**Maria Paula Ardila**

**Challenge 1 – Excel**

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

**CONCLUSION 1: BASED ON THE PARENT CATEGORY**

* 56.5% of the crowdfunding campaigns were successful, 36.4% failed, 5.7% were canceled and only 1.4% are still live.

**A screenshot of a spreadsheet

Description automatically generated**

* The theater category not only has the highest number of crowdfunding campaigns, 344 (34.4%), but it also has the highest rate of successful campaigns, 187 (18.7%).

**A graph of different colored bars

Description automatically generated**

* 76.3% of the crowdfunding campaigns happened in the US, and Theater was the most used campaign with 273 initiatives (35.78%) and the most successful with 149 initiatives.

**A screenshot of a computer

Description automatically generated**

**CONCLUSION 2: BASED ON THE PARENT SUB-CATEGORY**

* The sub-category analysis is only relevant for categories such as Film and Video, Games, Music, Publishing, and Technology.

**A screenshot of a computer

Description automatically generated**

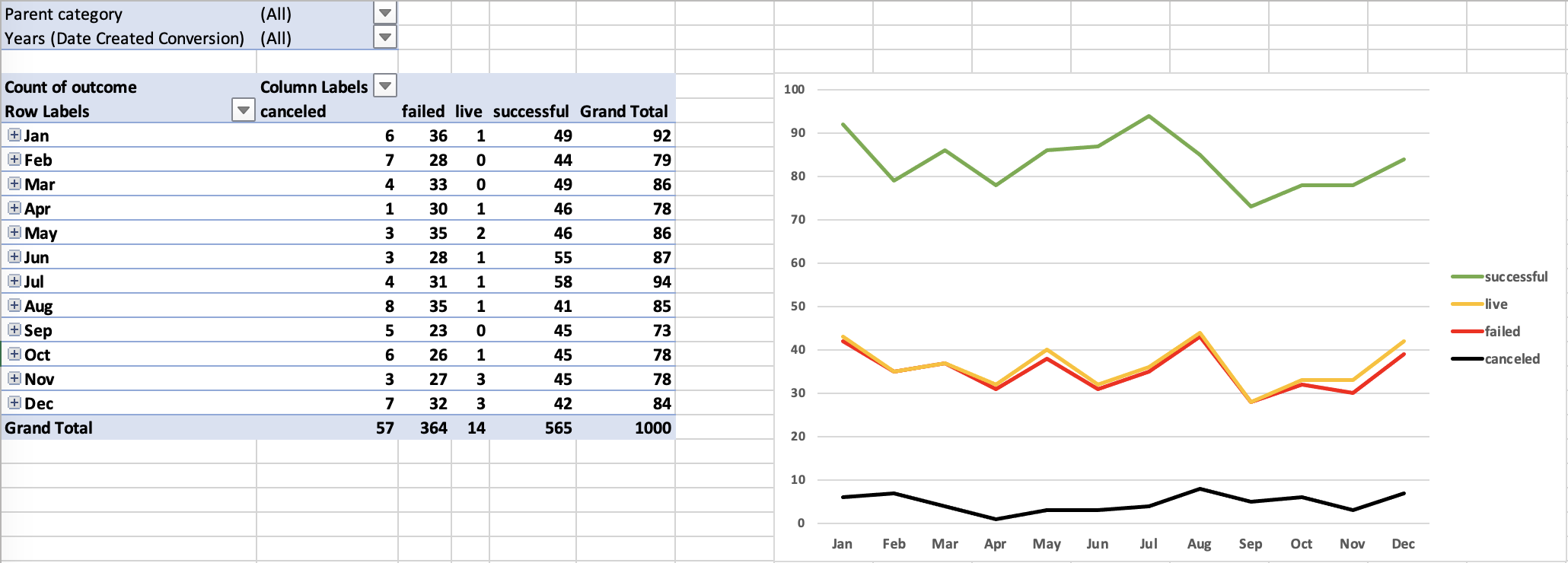
* The data matches the category analysis for the most relevant category, Theater, as no sub-categories are associated with it. 344 (34.4%) of the crowdfunding activities happened during a play, and the rate of success is the highest, with 187 plays (18.7%).

A screenshot of a graph

Description automatically generated

**CONCLUSION 3: BASED ON THE OUTCOME AND DATE**

* Throughout the year, the number of crowdfunding activities per month varies between 73 and 94, with September having the lowest number and July having the highest number. This shows a constant and continuous effort in crowdfunding.



* Throughout the year, the number of crowdfunding Theater activities per month varies between 23 and 35, with February having the lowest number of activities and October having the highest number. This shows a constant and continuous effort in Theater crowdfunding initiatives.

A graph on a sheet of paper

Description automatically generated

* The database contains data from 2010 until 2020. However, the data for 2020 is only for January. Analyzing the number of crowdfunding campaigns from 2010 until 2019, the number of annual initiatives went from 84 in 2012 to 108 in 2010.

A screenshot of a spreadsheet

Description automatically generated

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

* It will be interesting to know what the crowdfunding was for: to raise resources for the development of the product, to get an initial prototype, to do massive production, or to raise resources for marketing related to the product.
* It will be valuable to know the type of product related to the crowdfunding campaign. It could be related to technology, beauty, furniture, or food. Every category can show different behavior, and some can be more successful than others.
* It would be good to know the demographic attributes of the person issuing the crowdfunding initiative and those who donated to it.
* Nowadays, and considering that it is related to celebrities and content creators “only,” data about their followers on social media can strongly correlate with the success of the campaign.
* Not knowing the structured of the donation is a limitation of the dataset. For example, was it an open amount, or was it fixed to be multiples of 10, 25, 50, or 100? How open people are to donate a minimum of 5 or a minimum of 100 dollars can make a big difference in the success of crowdfunding.

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

* For the above analysis, I created a pivot table of category and subcategory to understand the subcategory's value better. My conclusion was that it is not that relevant to the analysis as Theater Crowdfunding, the most successful campaign type, can only happen in a play.
* I also created a year pivot table to see if the number of crowdfunding activities had a steady trend throughout the years. By doing so, the 2020 data, as it is only for a month and apparently for only part of the month, will be the outlier of the database.
* It will be important to run pivot tables and charts that include the goal amount; maybe Theater is only the most successful campaign type because the dollar amount is the lowest.
  + It will be interesting to understand how many donations made that campaign successful. The effort differs when the goal is achieved or exceeded with fewer than many donations.
* It is also important to analyze the time delta of the campaigns to understand which ones are successful. The amount of time you have to crowdfund the money and achieve or exceed the goal is a valuable metric for understanding success. Could the time frame of the initiative impact the level of success?

1. **Use your data to determine whether the mean or the median better summarizes the data.**

* The data for the number of bakers on successful and unsuccessful crowdfunding campaigns are right-skewed, which means they are not normally distributed, so the mean is not a good measure of central tendency, and the median should be used instead.
* One reason these two sets of data are right-skewed is that the number of donors cannot be less than zero, but there is no upper limit on the number of people donating to a campaign.
* Moreover, the histogram analysis shows that most of the records are toward the lower bound. For both sets, 67% or more of the records land in the first bar of the histogram, which also explains why both sets are right-skewed.

**A screenshot of a table

Description automatically generated**

**A graph with green bars and numbers

Description automatically generated** A graph with red bars and numbers

Description automatically generated

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

* The dataset of successful campaigns shows more variability than the dataset of failed campaigns.
  + The interquartile range (length of the box) and the lower and upper bound (length of the whiskers) give the box and whisker plots of the successful campaigns a wider data spread than those of the failed campaigns.
  + *The lower bound shown on the chart is based on the formula. However, you cannot have less than zero donors on a crowdfunding campaign.*

A screenshot of a table

Description automatically generated

A graph of a graph with red green and blue squares

Description automatically generated with medium confidence

* The greater variability of successful campaigns makes sense because, although most had a goal between $1,000 and $9,999, the next niche of success is for crowdfunding initiatives with a goal greater than $50,000. Not knowing if there was a minimum donation amount set or a limited number of donors, you would expect that the higher the dollar amount, the more bakers will be involved, which gives more variability to the data set.

A graph with lines and numbers

Description automatically generated