**Module 1 Crowdfunding**

**Written Report**

**Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

* There is no correlation between the Spotlight and Staff Pick tags affecting the success or failure of a campaign.
  + Spotlighted campaigns only made up 26% of the total of successful campaigns in the dataset, while 28% of the failed campaigns were also Spotlighted.
  + Staff Picked campaigns accounted for 5% of the successful and failed campaigns, respectively.
  + This is a reliable indicator that the marketing on the platform is not doing enough to generate interest in any campaigns, and the likelihood of a campaign succeeding or failing relies entirely on the signal boosting of the business running the campaign or the strength of the backers generating word of mouth.
* Plays (and by extension, theater) represent 34% of the grand total of campaigns, creating an outlier in the dataset. This draws several conclusions:
  + Overall public interest in plays and theater is higher than the other categories.
  + Public theaters have always had a presence in society since well before the technological revolution, and this dataset shows that the transition to an online marketplace for funding these ventures was a both a natural, and very successful transition. And in many cases, public theaters have long-established groups of donors and fans, making it easier to signal boost their initiatives.
  + People are more willing to use disposable income on backing these campaigns.
  + The crowdfunding platform is more vulnerable to shifts in the Theater sector than any other. A future competitor creating a better platform for Theater campaigns could jeopardize the business model.
* Crowdfunding activity is highest in the summertime. While not a guarantee, by any means, the yearly and monthly trends show that July is the best month to have your campaign take place.
  + 10% of all successful campaigns take place in July, which is the highest of all months.
  + Likely from the dedication of backer resources in the previous month, August has a 3% drop in successful campaigns, which is the highest decrease from month-to-month analysis.
* Crowdfunding goals that are over $50,000 are significantly less likely to be successful, suggesting that this threshold is a good indicator of where consumer confidence begins to wane.
  + The $10,000 to $14,999 range of crowdfunding goals also yielded a higher percentage of failed campaigns to successful ones, however only 9 total projects fell into that category, which is too small of a sample size to draw any conclusions that this is a potential dead-zone for campaign goals.

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

* The dataset does not provide any information on backer counts for people who funded multiple campaigns, backers that are unique to only one campaign, as well as the first campaign a backer funded.
  + This information would go a long way towards uncovering trends in potential market stagnation for crowdfunding campaigns that are either being bankrolled by a few well-off backers or have high counts of unique backers.
* The dataset does not provide the maximum donation a backer contributed to each campaign. This missing information is the primary reason why there are so many outliers in the backer count analysis, and why the mean is less useful than the median for most of the backer count analysis.

**What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

* To better visualize the breakdown of campaigns by category, using a pie chart gives a better view of some key concepts:
  + Theater plays a large role in where the money is going for crowdfunding campaigns. If we consider that the total number of backers on the platform only have a set amount of disposable income that they are willing to contribute to a crowdfunding campaign, then seeing how much of the pie chart is taken up by the Theater category makes it easier to visualize this concept.
  + Using a pie chart also gives a better insight into picking out the next potential “Theater-sized” category in the future.
    - Rock, Web, and Documentary campaigns all have strong representation on this style of visualization, showing they have the potential to grow into an even larger market-share over time.
* A Stacked Bar chart is better for visualizing the thresholds that campaigns are meeting over time.
  + Stacked Bar charts are great for handling multiple categories that are being weighed against the same statistics, which is ideal for monitoring overall progress.

**Statistical Analysis**

**Use your data to determine whether the mean or the median better summarizes the data.**

* Due to the high variance in backer counts, the median serves as a much more useful data point when determining the likelihood of a campaign succeeding or failing.
  + The standard deviation of the number of backers in successful and failed campaigns (1266 and 960 respectively) are both greater than the mean number of backers (851 and 586 respectively)
    - For successful and failed campaigns, the total distribution within one standard deviation of the mean is 75% and 80% respectively, indicating that this is not a normal distribution.
    - The median backer count gives a better threshold to indicate whether a campaign is likely to succeed or fail.
      * The median backer count of successful and failed campaigns (201 and 115, respectively) represents exactly 50% of the total successful and failed campaigns, respectively.
      * When analyzing the backer count of successful campaigns using the median from unsuccessful campaigns, only 21% of the total successful campaigns met that backer count threshold, signifying the median backer count of successful campaigns is a potential indicator that the campaign will be successful.

**Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

* Successful campaigns will inherently carry a higher variance than unsuccessful campaigns for two main reasons.
  + The core requirement of a successful campaign is the monetary goal being reached, rather than the number of backers it took to get there.
    - Campaigns have no upper limit on the contribution someone can make, creating situations where the overall backer count on a successful project might be lower than the median backer count of unsuccessful campaigns, but either the strength of the donations was much higher, or the goal was not as ambitious as they might have initially thought.
      * Harmon-Madden was one such example, where the project only had 76 backers, but a low target goal ($700) and high pledge amount ($8262) create a large variance in the overall picture.
  + The other reason for this variance is crowdfunding campaigns continue to accept contributions even after the goal has been met. This generates an ever-escalating upper limit to the backer count from successful campaigns, where unsuccessful campaigns inherently have less backers due to the condition of having failed to meet their goal.
* This variability makes sense because the backer counts are directly affected by the scale of the contributions, and even though the variance is larger on successful campaigns due to the lack of an implied upper limit, it is also present in the failed campaigns because campaigns with large goals can still have a high backer count with very low contribution values (and vice-versa)