**Linear regression model for iamb dataset -Final Report**

There are many factors that affect movies and make them unique from one another. In this project, I do analyze these factors and know which of them affect movie popularity.

First of all, I do cleaning the data set by removing null values from rows, removing duplicated values, and covert genre column from string to list, and assigning a genre to each movie.
Then I do some charts and plot them to know which variable is dependent on vote average or popularity as they were more distributed.
After that, I plotted some variables as shown above and found that movies with higher vote averages lead to higher budgets and higher revenue.
I also found that movies released between 1970-1985 lead to be voted higher than others.

**Limitations**

1- I found that limitations to drawing the last chart, so I grouped the movies by genre and drew the chart
2- Also, due to the columns representing categorical data and the type of selected questions, I used the above charts to visualize the data, such as scatter and bar
3- I found the score is low when I used linear regression, so I tried another algorithm to improve the score of the model