

```
In [3]: import pandas as pd
```

```
In [4]: data= pd.read_csv('/home/placement/Desktop/fiat500.csv')
```

```
In [19]: print(data)
```

	ID	model
0	1	lounge
1	2	pop
2	3	sport
3	4	lounge
4	5	pop
...
1533	1534	sport
1534	1535	lounge
1535	1536	pop
1536	1537	lounge
1537	1538	pop

```
[1538 rows x 2 columns]
```

In [6]: `data.head(10)`

Out[6]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.611560	8900
1	2	pop	51	1186	32500	1	45.666359	12.241890	8800
2	3	sport	74	4658	142228	1	45.503300	11.417840	4200
3	4	lounge	51	2739	160000	1	40.633171	17.634609	6000
4	5	pop	73	3074	106880	1	41.903221	12.495650	5700
5	6	pop	74	3623	70225	1	45.000702	7.682270	7900
6	7	lounge	51	731	11600	1	44.907242	8.611560	10750
7	8	lounge	51	1521	49076	1	41.903221	12.495650	9190
8	9	sport	73	4049	76000	1	45.548000	11.549470	5600
9	10	sport	51	3653	89000	1	45.438301	10.991700	6000

In [7]: `data.tail(10)`

Out[7]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
1528	1529	lounge	51	2861	126000	1	43.841980	10.51531	5500
1529	1530	lounge	51	731	22551	1	38.122070	13.36112	9900
1530	1531	lounge	51	670	29000	1	45.764648	8.99450	10800
1531	1532	sport	73	4505	127000	1	45.528511	9.59323	4750
1532	1533	pop	51	1917	52008	1	45.548000	11.54947	9900
1533	1534	sport	51	3712	115280	1	45.069679	7.70492	5200
1534	1535	lounge	74	3835	112000	1	45.845692	8.66687	4600
1535	1536	pop	51	2223	60457	1	45.481541	9.41348	7500
1536	1537	lounge	51	2557	80750	1	45.000702	7.68227	5990
1537	1538	pop	51	1766	54276	1	40.323410	17.56827	7900

In [8]: `data.describe()`

Out[8]:

	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
count	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000
mean	769.500000	51.904421	1650.980494	53396.011704	1.123537	43.541361	11.563428	8576.003901
std	444.126671	3.988023	1289.522278	40046.830723	0.416423	2.133518	2.328190	1939.958641
min	1.000000	51.000000	366.000000	1232.000000	1.000000	36.855839	7.245400	2500.000000
25%	385.250000	51.000000	670.000000	20006.250000	1.000000	41.802990	9.505090	7122.500000
50%	769.500000	51.000000	1035.000000	39031.000000	1.000000	44.394096	11.869260	9000.000000
75%	1153.750000	51.000000	2616.000000	79667.750000	1.000000	45.467960	12.769040	10000.000000
max	1538.000000	77.000000	4658.000000	235000.000000	4.000000	46.795612	18.365520	11100.000000

In [9]: `data.shape`

Out[9]: (1538, 9)

```
In [10]: data.sort_values("ID")
```

```
Out[10]:
```

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.611560	8900
1	2	pop	51	1186	32500	1	45.666359	12.241890	8800
2	3	sport	74	4658	142228	1	45.503300	11.417840	4200
3	4	lounge	51	2739	160000	1	40.633171	17.634609	6000
4	5	pop	73	3074	106880	1	41.903221	12.495650	5700
...
1533	1534	sport	51	3712	115280	1	45.069679	7.704920	5200
1534	1535	lounge	74	3835	112000	1	45.845692	8.666870	4600
1535	1536	pop	51	2223	60457	1	45.481541	9.413480	7500
1536	1537	lounge	51	2557	80750	1	45.000702	7.682270	5990
1537	1538	pop	51	1766	54276	1	40.323410	17.568270	7900

1538 rows × 9 columns

```
In [11]: data.columns
```

```
Out[11]: Index(['ID', 'model', 'engine_power', 'age_in_days', 'km', 'previous_owners',
               'lat', 'lon', 'price'],
              dtype='object')
```

```
In [12]: data.index
```

```
Out[12]: RangeIndex(start=0, stop=1538, step=1)
```

```
In [13]: data=data.loc[:,['ID','model']]  
data.head(10)
```

```
Out[13]:
```

	ID	model
0	1	lounge
1	2	pop
2	3	sport
3	4	lounge
4	5	pop
5	6	pop
6	7	lounge
7	8	lounge
8	9	sport
9	10	sport

```
In [14]: x=pd.DataFrame(data['ID'])  
y=pd.DataFrame(data['model'])  
x.size,y.size
```

```
Out[14]: (1538, 1538)
```

```
In [16]: data.ndim
```

```
Out[16]: 2
```

```
In [17]: data.sample()
```

```
Out[17]:
```

	ID	model
497	498	lounge

In [20]: data.dtypes

Out[20]: ID int64
model object
dtype: object

```
In [20]: import pandas as pd
import numpy as np
data= pd.read_csv('/home/placement/Desktop/fiat500.csv')
print(data)
```

	ID	model	engine_power	age_in_days	km	previous_owners	\
0	1	lounge	51	882	25000	1	
1	2	pop	51	1186	32500	1	
2	3	sport	74	4658	142228	1	
3	4	lounge	51	2739	160000	1	
4	5	pop	73	3074	106880	1	
...	
1533	1534	sport	51	3712	115280	1	
1534	1535	lounge	74	3835	112000	1	
1535	1536	pop	51	2223	60457	1	
1536	1537	lounge	51	2557	80750	1	
1537	1538	pop	51	1766	54276	1	

	lat	lon	price
0	44.907242	8.611560	8900
1	45.666359	12.241890	8800
2	45.503300	11.417840	4200
3	40.633171	17.634609	6000
4	41.903221	12.495650	5700
...
1533	45.069679	7.704920	5200
1534	45.845692	8.666870	4600
1535	45.481541	9.413480	7500
1536	45.000702	7.682270	5990
1537	40.323410	17.568270	7900

[1538 rows x 9 columns]

In [21]: data.dtypes

```
Out[21]: ID                int64
model                object
engine_power         int64
age_in_days          int64
km                   int64
previous_owners      int64
lat                  float64
lon                  float64
price                int64
dtype: object
```

In [22]: data.describe()

```
Out[22]:
```

	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
count	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000
mean	769.500000	51.904421	1650.980494	53396.011704	1.123537	43.541361	11.563428	8576.003901
std	444.126671	3.988023	1289.522278	40046.830723	0.416423	2.133518	2.328190	1939.958641
min	1.000000	51.000000	366.000000	1232.000000	1.000000	36.855839	7.245400	2500.000000
25%	385.250000	51.000000	670.000000	20006.250000	1.000000	41.802990	9.505090	7122.500000
50%	769.500000	51.000000	1035.000000	39031.000000	1.000000	44.394096	11.869260	9000.000000
75%	1153.750000	51.000000	2616.000000	79667.750000	1.000000	45.467960	12.769040	10000.000000
max	1538.000000	77.000000	4658.000000	235000.000000	4.000000	46.795612	18.365520	11100.000000

In [23]: data['previous_owners'].unique()

```
Out[23]: array([1, 2, 3, 4])
```

In [24]: `list(data.columns)`

Out[24]:

```
['ID',
 'model',
 'engine_power',
 'age_in_days',
 'km',
 'previous_owners',
 'lat',
 'lon',
 'price']
```

In [25]: `data.groupby(['model']).count()`

Out[25]:

	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
model								
lounge	1094	1094	1094	1094	1094	1094	1094	1094
pop	358	358	358	358	358	358	358	358
sport	86	86	86	86	86	86	86	86

In [26]: `data.groupby(['previous_owners']).count()`

Out[26]:

	ID	model	engine_power	age_in_days	km	lat	lon	price
previous_owners								
1	1389	1389	1389	1389	1389	1389	1389	1389
2	117	117	117	117	117	117	117	117
3	23	23	23	23	23	23	23	23
4	9	9	9	9	9	9	9	9


```
In [27]: data.iloc[0]
```

```
Out[27]: ID                1  
         model            lounge  
         engine_power      51  
         age_in_days       882  
         km              25000  
         previous_owners    1  
         lat             44.907242  
         lon             8.61156  
         price            8900  
         Name: 0, dtype: object
```

```
In [28]: data1= data.drop(['lat', 'ID'],axis=1)  
         data1= data.drop(['lon', 'ID'],axis=1)
```

In [29]: `print(data)`

	ID	model	engine_power	age_in_days	km	previous_owners	\
0	1	lounge	51	882	25000	1	
1	2	pop	51	1186	32500	1	
2	3	sport	74	4658	142228	1	
3	4	lounge	51	2739	160000	1	
4	5	pop	73	3074	106880	1	
...
1533	1534	sport	51	3712	115280	1	
1534	1535	lounge	74	3835	112000	1	
1535	1536	pop	51	2223	60457	1	
1536	1537	lounge	51	2557	80750	1	
1537	1538	pop	51	1766	54276	1	

	lat	lon	price
0	44.907242	8.611560	8900
1	45.666359	12.241890	8800
2	45.503300	11.417840	4200
3	40.633171	17.634609	6000
4	41.903221	12.495650	5700
...
1533	45.069679	7.704920	5200
1534	45.845692	8.666870	4600
1535	45.481541	9.413480	7500
1536	45.000702	7.682270	5990
1537	40.323410	17.568270	7900

[1538 rows x 9 columns]

```
In [30]: print(data1)
```

	model	engine_power	age_in_days	km	previous_owners	lat	\
0	lounge	51	882	25000	1	44.907242	
1	pop	51	1186	32500	1	45.666359	
2	sport	74	4658	142228	1	45.503300	
3	lounge	51	2739	160000	1	40.633171	
4	pop	73	3074	106880	1	41.903221	
...	
1533	sport	51	3712	115280	1	45.069679	
1534	lounge	74	3835	112000	1	45.845692	
1535	pop	51	2223	60457	1	45.481541	
1536	lounge	51	2557	80750	1	45.000702	
1537	pop	51	1766	54276	1	40.323410	

	price
0	8900
1	8800
2	4200
3	6000
4	5700
...	...
1533	5200
1534	4600
1535	7500
1536	5990
1537	7900

```
[1538 rows x 7 columns]
```

```
In [31]: data['price'].sum()
```

```
Out[31]: 13189894
```

```
In [32]: data2=data.loc[(data.model=='lounge')&(data.previous_owners==1)]
data2
```

```
Out[32]:
```

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.611560	8900
3	4	lounge	51	2739	160000	1	40.633171	17.634609	6000
6	7	lounge	51	731	11600	1	44.907242	8.611560	10750
7	8	lounge	51	1521	49076	1	41.903221	12.495650	9190
11	12	lounge	51	366	17500	1	45.069679	7.704920	10990
...
1528	1529	lounge	51	2861	126000	1	43.841980	10.515310	5500
1529	1530	lounge	51	731	22551	1	38.122070	13.361120	9900
1530	1531	lounge	51	670	29000	1	45.764648	8.994500	10800
1534	1535	lounge	74	3835	112000	1	45.845692	8.666870	4600
1536	1537	lounge	51	2557	80750	1	45.000702	7.682270	5990

993 rows × 9 columns

```
In [33]: data['km']<50000
```

```
Out[33]: 0      True
1      True
2     False
3     False
4     False
...
1533   False
1534   False
1535   False
1536   False
1537   False
Name: km, Length: 1538, dtype: bool
```

```
In [34]: data3=data[data['km']<50000]
data3
```

```
Out[34]:
```

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.61156	8900
1	2	pop	51	1186	32500	1	45.666359	12.24189	8800
6	7	lounge	51	731	11600	1	44.907242	8.61156	10750
7	8	lounge	51	1521	49076	1	41.903221	12.49565	9190
10	11	pop	51	790	43286	1	40.871429	14.43896	8950
...
1525	1526	lounge	51	790	41870	1	45.707249	11.47760	9500
1526	1527	lounge	51	1705	23600	1	38.122070	13.36112	9300
1527	1528	pop	51	517	3000	1	40.748241	14.52835	9999
1529	1530	lounge	51	731	22551	1	38.122070	13.36112	9900
1530	1531	lounge	51	670	29000	1	45.764648	8.99450	10800

900 rows × 9 columns

In [35]: data

Out[35]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.611560	8900
1	2	pop	51	1186	32500	1	45.666359	12.241890	8800
2	3	sport	74	4658	142228	1	45.503300	11.417840	4200
3	4	lounge	51	2739	160000	1	40.633171	17.634609	6000
4	5	pop	73	3074	106880	1	41.903221	12.495650	5700
...
1533	1534	sport	51	3712	115280	1	45.069679	7.704920	5200
1534	1535	lounge	74	3835	112000	1	45.845692	8.666870	4600
1535	1536	pop	51	2223	60457	1	45.481541	9.413480	7500
1536	1537	lounge	51	2557	80750	1	45.000702	7.682270	5990
1537	1538	pop	51	1766	54276	1	40.323410	17.568270	7900

1538 rows × 9 columns

In [39]: data7=data.drop(['model'],axis=1)

In [40]: data7

Out[40]:

	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	51	882	25000	1	44.907242	8.611560	8900
1	2	51	1186	32500	1	45.666359	12.241890	8800
2	3	74	4658	142228	1	45.503300	11.417840	4200
3	4	51	2739	160000	1	40.633171	17.634609	6000
4	5	73	3074	106880	1	41.903221	12.495650	5700
...
1533	1534	51	3712	115280	1	45.069679	7.704920	5200
1534	1535	74	3835	112000	1	45.845692	8.666870	4600
1535	1536	51	2223	60457	1	45.481541	9.413480	7500
1536	1537	51	2557	80750	1	45.000702	7.682270	5990
1537	1538	51	1766	54276	1	40.323410	17.568270	7900

1538 rows × 8 columns

In [41]: data7.corr()

Out[41]:

	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
ID	1.000000	-0.034059	-0.060753	-0.006537	0.007803	-0.058207	0.058941	0.028516
engine_power	-0.034059	1.000000	0.319190	0.285495	-0.005030	0.005721	-0.005032	-0.277235
age_in_days	-0.060753	0.319190	1.000000	0.833890	0.075775	0.062982	-0.042667	-0.893328
km	-0.006537	0.285495	0.833890	1.000000	0.097539	0.035519	0.004839	-0.859373
previous_owners	0.007803	-0.005030	0.075775	0.097539	1.000000	0.001697	-0.026836	-0.076274
lat	-0.058207	0.005721	0.062982	0.035519	0.001697	1.000000	-0.766646	-0.011733
lon	0.058941	-0.005032	-0.042667	0.004839	-0.026836	-0.766646	1.000000	-0.003541
price	0.028516	-0.277235	-0.893328	-0.859373	-0.076274	-0.011733	-0.003541	1.000000

In []: