Penn Data Science Bootcamp
Excel challenge
Clara Bucar

1. Given the provided data, what are three conclusions we can draw about crowdfunding campaigns?
   a. From 2010 to 2016 the proportion of failed and successful crowdfunding campaigns remained steady; starting 2017 more successful campaigns were identified and less failed ones.
   b. The three categories that dominate crowdfunding campaigns in the set are theater in first place with a large margin, more specifically plays, followed by music and film & video almost tied in second and third places. Journalism was the least prevalent category of crowdfunding. For subcategories, metal and radio & podcast were the least prevalent.
   c. The total dollar amount pleaded in successful campaigns is largely a reflection of the total amount of campaigns per subcategory, with the exception of documentary, videogames and foodtrucks. For those three subcategories the total dollar amount was, on average, below the average for the others.
   d. The US represented roughly 75% of total of campaigns, total of backers and total of pledged amount.
2. What are some limitations of this dataset?
   a. There is not enough data for the year of 2020 to be included in the analysis
   b. The data for journalism crowdfunding campaigns could have been insufficient due to low sampling
   c. The data gathered most likely focuses in the US, with 76,3% of all crowdfunding campaigns. The remaining 23,7% of the data is divided into six other countries. On a separate note, this could display a cultural difference and/or population size.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   a. Count of outcome per country: how the data spreads across countries
   b. Qty of backers per country: to rule out possibility of cultural difference influencing campaign outcomes
   c. Count of successful campaigns in comparison to total amount pledged: to understand which subcategories generate more cashflow on average than others

Bonus:

The mean represents the center of the data better than the median for this set

The failed campaigns have more variability than the successful campaigns

backers successful

backers failed