

# Data set 2

In [1]:

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

In [2]:

```
a=pd.read_csv(r"C:\Users\user\Downloads\Vehicle.csv")
```

To print top rows:

In [3]:

```
a.head()
```

Out[3]:

	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	8.61155986	11.4178
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.2418896	11.4178
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.4178	11.4178
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.6346092	11.4178
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.4956502	11.4178

To print Last rows:

In [4]:

```
a.tail()
```

Out[4]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
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

Statistical Summary:

In [5]:

```
a.describe()
```

Out[5]:

	ID	engine_power	age_in_days	km	previous_owners	la
count	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000	1538.000000
mean	769.500000	51.904421	1650.980494	53396.011704	1.123537	43.54136
std	444.126671	3.988023	1289.522278	40046.830723	0.416423	2.13351
min	1.000000	51.000000	366.000000	1232.000000	1.000000	36.85583
25%	385.250000	51.000000	670.000000	20006.250000	1.000000	41.80299
50%	769.500000	51.000000	1035.000000	39031.000000	1.000000	44.39409
75%	1153.750000	51.000000	2616.000000	79667.750000	1.000000	45.46796
max	1538.000000	77.000000	4658.000000	235000.000000	4.000000	46.79561

To print no of rows and columns

In [6]:

```
a.shape
```

Out[6]:

(1549, 11)

To print no of elements

In [7]:

```
a.size
```

Out[7]:

17039

Missing no of values

In [8]:

```
pd.isna(a)
```

Out[8]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
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 [9]:

```
a.dropna(axis=1,how='any')
```

Out[9]:

	lon	price
0	8.611559868	8900
1	12.24188995	8800
2	11.41784	4200
3	17.63460922	6000
4	12.49565029	5700
...	...	...
1544	length	5
1545	concat	lonprice
1546	Null values	NO
1547	find	1
1548	search	1

1549 rows × 2 columns

In [15]:

```
conda install matplotlib
```

```
Collecting package metadata (current_repodata.json): ...working... done  
Solving environment: ...working... done
```

```
# All requested packages already installed.
```

Note: you may need to restart the kernel to use updated packages.

```
==> WARNING: A newer version of conda exists. <==  
current version: 4.10.1  
latest version: 23.5.2
```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

In [16]:

```
import matplotlib.pyplot as pp
```

In [10]:

```
b=a[['model', 'km']]  
b
```

Out[10]:

	model	km
0	lounge	25000.0
1	pop	32500.0
2	sport	142228.0
3	lounge	160000.0
4	pop	106880.0
...	...	...
1544	NaN	NaN
1545	NaN	NaN
1546	NaN	NaN
1547	NaN	NaN
1548	NaN	NaN

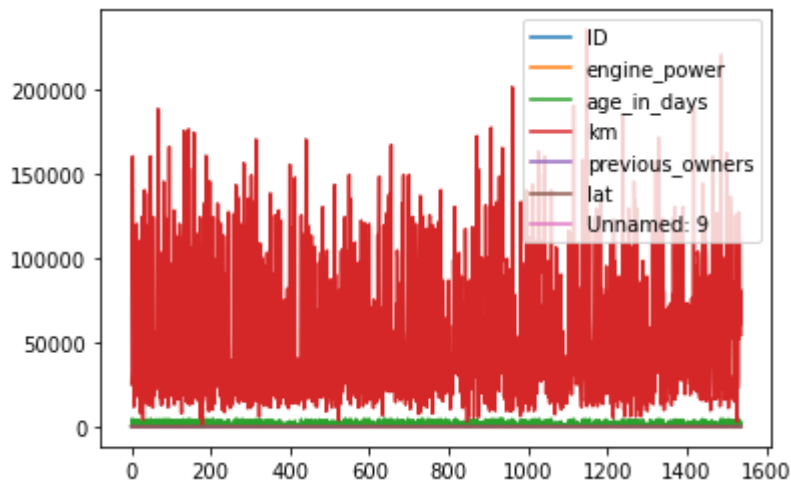
1549 rows × 2 columns

In [19]:

```
a.plot.line()
```

Out[19]:

&lt;AxesSubplot:&gt;

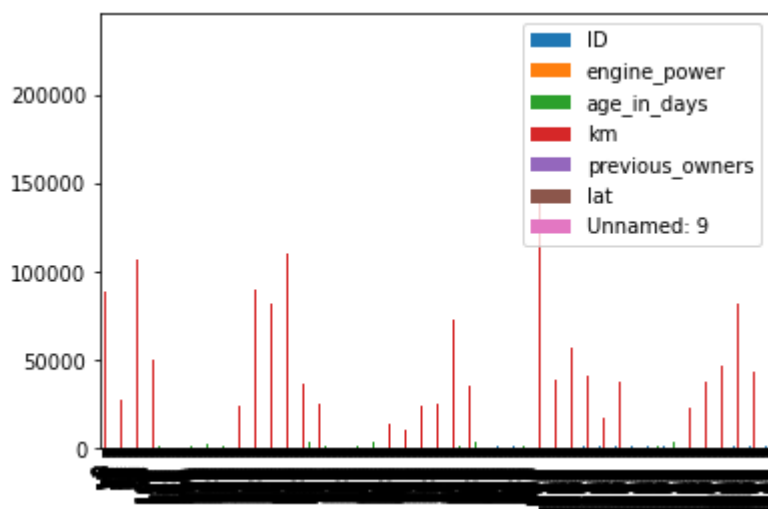


In [18]:

```
a.plot.bar()
```

Out[18]:

&lt;AxesSubplot:&gt;

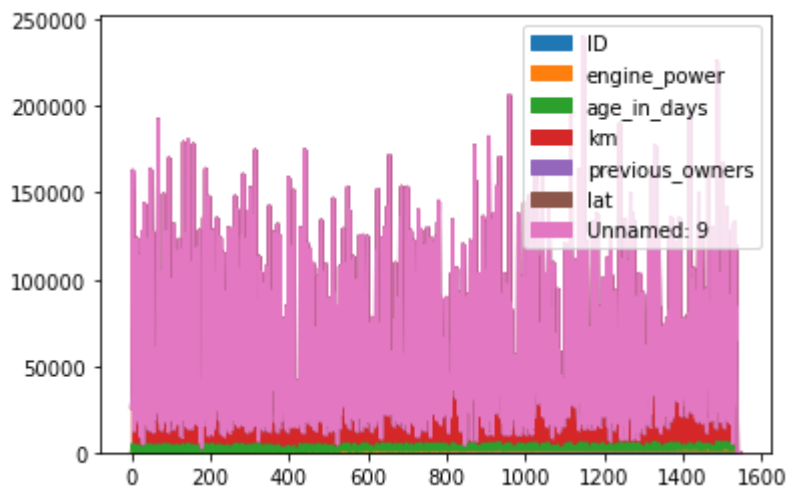


In [20]:

```
a.plot.area()
```

Out[20]:

<AxesSubplot:>

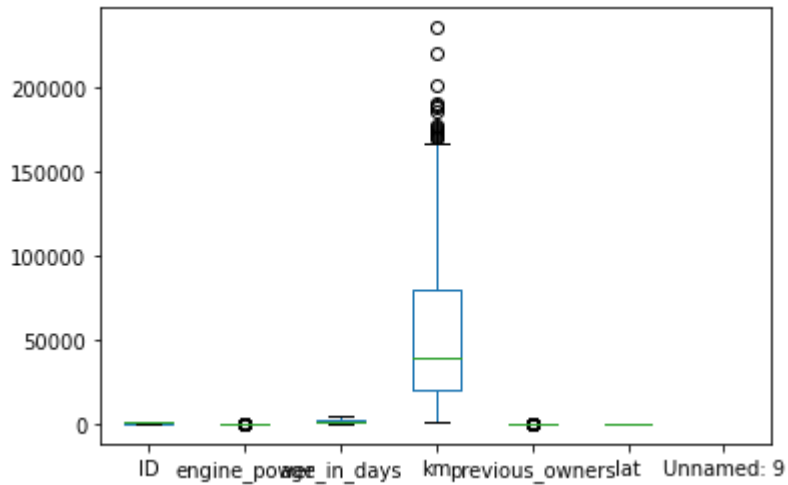


In [21]:

```
a.plot.box()
```

Out[21]:

<AxesSubplot:>

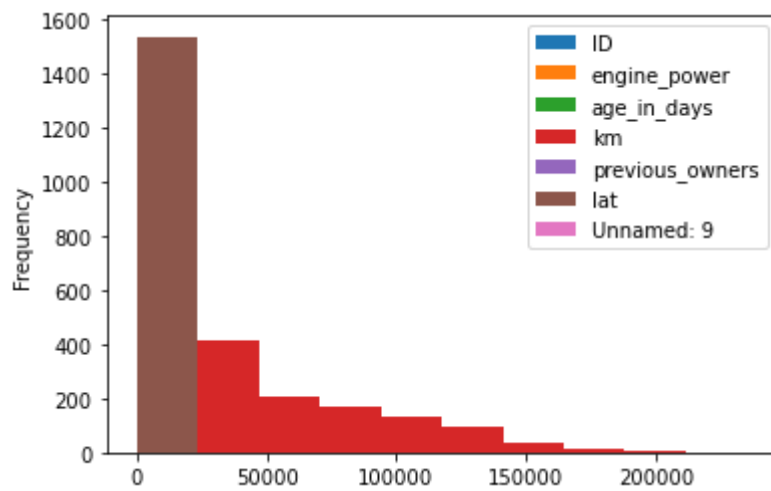


In [23]:

```
a.plot.hist()
```

Out[23]:

&lt;AxesSubplot:ylabel='Frequency'&gt;

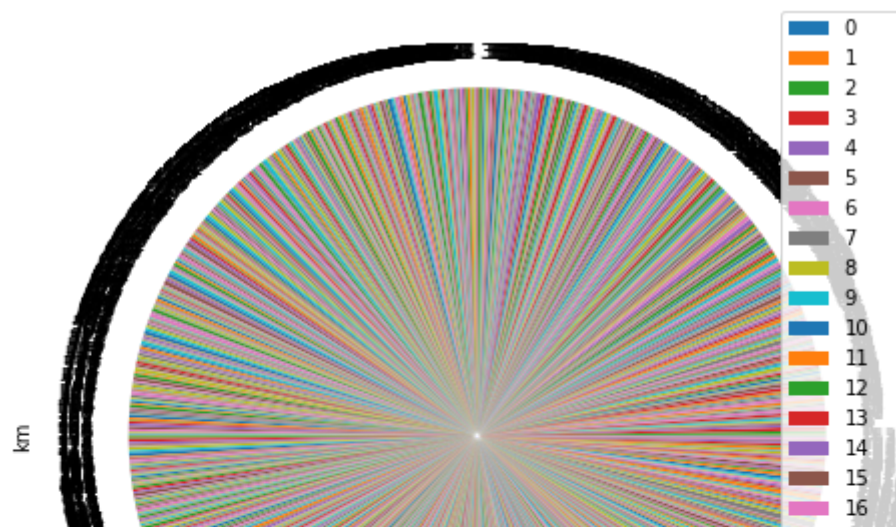


In [30]:

```
a.plot.pie(y='km',figsize=(8,8))
```

Out[30]:

&lt;AxesSubplot:ylabel='km'&gt;

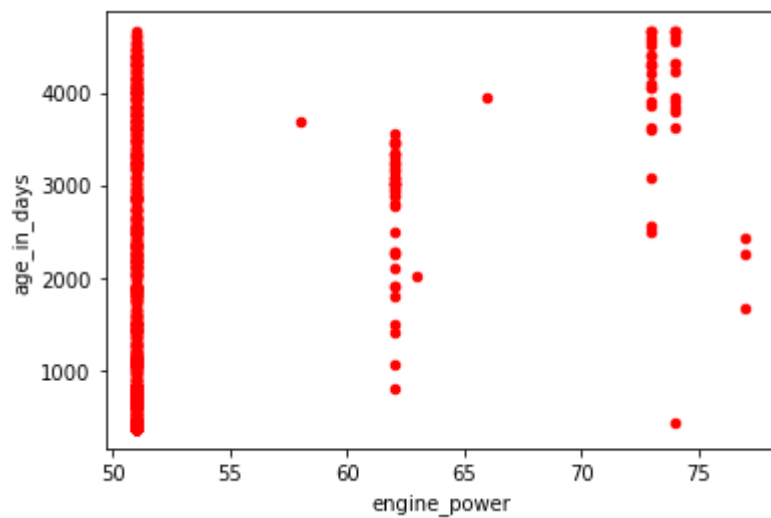


In [28]:

```
a.plot.scatter(x='engine_power',y='age_in_days',color='r')
```

Out[28]:

<AxesSubplot:xlabel='engine\_power', ylabel='age\_in\_days'>



In [ ]: