

dataset analysis using pandas

```
In [16]: import numpy as np
import pandas as pd
```

Import dataset

```
In [3]: data=pd.read_csv(r"C:\Users\user\Downloads\fiat500_VehicleSelection_Dataset(1).csv")
```

print data

```
In [4]: data
```

Out[4]:

	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
...
1544	NaN	NaN	NaN	NaN	NaN	NaN	NaN	length
1545	NaN	NaN	NaN	NaN	NaN	NaN	NaN	concat
1546	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Null values
1547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	find
1548	NaN	NaN	NaN	NaN	NaN	NaN	NaN	search

1549 rows × 11 columns



print first 10 rows using head

In [5]: `data.head(10)`

Out[5]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.611559868	8900
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.24188995	8800
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.41784	4200
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.63460922	6000
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.49565029	5700
5	6.0	pop	74.0	3623.0	70225.0	1.0	45.000702	7.68227005	7900
6	7.0	lounge	51.0	731.0	11600.0	1.0	44.907242	8.611559868	10750
7	8.0	lounge	51.0	1521.0	49076.0	1.0	41.903221	12.49565029	9190
8	9.0	sport	73.0	4049.0	76000.0	1.0	45.548000	11.54946995	5600
9	10.0	sport	51.0	3653.0	89000.0	1.0	45.438301	10.99170017	6000

print last 10 rows using tail

In [6]: `data.tail(5)`

Out[6]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price	Unnamed: 11
1544	NaN	NaN	NaN	NaN	NaN	NaN	NaN	length	5	NaN
1545	NaN	NaN	NaN	NaN	NaN	NaN	NaN	concat	lonprice	NaN
1546	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Null values	NO	NaN
1547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	find	1	NaN
1548	NaN	NaN	NaN	NaN	NaN	NaN	NaN	search	1	NaN

print describe of dataset

In [7]: `data.describe()`

Out[7]:

	ID	engine_power	age_in_days	km	previous_owners	lat	Unnamed: 0
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	N
std	444.126671	3.988023	1289.522278	40046.830723	0.416423	2.133518	N
min	1.000000	51.000000	366.000000	1232.000000	1.000000	36.855839	N
25%	385.250000	51.000000	670.000000	20006.250000	1.000000	41.802990	N
50%	769.500000	51.000000	1035.000000	39031.000000	1.000000	44.394096	N
75%	1153.750000	51.000000	2616.000000	79667.750000	1.000000	45.467960	N
max	1538.000000	77.000000	4658.000000	235000.000000	4.000000	46.795612	N

Number elements in dataset

In [9]: `data.size`

Out[9]: 17039

`print shape of dataset`

In [14]: `data.shape`

Out[14]: (1549, 11)

`print empty or not`

In [15]:

data.isna()

Out[15]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price	Unnamed: 9
0	False	False	False	False	False	False	False	False	False	True
1	False	False	False	False	False	False	False	False	False	True
2	False	False	False	False	False	False	False	False	False	True
3	False	False	False	False	False	False	False	False	False	True
4	False	False	False	False	False	False	False	False	False	True
...
1544	True	True	True	True	True	True	True	False	False	True
1545	True	True	True	True	True	True	True	False	False	True
1546	True	True	True	True	True	True	True	False	False	True
1547	True	True	True	True	True	True	True	False	False	True
1548	True	True	True	True	True	True	True	False	False	True

1549 rows × 11 columns

In [18]:

data.fillna(value=10)

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	12.24188995
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
...
1544	10.0	10	10.0	10.0	10.0	10.0	10.000000	length
1545	10.0	10	10.0	10.0	10.0	10.0	10.000000	concat lon
1546	10.0	10	10.0	10.0	10.0	10.0	10.000000	Null values
1547	10.0	10	10.0	10.0	10.0	10.0	10.000000	find
1548	10.0	10	10.0	10.0	10.0	10.0	10.000000	search

1549 rows × 11 columns

In [19]:

data.loc["ID": "km"]

Out[19]:

ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price	Unnamed: 9	Unnamed: 10
----	-------	--------------	-------------	----	-----------------	-----	-----	-------	------------	-------------

In [20]: data[10:20]

Out[20]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	pric
10	11.0	pop	51.0	790.0	43286.0	1.0	40.871429	14.43896008	895
11	12.0	lounge	51.0	366.0	17500.0	1.0	45.069679	7.704919815	1099
12	13.0	lounge	51.0	456.0	18450.0	1.0	45.426571	11.78812981	970
13	14.0	pop	51.0	3835.0	120000.0	1.0	40.531590	17.43615913	480
14	15.0	lounge	51.0	1035.0	40500.0	1.0	40.911362	14.21119976	930
15	16.0	lounge	51.0	1096.0	28200.0	1.0	45.697208	9.845970154	950
16	17.0	lounge	73.0	4200.0	110000.0	1.0	41.082352	14.25424957	525
17	18.0	pop	51.0	2223.0	96848.0	1.0	43.782372	11.25498962	799
18	19.0	lounge	51.0	2861.0	31000.0	1.0	45.069679	7.704919815	730
19	20.0	lounge	51.0	425.0	20030.0	1.0	45.354389	11.86925983	1050

In [21]: data[data["ID"]<=20]

Out[21]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	pric
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.611559868	890
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.24188995	880
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.41784	420
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.63460922	600
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.49565029	570
5	6.0	pop	74.0	3623.0	70225.0	1.0	45.000702	7.68227005	790
6	7.0	lounge	51.0	731.0	11600.0	1.0	44.907242	8.611559868	1075
7	8.0	lounge	51.0	1521.0	49076.0	1.0	41.903221	12.49565029	919
8	9.0	sport	73.0	4049.0	76000.0	1.0	45.548000	11.54946995	560
9	10.0	sport	51.0	3653.0	89000.0	1.0	45.438301	10.99170017	600
10	11.0	pop	51.0	790.0	43286.0	1.0	40.871429	14.43896008	895
11	12.0	lounge	51.0	366.0	17500.0	1.0	45.069679	7.704919815	1099
12	13.0	lounge	51.0	456.0	18450.0	1.0	45.426571	11.78812981	970
13	14.0	pop	51.0	3835.0	120000.0	1.0	40.531590	17.43615913	480
14	15.0	lounge	51.0	1035.0	40500.0	1.0	40.911362	14.21119976	930
15	16.0	lounge	51.0	1096.0	28200.0	1.0	45.697208	9.845970154	950
16	17.0	lounge	73.0	4200.0	110000.0	1.0	41.082352	14.25424957	525
17	18.0	pop	51.0	2223.0	96848.0	1.0	43.782372	11.25498962	799
18	19.0	lounge	51.0	2861.0	31000.0	1.0	45.069679	7.704919815	730
19	20.0	lounge	51.0	425.0	20030.0	1.0	45.354389	11.86925983	1050

In [22]: data["ID"]

Out[22]:

```

0      1.0
1      2.0
2      3.0
3      4.0
4      5.0
...
1544   NaN
1545   NaN
1546   NaN
1547   NaN
1548   NaN
Name: ID, Length: 1549, dtype: float64

```

```
In [25]: data["km"]
```

```
Out[25]: 0      25000.0  
         1      32500.0  
         2     142228.0  
         3     160000.0  
         4     106880.0  
         ...  
        1544      NaN  
        1545      NaN  
        1546      NaN  
        1547      NaN  
        1548      NaN  
        Name: km, Length: 1549, dtype: float64
```

```
In [ ]:
```