## **Challenge 1 – Crowdfunding Report – Yasar Sabir**

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  1. Crowdfunding campaigns in the arts (“theatre” and “film”) and “music” have more **absolute** successful campaigns, although
  2. Proportionally campaigns in “technology” have more **relative** success
  3. The subcategory of “plays” held both a large proportion of campaigns within the subcategory and also was the largest subcategory with the highest proportion of successful campaigns
* What are some limitations of this dataset?
  1. The largest flaw in this dataset is the simple assumption that “failure” is consistuted by not hitting the target goal of the funding campaign. For example data row 63 (value 61) (Romer-Hoffmans project called open source zero administration), raised over $185k USD just shy of their $199k target with over 2000 backers. That is a fantastic amount raised and a fantastic number of backers. Just because the funding goal was not reached, was the project itself successful. We do not know. Also was this a failure in itself in terms of a funding goal, it appears to be a great success. Perhaps the parameters for what constitute failure need to be adjusted it terms of the absolute target?
  2. The data includes crowdfunding from the United States, Australia and Europe. However, from Europe we only have representation from Italy (IT), Denmark (DK), Britain (GB)and Switzerland (CHF) with little representation from other European countries. The rest of the world is also not represented. There this data is not truly representative of the regions it represents.
  3. There is a paucity of data from 2020 and there is no data following this year, therefore the data must be interpreted within its historical context and may not represent findings currently.

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

* The average amount per backer is given in various difference currencies. We could use formulas to convert all this data into one single currency e.g. USD based on the historic exchange rate at the time of the particular project. This would allow us to truly compare the amount each project was given.
* We could correlate whether increased funding/donations were related to more successful campaigns using a scatter plot and regression analysis.
* We could look at how successful campaigns were based on how high their finding target was i.e are those that aimed for higher amounts more successful?
* We could analyse how many backers there were for successful vs unsuccessful campaigns.

## **Data analysis of successful vs failed campaigns.**

|  |
| --- |
| Which statistic summarises the data best? |
| This is a difficult question as there are arguments for and against both the mean and the median to represent the data. The mean shows us that on average around 800 currency is given per project, however because there were extreme values (outliers), which lifted the average, the median result of around 200 currency was more representative. In conclusion the median is a better summary, but I think in actual you need to look at both statistics together to better understand the data. |
|  |
|  |
| Is there more variability with successful or failed campaigns? |
| The standard deviation is slightly more with successful campaigns, this is probably because those projects that develop inertia are more likely to get more publicity and actually raise higher amounts thus increasing the variance and SD. It is actually a surprising result considering that many of the failed campaigns attracted no funding at all so I would have expected high variance in the failed campaigns however overall the failed campaigns didn’t manage to attract the higher amounts of funding that the successful campaigns did , hence why the latter had more standard deviation. |