

LINEAR REGRESSION MODEL - YEARLY AMOUNT SPENT ON APP & WEBSITE

In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
```

In [4]:

```
pwd
```

Out[4]:

```
'C:\\Users\\sidh'
```

In [5]:

```
df = pd.read_csv('Ecommerce Customers')
df.head(2)
```

Out[5]:

	Email	Address	Avatar	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
0	mstephenson@fernandez.com	835 Frank TunnelnWrightmouth, MI 82180-9605	Violet	34.497268	12.655651	39.577668	4.082621	587.951054
1	hduke@hotmail.com	4547 Archer CommonnDiazchester, CA 06566-8576	DarkGreen	31.926272	11.109461	37.268959	2.664034	392.204933

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Email                  500 non-null   object
1   Address                500 non-null   object
2   Avatar                 500 non-null   object
3   Avg. Session Length    500 non-null   float64
4   Time on App            500 non-null   float64
5   Time on Website        500 non-null   float64
6   Length of Membership    500 non-null   float64
7   Yearly Amount Spent    500 non-null   float64
dtypes: float64(5), object(3)
memory usage: 31.4+ KB
```

In [7]:

```
df.columns
```

Out[7]:

```
Index(['Email', 'Address', 'Avatar', 'Avg. Session Length', 'Time on App',
      'Time on Website', 'Length of Membership', 'Yearly Amount Spent'],
      dtype='object')
```

In [8]:

```
X = df[['Avg. Session Length', 'Time on App','Time on Website', 'Length of Membership']]
y = df['Yearly Amount Spent']
```

In [9]:

```
X_train,X_test,y_train,y_test = train_test_split(X ,y,test_size=0.3)
```

In [10]:

```
lr = LinearRegression()
```

In [11]:

```
lr.fit(X_train,y_train)
```

Out[11]:

```
LinearRegression()
```

In [12]:

```
lr.coef_
```

Out[12]:

```
array([26.1973113 , 39.06665892,  0.65186878, 61.74601314])
```

In [16]:

```
cdf = pd.DataFrame(lr.coef_ , X.columns , columns= ['coeff'])
cdf
```

Out[16]:

	coeff
Avg. Session Length	26.197311
Time on App	39.066659
Time on Website	0.651869
Length of Membership	61.746013

In [17]:

```
prede = lr.predict(X_test)
```

In [18]:

```
plt.scatter(y_test,prede)
```

Out[18]:

```
<matplotlib.collections.PathCollection at 0x2755725e3d0>
```

In [19]:

```
mae = metrics.mean_absolute_error(y_test,prede)
mse = metrics.mean_squared_error(y_test,prede)
rmse = np.sqrt(mse)
```

In [21]:

```
mae
```

Out[21]:

```
7.97543945199709
```

In [22]:

```
mse
```

Out[22]:

```
102.20732882221695
```

In [23]:

```
rmse
```

Out[23]:

```
10.109764033953361
```

In [ ]: