Brendan Thomas

Module 1 Challenge

9/21/2023

**Written Report**

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + The amount of backers plays a significant factor in Successful vs Unsuccessful crowdfunding campaigns, the more backers the higher success rate.
  + Theatre/Plays has the highest popularity of crowdfunding interest.
  + Time of year has little to no impact on outcome as rates tend to stay the same.
* What are some limitations of this dataset?
  + The category themes are fairly saturated (plays, theater, film & video, music, television, documentary, in result this leads to increased competition and lower success rates for crowdfunding.
  + The years span between 2010-2020, therefore economic environments can impact the influence on project outcome (Successful vs Unsuccessful)
  + While it does provide country location, city/region can have a major impact on crowdfunding and donor behavior, therefore audience and target location specific are an important factor that is not listed.
  + Donation motivation for some categories vs others will vary based on popularity.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + Scatter Plots (w/trend line): Could help identify trends (positive or negative) between certain variables, such as funding goals and number of backers
  + More Data Tables (w/conditional formatting): Create Data columns formatted to highlight certain values within a successful project that exceeds or canceled/failed projects that underwhelm donation amount. Detect trends within that data.
  + I would like to create a histogram that compares the data on campaign length and success rates. Felt that was a key detail that I didn’t delve into more.

**Statistical Analysis Write Up**

* Use your data to determine whether the mean or the median better summarizes the data.
  + In this instance, the median serves as a more accurate metric to encapsulate the data regarding backers. It offers a more illustrative depiction of the distributions center, remaining uninfluenced by extreme outlier values that are present in both successful and unsuccessful campaigns
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  + After reviewing the data, it is clear that both the variance and standard deviation are higher for successful campaigns as compared to unsuccessful ones. Therefore this signifies that there is a much larger variability in the number of backers among successful campaigns compared to unsuccessful campaigns.
    - This makes sense why: First successful campaigns might have some that are more popular, in demand, and attract a higher number of backers, while others just reach or barely exceed the original pledge goal. Second there are also more significant outliers in the extremely successful campaigns that