auto_mpg_data

November 8, 2022

1 Import required libraries

[72]:
```
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

[74]:
```
df_auto = pd.read_csv('auto_data.csv')
df_auto.head()
```

[74]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino

[75]:
```
print(df_auto.head())
print(df_auto.index)
print(df_auto.columns)
```

```
    mpg  cylinders  displacement horsepower  weight  acceleration  model_year  \
0  18.0          8         307.0        130    3504          12.0          70
1  15.0          8         350.0        165    3693          11.5          70
2  18.0          8         318.0        150    3436          11.0          70
3  16.0          8         304.0        150    3433          12.0          70
4  17.0          8         302.0        140    3449          10.5          70

   origin                       name
0       1  chevrolet chevelle malibu
1       1          buick skylark 320
2       1         plymouth satellite
3       1              amc rebel sst
4       1                ford torino
```

```
2       1         plymouth satellite
3       1              amc rebel sst
4       1                ford torino
RangeIndex(start=0, stop=398, step=1)
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
       'acceleration', 'model_year', 'origin', 'name'],
      dtype='object')
```

[77]:
```
df_auto.shape
```

[77]:
```
(398, 9)
```

[78]:
```
df_auto.isnull().any()
```

[78]:
```
mpg             False
cylinders       False
displacement    False
horsepower      False
weight          False
acceleration    False
model_year      False
origin          False
name            False
dtype: bool
```

[79]:
```
df_auto.dtypes
```

[79]:
```
mpg             float64
cylinders         int64
displacement    float64
horsepower       object
weight            int64
acceleration    float64
model_year        int64
origin            int64
name             object
dtype: object
```

[80]:
```
#write a user defined function for origin
# 1-USA, 2-Europe, 3-Asia

def origin(num):
    if num ==1:
        return 'USA'
    elif num ==2:
        return 'Europe'
    else :
```

```
        return  'Asia'

    auto_data['origin'] = auto_data['origin'].apply(origin)
```

[81]:
```
df_auto.head(30)
```

[81]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	\
0	18.0	8	307.0	130	3504	12.0	
1	15.0	8	350.0	165	3693	11.5	
2	18.0	8	318.0	150	3436	11.0	
3	16.0	8	304.0	150	3433	12.0	
4	17.0	8	302.0	140	3449	10.5	
5	15.0	8	429.0	198	4341	10.0	
6	14.0	8	454.0	220	4354	9.0	
7	14.0	8	440.0	215	4312	8.5	
8	14.0	8	455.0	225	4425	10.0	
9	15.0	8	390.0	190	3850	8.5	
10	15.0	8	383.0	170	3563	10.0	
11	14.0	8	340.0	160	3609	8.0	
12	15.0	8	400.0	150	3761	9.5	
13	14.0	8	455.0	225	3086	10.0	
14	24.0	4	113.0	95	2372	15.0	
15	22.0	6	198.0	95	2833	15.5	
16	18.0	6	199.0	97	2774	15.5	
17	21.0	6	200.0	85	2587	16.0	
18	27.0	4	97.0	88	2130	14.5	
19	26.0	4	97.0	46	1835	20.5	
20	25.0	4	110.0	87	2672	17.5	
21	24.0	4	107.0	90	2430	14.5	
22	25.0	4	104.0	95	2375	17.5	
23	26.0	4	121.0	113	2234	12.5	
24	21.0	6	199.0	90	2648	15.0	
25	10.0	8	360.0	215	4615	14.0	
26	10.0	8	307.0	200	4376	15.0	
27	11.0	8	318.0	210	4382	13.5	
28	9.0	8	304.0	193	4732	18.5	
29	27.0	4	97.0	88	2130	14.5	

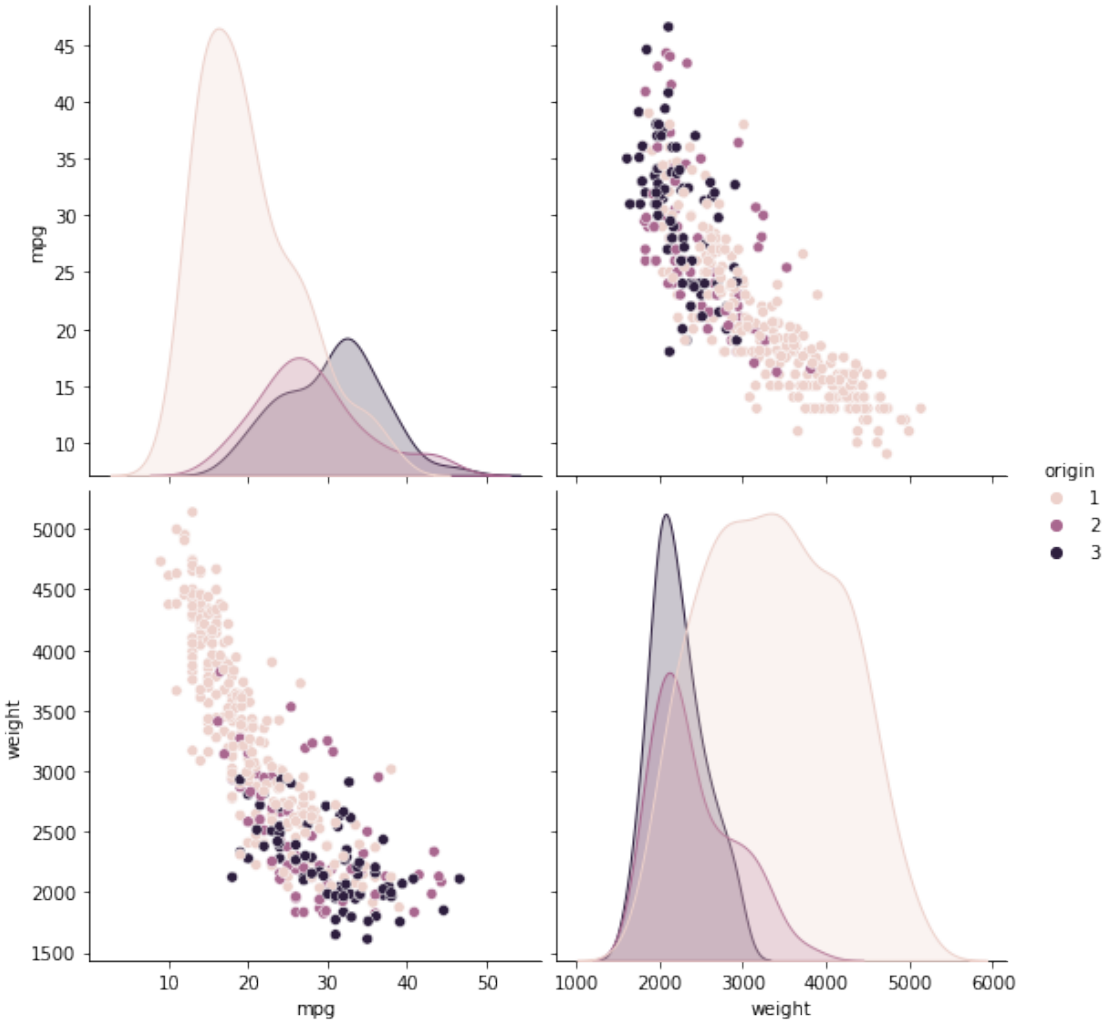
	model_year	origin	name
0	70	1	chevrolet chevelle malibu
1	70	1	buick skylark 320
2	70	1	plymouth satellite
3	70	1	amc rebel sst
4	70	1	ford torino
5	70	1	ford galaxie 500
6	70	1	chevrolet impala
7	70	1	plymouth fury iii

```
8           70       1              pontiac catalina
9           70       1            amc ambassador dpl
10          70       1           dodge challenger se
11          70       1            plymouth 'cuda 340
12          70       1         chevrolet monte carlo
13          70       1       buick estate wagon (sw)
14          70       3         toyota corona mark ii
15          70       1               plymouth duster
16          70       1                    amc hornet
17          70       1                ford maverick
18          70       3                 datsun pl510
19          70       2  volkswagen 1131 deluxe sedan
20          70       2                  peugeot 504
21          70       2                  audi 100 ls
22          70       2                     saab 99e
23          70       2                     bmw 2002
24          70       1                  amc gremlin
25          70       1                    ford f250
26          70       1                    chevy c20
27          70       1                   dodge d200
28          70       1                     hi 1200d
29          71       3                 datsun pl510
```

[86]:
```
sns.pairplot(df_auto[['mpg', 'weight', 'origin']], hue = 'origin', size=4)
```

```
/usr/local/lib/python3.7/site-packages/seaborn/axisgrid.py:2076: UserWarning:
The `size` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```

[86]:
```
<seaborn.axisgrid.PairGrid at 0x7f1f277e8c10>
```


mpg
weight
origin
1
2
3
10
20
30
40
50
1000
2000
3000
4000
5000
6000
1500
2000
2500
3000
3500
4000
4500
5000
10
15
20
25
30
35
40
45

[ ]: