**Overview of Project and Purpose of Analysis:**

A basic overview of the project and purpose of the analysis, aside from learning and practicing the functions of data analysis within Excel, is to find out or advise character Louise regarding a recommended goal amount and the best time to launch her Kickstarter campaign for a successful outcome. The workbook of data is based on historical trends gathered from a dataset from Kickstarter’s platform with data originating from 2009.

A more in-depth overview and purpose of the analysis teaches Excel practices and functions for data analysis, clean-up, error definitions and debugging, visual representations (i.e., charts, graphs, Pivot Charts), data organization (i.e., filtering conditional formatting, creation of Pivot Tables that allow for further data manipulations and uncovering of relationships/trends, etc.), and statistical modeling and analysis. All of the lessons aforementioned are taught in order to transform the raw data into digestible, organized outcomes, if existing, or not existing, in order to draw conclusions from relationships/non-relationships and patterns/non-patterns about Kickstarter campaigns. Louise is in Great Britain and wants to launch a Kickstarter campaign for her play. The end goal is for Louise to know a £ amount to set her campaign goal in Great Britain for her play, and what time period would be best to launch her campaign for a higher chance of success in meeting her goal.

After practicing the data analysis and function techniques in Excel, answering the needs of Louise, the student is left with a review of the big picture of the processes of data digestion and analysis. The student is also left with a basis to perform further analysis if needed. The basic introduction to receiving and starting to work with a large data set is ingrained in the student’s mind for future data approaches so that there is a process for future similar receivals. First data must be accounted for, or sized so that the user knows just how much information there is and the best approach for working with the data. Next the student must clean, organize, filter, format and breakdown the dataset. Then with Excel formulas and functions, Pivot Tables, etc. the student is able to pull and relate different ranges of data in order to find patterns, relationships, and answers in trends. Without Excel the ability to quickly arrange, tabulate, format, then pull and condense needed data into easily digestible visuals and recommendations would be a very long, inefficient, and arduous task with room for more human error and /or further subjectiveness.

**Analysis and Challenges**

1. Starting with a workbook and a spreadsheet of data, the initial challenge is to look at the size of the data to know the best approach. Is complete manual coding appropriate, are formulas and functions needed as a result of the size of the sample or population? Using the shift and arrow keys you can look at the range and size of your data. You can also select the data manually.
2. The next challenge is identifying types of data in the data set, getting to know your categories/sub-categories, and examining the data provided and what each variable and/or descriptor is trying to tell the reader. The challenge is reviewing the initial data provided and then understanding out of the data set what is important, useful and related in relation to the end goal of the user. The recipient of the data needs to make sure that the data is consistent as well to avoid errors in future functions across ranges of datasets. For example, if an entire column is in integers, however one cell is an accidental or purposeful text excel functions allow for programming or dismissing this error or variation with the IFERROR Fx (function). IFERRORs and debugging are mentioned further below (2d.).
3. Challenges begin with learning the importance and use of filtering data properly. Misuse of the filtering function can create roadblocks and confusion if there is a learning gap in how to apply the option in Excel. Through filtering the end user can reveal, organize, segment, and analyze nine main categories and 41 subcategories with data from 2009 on Kickstarter campaigns. The modules assist in blocking some of the data from the conscious view of the user by guiding the user or student through the modules and focusing on most, but not all of the data in the data set. Conditional formatting for successful, live, canceled, and active campaigns assist in narrowing down similar data in order to analyze it in relation to the goals and in quick filtering and visualization aid.
4. Timestamp conversions functions will probably have to be looked back up in the future as they were more difficult to remember than other functions. There was a disconnect in understanding the provided formula for converting timestamps manually in Excel. However, one can use the online Unix Timestamp Converters. Learning functions for debugging data and keeping values in cells through the use of proper programming in Excel assisted in further processing of data. VLOOKUPS and IFERRORs were not found to be challenging as the module provided ample practice for these functions and their uses.
5. Pivot Tables proved to be easier to make. The challenge of Pivot Tables lies in the movement of fields and categories into different sections and filters of the table in order to uncover or debunk relationships and patterns. Understanding the changes that one has made when moving and changing the filed settings in Pivot Tables can be challenging and though intensive. However, comfort exist in knowing that Excel is forgiving and allows the user to change or undo most functions prior to most auto saves. Through learning Pivot Tables and Pivot Charts, the participant after completing the module can take the analysis for Louise even further to try and specify recommendations through manipulating data in the Pivot Table.
6. Statistical analysis allows us to reveal relationships in data. Through learning statistical data and functions, and particularly graphing statistical functions the processor of the data can reveal conclusions, suggestions, and relationships. Challenges with this were making sure that functions were typed in correctly. Understanding central tendency, and the meaning behind outliers, what causes or caused them and is that important, identifying them, in order to validate your data to a perform more precise analysis was challenging however very interesting, essentially creating a fun learning environment.
7. The primary challenge of this assignment and module is the last analysis section regarding further analysis and data limitations. The application and thinking of material/s, information, data, basic programming order, logic and syntax are the largest personal challenges of this module. However, with practice, hopefully the mind will strengthen, adapt, and learn.

**Results:**

1. **Two Conclusions About Theater Outcomes based on Launch Date:**

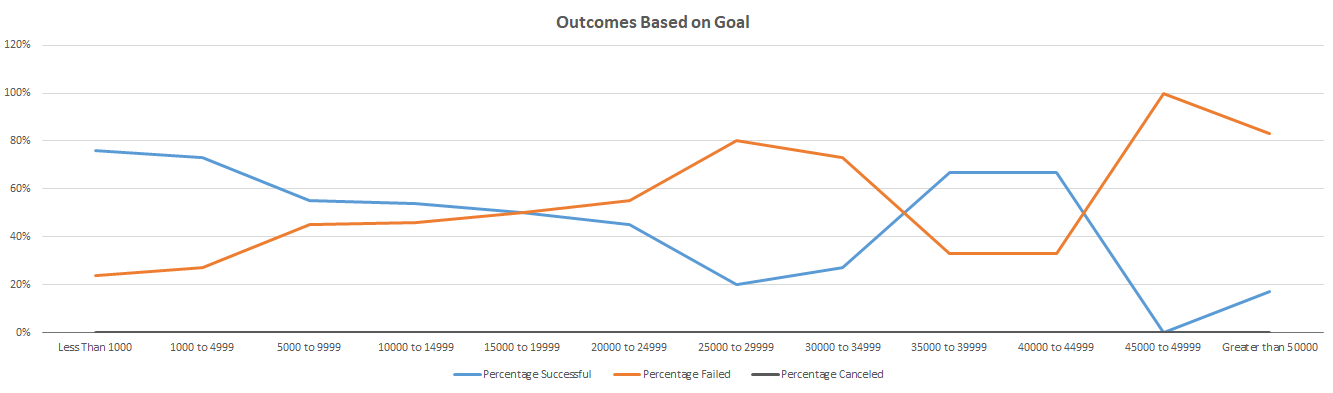
Resources/Theater\_Outcomes\_vs\_Launch.png

<https://github.com/sarahalonzo/kickstarter-analysis/blob/d41917ad86fa743341e7eb88e1d53125b88b919e/Resources/Theater_Outcomes_vs_Launch.png>

The first conclusion from Theater Outcomes Based on Launch Date Analysis is that most Kickstarter successful theater campaigns are launched in the summer. The second conclusion is more specific. The second conclusion is that May and June, with May being the first choice and June being the second, are the two months averaging since 2009 in Great Britain with Theater campaigns with the highest values of success. The conclusion or recommendation based on this trend is that Louise should launch her campaign in May and if not May then June. Additional conclusions (honestly, a stronger second conclusion) from the Pivot Table and analysis as it stands are that failed campaigns also follow a similar trajectory, however, they stay more consistent as compared to successful spikes in Summer, and do not spike as much as in summer. Both failed and successful campaigns spike around the time of the year when many international holidays begin, in October. This is also when people tend to have higher expenses and are more likely to spend on healthcare needs with deductibles met. Proving this effect from holidays and healthcare needs would take far more data, they are assumptions. The October successful spike is not considerable enough to risk launching a campaign especially with a higher or probability/historical risk of failure in the same month. Another recommendation for Louise is that she not launch her campaign in October because though there is a historical spike in successful campaigns, it is also the highest spike in launch dates for failed campaigns, with a subjective assumption that this could be due to the start of many international holidays. On a professional presentation the subjective assumption (s) could be included as a note/comment; but also, are irrelevant and could be discarded.

However, this pivot table and chart can be further filtered and manipulated to uncover more data. For example, the table can be filtered regarding the average amount pledged in successful campaigns and to the subcategory plays to set a goal amount at in the month of May and make sure that in Britain the month of May is still showing as the month correlated with the highest number of successful campaigns.

1. **One Conclusion of Outcomes Based on Goals:**



Resources/Outcomes\_vs\_Goals.png

<https://github.com/sarahalonzo/kickstarter-analysis/blob/d41917ad86fa743341e7eb88e1d53125b88b919e/Resources/Outcomes_vs_Goals.png>

One of the conclusions of Outcomes Based on Goals is a simple visual observation that successful goal ranges are contrary or contrasting to the failed goal ranges. The data displayed clearly visualizes that when goals are set in successful ranges, those same ranges reduce the failed campaigns and when goal ranges are set in ranges that result in more failed campaigns the same ranges for goals reduce the successful campaigns. Aside from outliers which are seen on the far right of the distribution, the successful and failed outcomes based on goals amount are an inverse relationship. This means, Louise in this instance, most consistent successful ranges for outcomes based on goals are the $1K-4.9K range, however closer to the 1K range.

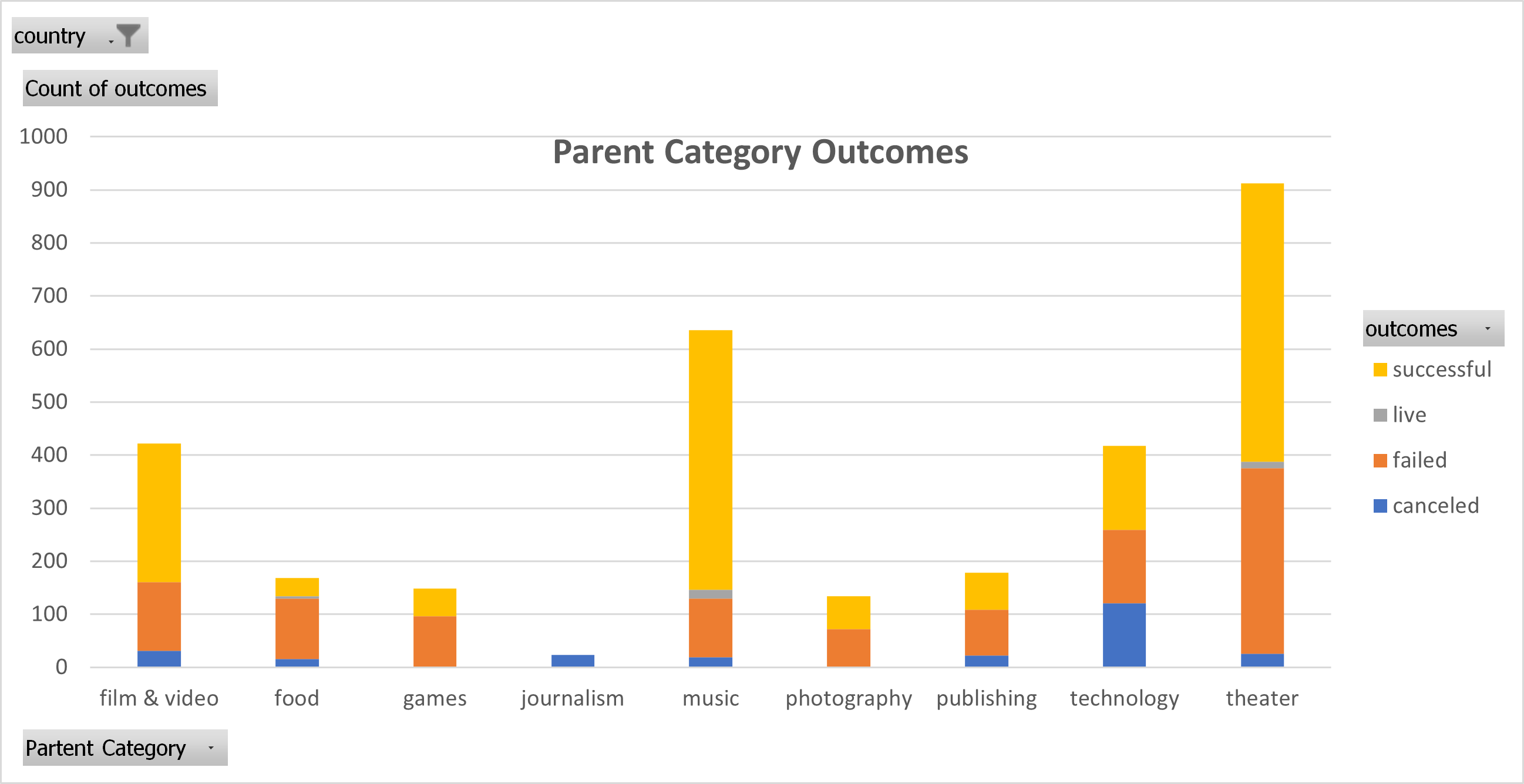
1. **Summary of Limitations of Dataset:**

Initial limitation of the dataset is that certain variables are not defined. For example, we assume that the country in the data set is representative of the origin of the campaign. However, one could ask or assume without a proper definition that the country also includes the scope or span of the campaign. Then the analyst would ponder if there is a relationship on successful campaigns v. failed campaigns based on size and economic strength, culture and interests, or even priorities of certain countries. If Louise lived in Great Britain but wanted to launch her campaign in the United States (U.S.) or a U.S. platform, would historical trends show that she would be more successful than just a launch in Great Britain, or a scope of potential contributors confined within Great Britain? Essentially, definitions of some of the data would be helpful. Same with number of backers. Does this mean the number of donors to each campaign? If yes, then utilizing this data in the data set could further clean the data of outliers. For example, why include data on a pledge and date amount for a play with only one backer when the average successful play campaign has for example 38 backers? If comparing successful to false campaigns both sets of data would have to be cleaned for the same or similar outliers for a more accurate representation. Lastly, what does staff pick mean and how does this relate to the data? Could it be important or hold value? I also think that the very large time span is a constraint. I have to ask if analyzing data from 2009 is a adequate time frame to represent a current goal.

The filtering of country seems to be tied to currency, though during the analysis we are looking at American dollars when, though convertible, the client Louise is trying to know how many British Pounds (£) to set her goal at. This is filtered later when the country is filtered to Great Britain. The analyst doesn’t really know if the amount reflected is in dollars across all currencies for comparison value or if the amounts of pledged and goal are simply formatted to a dollar amount because of formatting in Excel. At the end of the exercise, you want to advise Louise’s to stay I the £2.4K range for setting her goal, based on data filtering, pivot tables and manipulations. However, is the analyst advising Louise in American Dollars or £, even with the currency filter changed for Great Britain? Same questions can be applied to average donation column.

1. **Additional Graphs, Charts, and Analysis:**

As for additional charts and analysis, running a simple stacked bar chart to see what categories are the most successful over the entire data set would provide an idea of the market for Louise’s theater category play, and verify if Kickstarter were a good campaign platform for her to launch a play in Great Britain.



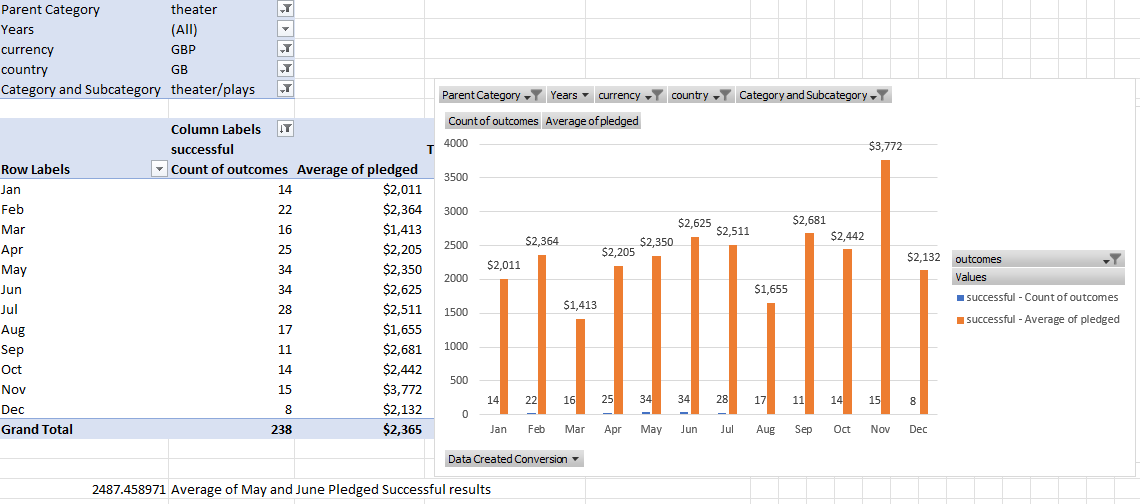
<https://github.com/sarahalonzo/kickstarter-analysis/blob/main/Deliverable_3_Addtl_Analysis/Parent_Category_Outcomes_Deliverable%20_3.png>

The bar chart validates that Kickstarter has a high chance of success for Louise’s venture. This could be narrowed down further to plays.

The analysis that I would want to provide Louise with would be a better assessment of the average amount of successful goals in plays with a pivot table and a visual aid. Outcomes based on goals includes outliers as a graph and if Louise or the client does not see or digest the actual spreadsheet numbers the graph without data points a false sense of best outcomes based on goal may occur.

The easieset way to create the pivot table showing average successful goals is to use one that already exists and play with it.

Theater Outcomes Based on Launch Date can be minipulated. First add currency and country to filters, just to be sure the data is representing Louise, though the initial data was filtered for this. After adding currency and country to filters filter currency to the British Pound and Country to Great Britian. Next place pledged in the Sum Values section and change the Vlaue Field Settings to Average of pledged and then format the results for currency without decimals. Next add Category/Subcategory to filters and filter for theater plays since the initial analysis was just plays. Sort outcomes to successful and the Pivot Table represents a grand total average gaol amount for all months since 2009 of 2364.72 British Pounds. The best month assumed to launch a campaign is May. The Pivot table now has a more specific presentation for Louise to consider. The highest total successful outcomes specifically for plays in Great Britian are in May and June, not just May as previously deduced. The average amount pledged in these successful months were 2.3K and 2.6K with exact numbers averaging 2487.46 British Pounds. This data leads to a recommendation that Louise start her goal in May or June, though May is better off for Theatree overall, and aroudn 2.4K -2.5K as a goal amount. I would then create a bar graph to go with the Pivot Table for better explanation that includes a visula and table for both of Louise’s questions at the same time.



<https://github.com/sarahalonzo/kickstarter-analysis/blob/14865911e217b5e8c5c27b21a4a7d7bd050d19b2/Deliverable_3_Addtl_Analysis/Plays_Pledge_Amount_and_Launch_Month.png>

Louise should also have an idea of how long she should run her campaign for or expect to based on length of campaigns compared to success and failure rates. In order to do this the launched and deadline dates would need to be formatted to successful and failed to limit some of the data size, Great Britain, and with the Category/Subcategory of Theater\_Plays. Next the dates for Launch Date and Deadline woud need to be converted to non Unix format. The difference between the launch and deadline would provide the length of the campaign. This could be turned into a pivot table where outcomes coud be sorted between successful and failed or shown at the same time and show the number of campaigns that were run for certain average lengths of time based on statistical analysis of the spread between the launch date and the deadline.