Ecommerce Customers

Project 2

Exploratory Data Analysis (EDA)

```
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
```

I see no missing values, outliers or duplicated values. Deleting Email, Address and Avatar columns.

```
df.isnull().sum()
```

```
Email                   0
Address                 0
Avatar                  0
Avg. Session Length     0
Time on App             0
Time on Website         0
Length of Membership    0
Yearly Amount Spent     0
dtype: int64
```

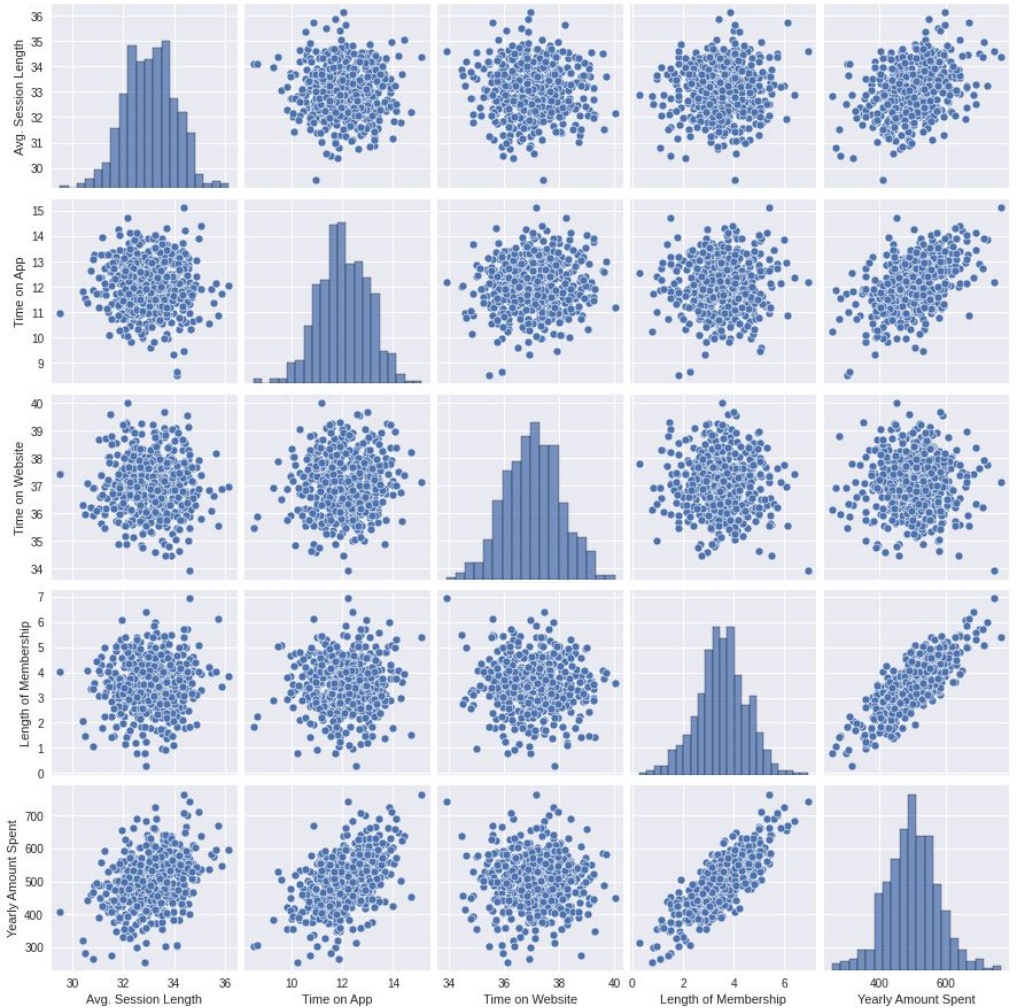
```
df.duplicated().sum()
```

```
0
```

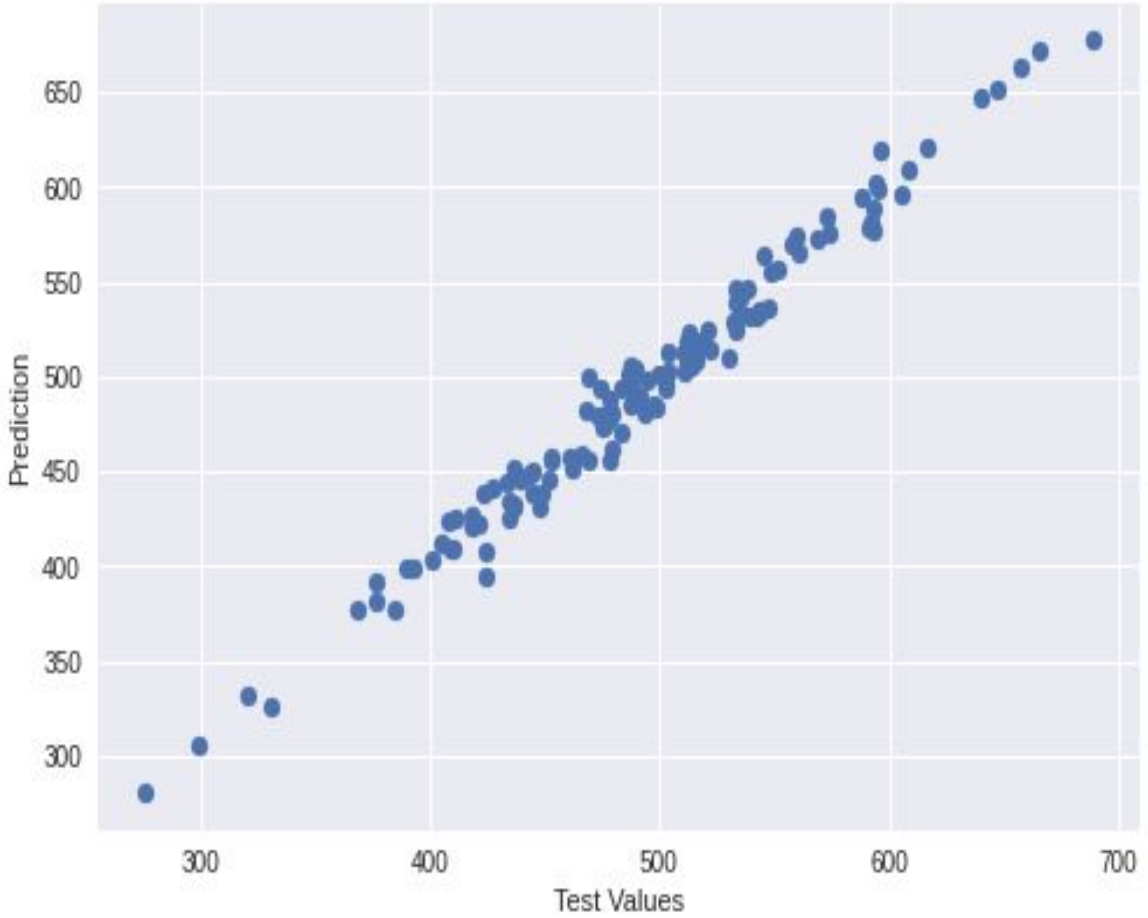
```
del df['Email']
del df['Address']
del df['Avatar']
```

Correlation HeatMap

PairPlot helps us to understand the correlations between last two rows between Yearly Amount Spent and Length of Membership as they are Linearly Correlated.



This diagram shows the linearity of the output



Coefficients indicates the direction of the relationship between a predictor variable and the response variable.

```
features = X.columns
coefficient = reg.coef_

pd.DataFrame({'Coefficient': coefficient}, index=features)
```

	Coefficient
Avg. Session Length	25.690832
Time on App	38.688156
Time on Website	0.452799
Length of Membership	61.710503

From this Linear Regression Model to predict the annual revenues whether to focus on the app or the website, I see both are not is a great choice to decide. The Length of Membership determines the Annual Revenues.