Report Homework 1

Jose Luis T. Duarte Alcántara

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

* The campaigns related to the "plays" subcategory are the most successful projects with only 33.7% of failed campaigns.
* The campaigns related to the sub-categories "documentary", "hardware" and "indie rock" have been 100% successful.
* The best success trend occurred in the first quarters of the year, the last quarters of the year the trend of failure of the campaigns rises.
* The categories of the least successful campaigns throughout the year are:
  + Food
  + Games
  + Publication
* The categories of the most successful campaigns throughout the year are:
  + Film and video
  + Music
  + Theater
  + Journalism
* This allows us to identify that the campaigns on the themes of entertainment and arts were quite successful and when they undertake campaigns of another type, most are unsuccessful and a few mixed (50% successful and 50% unsuccessful).

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

* To identify what elements, make a campaign successful or not, it would be convenient to have the following information:
  + Executive or head of the campaign
  + Number of impacts and frequency of the campaign
  + Target market of the approved campaign by country
  + Duration of each campaign

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

* Graphs showing correlations between the variables and future projections based on some regression that best fit the main variables would be useful.
* Table with only the main factors that determine the response variable "goal"

Bonus Statistical Analysis

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

In the Box Plot graph in the <StarterBook-JLDA.xlsx> *Sheet3*, we can see that de median is near to the low edge of the box not in the middle (that’s means the median is below the mean) in both cases, successful and unsuccessful campaigns.

We can confirm that the median has negative asymmetry, therefore is not representative to the data because these its concentrate in the upper part of the distribution.

Also, we can see in the same graph (box plot) a lot of outlier points, that’s explain why we have the median below to the mean

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

In the Box Plot graph, we can also compare the height of the boxes, for successful (blue) and unsuccessful (red) campaigns. The box for unsuccessful is shorter than the box for successful, that means there is less data variability for the unsuccessful data set compared to the other.

We also observe, although both data sets have outliers, 75% of the unsatisfactory data are closer to the mean and median (short box) compared to the successful data set, where 75% of them have a wider range (large box).

When we compare variances and the rest of the statistical measures (mean, median, maximum, minimum, etc.), we can see clearly obvious how different are both datasets, and explain why unsuccessful campaign data are more consistent.