Benoit Viollier

Data Analytics

Homework 1

03/08/2020

Questions:

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

Given the provided data, we can see that a little under 60% of startups between 2009 and 2017 succeeded rather than failed. A lot of these startups are part of the theater industry. Most of these startups are in the United States. 2015 was the year with the most kick starters that succeeded. As well as, it seems they do better at the beginning of the year rather than the end.

1. What are some limitations of this dataset?

Depending on where the information is coming from, the sources of this information could be stretching their truth a little. This could mean that some of the data is more inaccurate and therefore validity could be somewhat false. Another challenge this dataset faces is it could be missing data. After doing a little research, it seems that there were a lot more than approximately 4000 Kickstarters during that period of time. This could be a limitation since we are not getting accurate numbers.

1. What are some other possible tables and/or graphs that we could create?

We could create many different graphs that could represent this dataset. A bar graph could represent the difference in successful, cancelled, failed and live startups. So can a pie chart. The pie chart could let a user analyze the percentage that each section is composed of. A line chart can give you year by year comparison of the difference in successful and failed startups.

Statistical Bonus Question:

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

The mean summarizes the data more meaningfully. The mean gives the analyst the average backers it took a successful Kickstarter to succeed. The median simply gives the middle number in the data set. This means that the successful Kickstarter in the middle of the dataset had 62 backers.

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

There is more variability with successful campaigns. I believe that this makes sense as a dialed campaign will always have a lower number of backers. This means that there will be more of a congregation of data in the smaller numbers. Whereas, for a successful campaign, there is the chance that one person can back the full amount to a lot of backers backing the minimum amount. A lot more variability.