

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

```
In [2]: data=pd.read_csv("/home/placement/Desktop/venkatesh/fiat500.csv")
```

```
In [3]: data.describe()
```

Out[3]:

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

```
Out[4]:
```

	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
5	6	pop	74	3623	70225	1	45.000702	7.682270	7900
6	7	lounge	51	731	11600	1	44.907242	8.611560	10750
7	8	lounge	51	1521	49076	1	41.903221	12.495650	9190
8	9	sport	73	4049	76000	1	45.548000	11.549470	5600
9	10	sport	51	3653	89000	1	45.438301	10.991700	6000

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

```
Out[5]:
```

	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 [6]: data.tail()
```

```
Out[6]:
```

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
<b>1533</b>	1534	sport	51	3712	115280	1	45.069679	7.70492	5200
<b>1534</b>	1535	lounge	74	3835	112000	1	45.845692	8.66687	4600
<b>1535</b>	1536	pop	51	2223	60457	1	45.481541	9.41348	7500
<b>1536</b>	1537	lounge	51	2557	80750	1	45.000702	7.68227	5990
<b>1537</b>	1538	pop	51	1766	54276	1	40.323410	17.56827	7900

```
In [7]: data['previous_owners'].unique()
```

```
Out[7]: array([1, 2, 3, 4])
```

```
In [8]: data['engine_power'].unique()
```

```
Out[8]: array([51, 74, 73, 62, 63, 66, 77, 58])
```

```
In [9]: list(data.columns)
```

```
Out[9]: ['ID',  
         'model',  
         'engine_power',  
         'age_in_days',  
         'km',  
         'previous_owners',  
         'lat',  
         'lon',  
         'price']
```

```
In [10]: data.groupby(['previous_owners']).count()
```

```
Out[10]:
```

	ID	model	engine_power	age_in_days	km	lat	lon	price
previous_owners								
1	1389	1389	1389	1389	1389	1389	1389	1389
2	117	117	117	117	117	117	117	117
3	23	23	23	23	23	23	23	23
4	9	9	9	9	9	9	9	9

```
In [11]: data.groupby(['model']).count()
```

```
Out[11]:
```

	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 [12]: data1=data.drop(['lat','lon'],axis=1)
data1
```

Out[12]:

	ID	model	engine_power	age_in_days	km	previous_owners	price
0	1	lounge	51	882	25000	1	8900
1	2	pop	51	1186	32500	1	8800
2	3	sport	74	4658	142228	1	4200
3	4	lounge	51	2739	160000	1	6000
4	5	pop	73	3074	106880	1	5700
...	...	...	...	...	...	...	...
1533	1534	sport	51	3712	115280	1	5200
1534	1535	lounge	74	3835	112000	1	4600
1535	1536	pop	51	2223	60457	1	7500
1536	1537	lounge	51	2557	80750	1	5990
1537	1538	pop	51	1766	54276	1	7900

1538 rows × 7 columns

```
In [13]: data.head(5)
```

Out[13]:

	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 [14]: data1.head(5)
```

```
Out[14]:
```

	ID	model	engine_power	age_in_days	km	previous_owners	price
0	1	lounge	51	882	25000	1	8900
1	2	pop	51	1186	32500	1	8800
2	3	sport	74	4658	142228	1	4200
3	4	lounge	51	2739	160000	1	6000
4	5	pop	73	3074	106880	1	5700

```
In [15]: data.isnull()
```

```
Out[15]:
```

	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
...	...	...	...	...	...	...	...	...	...
1533	False	False	False	False	False	False	False	False	False
1534	False	False	False	False	False	False	False	False	False
1535	False	False	False	False	False	False	False	False	False
1536	False	False	False	False	False	False	False	False	False
1537	False	False	False	False	False	False	False	False	False

1538 rows × 9 columns

```
In [16]: data['price'].sum()
```

```
Out[16]: 13189894
```

```
In [17]: data2=data.loc[(data.model=='pop')]  
data2
```

```
Out[17]:
```

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
1	2	pop	51	1186	32500	1	45.666359	12.241890	8800
4	5	pop	73	3074	106880	1	41.903221	12.495650	5700
5	6	pop	74	3623	70225	1	45.000702	7.682270	7900
10	11	pop	51	790	43286	1	40.871429	14.438960	8950
13	14	pop	51	3835	120000	1	40.531590	17.436159	4800
...	...	...	...	...	...	...	...	...	...
1524	1525	pop	51	2192	53300	1	40.609531	14.980930	7900
1527	1528	pop	51	517	3000	1	40.748241	14.528350	9999
1532	1533	pop	51	1917	52008	1	45.548000	11.549470	9900
1535	1536	pop	51	2223	60457	1	45.481541	9.413480	7500
1537	1538	pop	51	1766	54276	1	40.323410	17.568270	7900

358 rows × 9 columns

```
In [18]: data3=data.loc[(data.km<50000)]
data3
```

Out[18]:

	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

900 rows × 9 columns



```
In [19]: data4=data.loc[(data.model=='sport')&(data.previous_owners==1)]
data4
```

Out[19]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
2	3	sport	74	4658	142228	1	45.503300	11.41784	4200
8	9	sport	73	4049	76000	1	45.548000	11.54947	5600
9	10	sport	51	3653	89000	1	45.438301	10.99170	6000
27	28	sport	51	3531	100000	1	40.976452	14.17228	4950
38	39	sport	51	790	43100	1	45.334080	11.37687	9950
...	...	...	...	...	...	...	...	...	...
1498	1499	sport	51	1400	60000	1	45.802021	9.18779	10800
1501	1502	sport	51	3774	161900	1	44.906120	7.67723	4500
1504	1505	sport	51	4322	95000	1	40.928501	14.20367	4500
1531	1532	sport	73	4505	127000	1	45.528511	9.59323	4750
1533	1534	sport	51	3712	115280	1	45.069679	7.70492	5200

69 rows × 9 columns

```
In [20]: data5=data.loc[(data.model=='sport')|(data.model=='pop')]
data5
```

Out[20]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
1	2	pop	51	1186	32500	1	45.666359	12.24189	8800
2	3	sport	74	4658	142228	1	45.503300	11.41784	4200
4	5	pop	73	3074	106880	1	41.903221	12.49565	5700
5	6	pop	74	3623	70225	1	45.000702	7.68227	7900
8	9	sport	73	4049	76000	1	45.548000	11.54947	5600
...	...	...	...	...	...	...	...	...	...
1531	1532	sport	73	4505	127000	1	45.528511	9.59323	4750
1532	1533	pop	51	1917	52008	1	45.548000	11.54947	9900
1533	1534	sport	51	3712	115280	1	45.069679	7.70492	5200
1535	1536	pop	51	2223	60457	1	45.481541	9.41348	7500
1537	1538	pop	51	1766	54276	1	40.323410	17.56827	7900

444 rows × 9 columns

```
In [21]: data.head(60)
```

```
Out[21]:
```

	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
5	6	pop	74	3623	70225	1	45.000702	7.682270	7900
6	7	lounge	51	731	11600	1	44.907242	8.611560	10750
7	8	lounge	51	1521	49076	1	41.903221	12.495650	9190
8	9	sport	73	4049	76000	1	45.548000	11.549470	5600
9	10	sport	51	3653	89000	1	45.438301	10.991700	6000
10	11	pop	51	790	43286	1	40.871429	14.438960	8950
11	12	lounge	51	366	17500	1	45.069679	7.704920	10990
12	13	lounge	51	456	18450	1	45.426571	11.788130	9700
13	14	pop	51	3835	120000	1	40.531590	17.436159	4800
14	15	lounge	51	1035	40500	1	40.911362	14.211200	9300
15	16	lounge	51	1096	28200	1	45.697208	9.845970	9500
16	17	lounge	73	4200	110000	1	41.082352	14.254250	5250
17	18	pop	51	2223	96848	1	43.782372	11.254990	7990
18	19	lounge	51	2861	31000	1	45.069679	7.704920	7300
19	20	lounge	51	425	20030	1	45.354389	11.869260	10500
20	21	lounge	51	397	19037	1	45.707249	11.477600	10500
21	22	lounge	51	1886	110000	1	40.835812	14.504410	6990
22	23	lounge	51	1035	8000	1	44.506088	12.044170	10600
23	24	lounge	51	790	27595	1	45.688259	8.731450	10200

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
24	25	lounge	51	1583	14900	1	45.069679	7.704920	9990
25	26	lounge	51	366	9218	1	45.438110	12.318150	10800
26	27	pop	51	3592	124000	1	40.966179	17.116480	6800
27	28	sport	51	3531	100000	1	40.976452	14.172280	4950
28	29	lounge	51	762	28900	1	45.131672	8.449170	10640
29	30	lounge	51	670	4000	1	41.349751	13.353320	9500
30	31	lounge	62	2769	59216	1	43.782372	11.254990	6990
31	32	lounge	51	4169	99477	2	40.550564	14.225625	5900
32	33	lounge	51	821	21730	2	41.903221	12.495650	10500
33	34	sport	51	3927	140000	2	40.755932	14.690190	5200
34	35	lounge	51	640	32033	2	44.283878	11.888140	9790
35	36	pop	51	3653	138116	2	40.633171	17.634609	5000
36	37	pop	51	852	17000	1	45.505161	8.939100	8990
37	38	pop	51	3013	58527	1	45.688259	8.731450	7200
38	39	sport	51	790	43100	1	45.334080	11.376870	9950
39	40	pop	51	1858	13373	1	41.903221	12.495650	9000
40	41	sport	51	4139	119000	1	45.349319	7.742600	4890
41	42	pop	51	609	28500	1	45.746021	9.049970	10900
42	43	pop	51	1096	83000	1	41.959721	12.798056	7900
43	44	lounge	73	4049	98000	1	38.218128	15.240240	5999
44	45	lounge	51	456	12693	1	45.393600	10.482240	10900
45	46	lounge	51	762	14586	1	42.341969	12.358490	10500
46	47	lounge	51	821	27640	1	41.903221	12.495650	10400
47	48	lounge	51	2039	49000	1	41.903221	12.495650	7500
48	49	sport	51	3684	160000	1	45.405472	10.278290	4900
49	50	sport	51	4596	107000	1	40.845901	14.369270	4300

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
<b>50</b>	51	lounge	51	2192	66000	1	40.840141	14.252260	6999
<b>51</b>	52	sport	51	790	35000	1	45.334080	11.376870	10200
<b>52</b>	53	pop	51	1186	24651	1	41.903221	12.495650	9300
<b>53</b>	54	sport	62	3288	71900	1	41.062401	14.273880	5990
<b>54</b>	55	lounge	51	701	11200	1	46.176498	9.903670	10600
<b>55</b>	56	lounge	51	3288	124050	1	43.869980	10.339830	5500
<b>56</b>	57	pop	51	1066	69395	1	45.231411	9.592890	7450
<b>57</b>	58	lounge	51	456	17961	1	45.393600	10.482240	10900
<b>58</b>	59	lounge	51	670	18669	1	45.000702	7.682270	10990
<b>59</b>	60	pop	51	1431	22700	1	41.903221	12.495650	8900

```
In [23]: datacor=data.drop(['model'],axis=1)
datacor
```

Out[23]:

	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

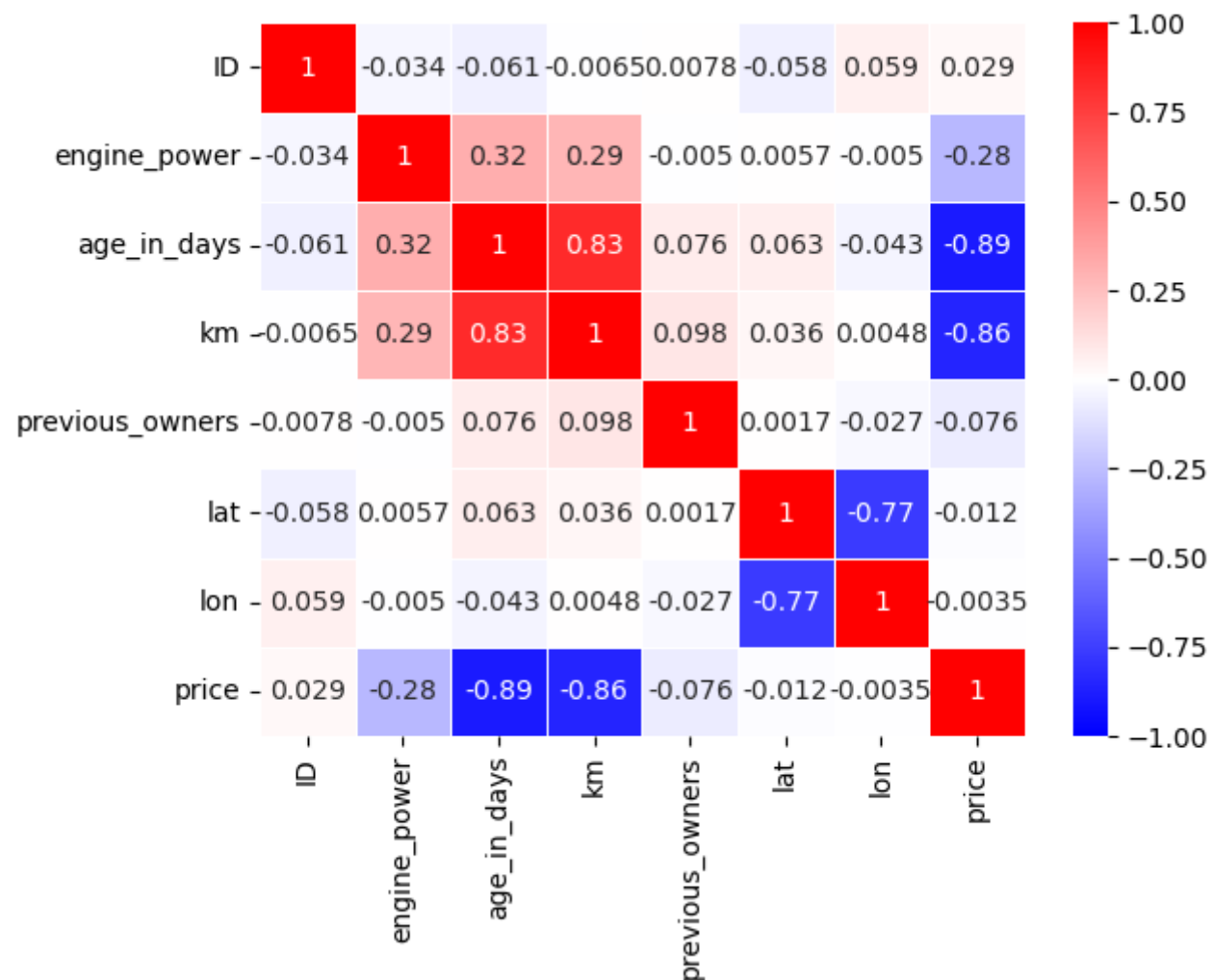
```
In [25]: cor=datacor.corr()  
cor
```

Out[25]:

	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
ID	1.000000	-0.034059	-0.060753	-0.006537	0.007803	-0.058207	0.058941	0.028516
engine_power	-0.034059	1.000000	0.319190	0.285495	-0.005030	0.005721	-0.005032	-0.277235
age_in_days	-0.060753	0.319190	1.000000	0.833890	0.075775	0.062982	-0.042667	-0.893328
km	-0.006537	0.285495	0.833890	1.000000	0.097539	0.035519	0.004839	-0.859373
previous_owners	0.007803	-0.005030	0.075775	0.097539	1.000000	0.001697	-0.026836	-0.076274
lat	-0.058207	0.005721	0.062982	0.035519	0.001697	1.000000	-0.766646	-0.011733
lon	0.058941	-0.005032	-0.042667	0.004839	-0.026836	-0.766646	1.000000	-0.003541
price	0.028516	-0.277235	-0.893328	-0.859373	-0.076274	-0.011733	-0.003541	1.000000

```
In [26]: import seaborn as sns
sns.heatmap(cor, vmax=1, vmin=-1, annot=True, linewidths=.5, cmap='bwr')
```

Out[26]: <Axes: >





In [ ]: