**Sravanthi Nallandula**

**21333737**

**Data Exploration & Analysis**

**Professor Neerja Sahu**

**Week 12 Term Project**

**Outcome of your EDA**

The outcome of my data analysis for best selling books is I was able to find how variables relate to one another. While plotting graphs, the impact of one variable changed the path of the graph. I was able to find the relation between variables is linear. Depending on the relationship, I was able to find out which regression analysis must be performed for my data. The main important thing I have learned is using different libraries for various functions.

**What do you feel was missed during the analysis?**

The most difficult part in my analysis is plotting probability mass function and cumulative distribution function. I have researched a lot in how to plot it and finally I did plot the graphs. But I feel, I might have missed understanding the concept fully and also in portraying it.

**Were there any variables you felt could have helped in the analysis?**

I might have used author variable for in-depth analysis. My analysis might be different because there will be a lot of chances to show variations in the data by using author variable. Even using the name variable will be as effective as using author variable.

**Were there any assumptions made you felt were incorrect?**

My assumption of using Year variable might be incorrect. I haven’t used it much as I used User Rating and Reviews. One more thing I felt was, I should have selected a dataset with more variables having numerical data.

**What challenges did you face, what did you not fully understand?**

I tried plotting a graph for multiple linear regression, but I couldn’t do it and whatever I tried to plot the graph threw errors. I was able to correct few of the errors but not all. Converting datatypes in my case gave me a tough time at last moments. I should concentrate on regression analysis, PMF, and CDF concepts.