#\*\*Kickstarting analysis\*\*

##Overview of Project  
###Purpose

###This kickstarting analysis will help Louis, our client, understand what are the inclinations that make a crowdfunding campaign succeed, fail or get canceled. Through data analysis of trends, based on launch data and goals. Throughout this process we will evaluate our challenges, limitations, alternative way to analyzing our data and draw conclusions from our analysis.

###Louis launched a fundraising campaign for her play ‘Fever’, unfortunately the pledged amount of $2,485 did not meet the goal of $2,885 needed. A kickstarted database has been provided to better understand the success and failure of previous campaigns. We will be placing our attention on the data for launch dates and goals.

## Analysis and Challenges

###For this analysis we will focus on the parent category ‘Theater’ and subcategory ‘Plays’. We will evaluate the measures that must be taken to guarantee the success or measures to avoid the failure of future projects. In addition, we will evaluate any trends that we see in the data. We will take her play ‘Fever’ as a benchmark, that was aired in the US, so we will observe the US market. Through the practice of excel functions and construction of tables, we will help Louis understand why previous crowdfunding campaign were successful and make assumptions as to what lead to their failures.

###The analysis will answer the following questions:

-What impact does the launch date have on the success or failure of a play?

-What conclusions can be drawn from the outcome of goals?

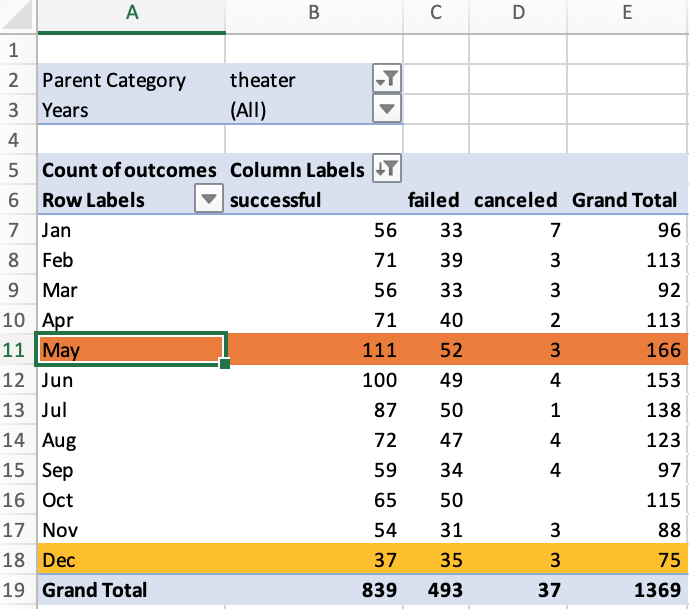
###The challenges with subjecting our analysis to only focusing on the subcategory of ‘Plays’ is that it will leave us with limited information, and we will be unable to fully understand what happened to other ‘Theater’ subcategories. In addition, this will not show us how launch dates may have affected the theater industry as a whole. We will be unable to know if there was a shortage just in the industry for that given year, of if the shortage only applied to plays. A second, limitation is narrowing our search to only US plays. Narrowing our data will limit our understanding of the impact that setting high or low goals can have on the rest of world. Third, the data for outcomes of goals seem to have outliers that may obstruct with our evaluation. Overall, we will see how the challenges play a role in our analysis.

###The Kickstarter dataset contains numerous categories and subcategories from across the globe, we will start by filtering our data. We will apply a filter on ‘Country’ and select to only view US options, for part of our analysis. Next, we need for the data to show us ‘Theater’ as a category and ‘Plays’ as a subcategory. However, both options are combined in Column N, to separate them we need to create subcategories. Once created we can narrow our data to show only the ‘Plays’ in the US.

## Analysis of Outcomes Based on Launch Date

###We will initiate our analysis of outcomes based on launch date. The launch dates can be found on our database However, the serial of numbers is in a Unix Timestamp format and need to be converted into a month/day/year. For this conversion we will create a new column one titled ‘Date Created Conversion’. The excel formatting will breakdown the epoch by telling it to give us the breakdown of the data into seconds, minutes and hours. Once this is done, we want to apply the Date Formula, and adding it to the baseline date of 1/1/1970. This will give us the format needed to analyze the impact launch date had on the campaign.

###Once we have the ‘Date Created Conversion’ column populated to all of our database we can go ahead and create a Pivot Table that evaluates the outcomes based on date created. Our table will be filtered on all the years available to us and categories, in our case the category will be filter by the parent category ‘Theater’ and now we can view the success, failure of cancelation outcomes based on the month it was launched. As we observe the data, a few things jump to us, first, the two months with the most launches are May and June. During these two months the success rate is over double the failure rate, and in May we only see 4 cancelations. In May 166 theater projects were launched, with 111 successes and 52 failures. The following month, June had 153 total projects, in which 100 succeeded, 49 failed and 4 cancelations. Meanwhile in December, fewer were launched, but the success rate is very close to the failed rate. A total of 75 projects were launched, but only 37 succeeded, meanwhile 35 failed and a total of 3 were canceled. This shows us that 38 projects had a negative result, when only 37 survived. Therefore, May is the most successes.



###The Pivot table is helpful, but let’s visualize the trends. For this we will create a line chart, staked lines with markers, and call it ‘Theater Outcomes Based on Launch Date’. Once the chart has been created, we can see some trends that we may have overlooked. The months of January, March and September have a similar number of successes and failures, about ~60% of projects successes, meanwhile ~35% failures. Also, January is the month with the most cancelations.

Chart, line chart

Description automatically generated

###Let’s go back to Louis, her play was launched in 6/13/16 in the US. Using the same method as above, let’s observe the relationship between launch date and outcomes for ‘Plays’ in the US during 2016. For this we will alter our Pivot Table a bit, we will add ‘Country’ and ‘Subcategory’ to our filters. Then filter ‘Country’ to US, ‘Year’ to 2016, ‘Parent Category’ to ‘Theater’, and ‘Subcategory’ to ‘Plays’. What changes? What stays the same? May continues to be a fairly successful month with about a success rate of 62% and a failure rate of 38%. However, that is not the most success launch month, we see that July has a success rate of 91% and a rate of 9% failures. Interesting enough, June has one of highest failure rate in comparison to successes, with 47% successes and 53% failures. June is tied with February, but December continues to be the month with the highest failure rate of 60% and 40% successes, but the number of projects is the lowest of all.

Table

Description automatically generated

###If we observe the success rate of all plays that were launched in 2016, we can see that 112 out of 177 succeeded, meanwhile 65 failed. Therefore, in the US in 2016, there were 63% successes and 37% failures. If we compare this data to the rest of the world, we see that in 2016, the subcategory of ‘Plays’ projects had a successes rate of 69% and the failure rate was 31%. Our observations, show that the year was fairly successful in the US and across the glove, however launching her campaign in June may have hurt her chanced of meeting her goal. Let’s evaluate the relationship of goal and outcomes.

Chart, table

Description automatically generated

## Analysis of Outcomes Based on Goals

###We will start our analysis of outcomes based on goals. For this we want to create a sheet for our \*\*Outcomes Based on Goals\*\* to store our data. We will start by creating a column for ‘Goals’ which contains rows with cells for a range of amounts. Once the ranges have been established, we want to find out what number of plays were successful, what number failed and what number got cancelled. For this, we will create three columns and populate them with the corresponding data using the \*COUNTIFS\* function.

###We can see that the number of canceled plays is cero, therefore we will disregard further analysis on canceled plays and conclude that no plays have been canceled in the US data set. At this point we have a count, but how does this relate to the number of total projects? What percentage of plays succeeded or failed in relation to the total amount? To get this information we need to create a column for ‘Total Projects’ and populate the cells by using the \*SUM\* function and add the number of failed and successful plays on each range. Based on the total project we can now find out what percentage of them did what.

###We will go ahead and create two new columns to gather this information, one titled ‘Percentage Successful’ and ‘Percentage Failed’. By looking at the percentages we can see that goals that were between less than $1,000 to $14,999 have a higher percentage of plays that were successful. Goals that ranged between $15,000 and $19,000 are split down the middle, meaning plays in that range have a 50/50 chance of succeeding or failing. Plays with range $20,000 to $34,999 have failed. The following information could potentially contain some outliers, given that is shows us that although plays under $34,999 failed, 6 plays between $35,000 and $44,999 seem to have succeeded in reaching their goal, meanwhile, 3 failed. Our last range of $45,000 to greater than $50,000 all failed.

A picture containing chart

Description automatically generated

###To better understand and visualize the relationship and trends between the percentage of success and failure of plays to the goal amount ranges, we will create a line chart.

Chart, line chart

Description automatically generated

###The line chart of titled ‘Outcomes Based on Goal’ shines light to new information. The success rate seems to be decreasing as the goal amount increases, however it picks up once the amount increases to about $25,000. From there forward the rate of success has a somewhat steady increase until it hits $35,000. From that point on, both the success and failure rate plateaus, the success at 65% meanwhile the failure stays at an about 35%. The failure has an inverse relation to the success rate, therefore the greater the goal amounts the higher the risk of failure. These stops being true once the goal hits the $30,000 range, where we see that the number of failed plays decrease as the goal amount increases. Another feature we neglect to see on the chart is that there are two additional interceptions between the percentage of successful plays and the failed ones. The plays have a 50% chance of succeeding or failing somewhere in the range $30,000 to $39,999, and $40,000 to $45,000. Both the table and the line chart will help us draw our conclusions as to what is the impact of goals on the outcome.

## Challenges and Difficulties Encountered

###We encountered challenges when it came to outcomes based on launch date. We saw that May and July were both months that had a higher rate of success in the ‘Theater’ parent category around the world; But somehow June had a higher rate of failure. To try to further understand this, we created two more Pivot tables that showed that this trend is consistent in 2016 and also consistent in the subcategory of ‘Plays’, in the US and around the world. We’re still unsure of the cause but the trend is consistent.

###We also found challenges on the analysis for outcomes based on goals. Given the trends, it’s hard to understand why Louis’s play ‘Fever’ did not meet its goal. Taking into account that at the goal range of $1,000 to $4,999 there was a 73% success rate and a 27% failure rate. Louis’s the goals outcome should had been successful. To try to better understand why this could’ve happened, a descriptive statistical analysis had to be done, where we can see that some of the limitations of this dataset are outliers. This is later explained on the section titled ‘Limitations of this Dataset’.

## Results

###The following conclusions from our kickstarting analysis will help Louis understand the trends we observed that made a crowdfunding campaign succeed, fail or get canceled based on launch date and goals. Also, it will hopefully shine some light on what may have caused the failure of her project ‘Fever’.

##Conclusions of Outcomes based on Launch Date

###In the theater category, we observed that the launch data had a significant impact on the success or failure of the project. For example, our Pivot table and line graph for ‘Theater Outcomes Based on Launch Date’ show us that the month of May has a higher number of projects and success rate, so the projects that were launched in May had a higher likelihood of succeeding. We also learned that the month of December has the least projects, but yet the highest rate of failure. When we change the filters and analyze the impact that launch dates had in the US, in 2016, for the subcategory of ‘Plays’. May continues to be the month with the most successful projects of the year. However, July is our winner for rate of success to projects. Also, December continues to be the least appealing month to launch a project. When we unfiltered for year and try to analyze which is the most successful month in the world in 2016, May continue months for this year. Therefore, with our data, it would be safe so assume that Louis would have had a better chance of having a successful outcome had her launch date been in May or July. However, June increased her likelihood of not having a successful outcome.

## Conclusions of Outcomes based on Goals

###When it comes to what we can conclude from ‘Outcomes based on Goals’ we see that the failure has an inverse relation to the success rate. Consequently, the greater the goal amounts the higher the risk of failure. Although, this conclusion stops being accurate when goal hits the $30,000 range, where we see that the number of failed plays decrease as the goal amount increases. We will need to a descriptive statistical analysis to understand if we have outliers, and to what degree they are obstructing with our analysis.

###That being said, we can explain to Louis that although the pledge amount of $2,485 should had met her goal of $2,885, given that at the goal range of $1,000 to $4,999 there was a 73% success rate and a 27% failure rate. There’s a likelihood that the launch date may had played a bigger role in her failed outcome and well as possible outliers that may have skewed the data.

## Limitations of this Dataset

###If we do a descriptive statistical analysis, we can see that some of the limitations of this dataset are outliers. For example, in the US, the mean of the successful outcomes for ‘Plays’ is $4,101 yet the medium is $2,500. Thus, the extreme goals for theater-building proposals result with a mean that’s located in the Upper Quartile. We can assume that this is because when the goal hits $30,000 range, is where we see that the number of failed plays decrease as the goal amount increases. We can also observe that the mean for successful pledged is also in the Upper Quartile. Outliers, play a big role in possibly changing the conclusions that were drawn for our analysis.

Table

Description automatically generated

##Other Possible Tables and Graphs That We Could Create

###Lastly, to observe what are the inclinations that make a crowdfunding campaign succeed, fail or get canceled. We have added additional pivot tables, filtering on the ‘Subcategory’ of ‘Plays’, the ‘Country’ US and ‘Year’ 2016, to further analyze the trends. Also, to better understand our challenges with the dataset, a table of descriptive statistical analysis has been created. We’re optimistic that this addition information can help us try to tell a broader story of the trends present in the outcomes based on goals and launch date.