

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

In [2]: a=pd.read_csv("C:/Users/Poonam/Desktop/dataset/Cars93 (2).csv")

In [3]: a.head()

Out[3]:
```

	id	Manufacturer	Model	Type	Min.Price	Price	Max.Price	MPGcity	MPGhighway	AirBags	...	Passengers	Length	Wheelbase	Width	Turn.circle	Rear.seat.room	Luggage.room
0	59	Mercedes-Benz	300E	Midsize	43.8	61.9	80.0	19	25	Driver & Passenger	...	5	187	110	69	37	27.0	15
1	48	Infiniti	Q45	Midsize	45.4	47.9	50.4	17	22	Driver only	...	5	200	113	72	42	29.0	15
2	4	Audi	100	Midsize	30.8	37.7	44.6	19	26	NaN	...	6	193	106	70	37	31.0	17
3	11	Cadillac	Seville	Midsize	37.5	40.1	42.7	16	25	Driver & Passenger	...	5	204	111	74	44	31.0	14
4	19	Chevrolet	Corvette	Sporty	34.6	38.0	41.5	17	25	Driver only	...	2	179	96	74	43	NaN	NaN

5 rows × 28 columns

```
In [4]: a.describe()

Out[4]:
```

	id	Min.Price	Price	Max.Price	MPGcity	MPGhighway	EngineSize	Horsepower	RPM	Rev.per.mile	FuelTankCapacity	Passengers	Length	Wheelbase
count	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000
mean	47.000000	17.125000	19.699177	21.869925	22.305591	29.086022	2.667742	143.827967	5280.645161	2332.204301	16.664516	5.066022	183.204301	103.946237
std	26.990739	8.748029	9.659430	11.030457	5.619812	5.331726	1.037363	52.374410	596.731690	496.506525	3.279370	1.038979	14.602382	6.819674
min	1.000000	6.700000	7.400000	7.900000	15.000000	20.000000	2.000000	95.000000	3800.000000	1320.000000	9.200000	2.000000	141.000000	90.000000
25%	24.000000	10.800000	12.200000	14.700000	18.000000	26.000000	1.800000	103.000000	4800.000000	1985.000000	14.500000	4.000000	174.000000	98.000000
50%	47.000000	14.700000	17.700000	19.600000	21.000000	28.000000	2.400000	140.000000	5200.000000	2340.000000	16.000000	5.000000	183.000000	103.000000
75%	70.000000	20.300000	23.300000	25.300000	25.000000	31.000000	3.300000	170.000000	5750.000000	2565.000000	18.800000	6.000000	192.000000	110.000000
max	93.000000	45.400000	61.900000	80.000000	46.000000	50.000000	5.700000	300.000000	6500.000000	3755.000000	27.000000	8.000000	219.000000	119.000000

```
In [5]: a.shape

Out[5]: (93, 28)

In [6]: a.isna()

Out[6]:
```

	id	Manufacturer	Model	Type	Min.Price	Price	Max.Price	MPGcity	MPGhighway	AirBags	...	Passengers	Length	Wheelbase	Width	Turn.circle	Rear.seat.room	Luggage.room
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	True	...	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	True	...	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	True	True
...
88	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
89	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
90	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
91	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
92	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False

93 rows × 28 columns

```
In [7]: a.isna().sum()

Out[7]:
```

id	0
Manufacturer	0
Model	0
Type	0
Min.Price	0
Price	0
Max.Price	0
MPGcity	0
MPGhighway	0
AirBags	4
DriveTrain	0
Cylinders	0
EngineSize	0
Horsepower	0
RPM	0
Rev.per.mile	0
Man.trans.avail	0
FuelTankCapacity	0
Passengers	0
Length	0
Wheelbase	0
Width	0
Turn.circle	0
Rear.seat.room	2
Luggage.room	11
Weight	0
Origin	0
Make	0
dtype	int64

```
In [8]: from preprocessor import replacer
replacer(a)

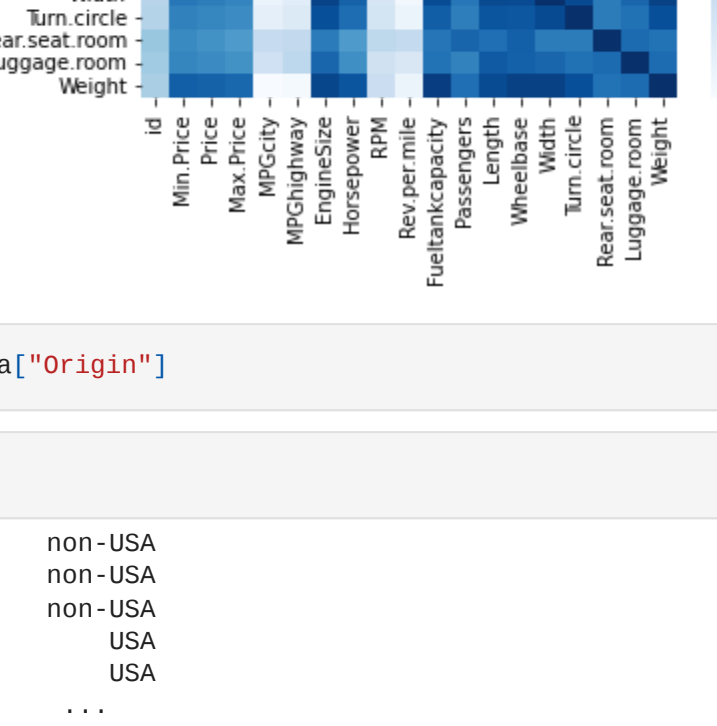
In [9]: a.isna().sum()

Out[9]:
```

id	0
Manufacturer	0
Model	0
Type	0
Min.Price	0
Price	0
Max.Price	0
MPGcity	0
MPGhighway	0
AirBags	0
DriveTrain	0
Cylinders	0
EngineSize	0
Horsepower	0
RPM	0
Rev.per.mile	0
Man.trans.avail	0
FuelTankCapacity	0
Passengers	0
Length	0
Wheelbase	0
Width	0
Turn.circle	0
Rear.seat.room	0
Luggage.room	0
Weight	0
Origin	0
Make	0
dtype	int64

```
In [10]: sb.heatmap(a.corr(),cmap='Blues')

Out[10]: <AxesSubplot:>
```



```
In [11]: y=a["Origin"]

In [12]: y

Out[12]:
```

0	non-USA
1	non-USA
2	non-USA
3	USA
4	USA
...	...
88	non-USA
89	non-USA
90	non-USA
91	non-USA
92	USA

Name: Origin, Length: 93, dtype: object

```
In [ ]:

In [13]: b=sb.drop(["RPM","Cylinders","DriveTrain","EngineSize","Luggage.room","Make","Wheelbase","EngineSize","Rev.per.mile","Man.trans.avail","Length","Rear.seat.room","Luggage.room","Weight"])

In [14]: b

Out[14]:
```

	Model	Type	Price	MPGcity	MPGhighway	Horsepower	Origin
0	300E	Midsize	61.9	19	25	217	non-USA
1	Q45	Midsize	47.9	17	22	278	non-USA
2	100	Midsize	37.7	19	26	172	non-USA
3	Seville	Midsize	40.1	16	25	295	USA
4	Corvette	Sporty	38.0	17	25	300	USA
...
88	Justy	Small	8.4	33	37	73	non-USA
89	Fox	Small	9.1	25	33	81	non-USA
90	Excel	Small	8.0	29	33	81	non-USA
91	323	Small	8.3	29	37	82	non-USA
92	Festiva	Small	7.4	31	33	63	USA

93 rows × 7 columns

```
In [15]: cat=[]
con=[]
for i in b.columns:
    if(b[i].dtypes=='object'):
        cat.append(i)
    else:
        con.append(i)

In [16]: cat

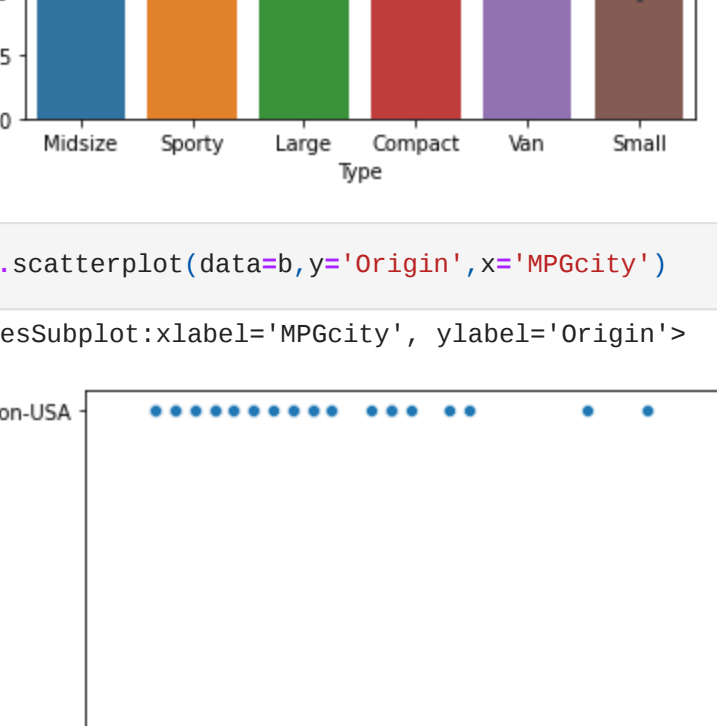
Out[16]: ['Model', 'Type', 'Origin']

In [17]: con

Out[17]: ['Price', 'MPGcity', 'MPGhighway', 'Horsepower']

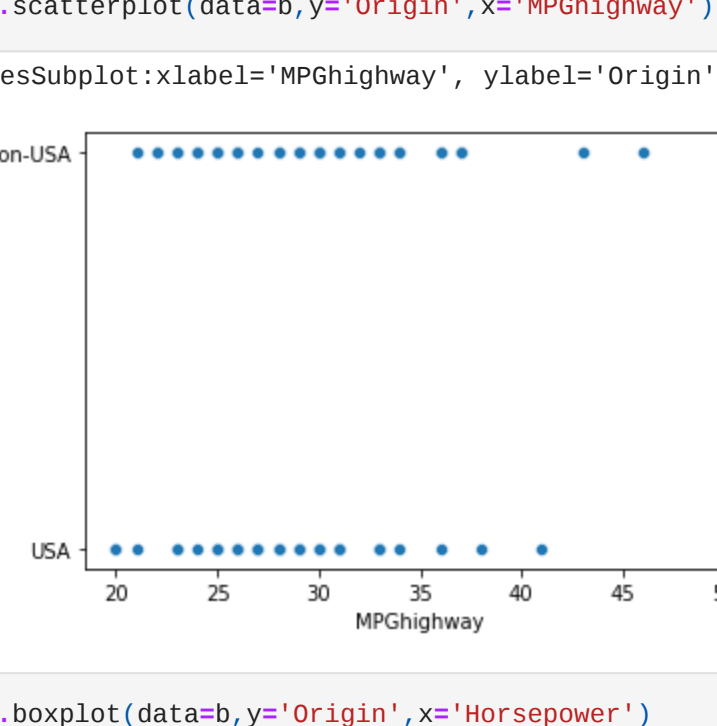
In [18]: sb.barplot(data=b,y='Price',x='Type')

Out[18]: <AxesSubplot:xlabel='Type', ylabel='Price'>
```




```
In [19]: sb.scatterplot(data=b,y='Origin',x='MPGcity')

Out[19]: <AxesSubplot:xlabel='MPGcity', ylabel='Origin'>
```



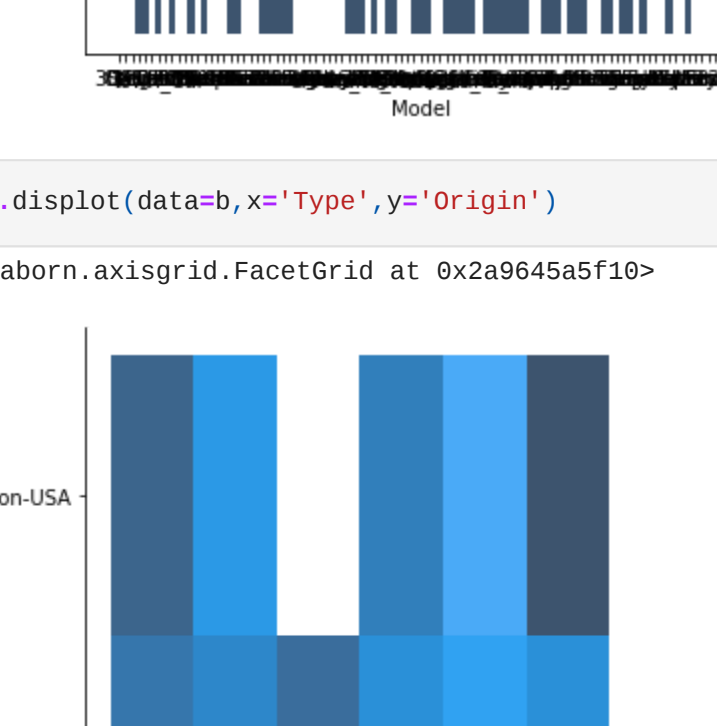
```
In [20]: sb.scatterplot(data=b,y='Origin',x='MPGhighway')

Out[20]: <AxesSubplot:xlabel='MPGhighway', ylabel='Origin'>
```



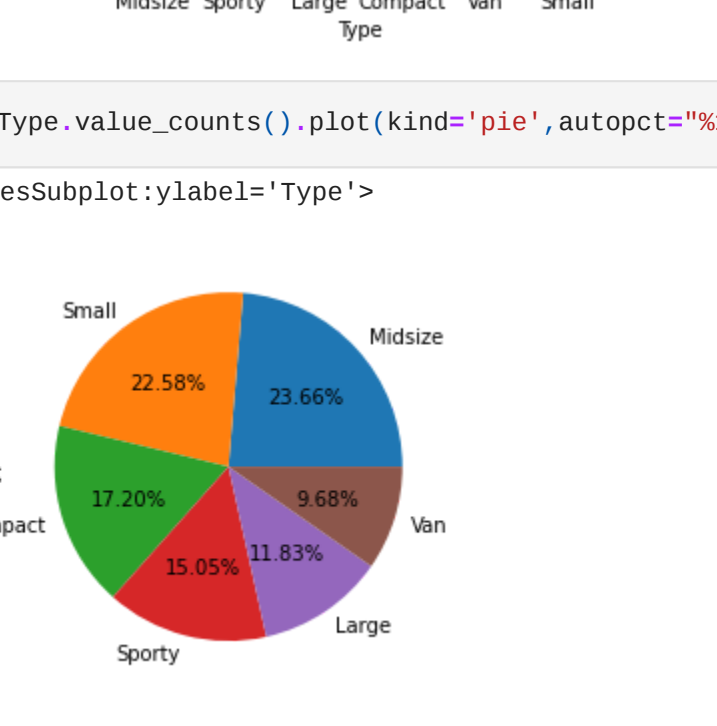
```
In [21]: sb.boxplot(data=b,y='Origin',x='Horsepower')

Out[21]: <AxesSubplot:xlabel='Horsepower', ylabel='Origin'>
```



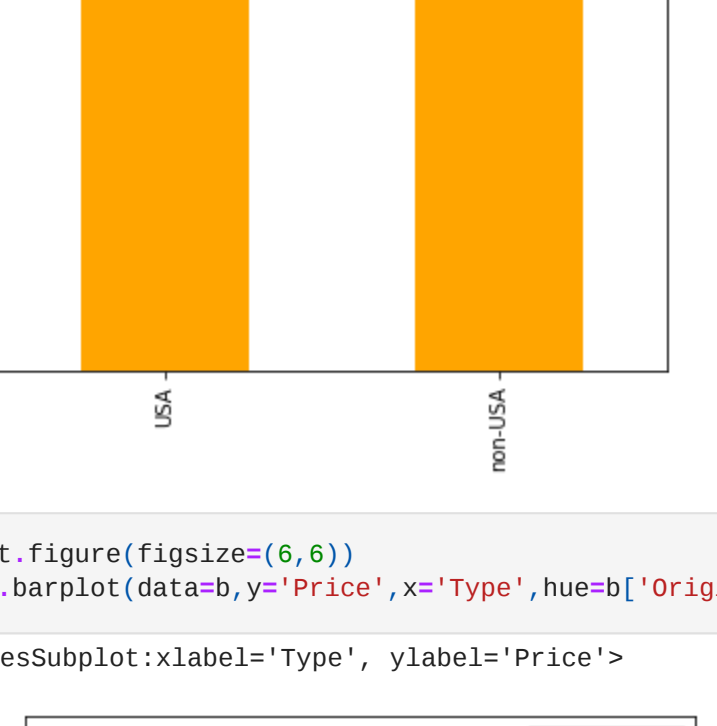
```
In [22]: sb.histplot(data=b,x='Model',y='Origin')

Out[22]: <AxesSubplot:xlabel='Model', ylabel='Origin'>
```



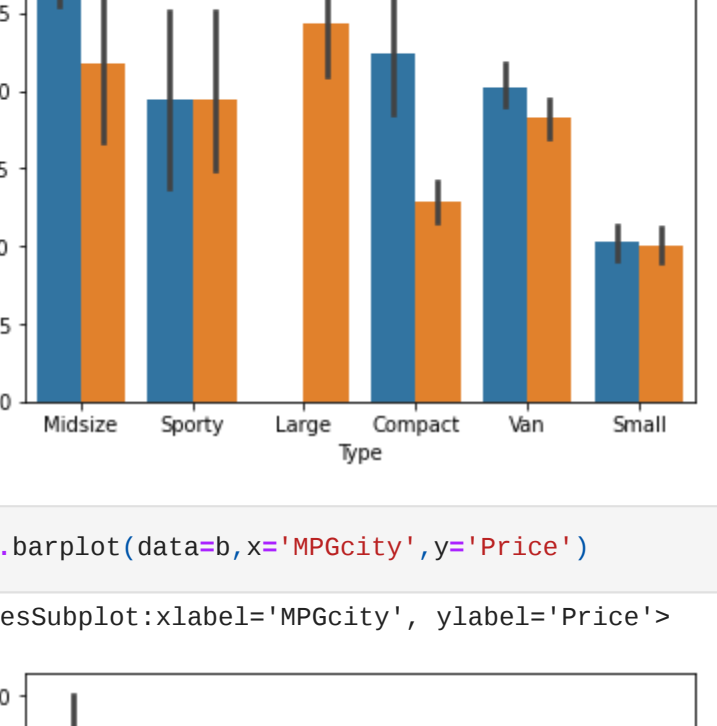
```
In [23]: sb.displot(data=b,x='Type',y='Origin')

Out[23]: <seaborn.axisgrid.FacetGrid at 0x2a9645a5f18>
```



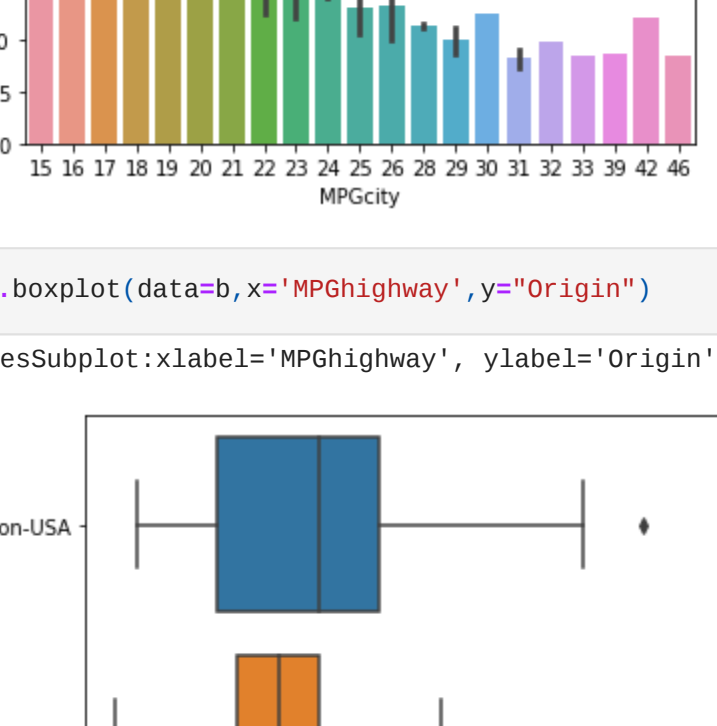
```
In [24]: b.Type.value_counts().plot(kind='pie',autopct="%1.2P%")

Out[24]: <AxesSubplot:ylabel='Type'>
```




```
In [25]: b.Origin.value_counts().plot(kind='bar',color='orange')

Out[25]: <AxesSubplot:>
```



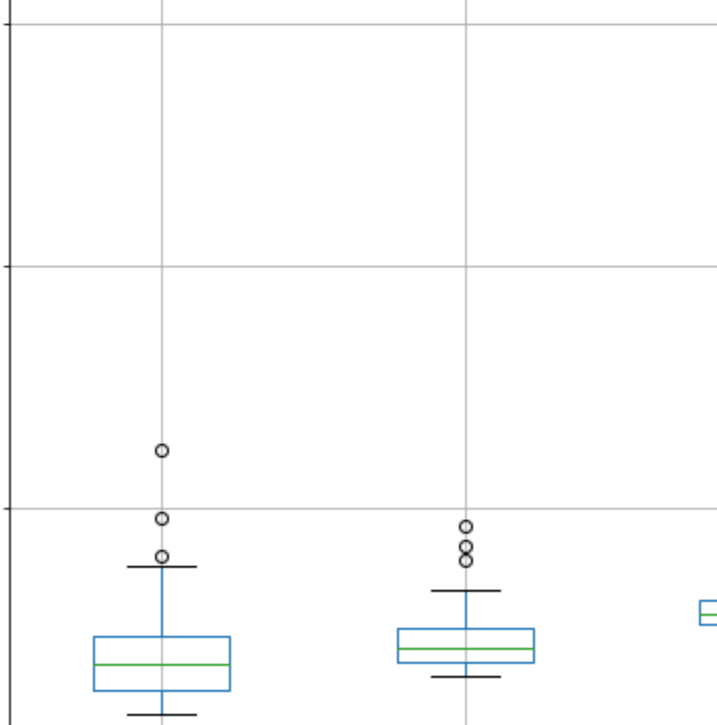
```
In [26]: plt.figure(figsize=(6,6))
sb.barplot(data=b,y='Price',x='Type',hue=b['Origin'])

Out[26]: <AxesSubplot:xlabel='Type', ylabel='Price'>
```



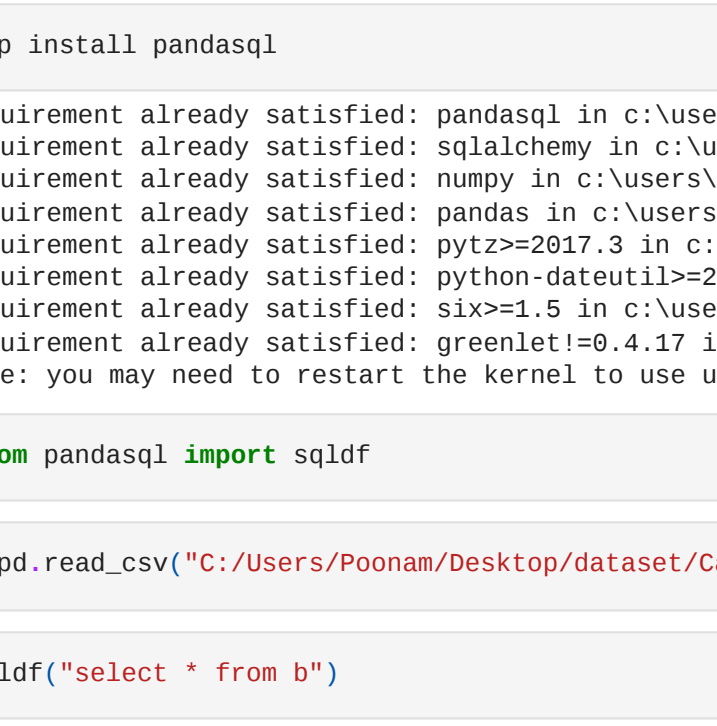
```
In [28]: sb.barplot(data=b,x='MPGcity',y='Price')

Out[28]: <AxesSubplot:xlabel='MPGcity', ylabel='Price'>
```




```
In [29]: sb.boxplot(data=b,x='MPGhighway',y='Origin')

Out[29]: <AxesSubplot:xlabel='MPGhighway', ylabel='Origin'>
```



```
In [31]: plt.figure(figsize=(50,100))
b.plot(kind='box')
plt.savefig("a.png")

In [32]:
```



```
In [33]: pip install pandasql

In [34]: bnpd.read_csv("C:/Users/Poonam/Desktop/dataset/Cars93 (2).csv")

In [35]: sqldf("select * from b")

Out[35]:
```

	id	Manufacturer	Model	Type	Min.Price	Price	Max.Price	MPGcity	MPGhighway	AirBags	...	Passengers	Length	Wheelbase	Width	Turn.circle	Rear.seat.room	Luggage.room
0	59	Mercedes-Benz	300E	Midsize	43.8	61.9	80.0	19	25	Driver & Passenger	...	5	187	110	69	37	27.0	1
1	48	Infiniti	Q45	Midsize	45.4	47.9	50.4	17	22	Driver only	...	5	200	113	72	42	29.0	1
2	4	Audi	100	Midsize	30.8	37.7	44.6	19	26	None	...	6	193	106	70	37	31.0	1
3	11	Cadillac	Seville	Midsize	37.5	40.1	42.7	16	25	Driver & Passenger	...	5	204	111	74	44	31.0	1
4	19	Chevrolet	Corvette	Sporty	34.6	38.0	41.5	17	25	Driver only	...	2	179	96	74	43	NaN	N
...
88	80	Subaru	Justy	Small	7.3	8.4	9.5	33	37	None	...	4	146	90	60	32	23.5	1
89	88	Volkswagen	Fox	Small	8.7	9.1	9.5	25	33	None	...	4	163	93	63	34	26.0	1
90	44	Hyundai	Excel	Small	6.8	8.0	9.2	29	33	None	...	5	168	94	63	35	26.0	1
91	53	Mazda	323	Small	7.4	8.3	9.1	29	37	None	...	4	164	97	66	34	27.0	1
92	31	Ford	Festiva	Small	6.9	7.4	7.9	31	33	None	...	4	141	90	63	33	26.0	1

93 rows × 28 columns

```
In [36]: sqldf("select Model,max(MPGcity)from b")

Out[36]:
```

	Model	max(MPGcity)
0	Metro	46

```
In [37]: sqldf("select Model,Price from b where Price between 17 and 25 order by Price ")

Out[37]:
```

	Model	Price
0	Accord	17.5
1	Firebird	17.7
2	Camry	18.2
3	Celica	18.4
4	Concorde	18.4
5	Grand_Prix	18.5
6	Caprice	18.8
7	Caravan	19.0
8	MPV	19.1
9	Quest	19.1
10	Vision	19.3
11	Legacy	19.5
12	Silhouette	19.5
13	Eurovan	19.7
14	Prelude	19.8
15	Aerostar	19.9
16	Pontiac	20.0
17	Taurus	20.2
18	Eighty-Eight	20.7
19	LeSabe	20.8
20	Crown_Victoria	20.9
21	Maxima	21.5
22	Previa	22.7
23	240	22.7
24	Corrado	23.3
25	Roadmaster	23.7
26	Bonneville	24.4

```
In [38]: sqldf("select MPGhighway from b order by MPGhighway desc limit 7")

Out[38]:
```

	MPGhighway
0	50
1	46
2	43
3	41
4	38
5	37
6	37

```
In [39]: sqldf("select Type,AirBags,Price from b where Type='Compact'and AirBags='None'order by Price < 20 ")

Out[39]:
```

	Type	AirBags	Price
0	Compact	None	20.0
1	Compact	None	13.4
2	Compact	None	13.5
3	Compact	None	11.1
4	Compact	None	11.3

```
In [40]: sqldf("select Model,Price from b where Type='Compact'and AirBags='None'order by Price < 20 ")

Out[40]:
```

	Model	Price
0	Pontiac	20.0
1	Cavalier	18.8
2	Achieva	21.5
3	Sunbird	22.7
4	Tempo	22.7

```
In [41]: sqldf("select MPGcity,Horsepower from b order by MPGcity desc limit 7")

Out[41]:
```

	MPGcity	Horsepower
0	46	55
1	42	102
2	39	70
3	33	73
4	32	82
5	31	74
6	31	63

```
In [ ]:

In [ ]:
```