Scott Frazier

SMU Data Science Writeup-Excel Data

1.Purpose and Data Description. (What are you doing with what data):

This data represents data from thousands of Kickstarter projects and was put together to look for patterns or trends within it that can help clarify why some Kickstarter projects are successful and others are not.

2.Procedure: Data Cleansing (How did you alter the raw data given to you):

The data was altered in many ways from having it visually altered with conditional formatting to give the successful, canceled, failed, and live projects different colors to help the viewer understand trends within the data quicker. The data was also sorted from greatest to least in terms of percent funded for each project.

3.Procedure: Analysis Steps (How did you analyze the data):

The data was analyzed through many means, including conditional formatting and creating various pivot tables which helped with categorizing the large data set into a much smaller and easier to navigate data set based on certain criteria. This criteria for the pivot tables ranged from the status of the project, to the countries that each of the data sets was in to help narrow down which countries have the most successful and least successful projects.

4.Findings: 3 Conclusions from the dataset:

From the dataset, it is clear that the Kickstarter projects that are in the theater category, and the subcategory of plays have the highest success rates. From the dataset , we can also conclude that the projects that were funded in the months of February and May had the highest rates of being successfully funded.

5.Findings: Recommendation/Call to Action (What can I do with this analysis?):

With this analysis, if an individual is looking to either fund or begin a kickstarted program, then their best chances of having a successful project, or investing their money successfully, is to choose to begin or fund projects that fall under the theater category, and to do so in the months of February and May.

6.Limitations/Data Bias:

A major limitation of this dataset is that a lot of these projects have a goal of only $1, and are funded far beyond the $1 they are asking for. These projects in return will present data that is not helpful in what we are trying to analyze. This is because the data becomes skewed when such a small goal is surpassed by massive amounts of donations. It presents a false picture that people really want to fund this type of project based on the percent funded, when in reality, the financial goal the project wanted to begin with was very small.

7.Future Work (extra vizzes, predictive model, get better data):

If granted more time to analyze the data, the data could have been presented with more visualizations that help paint a more in-depth picture of what is happening in the dataset.