Brendan Zhu

Module 1 Challenge

Written analysis

From the given data in the pivot table/chart excel sheets for the challenge assignment, one conclusion that can be drawn about crowdfunding campaigns is that theater and plays, which are paired together for a parent category and sub-category, have the most number of successful campaigns. This is indicated by looking at the pivot chart height for theater and plays respectively. The height for those categories is the tallest compared to the other categories. Another conclusion is that journalism and world music have the least number of successful campaigns. The pivot chart confirms this conclusion due to the smallest height. A third conclusion is July has the most number of successful campaigns at 58 and January has the most number of failed campaigns at 36. Overall, this data is telling us that a high number of campaigns in theater and plays for the month of July will very likely lead to success.

However, there are certain limitations to the dataset. This large dataset is mostly gathered from the US, which isn’t very broad. Therefore, this doesn’t give us a true representation of what campaigns are the most successful in a worldwide perspective. If other countries are included, it can possibly show differing results, which can change our conclusions. A second limitation is the launch and end dates. The dates shown in the crowdfunding dataset take place in the 2010s but don’t consider the late 2000s. Data can be collected in those years to confirm whether campaigns have grown in success.

Some other possible graphs we could use are scatterplot and a clustered column bar graph. With more data from other countries, we can use a clustered column bar graph to represent the number of successful, lived, failed, and canceled campaigns for those countries and see the change when combined with our initial data. In addition, a scatterplot can show various data points for the late 2000s. We can use a trendline and observe the outliers as well as the new data points that closely resemble the data already given.

Statistical analysis

The mean better summarizes the data as a whole because it takes into account the average of all values in the dataset. On the other hand, median is the middle value of the dataset so it doesn’t consider every data point. The data shows that there is more variability with successful campaigns than failed campaigns. The variance is higher for successful campaigns. This doesn’t make sense because crowdfunding campaigns have been gaining more success and popularity lately.