

DAY 6

kaviyadevi 20106064

```
importing libraries
```

```
In [1]: import numpy as np
import pandas as pd
```

importing dataset

```
In [2]: Users\user\Desktop\phase 2\fiat500_VehicleSelection_Dataset - fiat500_VehicleSele
```

Out[2]:

	er	age_in_days	km	previous_owners	lat	lon	price	Unnamed: 9	Unnamed: 10
	.0	882.0	25000.0	1.0	44.907242	8.611559868	8900	NaN	NaN
	.0	1186.0	32500.0	1.0	45.666359	12.24188995	8800	NaN	NaN
	.0	4658.0	142228.0	1.0	45.503300	11.41784	4200	NaN	NaN
	.0	2739.0	160000.0	1.0	40.633171	17.63460922	6000	NaN	NaN
	.0	3074.0	106880.0	1.0	41.903221	12.49565029	5700	NaN	NaN

	iN	NaN	NaN	NaN	NaN	length	5	NaN	NaN
	iN	NaN	NaN	NaN	NaN	concat	lonprice	NaN	NaN
	iN	NaN	NaN	NaN	NaN	Null values	NO	NaN	NaN
	iN	NaN	NaN	NaN	NaN	find	1	NaN	NaN
	iN	NaN	NaN	NaN	NaN	search	1	NaN	NaN

importing numeric values from data set

```
In [3]: df=data[["ID","engine_power","age_in_days","km","previous_owners","lat"]]
df
```

Out[3]:

	ID	engine_power	age_in_days	km	previous_owners	lat
0	1.0	51.0	882.0	25000.0	1.0	44.907242
1	2.0	51.0	1186.0	32500.0	1.0	45.666359
2	3.0	74.0	4658.0	142228.0	1.0	45.503300
3	4.0	51.0	2739.0	160000.0	1.0	40.633171
4	5.0	73.0	3074.0	106880.0	1.0	41.903221
...
1544	NaN	NaN	NaN	NaN	NaN	NaN
1545	NaN	NaN	NaN	NaN	NaN	NaN
1546	NaN	NaN	NaN	NaN	NaN	NaN
1547	NaN	NaN	NaN	NaN	NaN	NaN
1548	NaN	NaN	NaN	NaN	NaN	NaN

1549 rows × 6 columns

a. Find mean,median,mode and describe

```
In [4]: print(df.mean())
```

```
ID          769.500000
engine_power    51.904421
age_in_days    1650.980494
km          53396.011704
previous_owners    1.123537
lat           43.541361
dtype: float64
```

In [5]: `print(df.mode())`

```

      ID  engine_power  age_in_days      km  previous_owners      lat
0     1.0         51.0       366.0  17000.0           1.0  41.903221
1     2.0          NaN       790.0        NaN           NaN        NaN
2     3.0          NaN        NaN        NaN           NaN        NaN
3     4.0          NaN        NaN        NaN           NaN        NaN
4     5.0          NaN        NaN        NaN           NaN        NaN
...     ...         ...         ...         ...         ...         ...
1533 1534.0          NaN        NaN        NaN           NaN        NaN
1534 1535.0          NaN        NaN        NaN           NaN        NaN
1535 1536.0          NaN        NaN        NaN           NaN        NaN
1536 1537.0          NaN        NaN        NaN           NaN        NaN
1537 1538.0          NaN        NaN        NaN           NaN        NaN

```

[1538 rows x 6 columns]

In [6]: `print(df.median())`

```

ID                769.500000
engine_power      51.000000
age_in_days      1035.000000
km              39031.000000
previous_owners   1.000000
lat              44.394096
dtype: float64

```

In [9]: `print(df.describe())`

```

      ID  engine_power  age_in_days      km  previous_owners
\
count  1538.000000    1538.000000    1538.000000    1538.000000    1538.000000
mean    769.500000     51.904421    1650.980494    53396.011704     1.123537
std     444.126671     3.988023    1289.522278    40046.830723     0.416423
min       1.000000     51.000000     366.000000    1232.000000     1.000000
25%     385.250000     51.000000     670.000000    20006.250000     1.000000
50%     769.500000     51.000000    1035.000000    39031.000000     1.000000
75%    1153.750000     51.000000    2616.000000    79667.750000     1.000000
max    1538.000000     77.000000    4658.000000   235000.000000     4.000000

      lat
count  1538.000000
mean    43.541361
std      2.133518
min     36.855839
25%     41.802990
50%     44.394096
75%     45.467960
max     46.795612

```

b.) Find sum,cumsum,count,min and max values

In [10]: `print(df.sum())`

```
ID                1.183491e+06
engine_power      7.982900e+04
age_in_days       2.539208e+06
km                8.212307e+07
previous_owners   1.728000e+03
lat              6.696661e+04
dtype: float64
```

In [11]: `print(df.cumsum())`

	ID	engine_power	age_in_days	km	previous_owners	lat
0	1.0	51.0	882.0	25000.0	1.0	44.907242
1	3.0	102.0	2068.0	57500.0	2.0	90.573601
2	6.0	176.0	6726.0	199728.0	3.0	136.076900
3	10.0	227.0	9465.0	359728.0	4.0	176.710072
4	15.0	300.0	12539.0	466608.0	5.0	218.613293
...
1544	NaN	NaN	NaN	NaN	NaN	NaN
1545	NaN	NaN	NaN	NaN	NaN	NaN
1546	NaN	NaN	NaN	NaN	NaN	NaN
1547	NaN	NaN	NaN	NaN	NaN	NaN
1548	NaN	NaN	NaN	NaN	NaN	NaN

[1549 rows x 6 columns]

In [12]: `print(df.count())`

```
ID                1538
engine_power      1538
age_in_days       1538
km                1538
previous_owners   1538
lat              1538
dtype: int64
```

In [13]: `print(df.min())`

```
ID                1.000000
engine_power      51.000000
age_in_days       366.000000
km              1232.000000
previous_owners   1.000000
lat              36.855839
dtype: float64
```

In [14]: `print(df.max())`

```
ID          1538.000000
engine_power    77.000000
age_in_days    4658.000000
km          235000.000000
previous_owners    4.000000
lat           46.795612
dtype: float64
```

c.) Find covariance and correlation

In [17]: `df.cov()`

Out[17]:

	ID	engine_power	age_in_days	km	previous_owners	
ID	197248.500000	-60.325634	-3.479372e+04	-1.162613e+05	1.443071	
engine_power	-60.325634	15.904327	1.641481e+03	4.559580e+04	-0.008354	
age_in_days	-34793.715680	1641.480893	1.662868e+06	4.306313e+07	40.690115	1
km	-116261.337671	45595.798126	4.306313e+07	1.603749e+09	1626.593869	30
previous_owners	1.443071	-0.008354	4.069011e+01	1.626594e+03	0.173408	
lat	-55.153825	0.048676	1.732781e+02	3.034803e+03	0.001508	

In [18]: `df.corr()`

Out[18]:

	ID	engine_power	age_in_days	km	previous_owners	lat
ID	1.000000	-0.034059	-0.060753	-0.006537	0.007803	-0.058207
engine_power	-0.034059	1.000000	0.319190	0.285495	-0.005030	0.005721
age_in_days	-0.060753	0.319190	1.000000	0.833890	0.075775	0.062982
km	-0.006537	0.285495	0.833890	1.000000	0.097539	0.035519
previous_owners	0.007803	-0.005030	0.075775	0.097539	1.000000	0.001697
lat	-0.058207	0.005721	0.062982	0.035519	0.001697	1.000000

In []: