

```
#import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:

```
#import data
df=pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/MPG.csv')
```

In [4]:

```
df.head()
```

Out[4]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130.0	3504	12.0	70	usa	chevrolet chevelle malibu
1	15.0	8	350.0	165.0	3693	11.5	70	usa	buick skylark 320
2	18.0	8	318.0	150.0	3436	11.0	70	usa	plymouth satellite
3	16.0	8	304.0	150.0	3433	12.0	70	usa	amc rebel sst
4	17.0	8	302.0	140.0	3449	10.5	70	usa	ford torino

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 398 entries, 0 to 397
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   mpg           398 non-null    float64
 1   cylinders     398 non-null    int64  
 2   displacement  398 non-null    float64
 3   horsepower    392 non-null    float64
 4   weight        398 non-null    int64  
 5   acceleration  398 non-null    float64
 6   model_year    398 non-null    int64  
 7   origin        398 non-null    object 
 8   name          398 non-null    object 
dtypes: float64(4), int64(3), object(2)
memory usage: 28.1+ KB
```

In [6]:

```
df.nunique()
```

Out[6]:

```
mpg             129
cylinders         5
```

```
displacement     82
horsepower       93
weight          351
acceleration     95
model_year       13
origin            3
name            305
dtype: int64
```

In [7]:

```
df.describe()
```

Out[7]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year
count	398.000000	398.000000	398.000000	392.000000	398.000000	398.000000	398.000000
mean	23.514573	5.454774	193.425879	104.469388	2970.424623	15.568090	76.010050
std	7.815984	1.701004	104.269838	38.491160	846.841774	2.757689	3.697627
min	9.000000	3.000000	68.000000	46.000000	1613.000000	8.000000	70.000000
25%	17.500000	4.000000	104.250000	75.000000	2223.750000	13.825000	73.000000
50%	23.000000	4.000000	148.500000	93.500000	2803.500000	15.500000	76.000000
75%	29.000000	8.000000	262.000000	126.000000	3608.000000	17.175000	79.000000
max	46.600000	8.000000	455.000000	230.000000	5140.000000	24.800000	82.000000

In [8]:

```
df.corr()
```

Out[8]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year
mpg	1.000000	-0.775396	-0.804203	-0.778427	-0.831741	0.420289	0.579267
cylinders	-0.775396	1.000000	0.950721	0.842983	0.896017	-0.505419	-0.348746
displacement	-0.804203	0.950721	1.000000	0.897257	0.932824	-0.543684	-0.370164
horsepower	-0.778427	0.842983	0.897257	1.000000	0.864538	-0.689196	-0.416361
weight	-0.831741	0.896017	0.932824	0.864538	1.000000	-0.417457	-0.306564
acceleration	0.420289	-0.505419	-0.543684	-0.689196	-0.417457	1.000000	0.288137
model_year	0.579267	-0.348746	-0.370164	-0.416361	-0.306564	0.288137	1.000000

In [12]:

```
#remove missing value
df=df.dropna()
df.columns
```

Out[12]:

```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
       'acceleration', 'model_year', 'origin', 'name'],
      dtype='object')
```

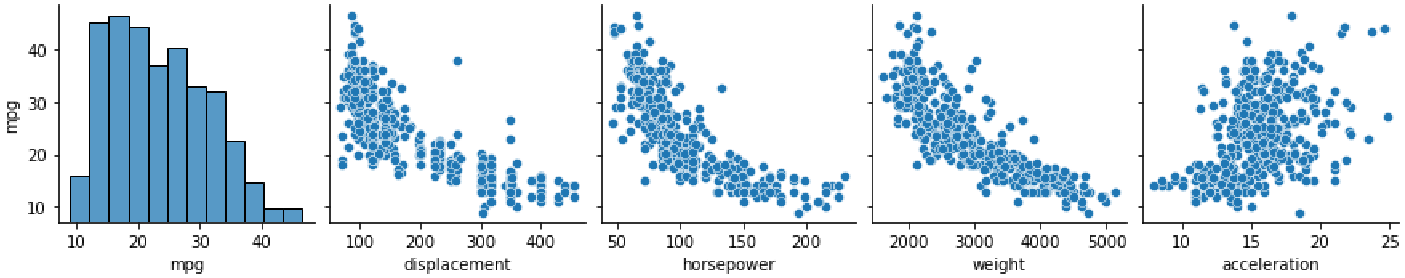
In [10]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 392 entries, 0 to 397
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   mpg           392 non-null    float64
 1   cylinders     392 non-null    int64  
 2   displacement  392 non-null    float64
 3   horsepower    392 non-null    float64
 4   weight        392 non-null    int64  
 5   acceleration  392 non-null    float64
 6   model_year    392 non-null    int64  
 7   origin        392 non-null    object 
 8   name          392 non-null    object 
dtypes: float64(4), int64(3), object(2)
memory usage: 30.6+ KB
```

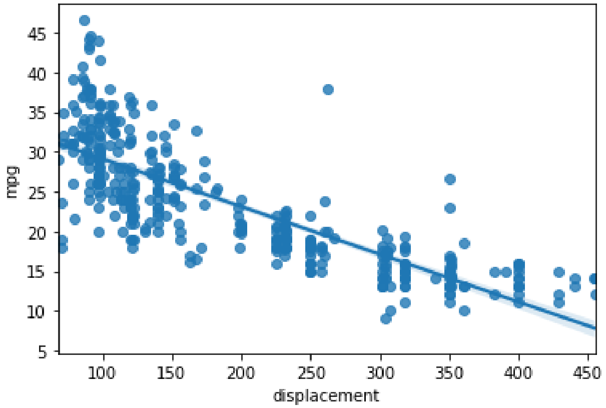
In [14]:

```
#Data visualization
sns.pairplot(df,x_vars=['mpg','displacement','horsepower','weight',
       'acceleration'],y_vars=['mpg']);
```



In [15]:

```
sns.regplot(x='displacement',y='mpg',data=df);
```



In [16]:

```
#define x and y
df.columns
```

Out[16]:

```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
       'acceleration', 'model_year', 'origin', 'name'],
      dtype='object')
```

In [17]:

```
y=df['mpg']
x=df[['displacement','horsepower','weight',
       'acceleration']]
```

In [18]:

```
#split
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=2529)
```

In [19]:

```
#model
from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

In [20]:

```
model.fit(x_train,y_train)
```

Out[20]:

```
LinearRegression()
```

In [21]:

```
model.intercept_
```

Out[21]:

```
45.84620249789292
```

In [22]:

```
model.coef_
```

```
Out[22]: array([-0.007859  , -0.05202824, -0.0048651 , -0.05998945])
```

```
In [23]: #prediction
         y_pred=model.predict(x_test)
         y_pred
```

```
Out[23]: array([18.45029029, 15.11872575, 14.25951901, 23.63777162, 29.77227939,
                23.78289678, 26.46274613, 24.63477759, 15.10361067, 11.92089347,
                24.03667612, 28.03774179, 31.7791986 , 31.04942136, 18.34939414,
                19.34562679, 28.14901371, 32.26833498, 31.23336778, 27.1706607 ,
                18.90264044, 22.69158865, 26.30616149, 32.53334114, 20.7455229 ,
                 8.43604922, 21.96939005, 18.16644283, 24.9187207 , 14.95041612,
                23.27573018, 17.10008397,  9.28416594, 30.02859334, 20.49341373,
                29.16402497, 24.1851619 , 21.82468561, 10.45764414, 12.99758931,
                21.55287965, 19.9763373 ,  5.81701795, 17.83479167, 22.69872144,
                29.39987303, 13.2638446 , 25.84303202, 29.29886179, 22.44116443,
                22.30857618, 16.57432268, 24.06827363, 30.19019859, 10.04817173,
                 9.3533171 , 28.14495274, 23.67665202, 20.07936568, 30.77322956,
                20.95405256, 26.72684739, 23.16157669, 14.10789682, 24.37223149,
                26.84731155, 15.26437637, 24.21355   , 30.81705563, 14.86794633,
                27.5428809 , 24.35148953, 10.75013125, 30.29658039, 30.95694009,
                27.35893598, 31.26808388, 10.29239165, 27.64504505, 16.41746006,
                25.5910977 , 29.48584659, 14.83958315, 32.76319208, 30.34965318,
                30.95305498, 14.61576534, 27.04413659, 26.74989971, 29.0983602 ,
                32.55952574, 29.50578249, 31.70671628, 31.69454341, 21.58369883,
                31.71427871, 26.19466037, 28.94617784])
```

```
In [25]: #accuracy
         from sklearn.metrics import mean_absolute_percentage_error
         mean_absolute_percentage_error(y_test,y_pred)
```

```
Out[25]: 0.14486145216628077
```

```
In [ ]:
```