

fiat500.csv

#7 vehicledata2

16june23

#saimohan

#16/6/2023

```
In [37]: import pandas as pd  
#import pickle
```

```
In [38]: #by using vehicle data ,i can do analysis.....  
data=pd.read_csv("/home/placement/Desktop/saimohan data/csv files/fiat500.csv")
```

```
In [39]: data.head()
```

Out[39]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|---|----|--------|--------------|-------------|--------|-----------------|-----------|-----------|-------|
| 0 | 1 | lounge | 51 | 882 | 25000 | 1 | 44.907242 | 8.611560 | 8900 |
| 1 | 2 | pop | 51 | 1186 | 32500 | 1 | 45.666359 | 12.241890 | 8800 |
| 2 | 3 | sport | 74 | 4658 | 142228 | 1 | 45.503300 | 11.417840 | 4200 |
| 3 | 4 | lounge | 51 | 2739 | 160000 | 1 | 40.633171 | 17.634609 | 6000 |
| 4 | 5 | pop | 73 | 3074 | 106880 | 1 | 41.903221 | 12.495650 | 5700 |

```
In [40]: data.describe()
```

```
Out[40]:
```

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|-------|-------------|--------------|-------------|---------------|-----------------|-------------|-------------|--------------|
| count | 1538.000000 | 1538.000000 | 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 | 11.563428 | 8576.003901 |
| std | 444.126671 | 3.988023 | 1289.522278 | 40046.830723 | 0.416423 | 2.133518 | 2.328190 | 1939.958641 |
| min | 1.000000 | 51.000000 | 366.000000 | 1232.000000 | 1.000000 | 36.855839 | 7.245400 | 2500.000000 |
| 25% | 385.250000 | 51.000000 | 670.000000 | 20006.250000 | 1.000000 | 41.802990 | 9.505090 | 7122.500000 |
| 50% | 769.500000 | 51.000000 | 1035.000000 | 39031.000000 | 1.000000 | 44.394096 | 11.869260 | 9000.000000 |
| 75% | 1153.750000 | 51.000000 | 2616.000000 | 79667.750000 | 1.000000 | 45.467960 | 12.769040 | 10000.000000 |
| max | 1538.000000 | 77.000000 | 4658.000000 | 235000.000000 | 4.000000 | 46.795612 | 18.365520 | 11100.000000 |

```
In [41]: list(data)
```

```
Out[41]: ['ID',  
          'model',  
          'engine_power',  
          'age_in_days',  
          'km',  
          'previous_owners',  
          'lat',  
          'lon',  
          'price']
```

```
In [42]: data1=data.loc[data.km<= 50000]
data1
```

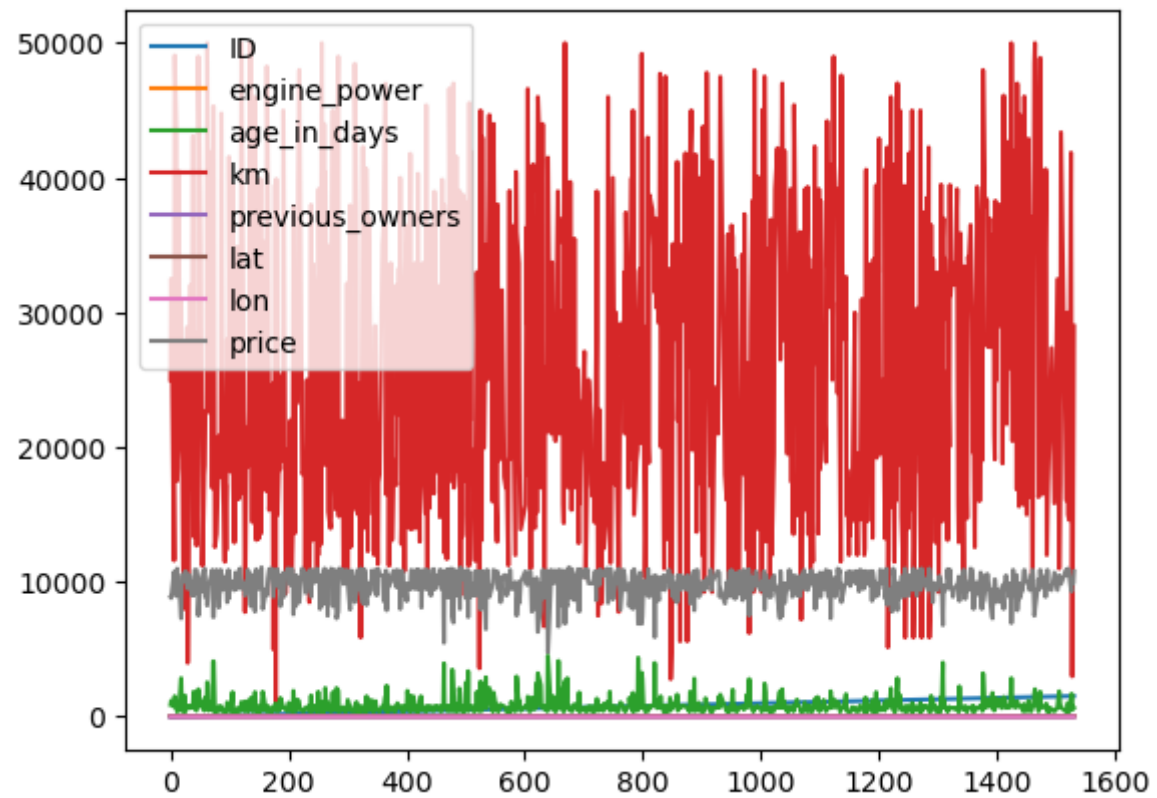
Out[42]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|------|------|--------|--------------|-------------|-------|-----------------|-----------|----------|-------|
| 0 | 1 | lounge | 51 | 882 | 25000 | 1 | 44.907242 | 8.61156 | 8900 |
| 1 | 2 | pop | 51 | 1186 | 32500 | 1 | 45.666359 | 12.24189 | 8800 |
| 6 | 7 | lounge | 51 | 731 | 11600 | 1 | 44.907242 | 8.61156 | 10750 |
| 7 | 8 | lounge | 51 | 1521 | 49076 | 1 | 41.903221 | 12.49565 | 9190 |
| 10 | 11 | pop | 51 | 790 | 43286 | 1 | 40.871429 | 14.43896 | 8950 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1525 | 1526 | lounge | 51 | 790 | 41870 | 1 | 45.707249 | 11.47760 | 9500 |
| 1526 | 1527 | lounge | 51 | 1705 | 23600 | 1 | 38.122070 | 13.36112 | 9300 |
| 1527 | 1528 | pop | 51 | 517 | 3000 | 1 | 40.748241 | 14.52835 | 9999 |
| 1529 | 1530 | lounge | 51 | 731 | 22551 | 1 | 38.122070 | 13.36112 | 9900 |
| 1530 | 1531 | lounge | 51 | 670 | 29000 | 1 | 45.764648 | 8.99450 | 10800 |

907 rows × 9 columns

```
In [53]: data1.plot()
```

```
Out[53]: <Axes: >
```



```
In [43]: data2=data.groupby(['model']).count()  
  
data2
```

Out[43]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|--------|------|--------------|-------------|------|-----------------|------|------|-------|
| model | | | | | | | | |
| lounge | 1094 | 1094 | 1094 | 1094 | 1094 | 1094 | 1094 | 1094 |
| pop | 358 | 358 | 358 | 358 | 358 | 358 | 358 | 358 |
| sport | 86 | 86 | 86 | 86 | 86 | 86 | 86 | 86 |

```
In [ ]:
```

```
In [44]: data['model']=data['model'].map({'lounge':1,'pop':2,'sport':3})
```

In [45]: data

Out[45]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|------|------|-------|--------------|-------------|--------|-----------------|-----------|-----------|-------|
| 0 | 1 | 1 | 51 | 882 | 25000 | 1 | 44.907242 | 8.611560 | 8900 |
| 1 | 2 | 2 | 51 | 1186 | 32500 | 1 | 45.666359 | 12.241890 | 8800 |
| 2 | 3 | 3 | 74 | 4658 | 142228 | 1 | 45.503300 | 11.417840 | 4200 |
| 3 | 4 | 1 | 51 | 2739 | 160000 | 1 | 40.633171 | 17.634609 | 6000 |
| 4 | 5 | 2 | 73 | 3074 | 106880 | 1 | 41.903221 | 12.495650 | 5700 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1533 | 1534 | 3 | 51 | 3712 | 115280 | 1 | 45.069679 | 7.704920 | 5200 |
| 1534 | 1535 | 1 | 74 | 3835 | 112000 | 1 | 45.845692 | 8.666870 | 4600 |
| 1535 | 1536 | 2 | 51 | 2223 | 60457 | 1 | 45.481541 | 9.413480 | 7500 |
| 1536 | 1537 | 1 | 51 | 2557 | 80750 | 1 | 45.000702 | 7.682270 | 5990 |
| 1537 | 1538 | 2 | 51 | 1766 | 54276 | 1 | 40.323410 | 17.568270 | 7900 |

1538 rows × 9 columns

```
In [46]: cor=data.corr()  
cor
```

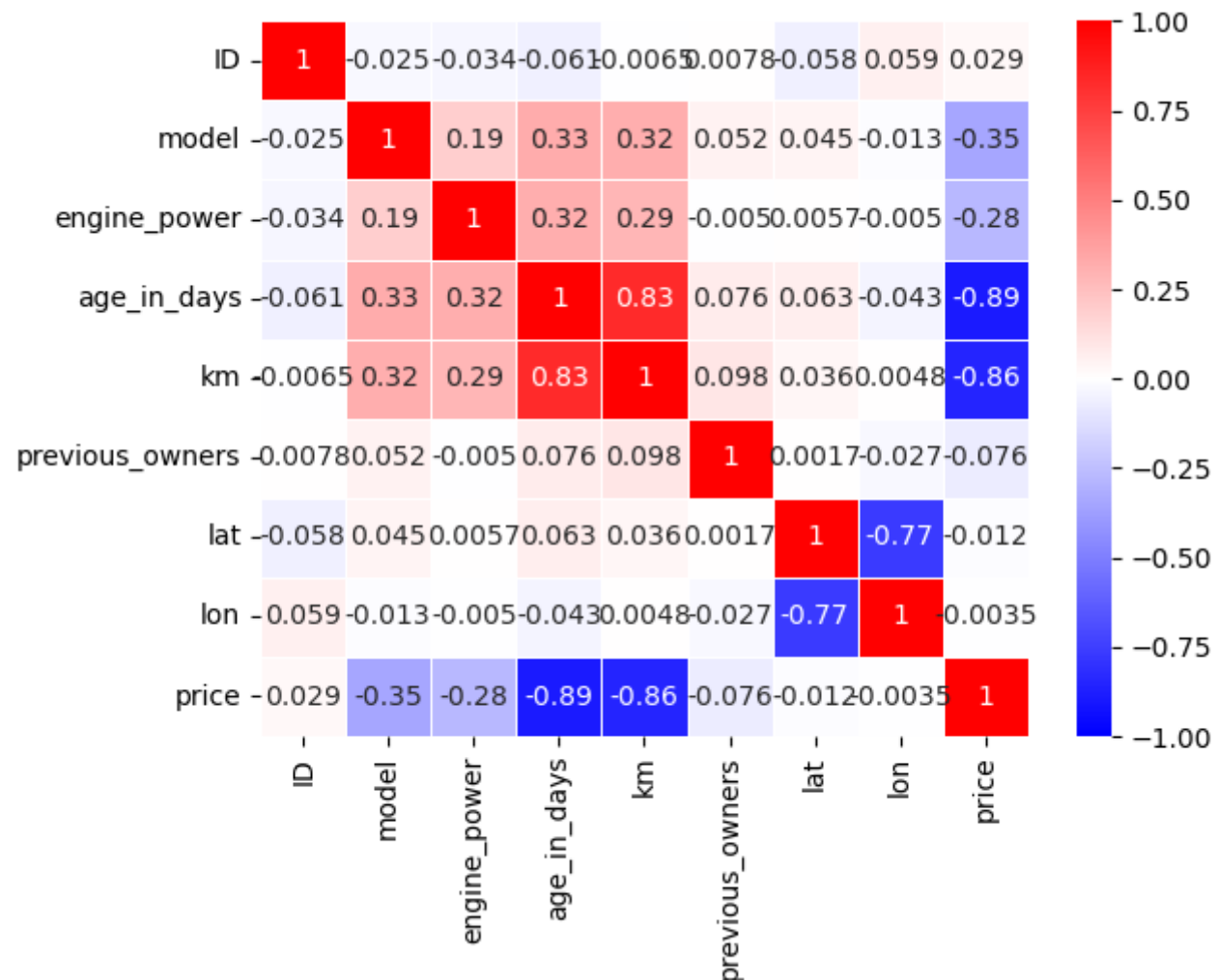
Out[46]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|------------------------|-----------|-----------|--------------|-------------|-----------|-----------------|-----------|-----------|-----------|
| ID | 1.000000 | -0.024740 | -0.034059 | -0.060753 | -0.006537 | 0.007803 | -0.058207 | 0.058941 | 0.028516 |
| model | -0.024740 | 1.000000 | 0.189906 | 0.326508 | 0.319580 | 0.052480 | 0.044901 | -0.013200 | -0.349885 |
| engine_power | -0.034059 | 0.189906 | 1.000000 | 0.319190 | 0.285495 | -0.005030 | 0.005721 | -0.005032 | -0.277235 |
| age_in_days | -0.060753 | 0.326508 | 0.319190 | 1.000000 | 0.833890 | 0.075775 | 0.062982 | -0.042667 | -0.893328 |
| km | -0.006537 | 0.319580 | 0.285495 | 0.833890 | 1.000000 | 0.097539 | 0.035519 | 0.004839 | -0.859373 |
| previous_owners | 0.007803 | 0.052480 | -0.005030 | 0.075775 | 0.097539 | 1.000000 | 0.001697 | -0.026836 | -0.076274 |
| lat | -0.058207 | 0.044901 | 0.005721 | 0.062982 | 0.035519 | 0.001697 | 1.000000 | -0.766646 | -0.011733 |
| lon | 0.058941 | -0.013200 | -0.005032 | -0.042667 | 0.004839 | -0.026836 | -0.766646 | 1.000000 | -0.003541 |
| price | 0.028516 | -0.349885 | -0.277235 | -0.893328 | -0.859373 | -0.076274 | -0.011733 | -0.003541 | 1.000000 |

```
In [47]: #-----
import seaborn as sns

sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=.5,cmap='bwr')
#finding good correlation between 2 columns by using heat map
```

Out[47]: <Axes: >




```
In [48]: #we can modify the column name with new name  
data3=data.rename(columns={'model':'model_name'})  
list(data3)
```

```
Out[48]: ['ID',  
          'model_name',  
          'engine_power',  
          'age_in_days',  
          'km',  
          'previous_owners',  
          'lat',  
          'lon',  
          'price']
```

```
In [49]: data3
# we can observe below table ,column name changed as (model) to (model_name)
```

```
Out[49]:
```

| | ID | model_name | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|------|------|------------|--------------|-------------|--------|-----------------|-----------|-----------|-------|
| 0 | 1 | 1 | 51 | 882 | 25000 | 1 | 44.907242 | 8.611560 | 8900 |
| 1 | 2 | 2 | 51 | 1186 | 32500 | 1 | 45.666359 | 12.241890 | 8800 |
| 2 | 3 | 3 | 74 | 4658 | 142228 | 1 | 45.503300 | 11.417840 | 4200 |
| 3 | 4 | 1 | 51 | 2739 | 160000 | 1 | 40.633171 | 17.634609 | 6000 |
| 4 | 5 | 2 | 73 | 3074 | 106880 | 1 | 41.903221 | 12.495650 | 5700 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1533 | 1534 | 3 | 51 | 3712 | 115280 | 1 | 45.069679 | 7.704920 | 5200 |
| 1534 | 1535 | 1 | 74 | 3835 | 112000 | 1 | 45.845692 | 8.666870 | 4600 |
| 1535 | 1536 | 2 | 51 | 2223 | 60457 | 1 | 45.481541 | 9.413480 | 7500 |
| 1536 | 1537 | 1 | 51 | 2557 | 80750 | 1 | 45.000702 | 7.682270 | 5990 |
| 1537 | 1538 | 2 | 51 | 1766 | 54276 | 1 | 40.323410 | 17.568270 | 7900 |

1538 rows × 9 columns

```
In [50]: #piecharts
'''we are using for percentage of share for any company'''
```

```
Out[50]: 'we are using for percentage of share for any company'
```

```
In [51]: data.shape
```

```
Out[51]: (1538, 9)
```

```
In [52]: #we can remove the columb  
data4=data3.drop(['model_name'],axis=1)  
data4
```

Out[52]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | lon | price |
|------|------|--------------|-------------|--------|-----------------|-----------|-----------|-------|
| 0 | 1 | 51 | 882 | 25000 | 1 | 44.907242 | 8.611560 | 8900 |
| 1 | 2 | 51 | 1186 | 32500 | 1 | 45.666359 | 12.241890 | 8800 |
| 2 | 3 | 74 | 4658 | 142228 | 1 | 45.503300 | 11.417840 | 4200 |
| 3 | 4 | 51 | 2739 | 160000 | 1 | 40.633171 | 17.634609 | 6000 |
| 4 | 5 | 73 | 3074 | 106880 | 1 | 41.903221 | 12.495650 | 5700 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1533 | 1534 | 51 | 3712 | 115280 | 1 | 45.069679 | 7.704920 | 5200 |
| 1534 | 1535 | 74 | 3835 | 112000 | 1 | 45.845692 | 8.666870 | 4600 |
| 1535 | 1536 | 51 | 2223 | 60457 | 1 | 45.481541 | 9.413480 | 7500 |
| 1536 | 1537 | 51 | 2557 | 80750 | 1 | 45.000702 | 7.682270 | 5990 |
| 1537 | 1538 | 51 | 1766 | 54276 | 1 | 40.323410 | 17.568270 | 7900 |

1538 rows × 8 columns

In []: