Robert Simeonoglou

Data Science Bootcamp

Homework 1

1. Three conclusions we can draw from the KickStarter campaigns are:
   1. Most KickStarter campaigns are in the theater category, especially in the plays subcategory.
   2. According to the sample size, there is a higher rate of successful campaigns than there are failed campaigns
   3. There are more successful campaigns in the beginning half of the year than there are in the second half of the year
2. A few limitations of the dataset are:
   1. Is this the whole population of the kickstarter campaigns or just a sample size?
   2. Are there requirements for projects to become campaigns?
   3. Marketing done by campaign for backers not shown
3. A few other graphs that would be helpful for analysis would be:
   1. line chart of state of campaigns compared to the amount of backers
   2. Map chart/Pie chart of state of campaigns vs country
   3. Line chart of state of campaigns vs Goal (lower the goal the more successful?)
   4. Column graph of state of campaigns compared by year

Bonus

1. In this dataset, the median is more meaningful, as there are very large outliers in both Successful backers and Failed backers that is affecting the mean value greatly, skewing the mean.
2. There is definitely more variability of successful campaigns when comparing to the failed campaigns, as you can tell from the standard deviation. This however makes sense, because the more backers, the more likely a campaign is to reach its goal. Successful campaigns ranged from a low number of backers with large contributions, to lots of backers, causing a large amount of variability. However, failed campaigns mostly had a low number of backers, including no backers at all. This kept the variability lower and in turn a smaller standard deviation.