Basic Analysis using Numpy and Pandas

Vehicle Dataset

```
To import library
```

In [1]:
```python
import numpy as np
```

In [5]:
```python
import pandas as pd
```

To import dataset

In [7]:
```python
data=pd.read_csv(r"C:\Users\user\Downloads\vehicle.csv")
data
```

Out[7]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	l
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.6115598
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.241889
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.4178
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.634609
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.495650
...	...	...	...	...	...	...	...	
1544	NaN	NaN	NaN	NaN	NaN	NaN	NaN	leng
1545	NaN	NaN	NaN	NaN	NaN	NaN	NaN	cond
1546	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Null valu
1547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	fi
1548	NaN	NaN	NaN	NaN	NaN	NaN	NaN	sear

1549 rows × 11 columns

To get Top 20 record

In [11]: `data.head(20)`

Out[11]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.611559868
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.24188995
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.41784
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.63460922
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.49565029
5	6.0	pop	74.0	3623.0	70225.0	1.0	45.000702	7.68227005
6	7.0	lounge	51.0	731.0	11600.0	1.0	44.907242	8.611559868
7	8.0	lounge	51.0	1521.0	49076.0	1.0	41.903221	12.49565029
8	9.0	sport	73.0	4049.0	76000.0	1.0	45.548000	11.54946995
9	10.0	sport	51.0	3653.0	89000.0	1.0	45.438301	10.99170017
10	11.0	pop	51.0	790.0	43286.0	1.0	40.871429	14.43896008
11	12.0	lounge	51.0	366.0	17500.0	1.0	45.069679	7.704919815
12	13.0	lounge	51.0	456.0	18450.0	1.0	45.426571	11.78812981
13	14.0	pop	51.0	3835.0	120000.0	1.0	40.531590	17.43615913
14	15.0	lounge	51.0	1035.0	40500.0	1.0	40.911362	14.21119976
15	16.0	lounge	51.0	1096.0	28200.0	1.0	45.697208	9.845970154
16	17.0	lounge	73.0	4200.0	110000.0	1.0	41.082352	14.25424957
17	18.0	pop	51.0	2223.0	96848.0	1.0	43.782372	11.25498962
18	19.0	lounge	51.0	2861.0	31000.0	1.0	45.069679	7.704919815
19	20.0	lounge	51.0	425.0	20030.0	1.0	45.354389	11.86925983

In [ ]: `To get last 20 record`

In [12]: `data.tail(20)`

Out[12]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	
1529	1530.0	lounge	51.0	731.0	22551.0	1.0	38.122070	13.3611:
1530	1531.0	lounge	51.0	670.0	29000.0	1.0	45.764648	8.9945(
1531	1532.0	sport	73.0	4505.0	127000.0	1.0	45.528511	9.59323(
1532	1533.0	pop	51.0	1917.0	52008.0	1.0	45.548000	11.5494(
1533	1534.0	sport	51.0	3712.0	115280.0	1.0	45.069679	7.704919
1534	1535.0	lounge	74.0	3835.0	112000.0	1.0	45.845692	8.666870
1535	1536.0	pop	51.0	2223.0	60457.0	1.0	45.481541	9.413479
1536	1537.0	lounge	51.0	2557.0	80750.0	1.0	45.000702	7.6822
1537	1538.0	pop	51.0	1766.0	54276.0	1.0	40.323410	17.5682(
1538	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1539	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1540	NaN	NaN	NaN	NaN	NaN	NaN	NaN	c
1541	NaN	NaN	NaN	NaN	NaN	NaN	NaN	cc
1542	NaN	NaN	NaN	NaN	NaN	NaN	NaN	s
1543	NaN	NaN	NaN	NaN	NaN	NaN	NaN	counta en
1544	NaN	NaN	NaN	NaN	NaN	NaN	NaN	le
1545	NaN	NaN	NaN	NaN	NaN	NaN	NaN	cc
1546	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Null va
1547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1548	NaN	NaN	NaN	NaN	NaN	NaN	NaN	se

Statistical Analysis

In [13]: `data.describe()`

Out[13]:

	ID	engine_power	age_in_days	km	previous_owners	lat	U
count	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	
std	444.126671	3.988023	1289.522278	40046.830723	0.416423	2.133518	
min	1.000000	51.000000	366.000000	1232.000000	1.000000	36.855839	
25%	385.250000	51.000000	670.000000	20006.250000	1.000000	41.802990	
50%	769.500000	51.000000	1035.000000	39031.000000	1.000000	44.394096	
75%	1153.750000	51.000000	2616.000000	79667.750000	1.000000	45.467960	
max	1538.000000	77.000000	4658.000000	235000.000000	4.000000	46.795612	

To get row and column

In [14]: `print(np.shape(data))`

```
(1549, 11)
```

Find Number of Elements

In [15]: `np.size(data)`

Out[15]: `17039`

Find Missing values

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

Out[17]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price	Un
0	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	
1544	True	True	True	True	True	True	True	False	False	
1545	True	True	True	True	True	True	True	False	False	
1546	True	True	True	True	True	True	True	False	False	
1547	True	True	True	True	True	True	True	False	False	
1548	True	True	True	True	True	True	True	False	False	

1549 rows × 11 columns

To fill the missing values

In [18]: `data.fillna(value=5)`

Out[18]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.611559868
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.241889995
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.41784
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.634609227
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.495650257
...	...	...	...	...	...	...	...	...
1544	5.0	5	5.0	5.0	5.0	5.0	5.000000	lengtl
1545	5.0	5	5.0	5.0	5.0	5.0	5.000000	conca
1546	5.0	5	5.0	5.0	5.0	5.0	5.000000	Null value:
1547	5.0	5	5.0	5.0	5.0	5.0	5.000000	fin
1548	5.0	5	5.0	5.0	5.0	5.0	5.000000	searcl

1549 rows × 11 columns

In [ ]: