Kickstarter of data analysis

**Introduction**

As a request of Louis’s concern how different campaigns fared relate to their goals and launch dates after she found out that her play Fever came close to its fundraising goal in a short amount of time. I visualized campaign outcome based on their launch dates and their funding goals to see is there a correlation.

**Method**

Data Collection was downloaded from Bootcamp Website with Excel

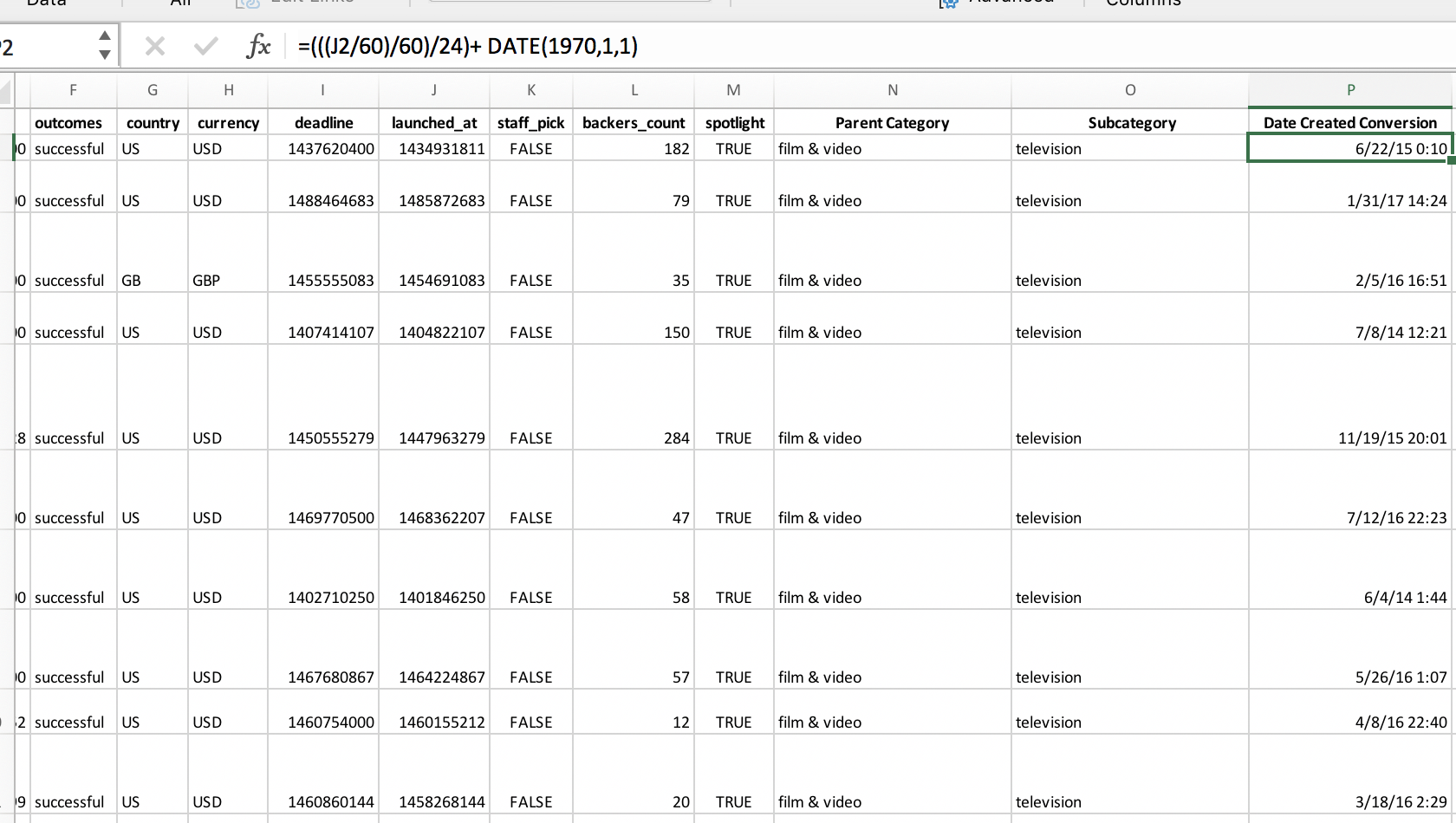
Used epoch to convert Unix time stamp to excel time

Applyed implanted excel functions such as sumif , countifs , pivot table, and line chart to manipulate and visualize data.

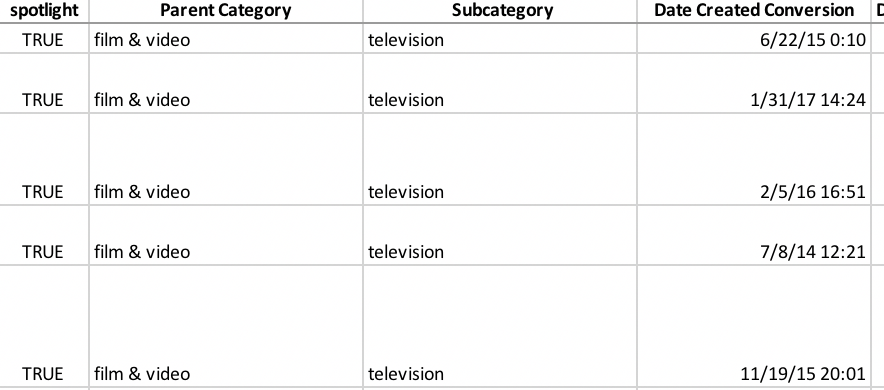
**Data Manipulation and visualization**

Analysis of Outcomes Based on Launch Date Chart

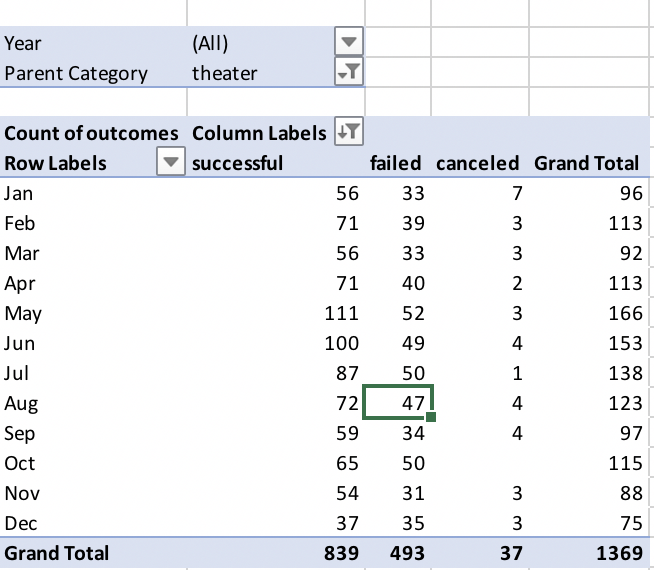
In order to figure out how time related to outcome, the first thing must be converted time into readable time series. Using epoch to convert Unix time Stamp to excel date. I also add one more column to show years so louis could be able to filter different year.

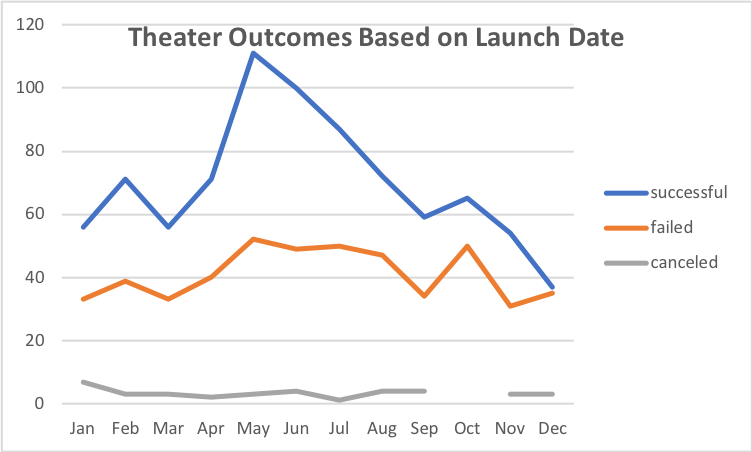


After converting into readable excel date, I split the category into parent category and subcategory by using implanted function called text to column so that plays could be pulled out when counting it.



After cleaning data, I moved to visualized data by using pivot table and line chart.

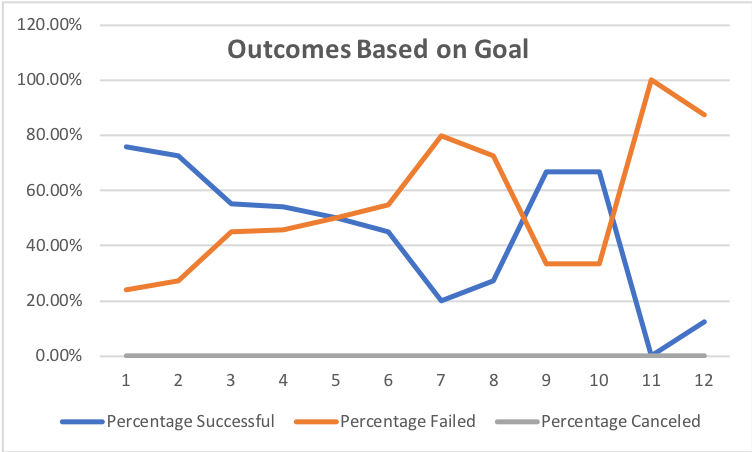




Analysis of Outcomes Based on Goals

After analyzing the relationships between Outcome and subcategory, I moved to identify the relationships between goals outcome by illustrating a line chart

In order to figure out how goals affect campaign fared, I divided goal into ten different ranges. I counted each range under every outcome result and make it to percentage. I also draw a chart line to visualize the relationships. Unfortunately, it seems like there is no correlation between these two variables. But we can conclude that as goal increase, the percentage of successful will decrease.



**Challenges and Difficulties Encountered**

I did not encounter any challenges or difficulties when I was doing this project, but I believed that there are some problems might be considered as challenges.

First, the dataset from bootcamp is obviously clean and is not necessary to clean it before analyzing it. But I think preprocessing data is most important part for data analytic. Usually more than 50 percent of time of data analyst used is to clean data such as watching out outlier or duplicates, checking missing value and inconsistencies or error existing.

Secondly, in order to identify the factor of campaigns fared, regression model should be utilized to figure out the correlations. If we apply Multiple regression, it might be violated because of multicollinearity.

**Results**

*Outcomes based on Launch Date*

we can concluded that there is a relationship between count of outcomes and launch data, as we can see from pivot table. Most campaigns happen round summer from April to July and it reaches peak during May and June. However, it did not identify the reason why this happened

*Outcomes based on Goals*

Based on the line chart of Outcomes based on Goals, we summarized that as goal increase, the percentage of successful will decrease under play category. Unfortunately, we don’t know what the reason behind it is.

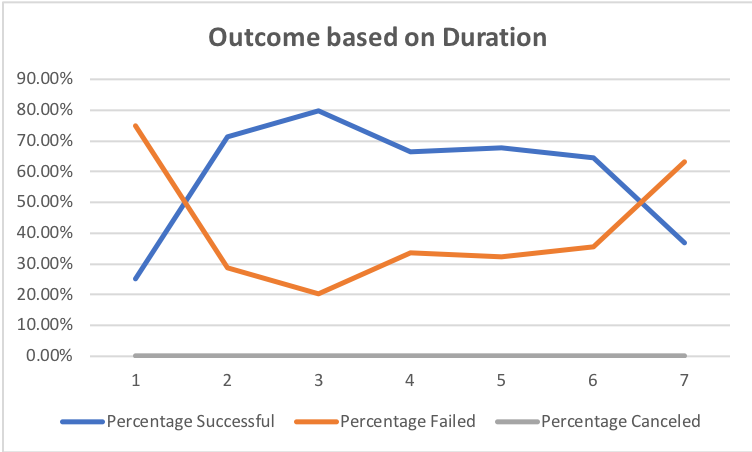
**Limitation of dataset**

The dataset is not relevant in terms of representing the whole population. For instance, dataset includes more than 15 countries, but the goals count with their own currency. This is not comparable between two different countries. Moreover, population should be considered one of most important factors that would affect outcome and goal.

On the other hand, geography is another factor that may affect the result as Australia is in both the Southern Hemisphere, which means the result is perfectly opposite.

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

Using both launch date and ending date will be more reasonable to illustrate Louis ’concern as the rate of successful is pretty much related to the duration of campaign.



As we can see from above chart that as duration campaign increase, the successful rate is higher, and it causes goal is higher.