SUMESH R - 20104169

In [1]:
```
import pandas as pd
import numpy as np
from numpy import cov
from scipy.stats import pearsonr
from scipy.stats import spearmanr
```

import data

In [2]:
```
df=pd.read_csv("1_fiat500_VehicleSelection_Dataset.csv")[0:1520]
df
```

Out[2]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	p
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.611559868	8
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.24188995	8
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.41784	4
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.63460922	6
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.49565029	5
...	...	...	...	...	...	...	...	...	
1515	1516.0	lounge	51.0	1917.0	124999.0	1.0	45.564491	10.11561012	6
1516	1517.0	pop	73.0	3865.0	80500.0	1.0	40.704109	17.34005928	8
1517	1518.0	pop	51.0	366.0	16100.0	1.0	44.692520	10.10396004	10
1518	1519.0	lounge	51.0	397.0	16053.0	1.0	38.122070	13.36112022	10
1519	1520.0	lounge	51.0	670.0	30000.0	1.0	45.764648	8.99450016	10

1520 rows × 11 columns

sum

In [3]:
```
df.sum()
```

Out[3]:
```
ID                                                    1155960.0
model             loungepopsportloungepoppoploungeloungesportspo...
engine_power                                           78866.0
age_in_days                                          2501801.0
km                                                  81027556.0
previous_owners                                        1710.0
lat                                                 66183.483574
lon               8.61155986812.2418899511.4178417.6346092212.49...
price             89008800420060005700790010750919056006000895O1...
Unnamed: 9                                                0.0
```

```
Unnamed: 10                                                    0
dtype: object
```

mean

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

Out[4]:
```
ID                 7.605000e+02
engine_power       5.188553e+01
age_in_days        1.645922e+03
km                 5.330760e+04
previous_owners    1.125000e+00
lat                4.354177e+01
price                       inf
Unnamed: 9                  NaN
Unnamed: 10                 NaN
dtype: float64
```

median

In [5]:
```
df.median()
```

Out[5]:
```
ID                  760.500000
engine_power         51.000000
age_in_days        1035.000000
km                39000.000000
previous_owners       1.000000
lat                  44.388222
lon                  11.869260
price              9000.000000
Unnamed: 9                 NaN
Unnamed: 10                NaN
dtype: float64
```

mode

In [6]:
```
df.mode()
```

Out[6]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	pr
0	1.0	lounge	51.0	366.0	17000.0	1.0	41.903221	12.49565029	105
1	2.0	NaN	NaN	790.0	NaN	NaN	NaN	NaN	N
2	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
3	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
4	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
...	...	...	...	...	...	...	...	...	
1515	1516.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1516	1517.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1517	1518.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	pr
1518	1519.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1519	1520.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

1520 rows × 11 columns

describe

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

Out[7]:

	ID	engine_power	age_in_days	km	previous_owners	lat	Unnamed: g
count	1520.000000	1520.000000	1520.000000	1520.000000	1520.000000	1520.000000	0.0
mean	760.500000	51.885526	1645.921711	53307.602632	1.125000	43.541766	NaN
std	438.930518	3.933121	1289.268008	40049.468144	0.418664	2.126313	NaN
min	1.000000	51.000000	366.000000	1232.000000	1.000000	36.855839	NaN
25%	380.750000	51.000000	670.000000	20000.000000	1.000000	41.802990	NaN
50%	760.500000	51.000000	1035.000000	39000.000000	1.000000	44.388222	NaN
75%	1140.250000	51.000000	2616.000000	79000.000000	1.000000	45.467960	NaN
max	1520.000000	77.000000	4658.000000	235000.000000	4.000000	46.795612	NaN

cumsum

In [8]:
```
df.cumsum()
```

Out[8]:

	ID	model	engine_power	age_in_days	k
0	1.0	lounge	51.0	882.0	2500
1	3.0	loungepop	102.0	2068.0	5750
2	6.0	loungepopsport	176.0	6726.0	19972
3	10.0	loungepopsportlounge	227.0	9465.0	35972
4	15.0	loungepopsportloungepop	300.0	12539.0	46660
...	...	...	...	...	
1515	1149886.0	loungepopsportloungepoppoploungeloungesportspo...	78640.0	2496503.0	8088490
1516	1151403.0	loungepopsportloungepoppoploungeloungesportspo...	78713.0	2500368.0	8096540
1517	1152921.0	loungepopsportloungepoppoploungeloungesportspo...	78764.0	2500734.0	8098150

	ID	model	engine_power	age_in_days	k
1518	1154440.0	loungepopsportloungepoppoploungeloungesportspo...	78815.0	2501131.0	8099755
1519	1155960.0	loungepopsportloungepoppoploungeloungesportspo...	78866.0	2501801.0	8102755

1520 rows × 11 columns

count

In [9]:
```
df.count()
```

Out[9]:
```
ID                1520
model             1520
engine_power      1520
age_in_days       1520
km                1520
previous_owners   1520
lat               1520
lon               1520
price             1520
Unnamed: 9           0
Unnamed: 10          0
dtype: int64
```

min

In [10]:
```
df.min()
```

Out[10]:
```
ID                      1.0
model                lounge
engine_power           51.0
age_in_days           366.0
km                   1232.0
previous_owners         1.0
lat                36.855839
lon                10.00240993
price                10000
Unnamed: 9             NaN
Unnamed: 10           None
dtype: object
```

max

In [11]:
```
df.max()
```

Out[11]:
```
ID                   1520.0
model                 sport
engine_power           77.0
age_in_days          4658.0
km                 235000.0
previous_owners         4.0
lat                46.795612
lon                9.980259895
```

```
price                   9999
Unnamed: 9               NaN
Unnamed: 10             None
dtype: object
```

covariance

In [12]:
```
cov(df["km"],df["ID"])
```

Out[12]:
```
array([[ 1.60395990e+09, -1.86550668e+05],
       [-1.86550668e+05,  1.92660000e+05]])
```

correlation

In [13]:
```
spearmanr(df["km"],df["ID"])
```

Out[13]:
```
SpearmanrResult(correlation=0.020403240138623772, pvalue=0.4266757050549993)
```

In [14]:
```
pearsonr(df["km"],df["ID"])
```

Out[14]:
```
(-0.010612171942226996, 0.6793063255303047)
```