Mahsa Bakhtiari Mehmandost Olia

Module one

weekly challenge

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
* The chance of success of all the campaigns is around %50.
* The majority of campaigns are in the Plays category, which is a sub-category of theater.
* Around summertime, there is a slight spike in the launch of new campaigns.
* What are some limitations of this dataset?

It would help our analysis if we could get these additional data points:

* Geographic distribution of backers.
* the percentage of people who visit the campaign page and then go on to pledge.
* Where backers found out about the campaign, such as social media, email marketing, or word-of-mouth.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
* A pivot table with a “year of campaign lunch date” row, & outcome column & value. So we can analyze the relationship between the launch year & outcome to see if there is a trend.
* A pivot table with a “year of campaign lunch date” in a row, & pledged in column & value. So we can analyze the relationship between the launch year & amount of money that has been raised.
* A pivot table with pledged row, an outcome column & outcome values to see how the amount of total pledged affects the outcome of a campaign.
* A pivot table with parent-category row & pledged in column and value so we can evaluate which categories can raise more money.
* We can add a column containing the percentage pledged to the goal. Then make a pivot table and see how the difference between a goal and a pledge affects the outcome of a campaign.
* Use your data to determine whether the mean or the median better summarizes the data.
* Since the data in both cases (failed/successful campaigns to the number of their backers) is skewed, and there are a lot of outliers, the mean cannot be a good factor in summarizing the data set because it is sensitive to the extreme values, But in a data set with these mentioned characteristics, the median is a better option to show a "typical" value of the data since it is not affected by outliers.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
* The variance in the number of backers of a successful campaign is slightly bigger than failed campaigns. It doesn't help to reach a clear relationship between the number of backers with the outcome of a campaign because the distribution of the number of backers for both successful and failed campaigns are close.