In [2]:
```
1 import pandas as pd
2 import numpy as np
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 df=pd.read_csv(r"C:\Users\P. VIJAY KUMAR\Downloads\fiat500_VehicleSelectio
6 print(df)
```

```
        ID   model  engine_power  age_in_days      km  previous_owners
0        1  lounge            51          882   25000                1  \
1        2     pop            51         1186   32500                1
2        3   sport            74         4658  142228                1
3        4  lounge            51         2739  160000                1
4        5     pop            73         3074  106880                1
...    ...     ...           ...          ...     ...              ...
1533  1534   sport            51         3712  115280                1
1534  1535  lounge            74         3835  112000                1
1535  1536     pop            51         2223   60457                1
1536  1537  lounge            51         2557   80750                1
1537  1538     pop            51         1766   54276                1

            lat        lon  price
0     44.907242   8.611560   8900
1     45.666359  12.241890   8800
2     45.503300  11.417840   4200
3     40.633171  17.634609   6000
4     41.903221  12.495650   5700
...         ...        ...    ...
1533  45.069679   7.704920   5200
1534  45.845692   8.666870   4600
1535  45.481541   9.413480   7500
1536  45.000702   7.682270   5990
1537  40.323410  17.568270   7900

[1538 rows x 9 columns]
```

In [3]:
```
#display top 10 rows
print(df.head(10))
```

```
   ID   model  engine_power  age_in_days      km  previous_owners        lat
0   1  lounge            51          882   25000                1  44.907242
\
1   2     pop            51         1186   32500                1  45.666359
2   3   sport            74         4658  142228                1  45.503300
3   4  lounge            51         2739  160000                1  40.633171
4   5     pop            73         3074  106880                1  41.903221
5   6     pop            74         3623   70225                1  45.000702
6   7  lounge            51          731   11600                1  44.907242
7   8  lounge            51         1521   49076                1  41.903221
8   9   sport            73         4049   76000                1  45.548000
9  10   sport            51         3653   89000                1  45.438301

         lon  price
0   8.611560   8900
1  12.241890   8800
2  11.417840   4200
3  17.634609   6000
4  12.495650   5700
5   7.682270   7900
6   8.611560  10750
7  12.495650   9190
8  11.549470   5600
9  10.991700   6000
```

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

Out[4]:

	ID	engine_power	age_in_days	km	previous_owners	lat
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

In [5]:
```
#check for Null Values
print(df.isna().any())
```

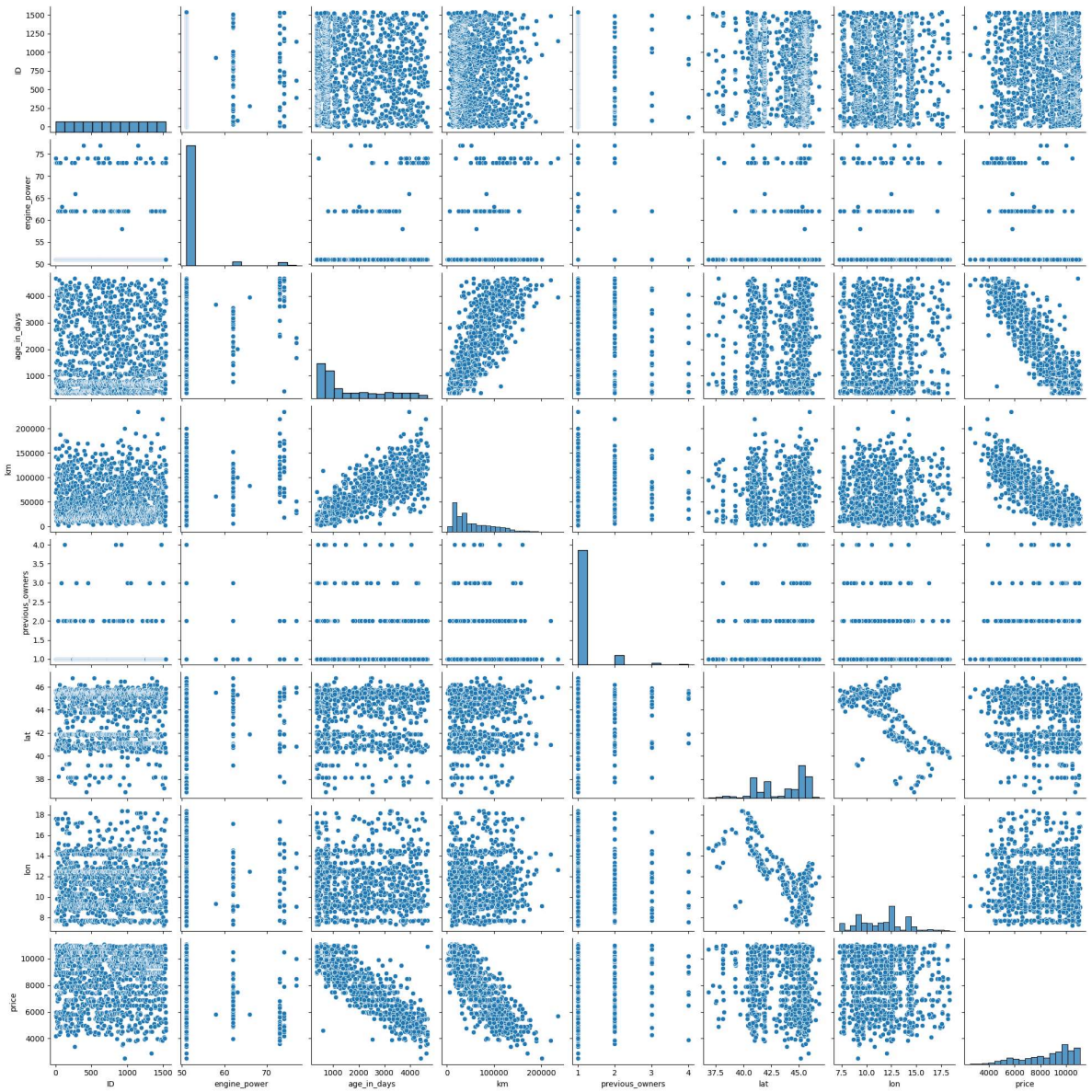
```
ID                False
model             False
engine_power      False
age_in_days       False
km                False
previous_owners   False
lat               False
lon               False
price             False
dtype: bool
```

In [6]:
```
df.columns
```

Out[6]:
```
Index(['ID', 'model', 'engine_power', 'age_in_days', 'km', 'previous_owners',
       'lat', 'lon', 'price'],
      dtype='object')
```

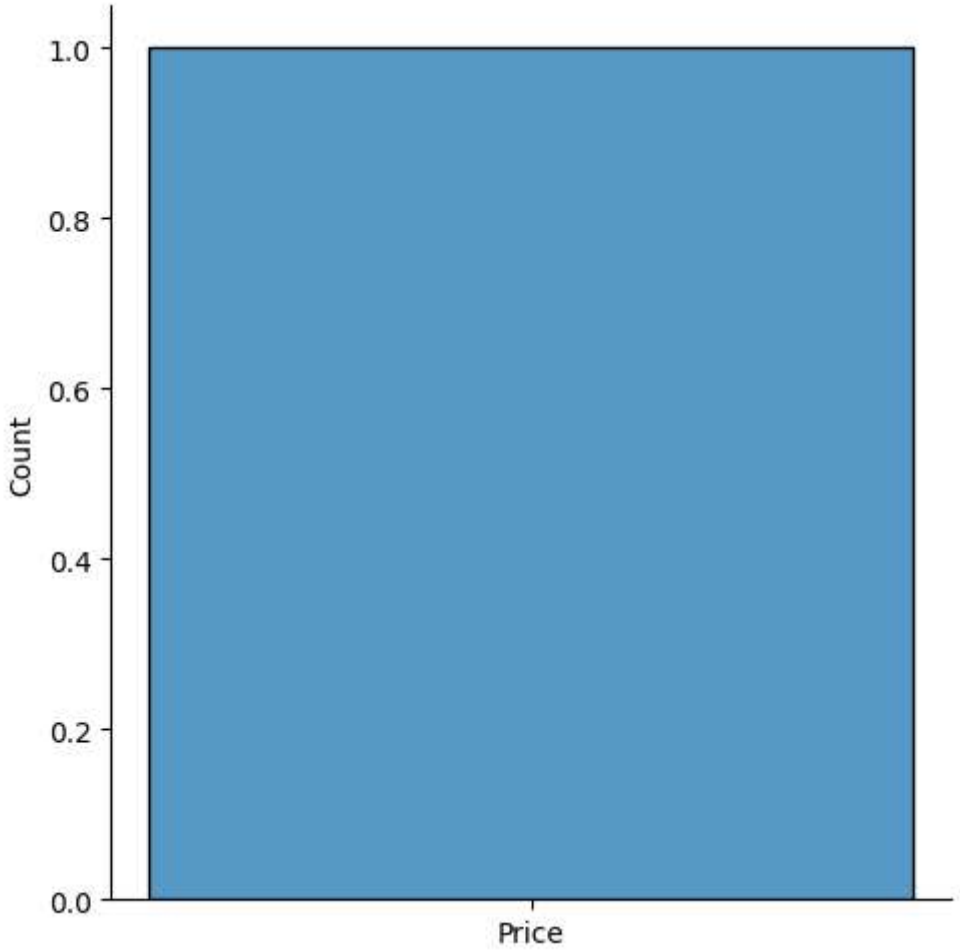
In [7]:
```
1 sns.pairplot(df)
```

Out[7]: <seaborn.axisgrid.PairGrid at 0x16567fb6e90>



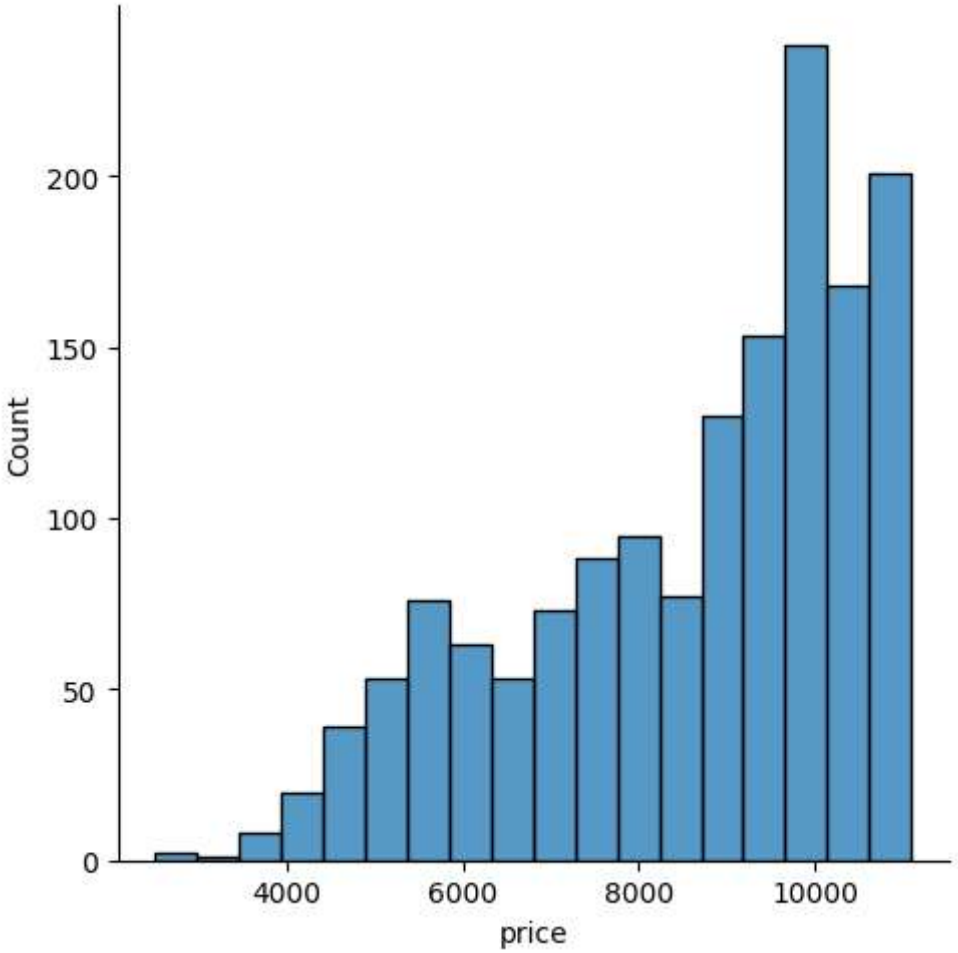
In [8]:
```
1 sns.displot(["Price"])
```

Out[8]: <seaborn.axisgrid.FacetGrid at 0x1650eb099c0>



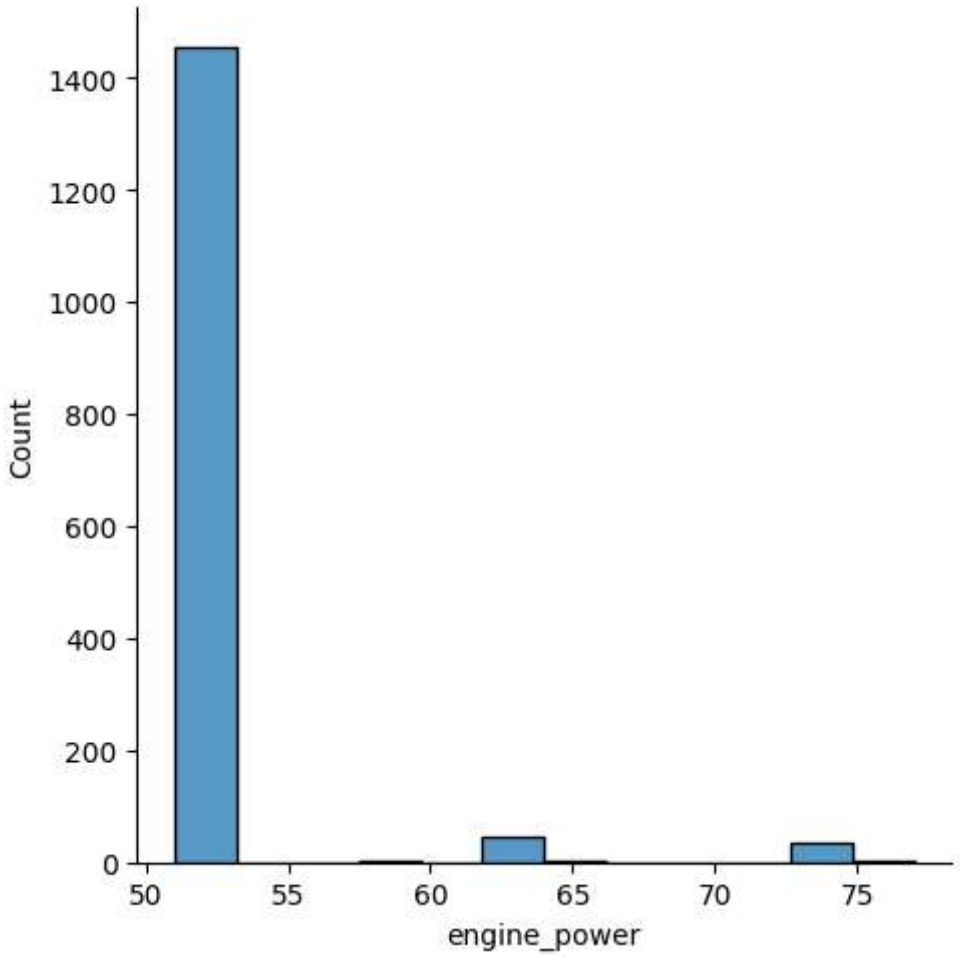
In [10]:
```
1 sns.displot(df["price"])
```

Out[10]: <seaborn.axisgrid.FacetGrid at 0x16512244970>



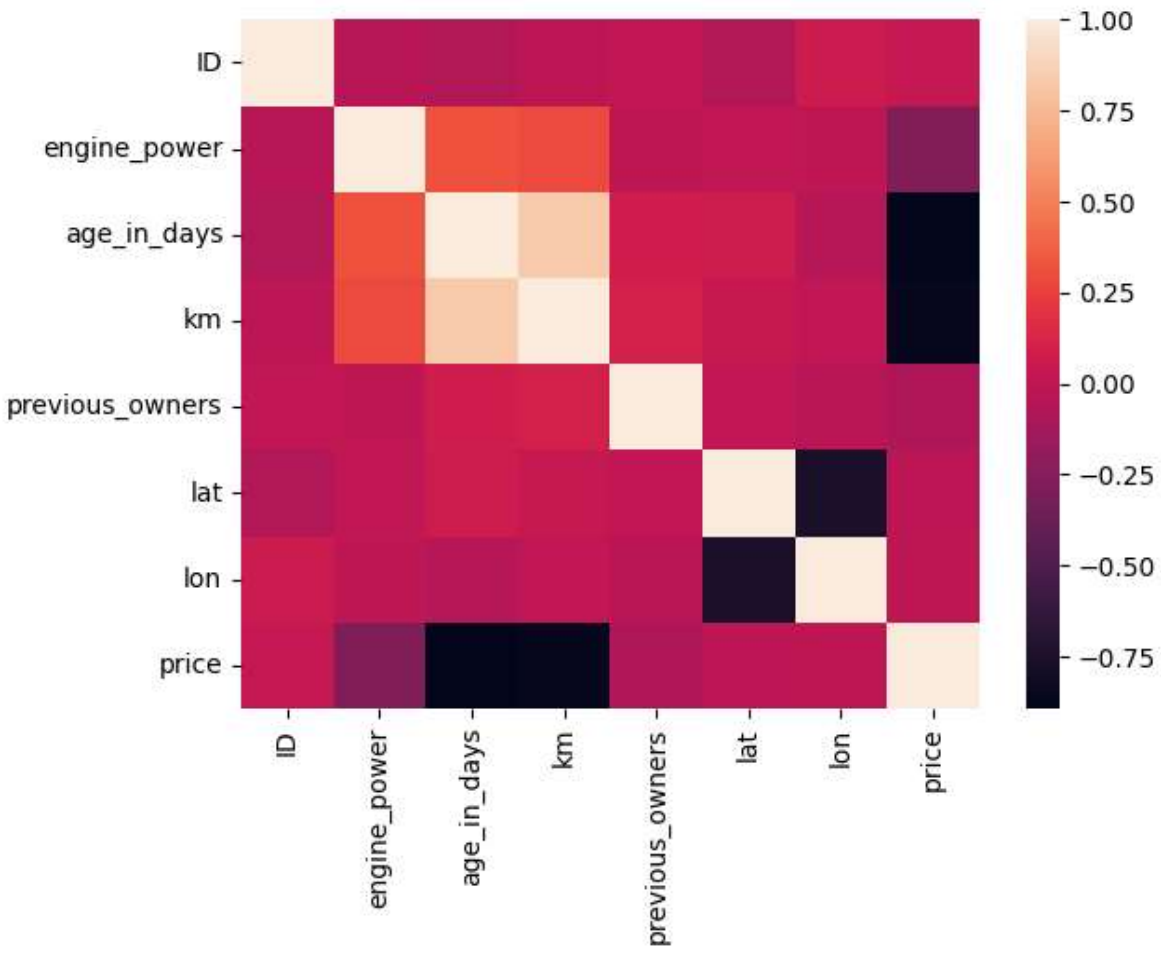
In [12]:
```
1 sns.displot(df["engine_power"])
```

Out[12]: <seaborn.axisgrid.FacetGrid at 0x1650e96e3b0>



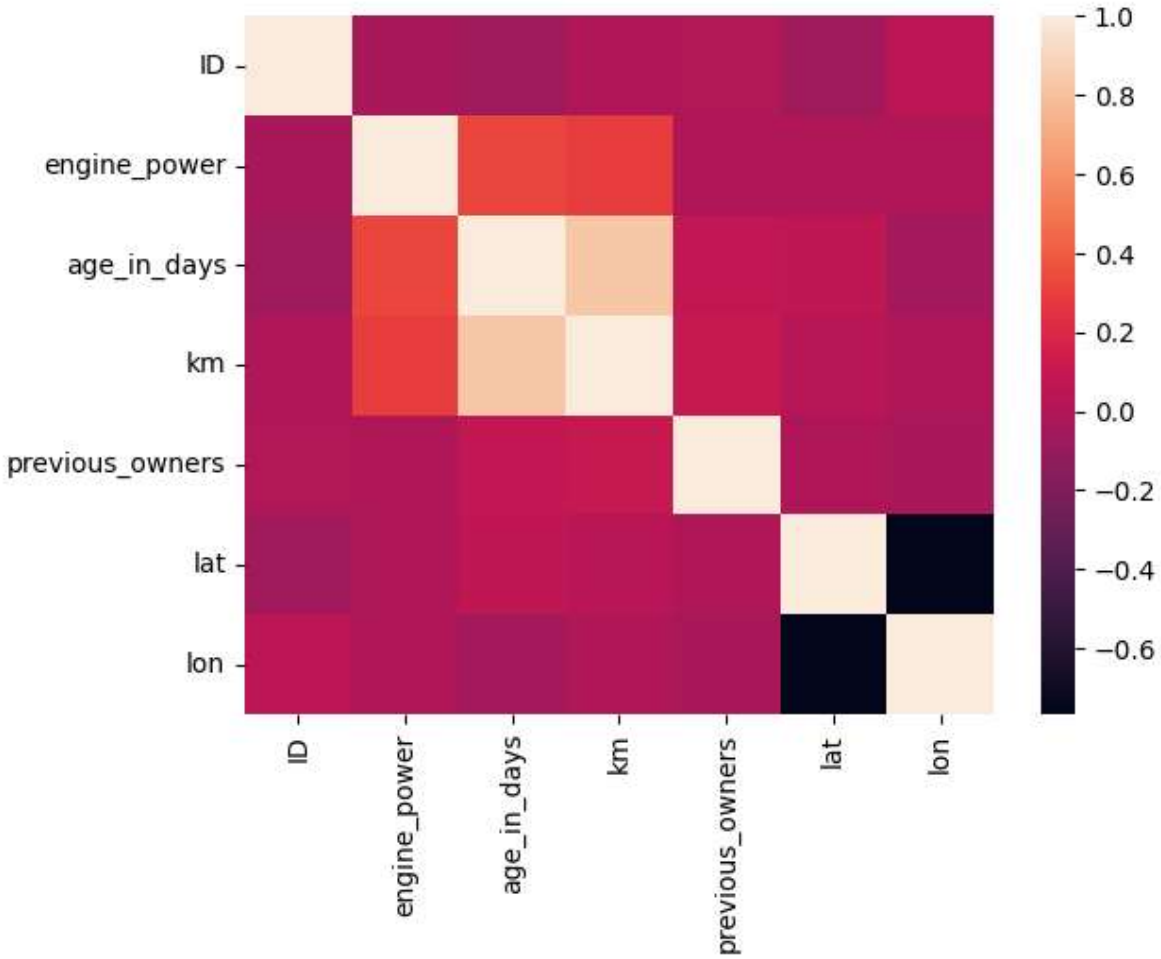
In [14]:
```
fiatdf=df[['ID', 'engine_power', 'age_in_days', 'km', 'previous_owners',
 'lat', 'lon','price']]
sns.heatmap(fiatdf.corr())
```

Out[14]: <Axes: >



In [15]:
```
fiatdf=df[['ID', 'engine_power', 'age_in_days', 'km', 'previous_owners',
 'lat', 'lon']]
sns.heatmap(fiatdf.corr())#without price
```

Out[15]: <Axes: >



In [16]:
```
X=fiatdf[['ID', 'engine_power', 'age_in_days', 'km', 'previous_owners',
 'lat', 'lon']]
y=df['price']
```

In [17]:
```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_s
from sklearn.linear_model import LinearRegression
regr=LinearRegression()
regr.fit(X_train,y_train)
print(regr.intercept_)

```

```
8971.195683500027
```

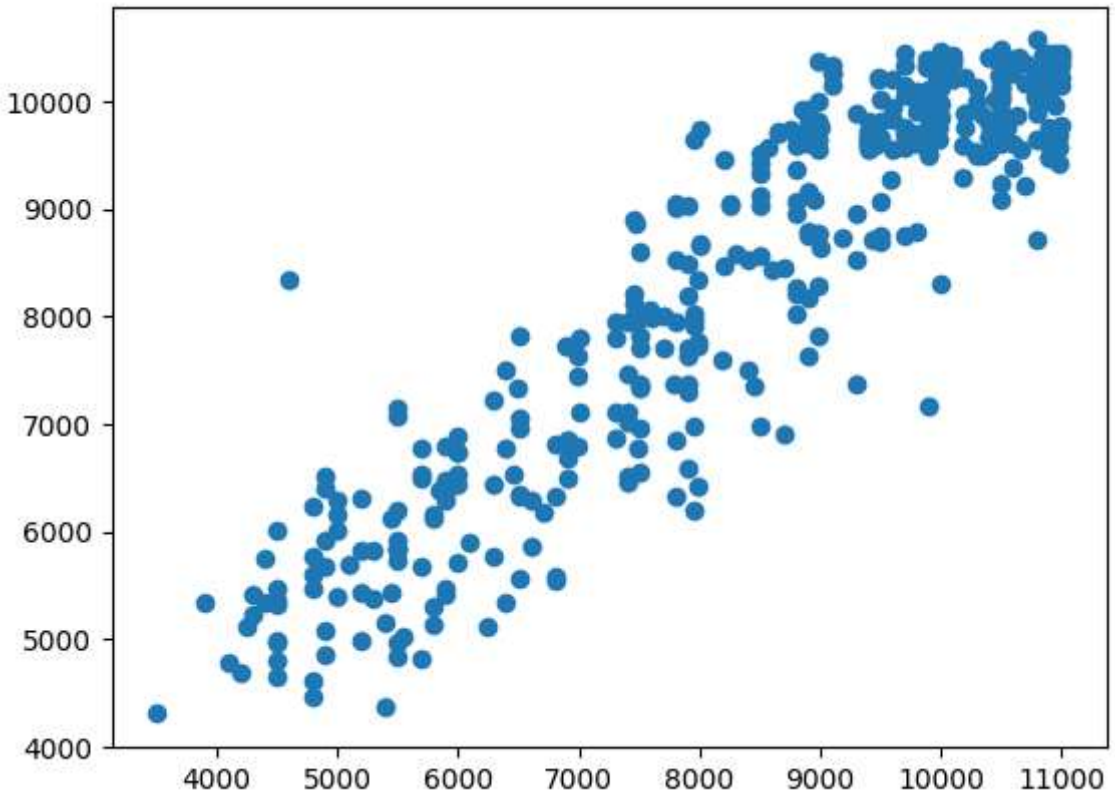
In [18]:
```
coeff_df=pd.DataFrame(regr.coef_,X.columns,columns=['coefficient'])
coeff_df
```

Out[18]:

	coefficient
ID	-0.046704
engine_power	11.646408
age_in_days	-0.898018
km	-0.017232
previous_owners	26.400886
lat	32.189709
lon	0.161073

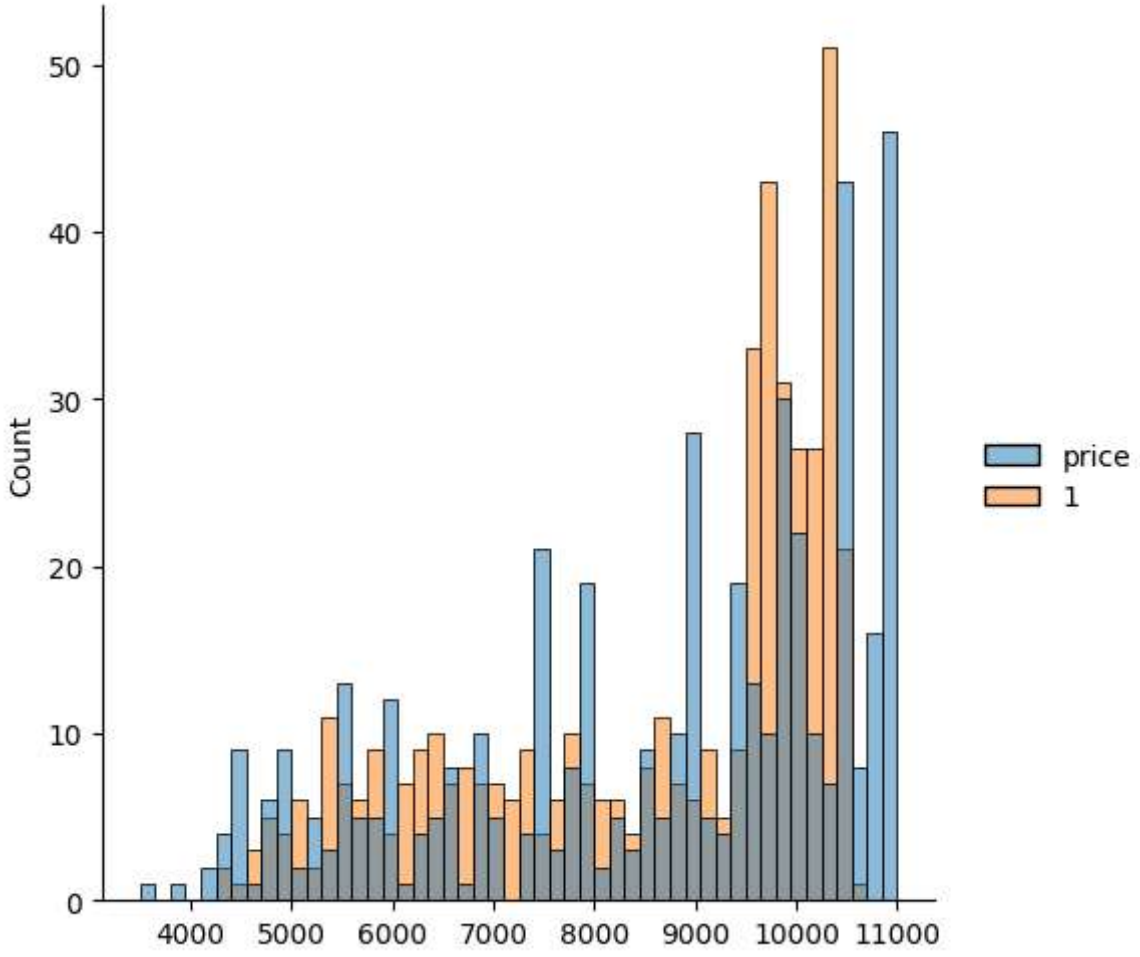
In [19]:
```
predictions=regr.predict(X_test)
plt.scatter(y_test,predictions)
```

Out[19]: <matplotlib.collections.PathCollection at 0x1651268b3a0>

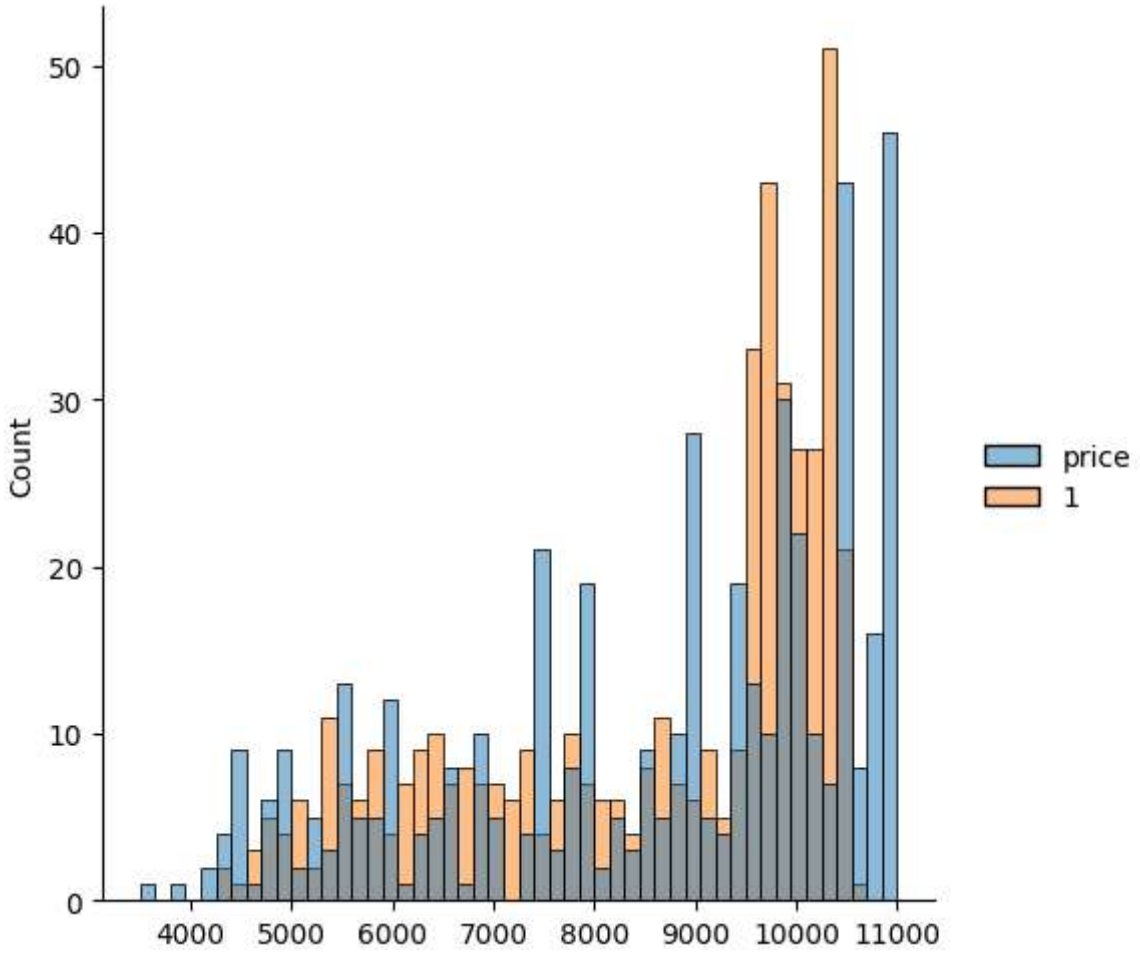


In [20]:
```
1 sns.displot((y_test,predictions),bins=50)#without semicolon
2
```

Out[20]: <seaborn.axisgrid.FacetGrid at 0x165126a0e80>



In [21]:
```
1 sns.displot((y_test,predictions),bins=50);#with semicolon
```



In [22]:
```
1 from sklearn import metrics
2 print('MAE:',metrics.mean_absolute_error(y_test,predictions))
3 print('MSE:',metrics.mean_squared_error(y_test,predictions))
4 print('MAE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))
5
```

```
MAE: 593.0876179519989
MSE: 551442.6799691883
MAE: 742.5918663500081
```

In [23]:
```
1 #accuracy
2 regr=LinearRegression()
3 regr.fit(X_train,y_train)
4 regr.fit(X_train,y_train)
5 print(regr.score(X_test,y_test))
6
```

```
0.8597136704308846
```

In [24]:
```
1 df.fillna(method='ffill',inplace=True)
```

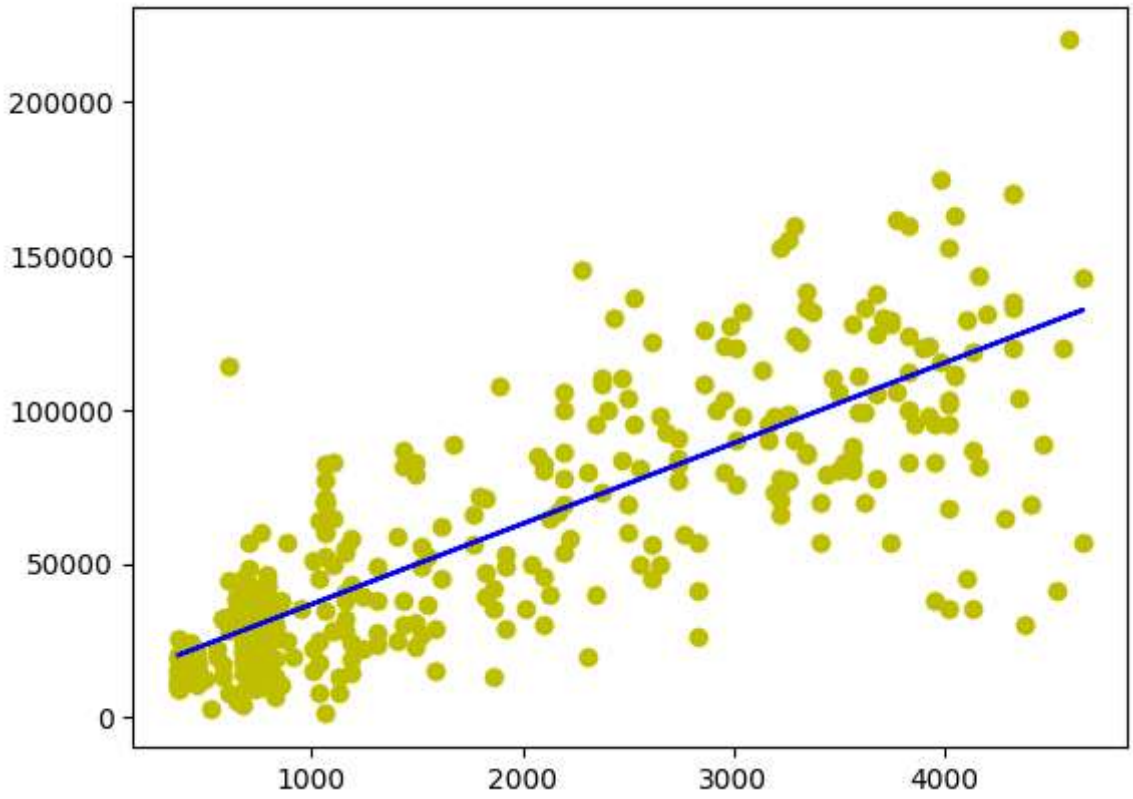
In [25]:
```
x=np.array(df['age_in_days']).reshape(-1,1)
y=np.array(df['km']).reshape(-1,1)
df.dropna(inplace=True)
```

In [26]:
```
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
regr.fit(X_train,y_train)
regr.fit(X_train,y_train)
```

Out[26]:
```
▾ LinearRegression

LinearRegression()
```

In [27]:
```
y_pred=regr.predict(X_test)
plt.scatter(X_test,y_test,color='y')
plt.plot(X_test,y_pred,color='b')
plt.show()
```



In [ ]:
```
1
```