Analyze E-commerce Customer Dataset Using Linear Regression

**CS306 Data Analysis And Visualization**

Problem Statement

➢ E-commerce Company sells clothes online. They try to identify if a company should focus on improving mobile app experience or on improving website experience for their customers.

➢ We used the E-commerce Company dataset of 500 different customers for the analysis. Dataset contains details like Email, Address, Avatar, Average Session length, Time on App, Time on Website, Length of Membership, Yearly Amount Spent.

➢ Our Main motive is to conclude that company should invest on improving website or app to get maximum profit from customers yearly amount spent. Using Linear Regression Technique we analyze the effects of different parameters like Avg. session length, time on app,time on website, length of membership on the yearly amount spent by the customers.

Exploratory Data Analysis

Histogram for "Yearly Amount Spent"

➢ No outliers are available
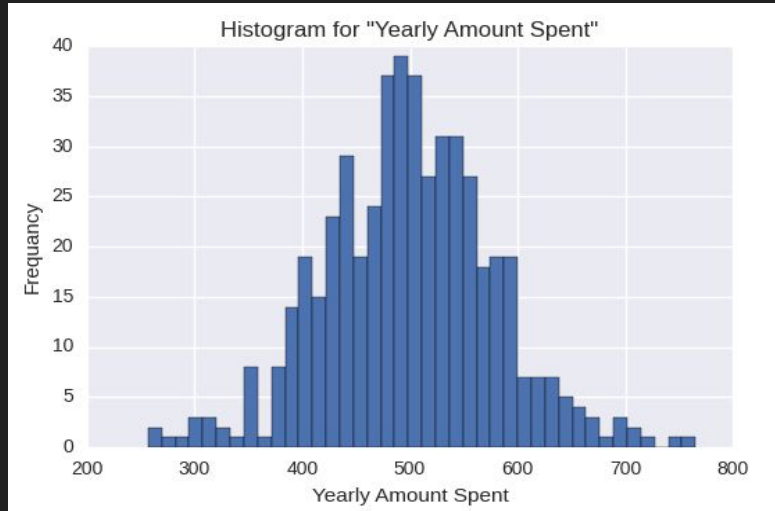
➢ ECDF gives us the sense of the probabilistic distribution of the data.

CDF
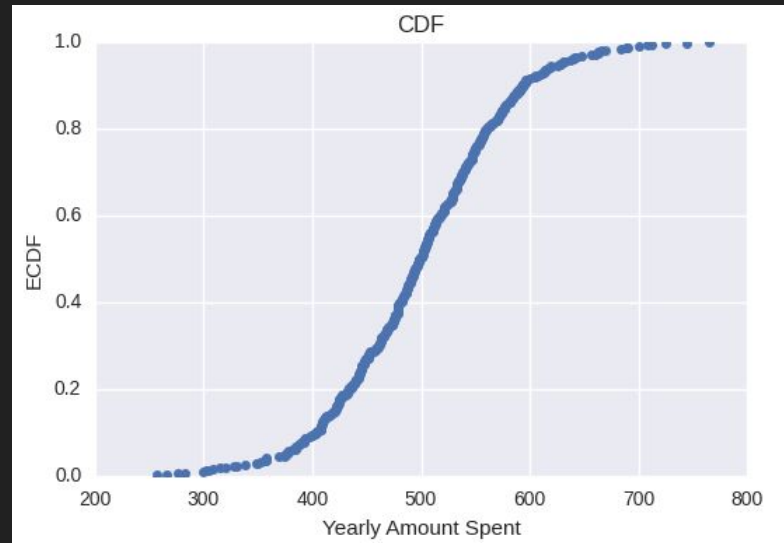
Correlation Matrix

	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
Avg. Session Length	1	-0.028	-0.035	0.06	0.36
Time on App	-0.028	1	0.082	0.029	0.5
Time on Website	-0.035	0.082	1	-0.048	-0.026
Length of Membership	0.06	0.029	-0.048	1	0.81
Yearly Amount Spent	0.36	0..5	-0.026	0.81	1
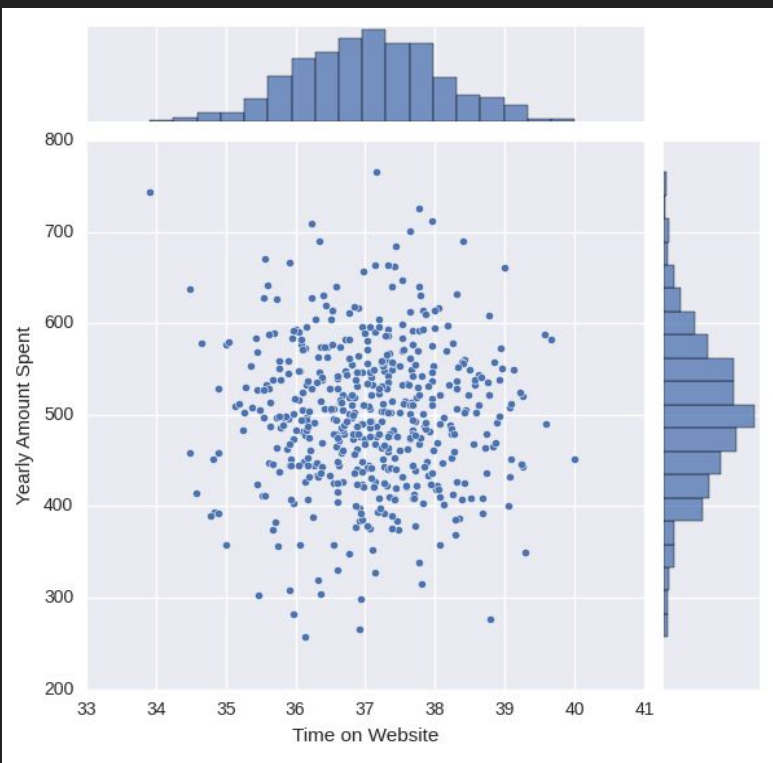
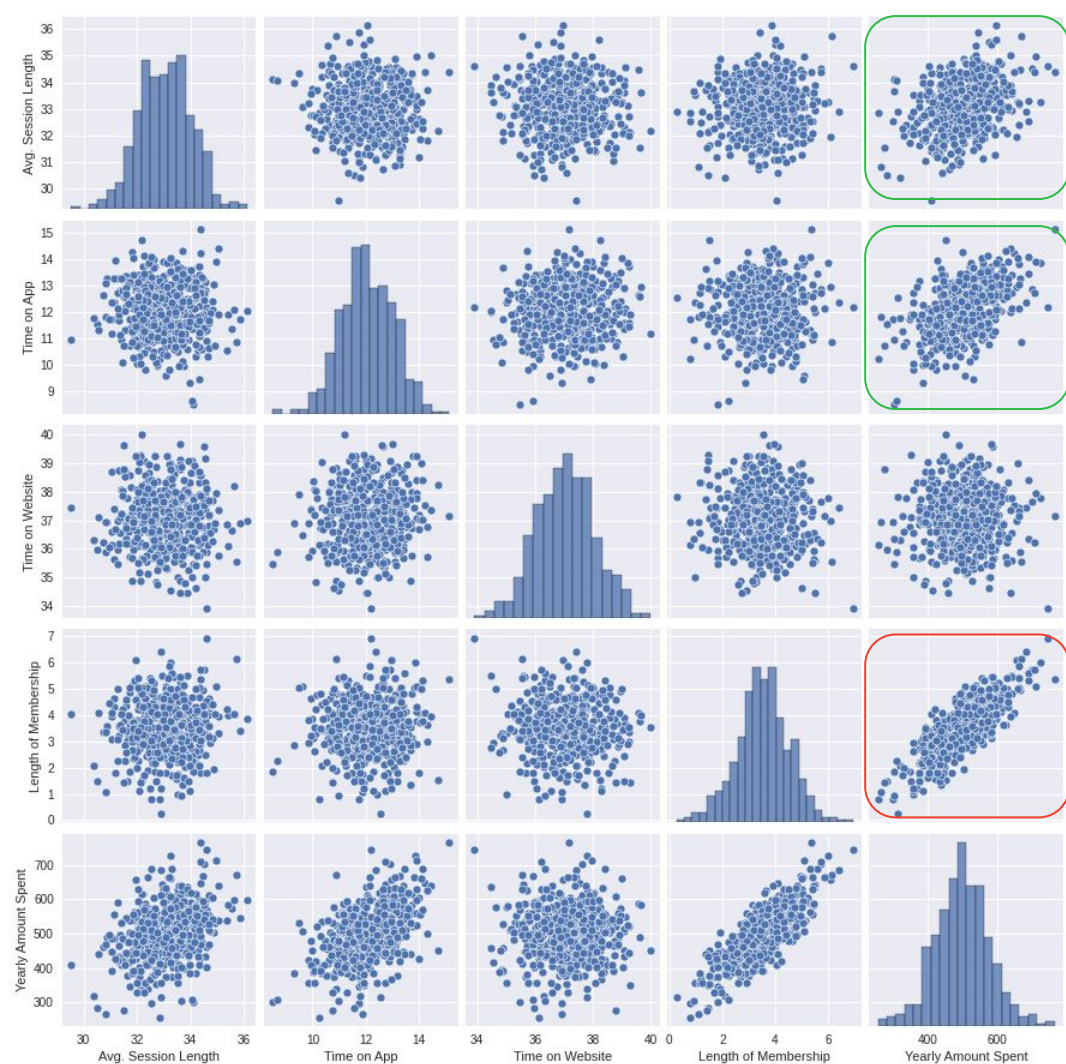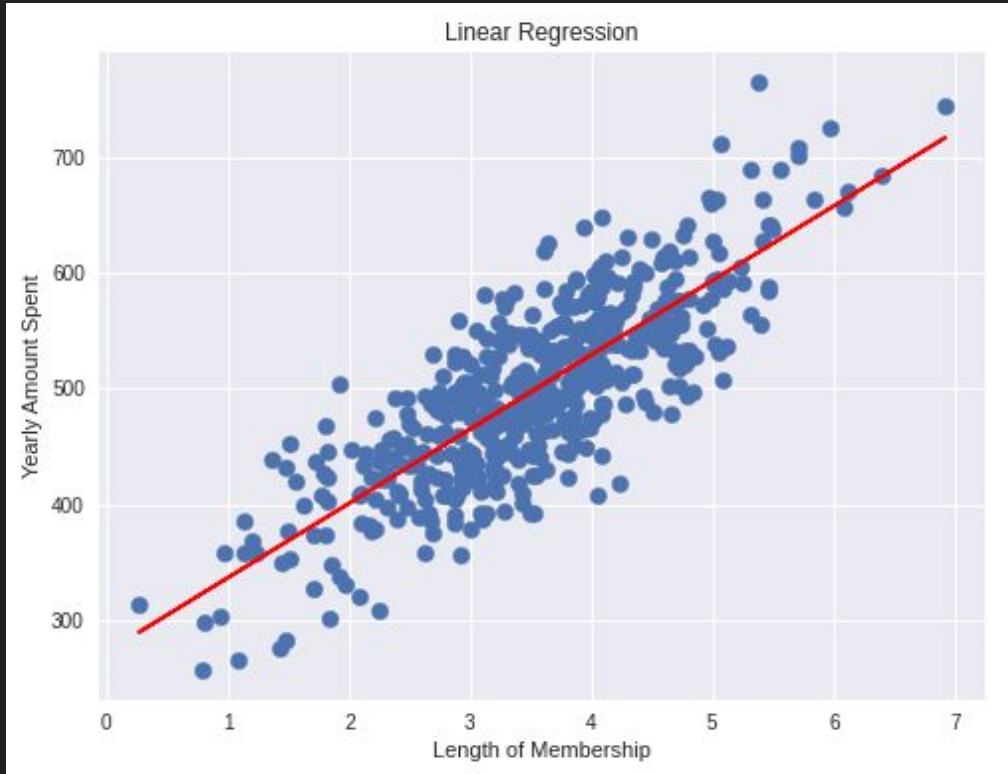➢ Correlation of Time on website and Yearly amount spent is very low .

➢ R = -0.026
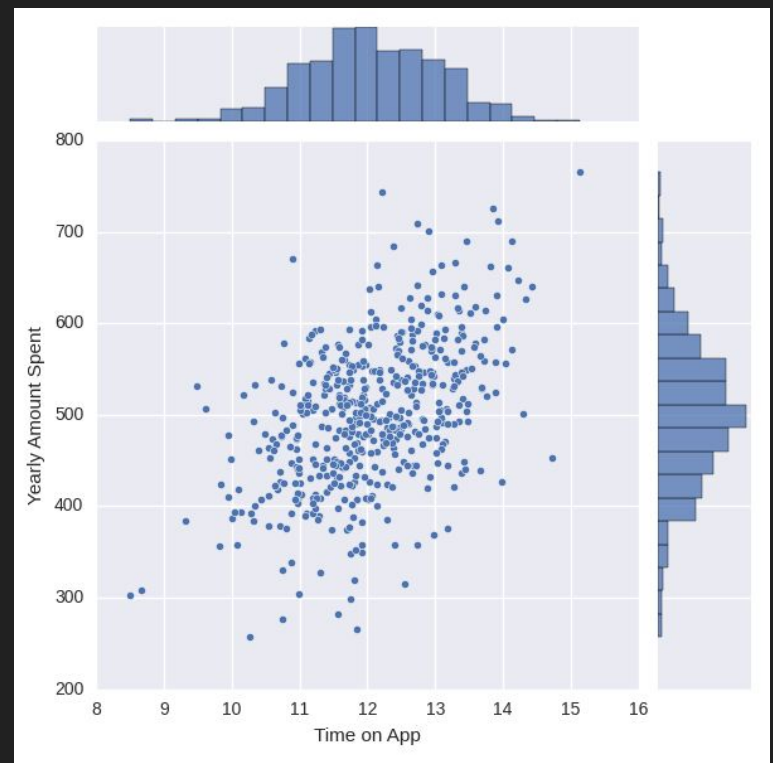
➢ Correlation of Time on App and Yearly amount spent is significant.

➢ R = 0.499

Pair Plot

➤ 'Length of Membership' is heavily correlated with 'Yearly Amount Spent' .

➤ 'Time on App' and 'Avg. Session Length' is slightly correlated with 'Yearly Amount Spent' .

Linear Regression of Yearly amount spent vs Length of Membership

➢ Length of membership looks like a linear relationship with the yearly amount spent So, we did Simple Linear Regression analysis for both parameter.

➢ The longer you stay a member, the larger your Yearly Amount Spent.

Y(Yearly Amount Spent) = 272.4 + 64.22 * X(Length of Membership)

Multiple Linear Regression Model
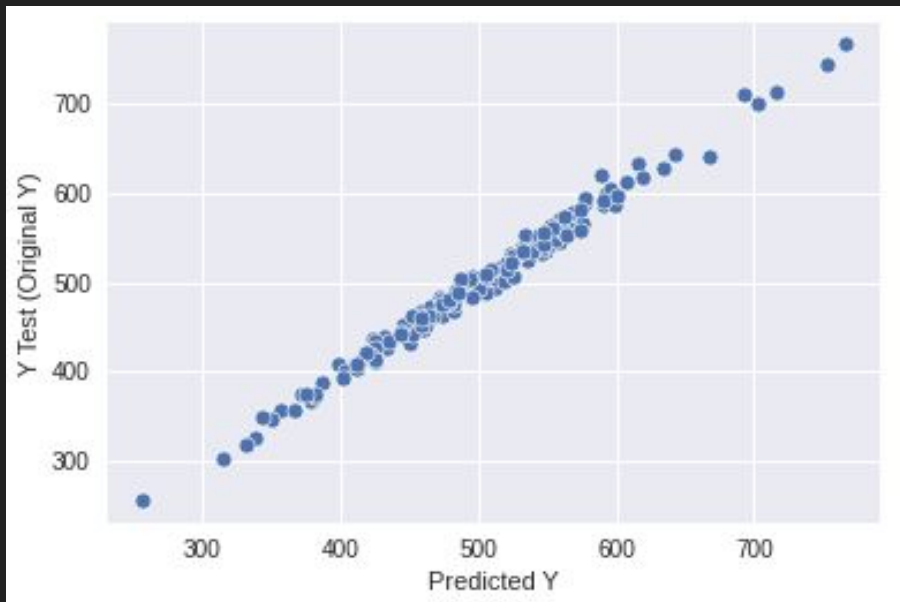
➢ In this multiple linear regression analysis we want to show that the Yearly Amount Spent feature is linearly depend on Avg. Session Length, Time on App, Time on Website, Length of Membership.

➢ β_0 = -1047.93
➢ β_1 = 25.98
➢ β_2 = 38.59
➢ β_3 = 0.19
➢ β_4 = 61.28

Y(Yearly Amount Spent) = β_0 + β_1*(Avg. Session Length) + β_2*(Time on App) + β_3*(Time on Website) + β_4 * X(Length of Membership)

Evaluating The Model

➢ The explained variance score (R^2) is used to determine how much variance the model explains.

➢ Other common calculations to measure the performance of a regression model is Root Mean Squared Error (RMSE). Here we used RMSE to calculate the error between Actual value and predicted value of the yearly amount spent.

➢ R^2 = 0.9891

➢ RMSE = 9.9338
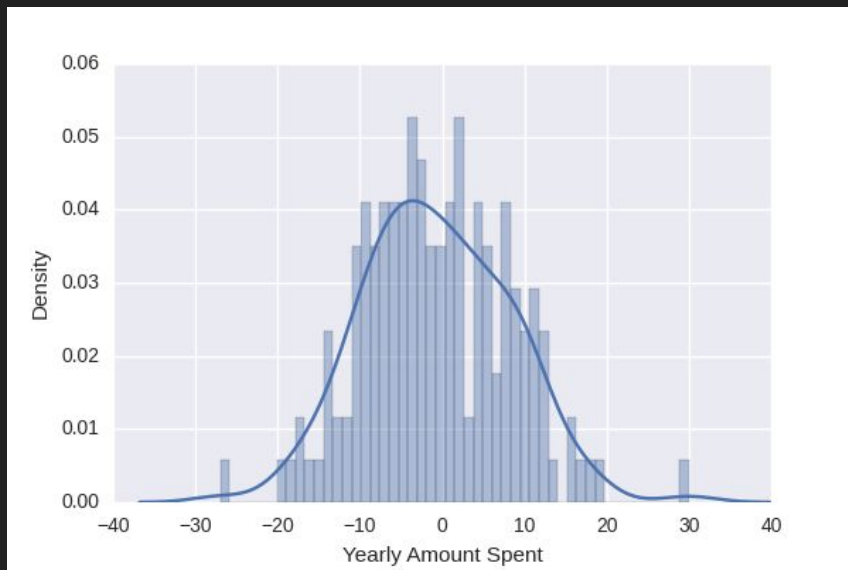
Conclusion

➢ Increase in the Length of Membership would affect more on the value Yearly Amount Spent.

➢ Time spent on App has more impact than Time spent on Website so, here we can say that company should focus on invest money on App rather than invest money on Website.

➢ Different economic factors would determine which course of action to take, but at least we now have knowledge of the state of our website and our app in terms of yearly spend per customer.

➢ With addition to this, however, Length of Membership was the greatest impact in the amount a customer spent yearly, meaning that the longer the customers stay with the company, the more money the company will make in the long run. So, length of Membership should be included as the economic factors to decide between focusing on their app experience or their website.

THANK YOU

Dishant Goti	201801449
Siddharth Moradiya	201801470