Crowdfunding - Module 1

Data is great at telling stories about a group of information, this is the story of a data set on crowdfunding projects. Firstly, according to the data, there seems to be more crowdfunding projects during the summer on average every year. Secondly, the largest number of crowdfunding projects are related to theater, music, or film and video. These three categories account for nearly 70% of the data given for all the crowdfunding projects. Lastly, the data tells us that projects below a $15,000 goal are typically the projects that fail the most, are almost the most canceled, and usually all around the least successful projects.

With every dataset there are limitations on the data that we collect. For example, in this data set we only get a blurb of information about the certain projects, so we do not get the full information on what could have made one project successful and another similar project fail. Another limitation would be the sites used to gather the information, since we have information off Kickstarter and Indiegogo, it excludes data from other crowdfunding websites such as Patreon or GoFundMe for example which could broaden our dataset.

Another table or graph that we can make for this dataset to provide more information would be a project funding length graph. This would be valuable information for the dataset since it can show how long it usually takes a similar project to either end successfully or end failed. This could also provide insight on the average amount of time a failed or successful projects were active for, providing a sleuth of more information.

# Statistical Analysis-

Median would fit this dataset better for the number of backers. This is because mean would take high backer count and average it with the lower backer counts, which is not necessarily bad but since there are some projects with a very large number of backers, it skews the data noticeably. So going with median gives a better picture to our backer count for our successful and unsuccessful projects.

There is more variability in the successful outcome of projects in the dataset. This is because there are a lot highly backed projects in the successful projects that act almost like outliers that skew the variation greatly. On the other hand, failed projects usually have lower backer counts and there are not too many outliers with a lot of backers, although there are a few. So with these outliers, it means that you must move further away from the mean to capture a larger percent of the projects.