**Week 1 Homework**

1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**
2. The less amount of goal is set, the higher percentage of successful is reached. Higher amount of goal leads to higher percentage of failed and cancelled Kickstarter campaigns.
3. The top 3 most successful categories/ sub-categories are: Theatre/ play, music/ rock, film&video/ documentary, however for play campaigns, the failed cases are also high, rock and documentary have no failed cases. Thus, people may prefer to support rock and documentary campaigns.
4. On average, the Kickstarter campaigns start in May have highest successful counts and start in December have lowest successful counts. Thus, if start campaign in May is more likely to success than start in December.
5. **What are some limitations of this dataset?**
   1. This dataset has a limitation of its sample size, only 4114 samples may cause bias of the results, if increase sample size, more reliable and accurate results can be obtained.
   2. This dataset does not include Asia/ Africa countries, due to culture differences, the successful results may change for different category and sub-categories.
   3. Most of the backers should have certain level of economic income, therefore for those young teenagers or some older people who will not pay attention to these Kickstart campaigns, their preference for different campaigns may cause different results.
   4. As all campaigns have a time duration, some campaigns maybe more popular in a certain year/ time period, and before or after that time, it is no longer popular and is more likely to fail or cancel.
6. **What are some other possible tables and/or graphs that we could create?**
   1. Can create bar chart for different category/ sub-category and the average amount of donation to see which kind of category is more popular.
   2. Line chart of years VS counts of different states of categories to see popularity of different categories in different time periods.
   3. Scatter plot to see the relation between years and successful counts.

**Bonus Statistic Analysis Question:**

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

Based on data, median is more meaningful to summarise the data, from the whisker chart, mean is out of the range for both successful and unsuccessful campaign, and the variance & standard deviation are both high which indicate mean is not reliable.

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 as variance and standard deviation are both higher than unsuccessful campaigns, this is making sense because there is more successful campaigns (higher sample number) than unsuccessful campaigns.