Excel Challenge HW

Amber Pizzo

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

* Conclusion 1: Theater has the most Kickstarter campaigns overall, and more successful campaigns than the number of total campaigns for any other category.
* Conclusion 2: The US alone is responsible for ~73% of all campaigns, and also has the most successful campaigns, at ~76%.
* The years for the highest numbers of campaigns were 2014 and 2015. However, while these 2 years showed the most campaigns total, only about half of the campaigns were actually successful. The years leading up to 2014 (2009 – 2013), there were significantly less campaigns launched each year, but about 75% of them were successful.

1. **What are some limitations of this dataset?**

* Location information is lacking – only 21 countries shown. Also, no cities are listed. For example, the US comprises about three quarters of the campaigns but there is no way to tell if there is a particular city/state component.
* No information on how quickly successful campaigns reached their goal. Over half the campaigns were over 100% funded. It would be interesting to see what types of categories reached their goal the fastest.
* No information about where the backers are located. This could be used to determine what categories backers in different countries are most/least interested in.

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

Some other graphs/tables that could be created are:

* Line chart - Average successful/failed campaigns by year (instead of by month).
* Bar graph – Average successful/failed campaigns by duration (launch – deadline).
* Pivot table showing number of backers (in sum field) for each category (in row field) with state (in column field). This can tell you what categories get the most/least attention from backers, as well as how many backers were part of successful, failed, or canceled campaigns.

**Bonus Statistical Analysis**

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

From the data, it can be determined that the median summarizes the data more accurately. The reasoning here is that the mean of any data set is very susceptible to outliers, while the median is robust to extreme outliers. From both the successful campaigns data set and the unsuccessful campaigns data set, it can be seen that there are a large number of outliers. This is shown through the use of a box and whisker plot for both populations of data. Due to this large number of outliers, the mean will be affected by this, while the median will not be as affected.

1. **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. This can be determined because the variance is much higher for the successful campaign data set. As variance gets farther from 0, it means the data is more spread out from the mean and other numbers of the data set. This does make sense. When looking at the successful campaigns, the number of outliers is higher than for unsuccessful campaigns, and the outliers are much more extreme.