In [31]:
```
import pandas as pd
import seaborn as sns
```

In [6]:
```
data=pd.read_csv("/home/placement/Desktop/yamuna/fiat500.csv")
```

In [7]:
```
data.describe()
```

Out[7]:

	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 [8]:
```
cor_mat=data.corr()
cor_mat
```

```
/tmp/ipykernel_6547/2947089049.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only valid columns or specify the value o
f numeric_only to silence this warning.
  cor_mat=data.corr()
```

Out[8]:

	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 [9]:
```
data.head(10)
```

Out[9]:

	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 [10]:
```
data1=data.drop(['model'],axis=1)
```

Type *Markdown* and LaTeX: α^2

In [11]:
```
cor=data1.corr()
cor
```

Out[11]:

	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 [12]:
```
data['previous_owners'].unique()
```

Out[12]:
```
array([1, 2, 3, 4])
```

In [13]:
```
data['engine_power'].unique()
```

Out[13]:
```
array([51, 74, 73, 62, 63, 66, 77, 58])
```

In [14]:
```
list(data.columns)
```

Out[14]:
```
['ID',
 'model',
 'engine_power',
 'age_in_days',
 'km',
 'previous_owners',
 'lat',
 'lon',
 'price']
```

In [15]:
```
data.groupby(['previous_owners']).count()
```

Out[15]:

previous_owners	ID	model	engine_power	age_in_days	km	lat	lon	price
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 [16]:
```
data.groupby(['model']).count()
```

Out[16]:

model	ID	engine_power	age_in_days	km	previous_owners	lat	lon	price
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 [17]:
```
data1=data.drop(['lat','ID'],axis=1)
```

In [18]:
```
data.head(5)
```

Out[18]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	price
0	1	lounge	51	882	25000	1	44.907242	8.611560	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 [19]:
```
data1.head(5)
```

Out[19]:

	model	engine_power	age_in_days	km	previous_owners	lon	price
0	lounge	51	882	25000	1	8.611560	8900
1	pop	51	1186	32500	1	12.241890	8800
2	sport	74	4658	142228	1	11.417840	4200
3	lounge	51	2739	160000	1	17.634609	6000
4	pop	73	3074	106880	1	12.495650	5700

In [20]:
```
data1=data.drop(['lon'],axis=1)
```

In [21]:
```
data1.head(5)
```

Out[21]:

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

In [22]:
```
data['price'].sum()
```

Out[22]: 13189894

In [23]:
```
data2=data.loc[(data.model=='lounge')]
data2
```

Out[23]:

	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
3	4	lounge	51	2739	160000	1	40.633171	17.634609	6000
6	7	lounge	51	731	11600	1	44.907242	8.611560	10750
7	8	lounge	51	1521	49076	1	41.903221	12.495650	9190
11	12	lounge	51	366	17500	1	45.069679	7.704920	10990
...	...	...	...	...	...	...	...	...	...
1528	1529	lounge	51	2861	126000	1	43.841980	10.515310	5500
1529	1530	lounge	51	731	22551	1	38.122070	13.361120	9900
1530	1531	lounge	51	670	29000	1	45.764648	8.994500	10800
1534	1535	lounge	74	3835	112000	1	45.845692	8.666870	4600
1536	1537	lounge	51	2557	80750	1	45.000702	7.682270	5990

1094 rows × 9 columns

In [24]:
```
data3=data.loc[(data.km<=25000)]
data3
```

Out[24]:

	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
6	7	lounge	51	731	11600	1	44.907242	8.61156	10750
11	12	lounge	51	366	17500	1	45.069679	7.70492	10990
12	13	lounge	51	456	18450	1	45.426571	11.78813	9700
19	20	lounge	51	425	20030	1	45.354389	11.86926	10500
...	...	...	...	...	...	...	...	...	...
1520	1521	lounge	51	1035	15000	1	41.903221	12.49565	10990
1522	1523	lounge	51	366	14618	1	45.707249	11.47760	10500
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

492 rows × 9 columns

In [25]:
```
data1=data.drop
```

In [ ]:
```

```

In [26]:
```
data2=data.loc[(data.model=='pop')&(data.previous_owners==1)]
data2
```

Out[26]:

	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

327 rows × 9 columns

In [27]:
```
data2=data.loc[(data.model=='lounge')|(data.model=='pop')]
data2
```

Out[27]:

	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
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
...	...	...	...	...	...	...	...	...	...
1532	1533	pop	51	1917	52008	1	45.548000	11.549470	9900
1534	1535	lounge	74	3835	112000	1	45.845692	8.666870	4600
1535	1536	pop	51	2223	60457	1	45.481541	9.413480	7500
1536	1537	lounge	51	2557	80750	1	45.000702	7.682270	5990
1537	1538	pop	51	1766	54276	1	40.323410	17.568270	7900

1452 rows × 9 columns

In [28]:
```
data4=data.loc[(data.model=='sport')|(data.model=='pop')]
data4
```

Out[28]:

	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 [29]:
```
cor=data.corr()
cor
```

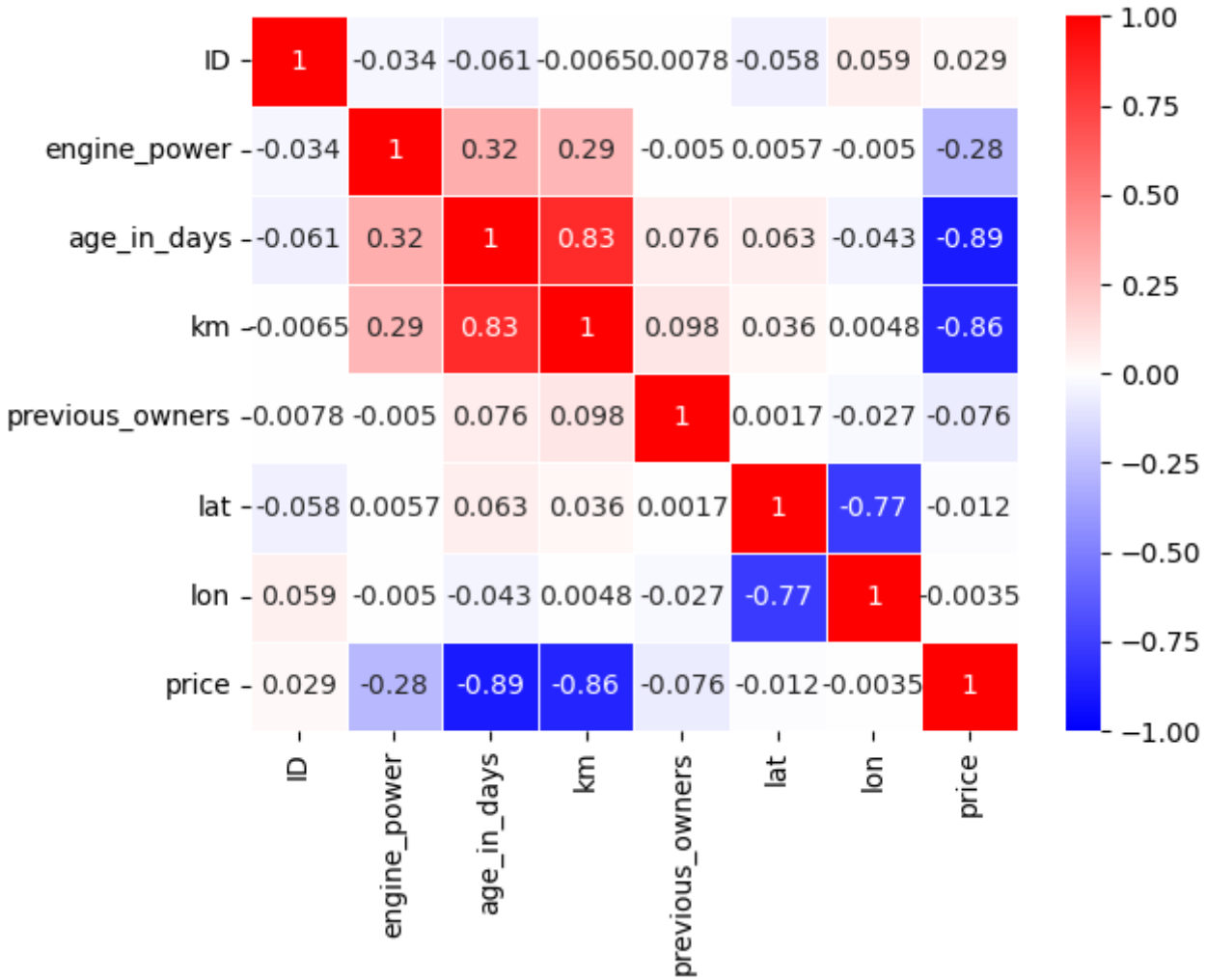
```
/tmp/ipykernel_6547/4173678507.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only valid columns or specify the value o
f numeric_only to silence this warning.
  cor=data.corr()
```

Out[29]:

	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 [32]:
```
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=.5,cmap='bwr')
```

Out[32]: <Axes: >



In [ ]: