In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\1_fiat500_VehicleSelection_Dataset.csv
df.fillna(0,inplace=True)
df
```

Out[2]:

	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	8.61155986
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.2418899
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	17.6346092
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.4956502
...	...	...	...	...	...	...	...	...
1544	0.0	0	0.0	0.0	0.0	0.0	0.000000	lengt
1545	0.0	0	0.0	0.0	0.0	0.0	0.000000	conca
1546	0.0	0	0.0	0.0	0.0	0.0	0.000000	Null value
1547	0.0	0	0.0	0.0	0.0	0.0	0.000000	fin
1548	0.0	0	0.0	0.0	0.0	0.0	0.000000	searc

1549 rows × 11 columns

In [3]:
```
df.head()
```

Out[3]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	lon	p
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	

In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1549 entries, 0 to 1548
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   ID               1549 non-null   float64
 1   model            1549 non-null   object 
 2   engine_power     1549 non-null   float64
 3   age_in_days      1549 non-null   float64
 4   km               1549 non-null   float64
 5   previous_owners  1549 non-null   float64
 6   lat              1549 non-null   float64
 7   lon              1549 non-null   object 
 8   price            1549 non-null   object 
 9   Unnamed: 9       1549 non-null   float64
 10  Unnamed: 10      1549 non-null   object 
dtypes: float64(7), object(4)
memory usage: 133.2+ KB
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1549 entries, 0 to 1548
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   ID               1549 non-null   float64
 1   model            1549 non-null   object 
 2   engine_power     1549 non-null   float64
 3   age_in_days      1549 non-null   float64
 4   km               1549 non-null   float64
 5   previous_owners  1549 non-null   float64
 6   lat              1549 non-null   float64
 7   lon              1549 non-null   object 
 8   price            1549 non-null   object 
 9   Unnamed: 9       1549 non-null   float64
 10  Unnamed: 10      1549 non-null   object 
dtypes: float64(7), object(4)
memory usage: 133.2+ KB
```

In [6]:
```
df=pd.read_csv("1_fiat500_VehicleSelection_Dataset.csv")
```

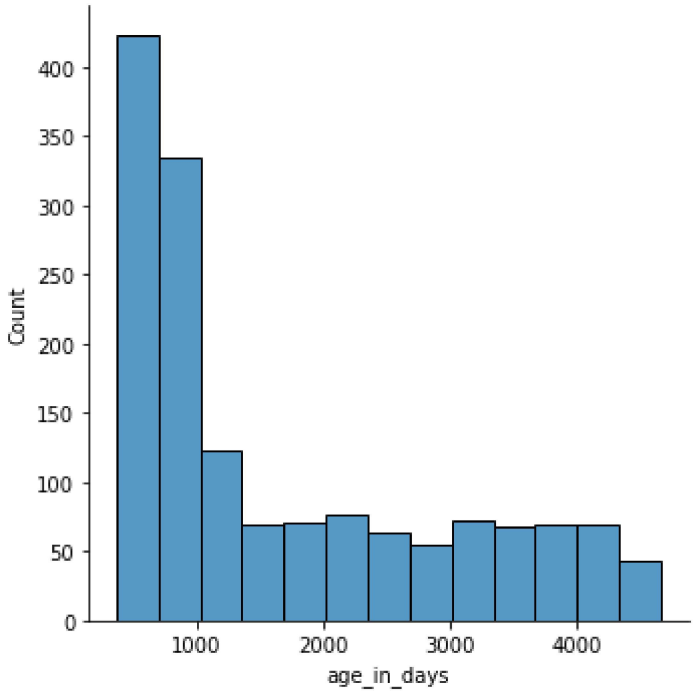
In [8]:
```
import seaborn as sns
```

In [9]:
```
sns.pairplot(df)
```

Out[9]: <seaborn.axisgrid.PairGrid at 0x16d214fa460>

In [10]:
```
sns.displot(df['age_in_days'])
```

Out[10]:
```
<seaborn.axisgrid.FacetGrid at 0x16d23d940a0>
```

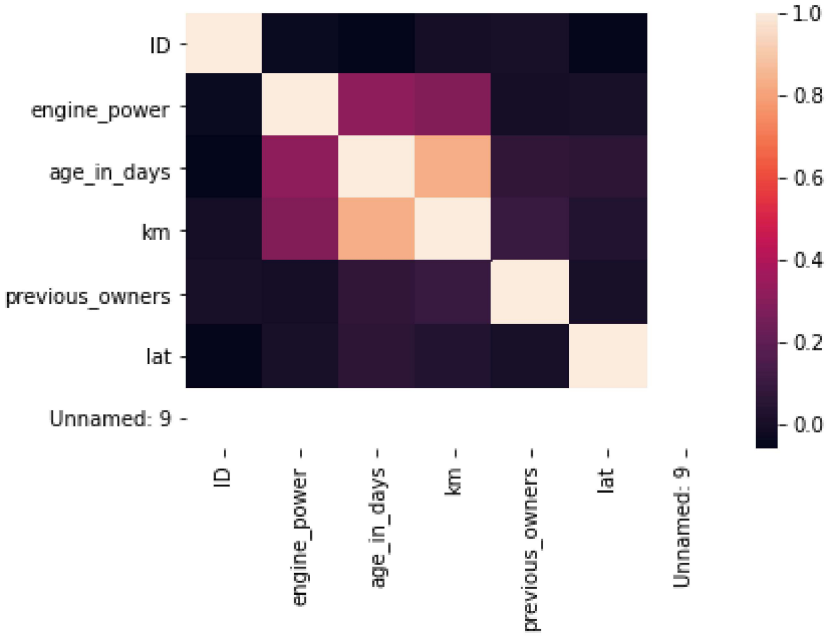


In [11]:
```
df1=df.drop(['Unnamed: 10'],axis=1)
df1
df1=df1.drop(df1.index[1537:])
df1.isna().sum()
```

Out[11]:
```
ID                   0
model                0
engine_power         0
age_in_days          0
km                   0
previous_owners      0
lat                  0
lon                  0
price                0
Unnamed: 9        1537
dtype: int64
```

In [12]:
```
sns.heatmap(df1.corr())
```

Out[12]: <AxesSubplot:>



In [13]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [14]:
```
df1.isna().sum()
```

Out[14]:
```
ID                   0
model                0
engine_power         0
age_in_days          0
km                   0
previous_owners      0
lat                  0
lon                  0
price                0
Unnamed: 9        1537
dtype: int64
```

In [15]:
```
y=df1['price']
x=df1.drop(['price','Unnamed: 9','model'],axis=1)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
print(x_train)
```

```
          ID  engine_power  age_in_days        km  previous_owners        lat
\
662    663.0          51.0       1492.0   23016.0              1.0  45.131672
252    253.0          51.0        790.0   27595.0              2.0  45.688259
1335  1336.0          51.0        366.0   12950.0              1.0  44.512428
151    152.0          51.0       2892.0   67000.0              1.0  45.674839
1477  1478.0          51.0       1917.0  126426.0              1.0  41.769051
...      ...           ...          ...       ...              ...        ...
456    457.0          51.0        456.0   14800.0              1.0  44.063129
730    731.0          51.0        762.0   12337.0              1.0  43.782372
1412  1413.0          51.0       1431.0   38000.0              1.0  45.215408
1361  1362.0          51.0       3227.0   73000.0              1.0  38.122070
1392  1393.0          51.0       1766.0   29400.0              2.0  41.741779

              lon
662   8.449170113
252   8.731450081
1335  11.12837982
151   9.687859535
1477  12.66281033
...           ...
456   12.44701004
730   11.25498962
1412  7.633669853
1361  13.36112022
1392   12.6441803

[1075 rows x 7 columns]
```

In [16]:
```
model=LinearRegression()
model.fit(x_train,y_train)
model.intercept_
```

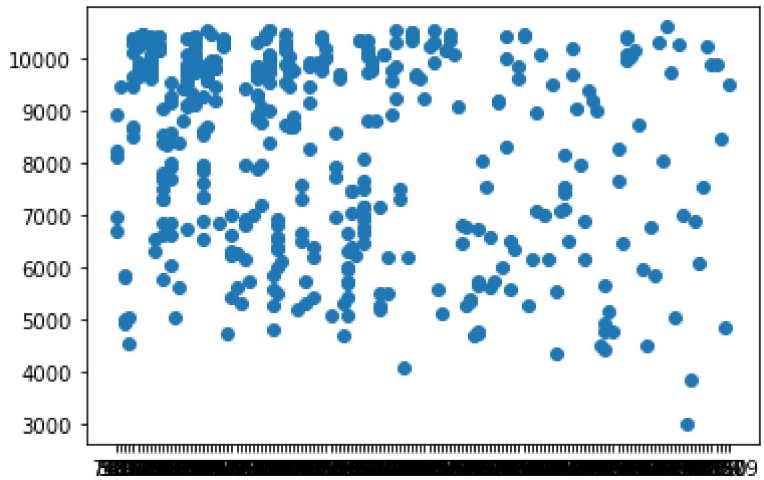
Out[16]:
```
9390.660847702526
```

In [17]:
```
model.coef_
```

Out[17]:
```
array([-5.11101287e-02,  6.39840119e+00, -9.22719101e-01, -1.70984147e-02,
        4.75965035e+00,  3.20330098e+01, -5.90726736e+00])
```

In [18]:
```
prediction=model.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[18]: <matplotlib.collections.PathCollection at 0x16d25c92e50>



In [19]:
```
model.score(x_test,y_test)
```

Out[19]: 0.8552727844290008

In [ ]: