Erik Dubec

Homework 01 Report

1. What are three conclusions we can make about Kickstarter campaigns given the provided data?
   1. The percentage of successful vs failed/canceled Kickstarter campaigns varies heavily by sub-category, and somewhat less by parent-category. To me this suggests that parent-category is not particularly useful for an evaluation except as a control.
   2. The percentage of successful vs failed Kickstarter campaigns is very similar between a goal range of 10000 and 45000. Before that point, successes are more likely, and afterward failures are more likely. Canceled projects increase (appears linear) as the goal range increases.
   3. There is a larger difference in the number of successful and failed Kickstarter campaigns created near the beginning of the year than there is in those created near the end of the year. In general, there are more successful campaigns than failed campaigns throughout the year.
2. What are some of the limitations of this dataset?
   1. This dataset is not complete. We are only working with around 4000 Kickstarter campaigns, whereas the Kickstarter website states that there have been over 425000 projects launched since its inception. This effectively means that I have no idea whether this dataset is even representative.
   2. We don’t have a way to easily check how much attention each Kickstarter campaign received. That is, some campaigns probably had far more people viewing them than did others and that may not translate directly into number of backers. Additionally, some campaigns were likely advertised outside the platform, leading to an increased number of backers, or again, simply people looking into the campaign.
   3. We don’t have access to any information regarding Kickstarter backer bonuses. I expect that many Kickstarter backers will donate at one backing tier or another in order to both help fund the product and obtain whatever sort of reward is listed for that tier. Having access to these would prove helpful in determining what sort of rewards should be considered when creating a Kickstarter campaign (incidentally I would also like the number of backers at each backing tier, otherwise we would have to do some weird guesswork with the average donation and number of backers).
3. What are some other possible tables/graphs that we could create?
   1. We absolutely need a graph of Campaigns by Goal Amount vs Sub-Category, a scatter plot to show us generally which sub-categories we would expect to see at each goal range. It would be even better if it could be made to color successes, failures, and cancels differently, ideally having a filter so we can display them as needed.
   2. Some graphs that I would like to have but I don’t feel as strongly about would be something like a Outcomes by Campaign Length table and an Outcomes by Number of Backers table, where the campaign length and number of backers are each in ranges so as not to have a massive unreadable table.