In [8]:
```
import pandas as pd
from sklearn import linear_model
import seaborn as sns
```

In [2]:
```
df=pd.read_csv("carprices.csv")
```

In [3]:
```
df
```

Out[3]:

	Mileage	Age(yrs)	Sell Price($)
0	69000	6	18000
1	35000	3	34000
2	57000	5	26100
3	22500	2	40000
4	46000	4	31500
5	59000	5	26750
6	52000	5	32000
7	72000	6	19300
8	91000	8	12000
9	67000	6	22000
10	83000	7	18700
11	79000	7	19500
12	59000	5	26000
13	58780	4	27500
14	82450	7	19400
15	25400	3	35000
16	28000	2	35500
17	69000	5	19700
18	87600	8	12800
19	52000	5	28200

In [4]:
```
'''change the column name'''
df = df.rename(columns={'Mileage': 'mileage',"Age(yrs)":"age","Sell Price($)":"price"})
```

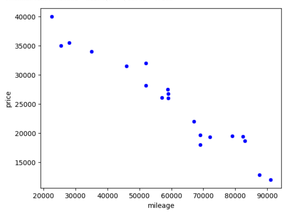
In [5]:
```
df
```

Out[5]:

	mileage	age	price
0	69000	6	18000
1	35000	3	34000
2	57000	5	26100
3	22500	2	40000
4	46000	4	31500
5	59000	5	26750
6	52000	5	32000
7	72000	6	19300
8	91000	8	12000
9	67000	6	22000
10	83000	7	18700
11	79000	7	19500
12	59000	5	26000
13	58780	4	27500
14	82450	7	19400
15	25400	3	35000
16	28000	2	35500
17	69000	5	19700
18	87600	8	12800
19	52000	5	28200

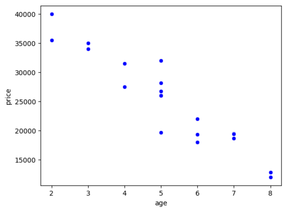
In [10]:
```
sns.scatterplot(x="mileage",y="price",data=df,color="blue")
# here we see the relation between mileage and price linear
```

Out[10]:
```
<AxesSubplot: xlabel='mileage', ylabel='price'>
```



In [11]:
```
sns.scatterplot(x="age",y="price",data=df,color="blue")
# here we can consider this as linear as well
# so we biuld the linear model
```

Out[11]:
```
<AxesSubplot: xlabel='age', ylabel='price'>
```



In [12]:
```
x=df[['mileage','age']]
```

In [13]:
```
x
```

Out[13]:

	mileage	age
0	69000	6
1	35000	3
2	57000	5
3	22500	2
4	46000	4
5	59000	5
6	52000	5
7	72000	6
8	91000	8
9	67000	6
10	83000	7
11	79000	7
12	59000	5
13	58780	4
14	82450	7
15	25400	3
16	28000	2
17	69000	5
18	87600	8
19	52000	5

In [14]:
```
y=df.price
```

In [15]:
```
y
```

Out[15]:
```
0     18000
1     34000
2     26100
3     40000
4     31500
5     26750
6     32000
7     19300
8     12000
9     22000
10    18700
11    19500
12    26000
13    27500
14    19400
15    35000
16    35500
17    19700
18    12800
19    28200
Name: price, dtype: int64
```

In [16]:
```
from sklearn.model_selection import train_test_split
# here we are going to divide our dataset for training and testing
```

In [31]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

In [32]:
```
x_train
```

Out[32]:

	mileage	age
1	35000	3
3	22500	2
15	25400	3
14	82450	7
10	83000	7
4	46000	4
2	57000	5
19	52000	5
16	28000	2
11	79000	7
13	58780	4
9	67000	6
5	59000	5
17	69000	5
8	91000	8
12	59000	5

In [33]:
```
y_train
```

Out[33]:
```
1     34000
3     40000
15    35000
14    19400
10    18700
4     31500
2     26100
19    28200
16    35500
11    19500
13    27500
9     22000
5     26750
17    19700
8     12000
12    26000
Name: price, dtype: int64
```

In [34]:
```
x_test
```

Out[34]:

	mileage	age
7	72000	6
18	87600	8
6	52000	5
0	69000	6

In [35]:
```
reg=linear_model.LinearRegression()
```

In [36]:
```
reg.fit(x_train,y_train)  # here provide 80% for training
```

Out[36]:
```
LinearRegression()
```

In [37]:
```
reg.predict(x_train)
```

Out[37]:
```
array([34051.48423914, 38359.60199786, 36935.52522528, 17585.08659119,
       17419.85507636, 30193.9978845 , 26336.51152906, 27838.61621015,
       36707.28684955, 18621.53882059, 26354.61832169, 22779.44611128,
       25735.66965775, 22731.46029718, 14463.63152989, 25735.66965775])
```

In [38]:
```
reg.predict(x_test)  # this is the predicted  price of cars of 20%
```

Out[38]:
```
array([21277.341431  , 15485.06271249, 27838.61621015, 22178.60423917])
```

In [39]:
```
y_test # this is the price of cars of 20% dataset
```

Out[39]:
```
7     19300
18    12800
6     32000
0     18000
Name: price, dtype: int64
```

In [40]:
```
reg.score(x_test,y_test)
# this is the accuracy of model
```

Out[40]: 0.7696237249437479

In [ ]: