**Table of Contents:**

1. **Conclusions** 
   1. **From initial charts**
   2. **From outcomes by goal**
   3. **From Summary Statistics**
2. **Limitations of the dataset**
3. **Suggestions for additional graphs**

**Conclusions (from basic charts):**

1. Passively-consumed media is by far the most common campaign category (theatre, music, and film&video), so more competition for these categories, but no obvious correlation to success rate.
   1. Category largely doesn’t seem to matter to success rate.
2. Plays are the most common sub-category, so more competition in that space, but again this does not seem to correlate to success rate.
   1. Sub-category doesn’t seem to matter to success rate.
3. Campaigns launched in June and July are more successful than other months, while campaigns launched in August have the lowest success rate. Some of this is likely due to the fact that more campaigns start in June and July than other months, but this trend is still present on the combination graph of percent success by month and number of campaigns by month. This trend is more apparent for June than for July.
   1. For the best chance of success, June is the optimal launch time for a new campaign.

**Conclusions (from bonus):**

1. Outcomes by goal:
   1. Upon initial inspection, goals of $15,000 – $49,999 appear to have the highest success rate and as such the optimal campaign goal should be in this range. However, if you look more closely at the number of campaigns run with each goal, it is clear that there is not enough data to make appropriate analyses for goals in this range (the total campaigns run at each bin from $15,000 - $49,999 range from 7-14 campaigns, in comparison to 51 campaigns at the under $1,000 bin and between 231 and 315 campaigns at the remaining bins). The goal ranges for this dataset that provide each bin with enough data to perform a meaningful analysis are:
      1. Less than $1,000
      2. $1,000 - $4,999
      3. $5000 - $9,999
      4. Greater than or equal to $10,000
   2. Using these updated goal ranges, the success rate is highest at a goal of $1,000 - $4,999, making this the optimal goal range for a new campaign.
2. Summary statistics:
   1. The median summarizes the number of backers of successful and unsuccessful campaigns better than the mean, because the large number of upper outliers skews the mean to a higher number than is representative of the entirety of the dataset. The median is not impacted by these outliers and is therefore a better summary of the dataset.
   2. There is more variability in number of backers among successful campaigns. This makes sense because successful campaigns can be wildly successful, raising anywhere from 100% to, in this dataset alone, 2339% of their goal. However, failed campaigns can only range from 0% to 99.9% of their goal.
   3. Quantity of backers may not guarantee success; unsuccessful campaigns can still have many backers comparable to successful ones.

**Limitations of the dataset:**

* We don’t have any marketing information for successful or failed campaigns (how much spent, if any, on what platforms, and at what time relative to the start of the campaign). There is likely a large correlation between marketing and success.
* We don’t know how much of an audience the campaign organizers have. This is similar to marketing, but an organizer with a large following (such as Brandon Sanderson) is likely to be able to get a campaign funded without spending as much for marketing as they would without that pre-obtained audience. This likely also correlates to success, though may be inversely correlated to marketing spend.
* We don’t know which campaigns, if any, hired consulting support to run their campaigns. A consultant may have optimized campaigns before and be able to increase campaign success.
* We don’t know the history of these campaigns and their organizers. An initial campaign from an unknown organizer may perform worse than one run by an organizer who has run and been successful in previous campaigns. Similarly, an organizer who has failed previous campaigns may continue to be unsuccessful.
* We don’t know when within the campaign successful campaigns were fully funded, or how that relates to the percent funded above 100%. This data would allow for deeper analysis and stratification of successful campaigns and potentially provide insight into how to improve success rate.

**Suggestions for additional graphs and the value they add:**

* As noted in the bonus analysis section, a combination graph with percent outcome by goal and total campaigns by goal.
  + This graph would tell you which “percent successful” per goal data points are most meaningful and provide a better idea of the optimal campaign goal.
* A graph of the length of campaign vs success rate.
  + This should tell you whether there’s a correlation between length of campaign and success rate, and indicate an optimal length of campaign if so.
* A graph of average spend vs success rate.
  + This should tell you whether there’s a correlation between cost to back a campaign and success rate, and indicate an optimal price point for rewards if so.
* A graph of campaign end date vs success rate.
  + This should tell you whether there’s a correlation between end date and success rate, and indicate whether a campaign should ideally end by a certain time or month (i.e. whether a campaign should finish before August starts, since campaigns that start in August perform relatively poorly). This, in combination with optimal campaign length and start date, could narrow down the details of a new campaign.
* Combination graphs of the proportion of successful campaigns by category with the number of campaigns by category, and the proportion of successful campaigns by sub-category with the number of campaigns by sub-category.
  + This would provide a clearer idea of how success relates to category and sub-category.
  + This info is approximated by the stacked column graphs, but this particular configuration would give actual numbers to success percentage for each category.
* A combination graph of percent success, failure, and cancellation outcomes by start date and total number of campaigns by start date.
  + This provides a clearer picture of the correlation between campaign outcome and start date, removing the bias caused by the launch of different numbers of campaigns each month.