Module 1 Challenge – Analysis

Date: 09/25/2020

Student Name: Soona Cheon

1. Overview of the project:

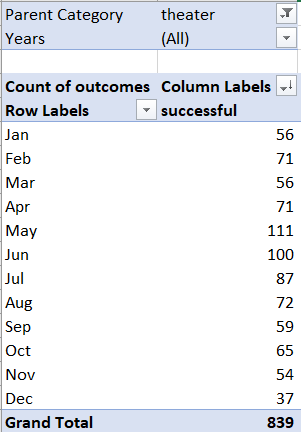
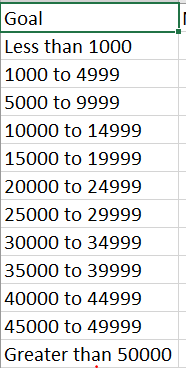
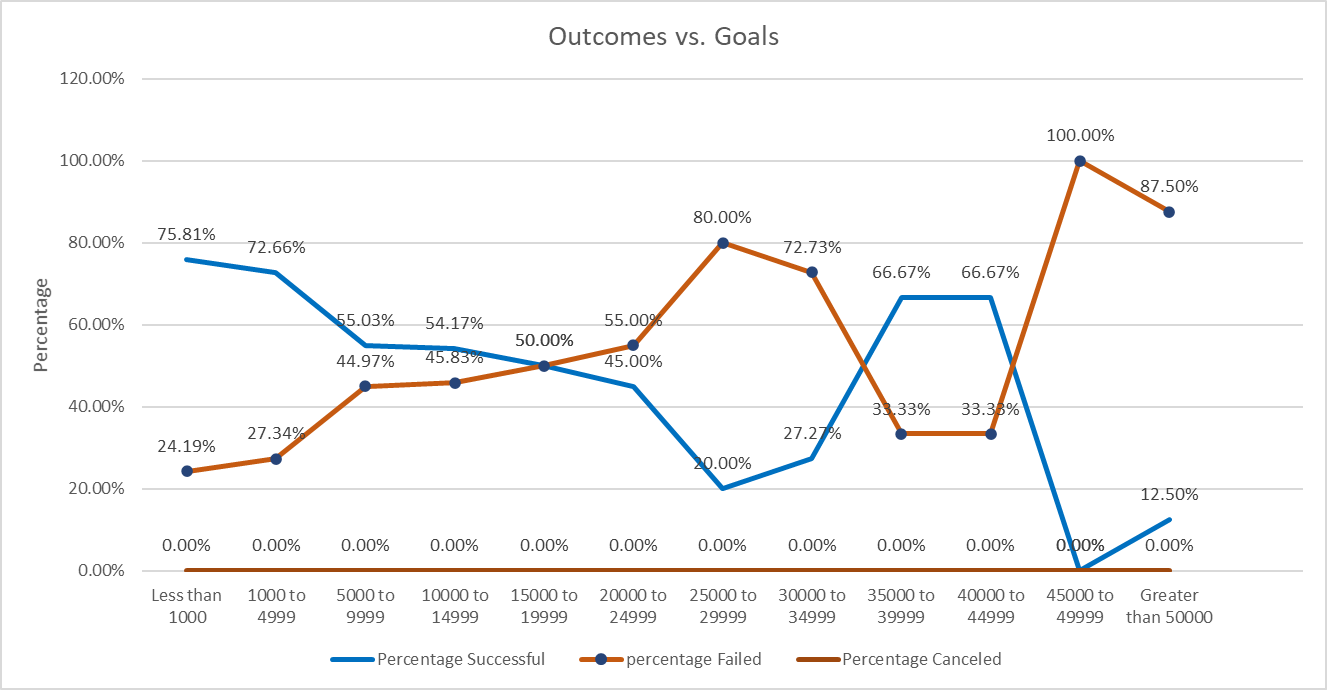
The purpose of the project is analyzing the Kickstarter Fund Raising Project data. The Kickstart project is held globally in 21 countries between 2009 and 2017 to raise fund using various methods (parent category and subcategory), which contains 4115 records.

The Challenge 1 is specifically analyzing the campaign results data based on the 2 elements: (1) Launched Date - if the outcome is related with the campaign Launched month cross years.

(2) Goal – if the fund-raising goal range is related to the actual outcome of the project.

1. Analysis and Challenges

The challenge of the Analysis is the below:

1. The Kickstart data is a real word data, which was not controlled by specific timeline, project time or designed to remove the geographical, socio-economic factors or participants, number of projects category/subcategory or exact number of projects. It means even though there were same campaigns (e.g. “Theater” as category, “Plays” as subcategory) held in 2 country, the quality, content of the event or social economic status of the participants are unknow.
2. The Kickstart data is a real word data, which is not weighted. In each country, the number of events & projects are all different, we will analyze it as is.
3. Thus, the assumption is geographical, socio-economic factors are neutral cross country & participants.
4. Results
5. **Outcome vs. Initiation Date (Month)**
   * Condition: The data was analyzed after filtered by Parent Category is “Theater” cross year.
   * Outcome: When we see the number of success, the most successful campaign happened between May (111), June (100), and July (87). So, when we initiate the campaign in the month, it will be most likely successful, when you choose to campaign in theater.
   * The pivot table shows the number of success as below
   * The line pivot chart also shows the trend, which the campaigns are most likely successful when the initiation date is between May to July. It makes a sense, when spring and summer comes, weather is pleasant, people have tendency to go to theater, meantime, when weather is cold, people do not go to theater less.
   * 
   * Conclusion
     + Conclusion 1: Theater campaign will be most likely successful when it is initiated between April and September, and it is the most successful when it is initiated in May.
     + Conclusion 2: Theater campaign will be most likely failed when it is initiated in October, since the weather is getting cold, and people will not go to the theater.
   * Limitations
     + We only analyzed the “theater” as category. If we choose a different category, the analysis result may be different.
6. **Outcome based on Goals**
   * Condition 1: We filtered the data for “**Plays” in Subcategory**.
   * Condition 2: We analyzed the data based on the below Goal range.
   * 
   * How to: We calculated the number of success, number of failed, and number of canceled outcomes first. Then, we calculated the percentage of success, percentage of failed, and percentage of canceled outcomes using the total number.
   * 
   * Conclusions: The above Pivot Chart shows that
     + Conclusion 1 In Low Goal Range (less than 1000, 1000 to 4999, 5000 to 9999)
       - The percentage of successful is much higher when the goal is low.
       - The percentage of failed is much higher when the goal is high.
       - It means when the goal is low, it is easy to make the fund-raising campaign successful
     + Conclusion in High Goal Range (40000 to 44999, 45000 to 49999, above 50000)
       - The percentage of successful is much low, and the percentage of failed is much higher. It means, when the goal is too high, it is very difficult to make the fund-raising campaign successful, and often failed.
     + Conclusion In mid Goal Range (35000 to 39999, 40000 to 44999)
       - Surprisingly, the Goal level is not impacting the percentage of successful rate in this range.
       - It seems, regardless of the level of goal, donators, who can donate this much fund will donate anyway. Perhaps, this level of donators can be labeled as “solid and steady” donators.
   * Limitations of the Dataset:
     + It is real data, and we did not weight the dataset.
     + Outcomes based on the Launch Date – The dataset is analyzed only for Theater category.
     + Outcomes based on Goal – The data set is analyzed only for Plays in subcategory.
     + Assumption: If we analyze the dataset for different category or subcategory.
   * What are some other table and/or graphs that we could create?
     + Outcomes based on Launch Date – We could create T-value or R residual table/chart to find out the correlation between Outcomes and Launch Date per Month.
     + Outcomes based on Goals – We could create PivotTable to collect the outcomes (e.g. number of successful, failed, canceled, live) instead of using the =COUNTIFS formula.

Thank you!