**Excel Assignment Report – Anh Huong**

**Questions & Answers:**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   * There are totally 9 main categories and 42 sub-categories, and 53% of those projects are successful. Only US participates in everything with the highest total pledged of $36,182,653 USD amongst 21 participated countries.
   * “Theatre” is the highest category with 1393 projects, but the success rate is only 60%. Top 4 countries show 100% success rate in “Theatre” are: CH, ES, LU and SG. The second highest category is “Music” with 700 projects, and it has the highest success rate of 77%. The top countries with a 100% success rate in “Music” are: AT, AU, DK, ES, IT and SE.
   * Even though US “Music” only has 77% success rate but US “Rock” seems be dominated around the world with the highest pledged. 92% US “Rock” projects in category “Music” show 100% success rate with the total pledged of $1,502,511 USD.
2. What are some limitations of this dataset?

* The data for the currency of the pledges and goals are based on the local country’s currency. It does not tell the true story when calculating the percentage funding and donation as there’s assumption that all currencies are having the same value.
* Not all countries are on the list

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

* Convert all the currencies into the same unit, e.g., USD, by using the latest currency exchange rate (refer to worksheet “Latest Currency Rate” & worksheet “Compare Pledged”)
* Each country pledged for every category & sub-category with its status
* Make/loss money (i.e., good/bad investment) based on the “state” for each country, category and its sub-category

**Bonus**

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

The median is more preferred to summarise the data than the mean as there are too many outliers around the mean (refer to the graph on sheet “Bonus Stats Analysis”). There are significant gaps between the minimum, mean and maximum values for both successful and unsuccessful campaigns.

1. 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 observed that excel treats both sorted and unsorted data the same. The calculation of the median, standard deviation, variance and interquartile range (IQR) give the same results.

IQR for the unsuccessful campaign is 112. The value spans from 1 (minimum) to 26457 (maximum) with median is 64, variance is 712841 and standard deviation is 844.

IQR for the unsuccessful campaign is 11. The value spans from 0 (minimum) to 1293 (maximum) with median is 4, variance is 3773 and standard deviation is 61.

Both datasets have very high variability, values in the datasets are spreading out too much. The number of outliers in the graphs in the worksheet “Bonus Stats Analysis” also visually demonstrate the data spreading.