Day 6

Vehicle Dataset

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
d=pd.read_csv(r"c:\Users\user\Downloads\ve.csv")
d
```

Out[2]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.611
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.241
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.634
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.495
...	...	...	...	...	...	...	...	
1544	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1545	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1546	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Null
1547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1548	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

1549 rows × 11 columns

Mean,median,mode,describe

In [17]:

```
data=pd.DataFrame(d[['engine_power','km']][0:500])
data
```

Out[17]:

	engine_power	km
0	51.0	25000.0
1	51.0	32500.0
2	74.0	142228.0
3	51.0	160000.0
4	73.0	106880.0
...	...	...
495	51.0	15003.0
496	51.0	38718.0
497	51.0	17488.0
498	51.0	24281.0
499	51.0	25076.0

500 rows × 2 columns

In [4]:

```
print(data.mean())
```

```
engine_power    51.908
price              inf
dtype: float64
```

In [5]:

```
print(data.median())
```

```
engine_power      51.0
price           9145.0
dtype: float64
```

In [18]:

```
data.fillna(value=1)
```

Out[18]:

	engine_power	km
0	51.0	25000.0
1	51.0	32500.0
2	74.0	142228.0
3	51.0	160000.0
4	73.0	106880.0
...	...	...
495	51.0	15003.0
496	51.0	38718.0
497	51.0	17488.0
498	51.0	24281.0
499	51.0	25076.0

500 rows × 2 columns

In [7]:

```
print(data.mode())
```

```
   engine_power  price
0          51.0  10500
```

In [8]:

```
print(data.describe())
```

```
       engine_power
count     500.00000
mean       51.90800
std         4.03337
min        51.00000
25%        51.00000
50%        51.00000
75%        51.00000
max        77.00000
```

Sum,cumsum,count,min,max

In [9]:

```
print(data.sum())
```

```
engine_power                                                 25954.0
price           8900880042006000570079001075091905600600089501...
dtype: object
```

In [10]:

```
print(data.cumsum())
```

```
     engine_power                                              price
0            51.0                                               8900
1           102.0                                           89008800
2           176.0                                         890088004200
3           227.0                                      8900880042006000
4           300.0                                   89008800420060005700
..            ...                                                  ...
495       25750.0  8900880042006000570079001075091905600600089501...
496       25801.0  8900880042006000570079001075091905600600089501...
497       25852.0  8900880042006000570079001075091905600600089501...
498       25903.0  8900880042006000570079001075091905600600089501...
499       25954.0  8900880042006000570079001075091905600600089501...

[500 rows x 2 columns]
```

In [11]:

```
print(data.count())
```

```
engine_power    500
price           500
dtype: int64
```

In [12]:

```
print(data.min())
```

```
engine_power     51.0
price           10000
dtype: object
```

In [13]:

```
print(data.max())
```

```
engine_power    77.0
price           9999
dtype: object
```

covariance and correlation (spearman and pearsons)

In [14]:

```
data1=data['engine_power'][0:10]
data1
```

Out[14]:

```
0    51.0
1    51.0
2    74.0
3    51.0
4    73.0
5    74.0
6    51.0
7    51.0
8    73.0
9    51.0
Name: engine_power, dtype: float64
```

In [19]:

```
data2=data['km'][0:10]
data2
```

Out[19]:

```
0     25000.0
1     32500.0
2    142228.0
3    160000.0
4    106880.0
5     70225.0
6     11600.0
7     49076.0
8     76000.0
9     89000.0
Name: km, dtype: float64
```

In [20]:

```
from numpy import cov
print(cov(data1,data2))
```

```
[[1.35111111e+02 2.27466444e+05]
 [2.27466444e+05 2.44032836e+09]]
```

In [21]:

```
from scipy.stats import pearsonr
print(pearsonr(data1,data2))
```

```
(0.39613906530125964, 0.25710544510156774)
```

In [22]:

```
from scipy.stats import spearmanr
print(spearmanr(data1,data2))
```

```
SpearmanrResult(correlation=0.4128614119223852, pvalue=0.2357037774356011
1)
```

In [ ]:

```
```