# Kickstarter Analysis: Theater outcomes, goals, and launch date data

## Overview of Project

### For this data set there are two comparisons in focus. The first looks at the project outcome based on which month it is launched. The second being the percentage of success, failed, and canceled projects depending on the fundraising goals.

### Purpose:

This analysis is designed to draw direct comparisons regarding kickstarter campaigns for plays. Giving guidance and setting expectations for the upcoming play *Fever*. Specifically, regarding establishing a realistic goal for fundraising and determining a launch date. In hopes for a higher likelihood of success.

## Analysis and Challenges

### Analysis of Outcomes Based on Launch Date

I started with pulling data from the “kickstarter” sheet, looking at outcomes, parent category, and date created conversion. With those values I created a pivot chart, they were organized in the pivot table fields as below.

A screenshot of a computer

Description automatically generated with medium confidence

With the filters being Parent category and years, we can look at only plays between 2010-2017. The columns which contained the outcomes were specified down to successful, failed, and successful (seen below). The rows also contained data from the date created conversion you can pinpoint it down to just months. The second repeated set being outcomes which also into the values section. That’s what populates the values in the respective cells.

Table

Description automatically generated

Highlighting the pivot chart I was able to create a line graph with markers which shows the amount of successful, failed, and canceled plays per month.

Chart, line chart

Description automatically generated

### Analysis of Outcomes Based on Goals

This comparison is function heavy, it requires the analyst to have a high attention to detail. Here we want to compare the likelihood of success based on the financial goal amount.

Table

Description automatically generated

For the columns “Number Successful”, “Number Failed”, and “Number Canceled” are using a “=COUNTIFS” function. It is the same formula used in each cell with only minor changes. But with out those minor changes it can throw the whole analysis off. Here we are looking at the specific ranges vs outcomes so the way the function changes would be as followed.

Number Successful with a goal between 15000 to 19999: =COUNTIFS(Kickstarter!D:D,">=15000",Kickstarter!F:F,"Successful",Kickstarter!D:D,"<=19999",Kickstarter!R:R,"Plays")

Number Failed with a goal between 1000 to 4999: =COUNTIFS(Kickstarter!D:D,">=1000",Kickstarter!F:F,"Failed",Kickstarter!D:D,"<=4999",Kickstarter!R:R,"Plays")

Number Canceled with a goal between 45000 to 49999: =COUNTIFS(Kickstarter!D:D,">=40000",Kickstarter!F:F,"canceled",Kickstarter!D:D,"<=44999",Kickstarter!R:R,"Plays")

The layout of the formulas is the same in each cell, the main differences being the ranges listed in the front and end. And the outcome being either successful, failed, or canceled.

The last four columns were easy to figure out at that point. “Total Project” was found using =SUM function.

Total project with a goal between 5000-9999:

=SUM(B4:D4)

Which was repeated for the whole column.

The last three columns “Percentage Successful”, “Percentage Failed”, and “Percentage Canceled” were populated using simple division. I would take one of the outcomes from the previous columns and divide that by the total projects of that row.

Giving the percentage based off that outcome.

Percentage of successful outcomes with a goal between 25000 to 29999:

=B8/E8

Since the values under number canceled projects is zero the percentage canceled also had a consistent value of 0%.

Highlighting the table of goal ranges and percentages I was able to create a line graph that showed the percentages of successful, failed, and canceled plays in relation to their goal amount.(see below)

Chart, line chart

Description automatically generated

### Challenges and Difficulties Encountered

For outcomes based on launch date it did not result in any challenges for me. That being said, it does not mean it would be the same for everyone. The important thing to remember when trying to organize the layout of the chart is that some data sets may need to be placed in multiple fields. I.e., Outcomes-> columns and values. Date Created Conversion-> filters and rows. If I had forgotten that outcomes also needed to go into values the table would have been empty.

For outcomes based on goals, this section I did have a challenge. When it was time to populate values for the “Number canceled” column I used the same functions as the others and immediately thought I had made an error when I noticed the whole column was “0” I went through each cell carefully making sure I had not forgotten a punctuation, misspelled a word, or put the wrong goal ranges. After I realized it wasn’t any of those things I went to the internet. Seeing if there was anyone with issues similar to mine. They suggested things like putting asterisks(\*), dollar signs($), or typing out “greater or equal to” instead of >=. None of those working as well I knew I needed to ask someone more familiar with this function. During my class office hors I asked the instructor and TA to review my formula and help me troubleshoot the issue I was having. After multiple suggestions (which also didn’t produce any different results) my instructor came to the realization that I was right the first time. The column was supposed to have a consistent value of zero. The example “Outcomes Based on Goals” graph attached to the challenge also confirmed that realization. It shows that the percentage canceled line was at zero for the whole graph. Although the experience was a bit frustrating it was a great example of a good flow for debugging. First check for and errors in the phrasing for the function, google it to see if there are any recommended solutions, and if al else fails sometimes peer review or a fresh set of eyes can see what you may have looked over. Or validate that you’re doing things correctly and maybe it’s the data itself that is giving unusual outcomes.

## Results

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

* Between 2010-2017 the month with the highest success rate is May.
* The only month that reflected an almost 50/50 chance of success/failure is December.

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

* The highest percentage of success (%72-%74) in plays had a fundraising goal between $0-$4,999.

- What are some limitations of this dataset?

The spreadsheet given is very well compiled, but that doesn’t mean it couldn’t be improved. Things that I think should be included to have a more in-depth guidance would include.

* Location breakdown by City/Region

Logic being that a play showing in Indianapolis with have different outcome vs a play showing in New York City.

* Genre

With breaking it down to the genre you can have a more direct comparison with those that have shown previously. You can’t expect a drama and comedy to be received exactly the same.

* More recent plays included.

The data set only has ranges between 2010-2017. It should be updated to at least the most recent year. Markets and trending interests change quickly after all. What was trending in 2017 could be significantly different than what was trending last year in 2021.

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

If the suggestions above were added, the first recommended chart would be a scatter point chart. It would focus on a relation of Goal amount vs Launch date within a specific City/Region. There are times of the year where people are more/less likely to donate. Example people are more likely to have more expendable funds in May/June vs Nov/Dec. But depending on local city/state holidays that might sway people to stay in town or leave. Example, most of the New York City wealthy tend to leave in the summer months. ETC..

A bar graph comparing successful, failed and canceled plays vs genre in a specific City/Region. A comedy set in a Chicano culture might have a higher % of success vs Kansas City.

## Conclusion

The data sheets and results from the tables/graphs above are very helpful and give great guidance on the risks and expectations that one should have if they are wanting to put on a production funded through a kickstarter. However great isn’t perfect. The more details you can filter down to the more clarity you van give a client for making such important decisions.