data visualization

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

Import dataset

In [2]:
```python
data=pd.read_csv(r"C:\Users\user\Downloads\fiat500_VehicleSelection_Dataset(1).
```

print data

In [3]:
```python
data
```

Out[3]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lo
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	conc
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

print first 10 rows using head

In [4]: `data.head(10)`

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
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

print last 10 rows using tail

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

Out[5]:

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

print describe of dataset

In [6]: `data.describe()`

Out[6]:

	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	

Number elements in dataset

In [7]: `data.size`

Out[7]: 17039

print shape of dataset

In [8]: `data.shape`

Out[8]: (1549, 11)

print empty or not

In [9]: `data.isna()`

Out[9]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price	Unn
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

In [38]: `data.isnull().sum()`

Out[38]:
```
ID                 11
model              11
engine_power       11
age_in_days        11
km                 11
previous_owners    11
lat                11
lon                 0
price               0
Unnamed: 9       1549
Unnamed: 10      1548
dtype: int64
```

In [39]:
```python
data1 = data.fillna(value=10)
data1
```

Out[39]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lo
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	11.417
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	10.0	10	10.0	10.0	10.0	10.0	10.000000	leng
1545	10.0	10	10.0	10.0	10.0	10.0	10.000000	conc
1546	10.0	10	10.0	10.0	10.0	10.0	10.000000	Null valu
1547	10.0	10	10.0	10.0	10.0	10.0	10.000000	fi
1548	10.0	10	10.0	10.0	10.0	10.0	10.000000	sear

1549 rows × 11 columns

In [40]:
```python
data1[10:20]
```

Out[40]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon
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 [13]: `data[data["ID"]<=20]`

Out[13]:

	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 [41]: `data1["ID"]`

Out[41]:
```
0       1.0
1       2.0
2       3.0
3       4.0
4       5.0
        ...
1544    10.0
1545    10.0
1546    10.0
1547    10.0
1548    10.0
Name: ID, Length: 1549, dtype: float64
```

In [42]:
```python
data1["km"]
```

Out[42]:
```
0          25000.0
1          32500.0
2         142228.0
3         160000.0
4         106880.0
            ...
1544          10.0
1545          10.0
1546          10.0
1547          10.0
1548          10.0
Name: km, Length: 1549, dtype: float64
```

In [43]:
```python
data1 = data[["engine_power","km"]]
data1
```
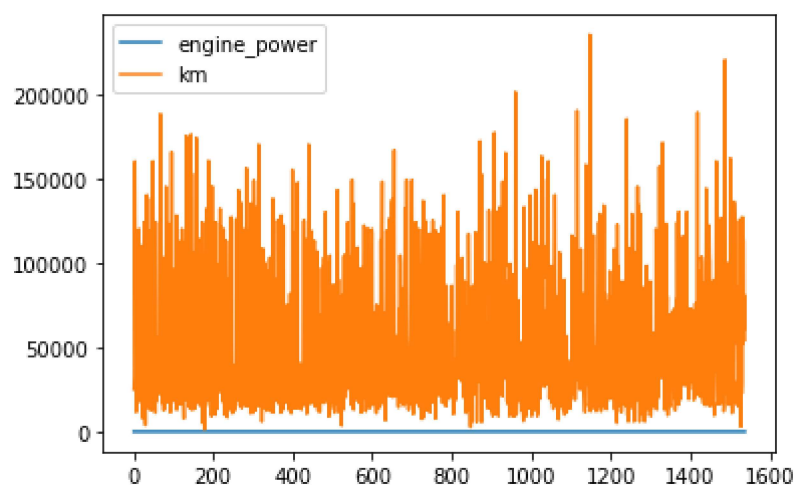
Out[43]:

	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
...	...	...
1544	NaN	NaN
1545	NaN	NaN
1546	NaN	NaN
1547	NaN	NaN
1548	NaN	NaN

1549 rows × 2 columns

In [44]: `data1.plot.line()`

Out[44]: <AxesSubplot:>

In [45]: `data1.plot.bar()`

Out[45]: <AxesSubplot:>
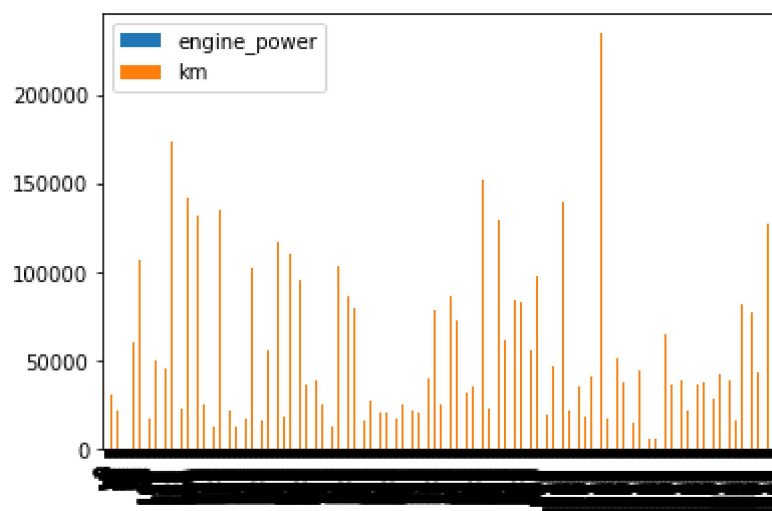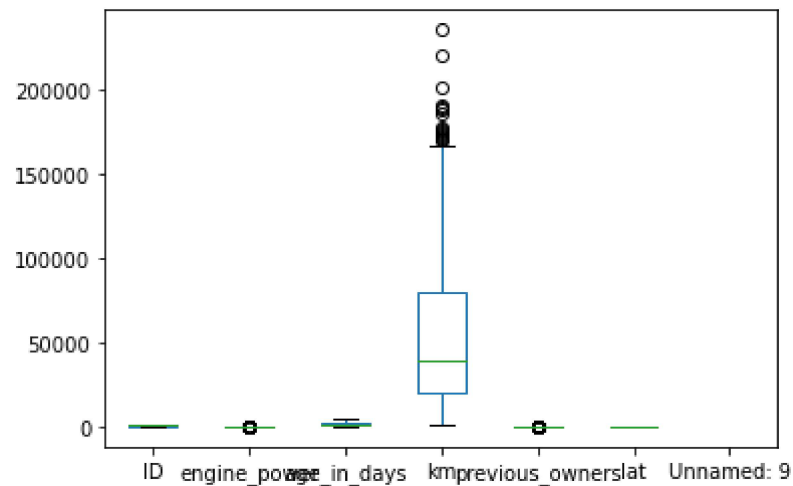
In [46]: `data1.plot.hist()`

Out[46]: `<AxesSubplot:ylabel='Frequency'>`
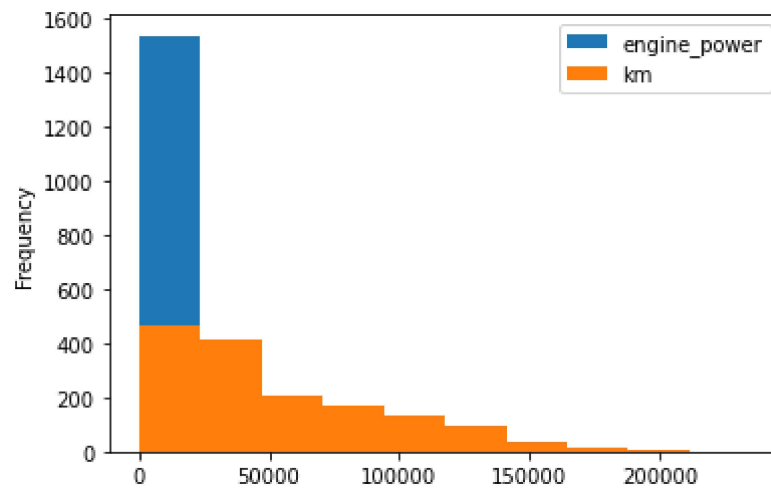
In [47]: `data.plot.box('km')`
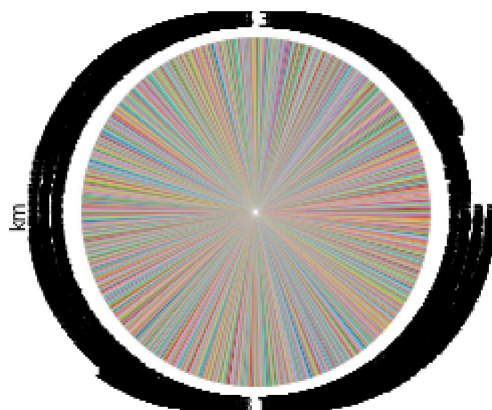
Out[47]: `<AxesSubplot:>`

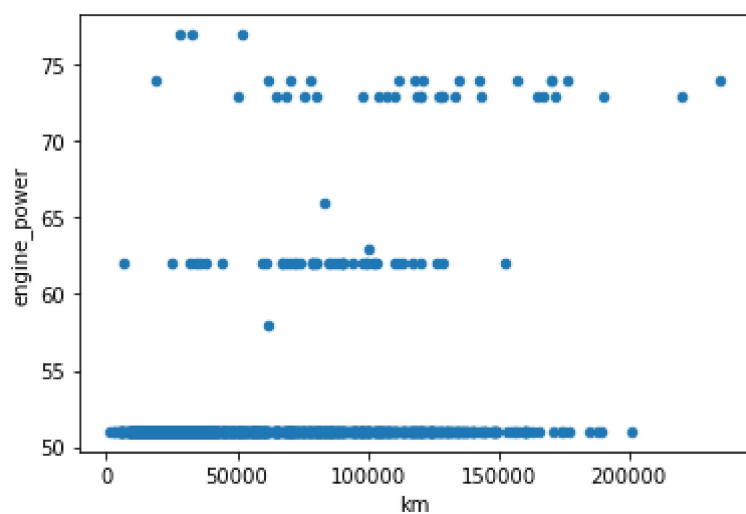In [48]: `data2 = data1["km"]`

In [49]: `data2.plot.pie()`

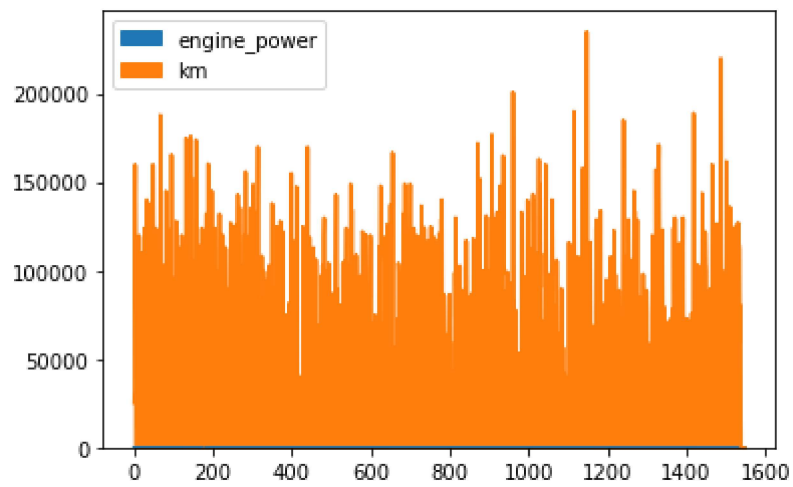Out[49]: `<AxesSubplot:ylabel='km'>`

In [50]: `data1.plot.scatter('km','engine_power')`

Out[50]: `<AxesSubplot:xlabel='km', ylabel='engine_power'>`

In [51]: `data1.plot.area()`

Out[51]: `<AxesSubplot:>`

In [ ]: