Matt Garcia

Data Analytics

Cohort 4

Homework 1

1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**

While campaigns categorized as Theater had the highest number of successes overall, campaigns categorized as Music had a greater proportion of successful campaigns. Relatively speaking the Music category was more successful than any other category.

Considering all categories and all years, campaigns starting in May were more successful than campaigns started than any other month. Campaigns starting in April had the least amount of campaigns ending in a failed status.

The size of the average donation is not a good predictor of the amount of money the campaign has raised relative to its goal.

1. **What are some limitations of this dataset?**

It is difficult to compare success across categories due to the varying number of campaigns in each category. While we can create a field and chart to look at the status of each category as a percent of the total number of campaigns within that category, we cannot make predictions as to which type of campaign will be most successful in the future. For example, though the Journalism category has the fewest successful campaigns, it also had no failures as well as the fewest number of campaigns in total. We cannot reliably say that a campaign categorized as Journalism will have the lowest chance of having a successful outcome because of the small sample size. Furthermore, the data does not contain reasons why the some of the campaigns were cancelled. If that information could be standardized in the data, it would make cancelled status more useful when analyzing the data set.

The dollar values of the campaign goal, pledge and average donation amounts are not in a single currency making comparisons using funding values difficult. However, this could be easily converted to a single currency.

1. **What are some other possible tables and/or graphs that we could create?**

We could create a table that displays the status of each category and subcategory as a percentage of the total number of campaigns in each subset. This could better illustrate the success of campaign categories relative to each other as it stands in the data. However, we would not be able to make reliable predictions as to which category will be most successful in the future due to the varying sample size of each category. This would best be represented by a stacked bar chart.

We could chart the percentage funded for cancelled campaigns and compare that with the percentage funded of failed accounts

We could create a bar chart to compare the campaign outcome status between countries.

We could create a table showing status outcome counts by average donation dollar ranges. This could help to answer if a higher average predicts campaign status outcome.