### **Crowdfunding Solutions**

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

* We can conclude that majority of crowdfunding campaigns are likely to be successful given that >50% have an outcome of successful – basis of this conclusion is that out of the 1000 campaigns, around 565 had successful outcomes and only 364 failed, the rest were either live or cancelled. Of all completed campaigns (except live i.e. 986) 57% of completed campaigns were successful, 37% failed and 6% were canceled.
  + We can also observe that for campaigns with goal of >$50,000, the fail rate was higher than success rate.
* Parent Category wise, we can conclude
  + success rate is highest in journalism (100%), followed by technology (62%) and film & Video (57%).
  + Failure rate is highest in games where nearly half the campaigns (48%) failed, followed by food 43% and theater and music both at 38%.
* Sub-category wise, we can conclude
  + Success rate is highest for audio and world music, however the number of projects is too small to conclude with confidence. Web projects are successful 71% of times, which is a good insight for backers.
  + Failure rate is more than or equal to 50% in some sub categories like science fiction at 64%, mobile games 62% and radio and podcasts at 50%, making these categories very high risk for backers.
* We can also conclude that number of backers is not an appropriate indicator of success or failure of a project based on statistical analysis. The variance of the data set as well as the standard deviation is quite high, there is an abundance of outliers as well based on the box plot.

1. What are some limitations of this dataset?
   * + - The fact that some projects have been cancelled or are live could skew the data. Reasons for cancellation are not mentioned, it could also include data of campaigns which were unlikely to succeed and were withdrawn preemptively to avoid being marked as a failed projects – this can potentially skew the category wise data. Some projects are still live and that leaves us with room for interpretation. The outcome from these live campaigns could again change the summary of the sub-category wise data set, if not the entire data set.
       - The data doesn’t share any information on drivers of success or failure. If there were additional data points like marketing budgets, social media platforms targeted, or budgets by platforms, or the management team behind the campaign, more conclusive insights could have been drawn.
       - The two fields of staff pick and spotlight are Boolean, however, there is lack of description of what these fields refer to for the data analyst and makes it a set of redundant data which is not adding any value to analysis due to lack of context behind the data provided.
       - The goal and pledged data for campaigns in this data set are in different currencies and does not make comparisons like to like, which limits the value that can be drawn from analysis.
2. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   * + - A table with duration of the campaign can give additional insights into success or fail rate. This can be done using date created and date ended conversion data. It may be possible to see some relationship between duration of campaign with success, a campaign that went on for too long or too less time could point to success or failure. Mean and median of such data could also point to what should be the duration of campaigns for the teams launching the campaign.
       - Creating tables on percent successful/fail/can/live for each category and sub-category can be done. This will yield insights on which categories of campaigns are likely to be successful or fail. *(I have already used this data for my conclusions and analysis).*
       - We could draw a table and plot a chart which compares number of backers with percent funded for each campaign. This could help establish a relationship with whether campaigns which are higher funded backed by more investors or not. A gut assumption could be that highly funded projects (say >200% funded) may have more backers, but data analysis may sometimes give opposite results and prove such gut assumptions wrong or right.
       - We could add a table to convert all data in same currency and then compare the goal ranges with the outcome, this could also point to better relationship between goal ranges and outcomes as it will be a like to like comparison.
       - We could use box plot of number of backers vs. successful campaigns and failed campaigns (already included in my excel sheet under summary table), this could help establish relationship between whether the number of backers is a useful data point for a campaign to be a success or not.

Bonus Solution:

### Mean seems to provide a better summary for the data set here compared to median, given z score analysis, most of the data points have a z-score less than 1, which means that the distance of these data points from mean is less than one std deviation away to the right or left of the mean. Having said that, we cannot ignore the reasonable number of outliers in this data, which suggests that neither mean nor median are a great explanation for this dataset and should be used with caution and in conjunction with deeper analysis if at all.