Excel Challenge Notes & Analysis:

**The spreadsheet has been broken up into following sheets:**

* *Kickstarter Data*
  + This contains the raw data from Kickstart, we have also spilt out other information from the raw information in columns P to T
* *Category Analysis*
  + This contains pivot table from data tab show a break of information by broad category and country Slicer has been added
* *Sub Category Analysis*
  + This contains pivot table from data tab show a break of information by specific category and country & main category slicer has been added
* *Date Created Conversion* 
  + This show a down of success and failure on a annual basis and a Year slicer has been added
* *Goals Analysis* 
  + *Contains bonus analysis answers*
* *Stat Analysis* 
  + Shows mean,median, stdev info

**Conclusions:**

* We can see that annual activity for Kickstart was very low for years 2010 to 2014. It saw its active peak in 2015 and progressively decline year on year afterwards.
* The most successful campaigns were:
  + Theatre total Campaigns 1393 with success rate 60%
  + Music total Campaigns 700 with success rate 77%
  + Technology total Campaigns 600 with success rate 34%
  + Film total Campaigns 520 with success rate 57%

Further breakdown of the sub-categories shows that there have not been any campaigns labels as Rock Music that have been unsuccessful by either cancellation or failure. If you were to target a new campaign you would have best chance of success if it can be categorised as Rock Music based on the data at hand.

* Data shows that in the from 2010 to 2017 the have been no successful journalism campaigns. It may be that Kickstart demographic are not respective to invest or see no value this medium/category. We say this as review of sub-categories for Film (the 3rd most success by %) shows that Documentaries we the most successful funded. While at the same time Pop Culture type such as Scifi, Drama and animation saw 0 success. If you were wish to launch a Journalistic campaigns our data suggests that you should reframe into a documentary to maximise the chance of success

**Limitations:**

* The data does not provide any information is the campaigns was successful delivered to the backers
* Data does not identify individual group the create campaigns we therefore cannot see one off success or multiple successes.
* Data on Backers is aggregated therefore cannot glean any information type of backers or their possible drivers.

**Other Analysis:**

Some other analysis that could be undertaken with data at hand. Is that we could review the average donation size against its success with an category. To determine how much a back will likely to invest in a given campaign and you could therefore use this information to modify any future campaign to make them more attractive

The same information could also be used to evaluate is backer is different country are happier to pay more and another country

Also we have not compared or analysis any information about the relationship of success to if the campaign were marked “Staff Pick” or “Spotlighted’ and if this had any bearing of success or failure

Bonus Statistical Commentary:

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

|  |  |  |  |
| --- | --- | --- | --- |
|  | Overall | Successful | unsuccessful |
| Median Backer | 25 | 62 | 3 |
| Mean Backers | 112.60 | 194.43 | 19.49 |
| Min Backers | 0 | 1 | 0 |
| Max Backer | 26457 | 26457 | 1501 |
| STD Dev Backer | 623.51 | 844.30 | 72.41 |
| Variance of Backers | 388861.88 | 713167.38 | 5246.47 |

The distribution for Successful campaigns is skewed by small number of significantly large backers. On spreadsheet a chart has also been created that clearly demonstrates this. But indicator and the large Standard Dev and Variance score shown. This show that the median would be a better measure to summaries the data than mean.

The distribution of unsuccessful campaigns while also appearing to show an exponential trend. It is a more confined range/variance and it is also noted that this data sees a number of 0 backers scores. This contributed to a very low Median and we therefore believe the mean is a better measure to summaries this data.

*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 significantly more variability in the successful data verse the unsuccessful data. That does make sense because once a campaign is successfully there is not upper limit on backers and project can be overfunded many times. But on the other side Unsuccessful campaigns will always be confined and are bound to minimum number of backers being 0 and the maximum number of backer being one dollar less that their target. While is might be useful in successful campaign data to under how many where wildly successful or couldn’t be applied in the same way to unsuccessful data.