Module 1 Challenge

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. Using the data sample, it is clear to draw up three conclusions in regard to crowdfunding campaigns. Firstly, it is clear to see from the “Parent Category Chart” that there are far greater successful campaigns (565) as opposed to failed campaigns (364). This shows that businesses have had success using crowdfunding campaigns.
2. The data shown in the “Sub-Category Chart” shows that campaigns involving the subcategory “plays” have had the most success as they hold the highest amount of successful campaigns (93,360) as opposed to “mobile games” which has had a significantly low success rate (824) with a high amount of cancelled campaigns (4,595)
3. The Yearly line graph suggests that March was the most successful month for campaigns on crowdfunding campaigns as it received the highest amount of success (59), The lowest performing month being November at 33 successful campaigns with 30 cancelled campaigns.

* What are some limitations of this dataset?

1. A limitation is the sample size. With only 1000 samples we are restricted to the information that is given to us, potentially resulting in a prediction that is not entirely accurate.
2. Another limitation is the validity of the data, we do not find out how some campaigns fail and so can only make assumptions as to why they failed or cancelled, this disables us from being able to make predictions to help future campaigns in not making the same mistakes.

* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

1. A way we could present this data set is in the form of scatter plots. A scatter plot could be drawn with the campaign goal on the x-axis and the pledged amount on the y-axis. The outcomes could then be color coded in relation to whether it was “successful”, “failed”, “live” or “cancelled”. This could help us better determine the relationship between the outcomes and the how much was pledged for that particular campaign.
2. Another way the data could have been presented is by using a bar chart, for this we start by calculating the average amount pledged per backer for the amount of successful campaigns compare that to the average of unsuccessful campaigns. A bar chart allows us to see a side by side comparison of the data and determine how much backers affected the data.

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

The average number of backers for successful campaigns is 851, whilst the mean number of unsuccessful backers is 586. This suggests that the campaigns with more backers proved to be more successful. This is also reflected in the median results as the median for successful backers is 201, which is significantly higher than the unsuccessful median which is 114.5. From this I can determine that the mean is the better representative of the data as it collates all the backers who were both successful and unsuccessful as opposed to the median as it only obtains the value in the middle which does not represent the data.

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

It is clear to determine the difference in variability between the successful campaigns (1,603,373) and unsuccessful campaigns (1603373) as successful campaigns are significantly higher than the unsuccessful campaigns. The minimum number of backers in successful campaigns is 16 and the maximum number of backers is 7295 resulting in a very high range of backers. The standard deviation of the successful campaigns is 1266.243, which indicates that data points are spread out over a large range of values. We see a change when looking at the data for unsuccessful campaigns as its variability is considerably lower at 921574. This makes sense as we know from looking at the data, there were a considerably greater amount of backers for successful campaigns as opposed to unsuccessful campaigns resulting in greater variability