# Kickstarting with Excel

## Overview of Project

### Purpose

Louise, a promising play writer, is looking to kickstart a crowd funding campaign to obtain enough capital to finance her latest play called “Fever”. She is approximating her expenses at nearly $10,000-$12,000 and is seeking assistance to help launch a successful campaign using public crowd funding data. Our data analytics team collected, organized, and analyzed the data to provide Louise with the necessary tools to create a successful campaign. Included in this report is an analysis of the public crowd funding data, encountered challenges, visual aids, and findings for Louise to make proper decisions regarding the “Fever” campaign.

## Analysis and Challenges

### Analysis of Outcomes Based on Launch Date

The data analytics team created a pivot line chart to illustrate how the key metrics, launch date and outcome type, relate. Here is the link: xxx. To create this pivot line chart the following steps had to occur:

1. Organize and format data in the “Kickstarter Data” tab of the EXCEL.
2. Create a pivot table from the “Kickstarter Data” tab and copy it into a new tab called “Theater Outcomes by Launch Date”.
3. Adjust pivot table to show appropriate fields including filtering “Parent Category” by “theater” and “Years”, adding row labels by “Month”, and adding descending column labels by “Outcome”.
   1. To create the “Years” filter we had to add a column to the “Kickstarter Data” tab and use the following formula to extract the year from the “Date Created Conversion” column: =YEAR(S2)
   2. To create the “Month” row labels we had to add a column to the “Kickstarter Data” tab and use the following formula to extract the month from the “Date Created Conversion” column: =TEXT(S2, "mmm")
4. Create a pivot line chart with markers from the pivot table data.
5. Add a title and save the pivot line chart into a PNG file.

### Analysis of Outcomes Based on Goals

The data analytics team created a pivot line chart to illustrate how the key metrics, goal and outcome type, relate. Here is the link: xxx. To create this pivot line chart the following steps had to occur:

1. Organize and format data in the “Kickstarter Data” tab of the EXCEL.
2. Create a new tab called “Outcomes Based on Goals” and create a table with the following:
   1. Columns: goal, number successful, number failed, number canceled, total projects, percentage successful, percentage failed, percentage canceled.
   2. Rows: less than 1000, 1000 to 4999, 5000 to 9999, 10000 to 14999, 15000 to 19999, 20000 to 24999, 25000 to 29999, 30000 to 34999, 35000 to 39999, 40000 to 44999, 45000 to 49999, 50000 or more.
3. Use a countifs formula for the “Number Successful”, “Number Failed”, and “Number Canceled” columns to include the following criteria “Outcome”, goal amount (numeric row range, subcategory of “Plays”. Here is an example of the countifs formula: =COUNTIFS('Kickstarter Data'!$F:$F,"successful",'Kickstarter Data'!$D:$D,"<1000",'Kickstarter Data'!$R:$R, "plays")
4. Sum the “Total Projects” column using the values in the “Number Successful”, “Number Failed”, and “Number Canceled” rows using the following formula: =SUM(B2:D2)
5. Calculate the percentage of the “Number Successful”, “Number Failed”, and “Number Canceled” rows when compared to total projects using the following formula: =(B2/E2)\*100
6. Create a pivot line chart from the “Outcomes Based on Goals” tab data range A1:H13
7. Add a title to the chart, add a legend to the bottom of the chart showing the percent column labels and the goal range being the x-axis labels.
8. Save the image as a PNG file.

### Challenges and Difficulties Encountered

There were no major challenges or difficulties encountered while performing this data analysis or visualization, as the data analytics team is well versed in Excel, a common software program.

It is important to note to update the data range of any pivot table if new rows or columns are added to the data source and subsequently refreshing the data for the pivot table and chart to update based on changes.

## Results

### What are two conclusions you can draw about the Outcomes based on Launch Date?

Once the Kickstarter data was groomed and the “Outcomes based on Launch Date” visualization was created, the data analytics team observed the following conclusions:

1. The month of May, June, and July had the highest count of successfully launched campaigns. Campaigns followed a negative success trend in the holiday season starting in September through January.
2. Although the amount of failed Kickstarters also peaked in the summer months, there is not much variance in the data points during a calendar year. The mean of the “Failed” outcome data points is 41 and only one month, November, falls outside the standard deviation of 8, illustrating nearly 92% of calendar year data points are alike. Similar results were found for Kickstarters that had the “Canceled” outcome.
3. It is recommended that Louise launch her crowd fund Kickstarter for her play “Fever” in the month of May or subsequently June or July for a successful outcome.

### What can you conclude about the Outcomes based on Goals?

Once the Kickstarter data was groomed and the “Outcomes based on Goal” visualization was created, the data analytics team observed the following conclusions:

1. There were zero Kickstarters that were classified with a “Failed” outcome in the “Play” subcategory. Alluding to the fact that all “Play” Kickstarters either failed or succeeded.
2. We found that the campaigns with the highest success rates were on either end of the goal spectrum with 76% success rate for a goal of “Less Than 1000”, 73% success rate for goal of “1000 to 4999”, 67% success rate for both “35000 to 39999” and “40000 to 44999”. The successful campaigns are all in the upper quartile range of 67 or higher and the median is bigger than the average illustrating a negative skew of extreme data points pulling the average down. All but two goal data points, “50000 or More” & “45000 to 49999”, are within the interquartile range (IQR). Illustrating a higher IQR and a large spread from lower quartile to lowest point of “0”.
3. We found that the campaigns with the highest fail rates are “25000 to 29999”, “30000 to 34999”, “45000 to 49999”, and “50000 or More”. The median is smaller than the average data point demonstrating a positive skew of data points bringing up the average of “Failed” outcomes. Majority of campaigns were below the upper quartile range and within the interquartile range and with a small range in the lower quartile.
4. It is recommended that Louise chooses a goal amount based on the pivot chart table when the line of “Percentage Successful” is above the “Percent Failed” line with as much spread as possible. For example, a range from 0 to 14999 or 3500 to 44999.

### What are some limitations of this dataset?

The data analytics team identified the limitations of this dataset for Louise to understand possible inconsistencies and/or errors with the data:

1. There could be discrepancies if the campaign representatives self-reported the data causing a false reality of metric values.
2. The sample size can cause concern if our findings are representative of too small of a population. Data based on a small sample size reveals low statistical power and can be mitigated through a more appropriate size of the population. For example, the data analytics team could collect more Kickstarter data specifically from “Theater” or “Play” category campaigns and exclude those that are not relevant to Louise’s needs.
3. There is always concern for bias in analyzing the data, as it can be natural to strive for a higher goal dollar amount even if the data is indicating otherwise. It is important to use the data to make educated decisions and remove as much bias and emotion from the equation as possible to ensure a successful campaign.

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

The data analytics team expressed the idea to create a visualization that would combine the “Theater Outcomes Based on Launch Date” and “Outcomes Based on Goals” pivot chart to simplify the best time and goal for Louise. Instead of synthesizing the two graphs independently, the team could create a visualization that shows the optimal time to launch and at the most appropriate goal dollar amount. This will remove any end-user’s (Louise) human error in determining the best strategy. It is determined by the data analytics team that these two best scenarios for Louise to implement for her Kickstarter crowd funding campaign for “Fever”:

1. Conservative: launching between May and June at $0-$10,000 goal.
2. Risky: launching between April and August at $35,000-$45,000 goal.