```python
In [22]:  import pandas as pd
          import numpy as np
          import matplotlib.pyplot as plt
          %matplotlib inline
          from sklearn.model_selection import train_test_split
```
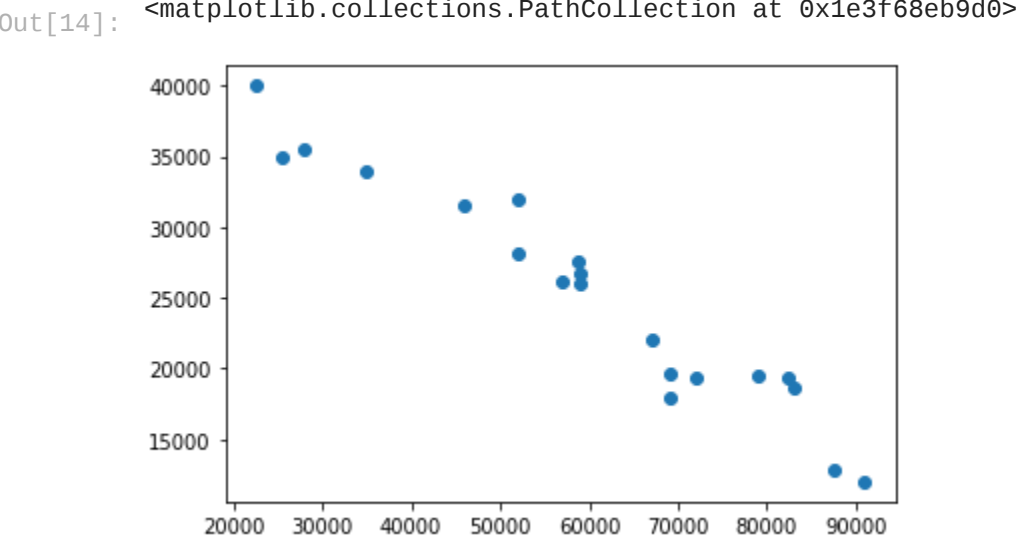
```python
In [12]:  df=pd.read_csv('E:/machine learning/train-test split.csv')
```

```python
In [13]:  df.head()
```

Out[13]:

	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

```python
In [14]:  plt.scatter(df['Mileage'],df['Sell Price($)'])
```

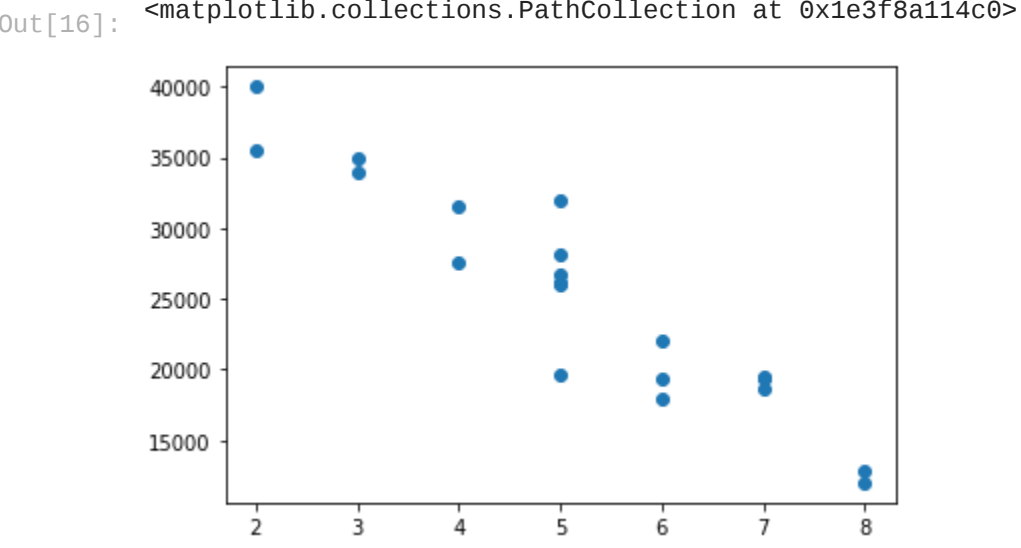Out[14]: <matplotlib.collections.PathCollection at 0x1e3f68eb9d0>


```python
In [15]:  df.tail()
```

Out[15]:

	Mileage	Age(yrs)	Sell Price($)
15	25400	3	35000
16	28000	2	35500
17	69000	5	19700
18	87600	8	12800
19	52000	5	28200

```python
In [16]:  plt.scatter(df['Age(yrs)'],df['Sell Price($)'])
```

Out[16]: <matplotlib.collections.PathCollection at 0x1e3f8a114c0>


```python
In [17]:  x=df.drop(['Sell Price($)'],axis='columns')
```

```python
In [18]:  x
```

Out[18]:

	Mileage	Age(yrs)
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

```python
In [19]:  y=df['Sell Price($)']
```

```python
In [20]:  y
```

Out[20]:
```
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: Sell Price($), dtype: int64
```

```python
In [24]:  x_train, x_test, y_train, y_test=train_test_split(x,y,test_size=0.2) #0.2 specifies that 20% data is for testing
```

```python
In [25]:  len(x_train)
```
Out[25]: 16

```python
In [26]:  len(x_test)
```
Out[26]: 4

```python
In [27]:  from sklearn.linear_model import LinearRegression
```

```python
In [28]:  modelc=LinearRegression()
```

```python
In [30]:  modelc.fit(x_train,y_train)
```
Out[30]: LinearRegression()

```python
In [31]:  modelc.predict(x_test)
```
Out[31]: array([37115.56903286, 13665.29353348, 38878.41124726, 18054.44793276])

```python
In [32]:  y_test
```
Out[32]:
```
16    35500
8     12000
3     40000
11    19500
Name: Sell Price($), dtype: int64
```

```python
In [34]:  modelc.score(x_test,y_test)
```
Out[34]: 0.9832822442087942
```