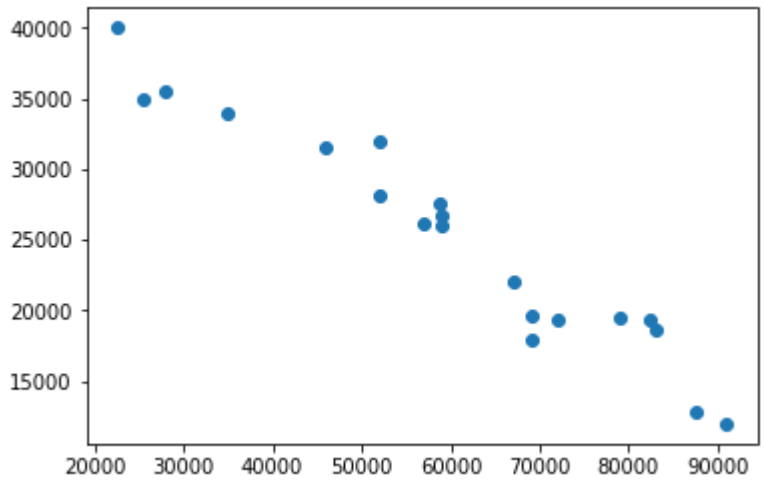
In [2]:
```
# import the data from file

import pandas as pd
df = pd.read_csv(r"C:\Users\309962\Desktop\carprices.csv")
df.head()
```

Out[2]:

	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

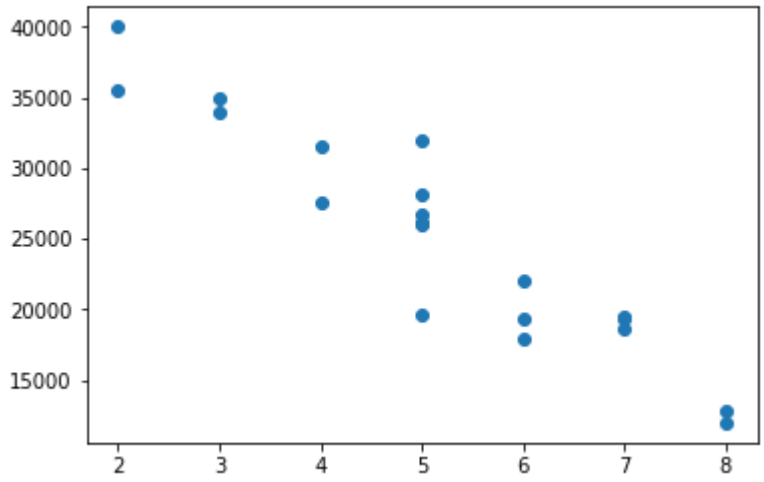
In [3]:
```
# import matplotlib to plot the data between Mileage vs SellPrice
import matplotlib.pyplot as plt
%matplotlib inline

plt.scatter(df['Mileage'],df['Sell Price($)'])
```

Out[3]:
```
<matplotlib.collections.PathCollection at 0x19bc86616d8>
```



In [4]:
```
# Plot the data between Age vs Sell Price
plt.scatter(df['Age(yrs)'],df['Sell Price($)'])
```

Out[4]: <matplotlib.collections.PathCollection at 0x19bc8bc55c0>



In [5]:
```
# Split the Data with Indepdendent variables as age and Mileage and dependent var

X = df[['Mileage','Age(yrs)']]
y = df['Sell Price($)']
```

In [6]:
```
# import sklearn to split the data between test and train

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)
```

In [7]:
```
X_train
```

Out[7]:

	Mileage	Age(yrs)
10	83000	7
0	69000	6
9	67000	6
1	35000	3
6	52000	5
11	79000	7
14	82450	7
17	69000	5
19	52000	5
15	25400	3
18	87600	8
13	58780	4
16	28000	2
5	59000	5

In [8]:
```
X_test
```

Out[8]:

	Mileage	Age(yrs)
7	72000	6
4	46000	4
8	91000	8
2	57000	5
12	59000	5
3	22500	2

In [9]:
```
y_train
```

Out[9]:
```
10    18700
0     18000
9     22000
1     34000
6     32000
11    19500
14    19400
17    19700
19    28200
15    35000
18    12800
13    27500
16    35500
5     26750
Name: Sell Price($), dtype: int64
```

In [10]:
```
y_test
```

Out[10]:
```
7     19300
4     31500
8     12000
2     26100
12    26000
3     40000
Name: Sell Price($), dtype: int64
```

In [11]:
```
# now create a linear model

from sklearn.linear_model import LinearRegression
clf = LinearRegression()
clf.fit(X_train, y_train)
```

Out[11]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

In [12]:
```
X_test
```

Out[12]:

	Mileage	Age(yrs)
7	72000	6
4	46000	4
8	91000	8
2	57000	5
12	59000	5
3	22500	2

In [13]:
```
# Predict the sell price now

clf.predict(X_test)
```

Out[13]:
```
array([21027.17380314, 29927.68743754, 14378.61246857, 26120.84556315,
       25477.43062034, 38023.93239343])
```

In [14]:
```
y_test
```

Out[14]:
```
7     19300
4     31500
8     12000
2     26100
12    26000
3     40000
Name: Sell Price($), dtype: int64
```

In [15]:
```
# Do a comparison with Score(Actual Vs predicted output)
clf.score(X_test, y_test)
```

Out[15]:
```
0.9672523728982899
```

In [16]:
```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3,random_stat
X_test
```

Out[16]:

	Mileage	Age(yrs)
7	72000	6
10	83000	7
5	59000	5
6	52000	5
3	22500	2
18	87600	8

In [ ]:
```

```