UWA Data Analytics Bootcamp Challenge 1 Report

1. **Introduction**

Crowdsourcing is a collaborative method where individuals or organisations obtain ideas, services, or content by soliciting contributions from a large group of people, typically from the online community. This approach leverages collective intelligence and creativity from a vast pool of participants to solve problems, create content, develop new technologies, or conduct research. Crowdsourcing can be applied in various domains, including but not limited to, innovation, problem-solving, fundraising (crowdfunding), data analysis, and content creation.

According to Aitamurto (2015) the term "crowdsourcing" was coined by Jeff Howe in a 2006 article for "Wired" magazine, where he explored how businesses were using the internet to outsource work to individuals. The concept has since evolved and expanded, influencing numerous fields and leading to the creation of platforms dedicated to crowdsourcing efforts, such as Waze and Kickstarter for crowdsourced traffic data, and Wikipedia for crowdsourced knowledge compilation.

Crowdsourcing is praised for its ability to gather a wide range of ideas and solutions, democratise participation, and accelerate research and development. However, it also faces criticism related to quality control, fairness to contributors, and potential exploitation of unpaid or underpaid labour.

As part of a University of Western Australia Data Analytics Bootcamp, a crowdsourcing dataset provided for review and analysis in Excel. The dataset comprises a comprehensive overview of various crowdsourced projects, capturing a snapshot of the dynamic landscape of crowdfunding initiatives. It includes essential metrics and attributes that define each project's journey, such as project names, brief descriptions (blurbs), funding goals, amounts pledged, percentage funded, outcomes (e.g., successful, failed), number of backers, average donation per backer, country of origin, and key dates related to project creation and completion. Additionally, the dataset categorizes projects into main and sub-categories, offering insights into the diversity of sectors that engage in crowdfunding efforts, ranging from technology, music, and food to theatre. This categorisation not only reflects the wide array of ideas that leverage crowdsourcing for funding and support but also facilitates an analysis of trends and success factors across different domains. The presence of fields like 'staff pick' and 'spotlight' further enriches the dataset by indicating projects that have received special recognition or visibility, potentially influencing their success rates.

This report provides a brief overview of the data analysis process, results and conclusions.

1. **Data Analysis Process**

The data was reviewed over a period of two days, following a methodology defined by the programme coordinators. Excel was used as the tool for this analysis process and the results are provided in a separate Excel file called “20240314 Crowdsourcing Book WWare 002.xlsx”. The file comprises multiple worksheets, and includes the instructions (Appendix A), the original data, and the various workings.

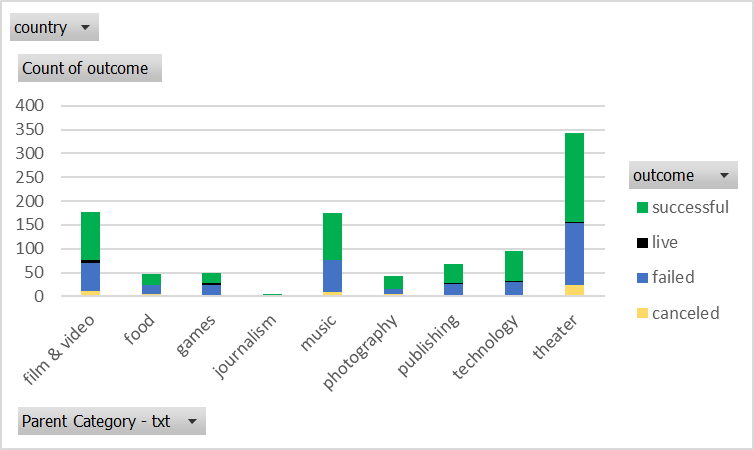
The dataset was modified to include calculated fields such as 'Percent Funded' and 'Average Donation', which provided insights into campaign performance relative to their funding goals and backer engagement. Projects were categorised into 'Parent Category' and 'Sub-Category' for a granular analysis. Conditional formatting highlighted campaign outcomes and funding success visually.

Pivot tables and charts were constructed to analyse the success rates per category and sub-category, filtered by country and parent category. Unix timestamps in 'deadline' and 'launched\_at' columns were converted to normal dates for better readability and to facilitate temporal analysis.

1. **Findings**

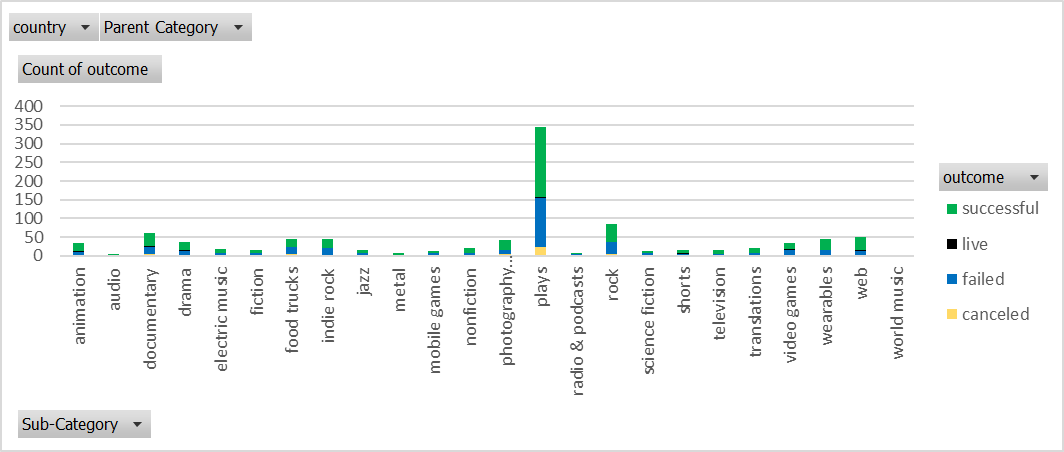
The key observations are as follows:

**Campaign Success by Category:** The distribution of outcomes across different categories shows a notable concentration of successful campaigns in the 'theater' and 'music' categories, whereas 'technology' has a significant number of failed campaigns (Figure 1). This suggests that projects related to performing arts may resonate more with crowdfunding audiences or that these domains may have more effective campaign strategies.



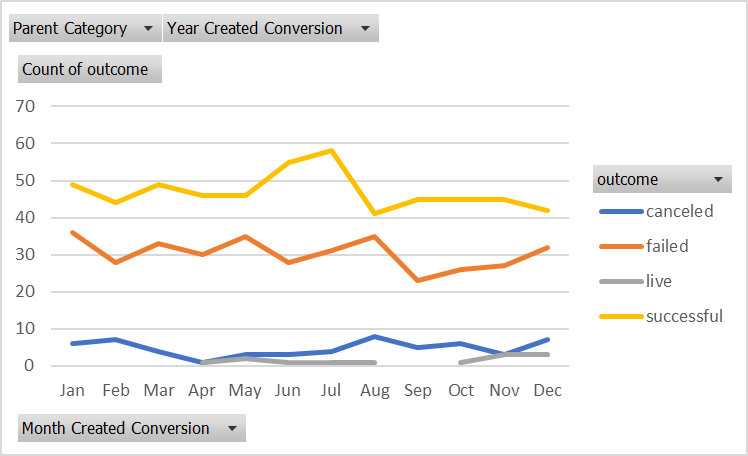
**Figure 1.** The distribution of outcomes across different categories.

**Sub-category Outcome Analysis:** Looking at the sub-categories, the data shows that specific niches like 'rock' within the music category have a higher success rate compared to others like 'web' within the technology category (Figure 2). These granular insights could guide creators in tailoring their projects to align with proven successful themes or identifying underserved niches with potential for growth.



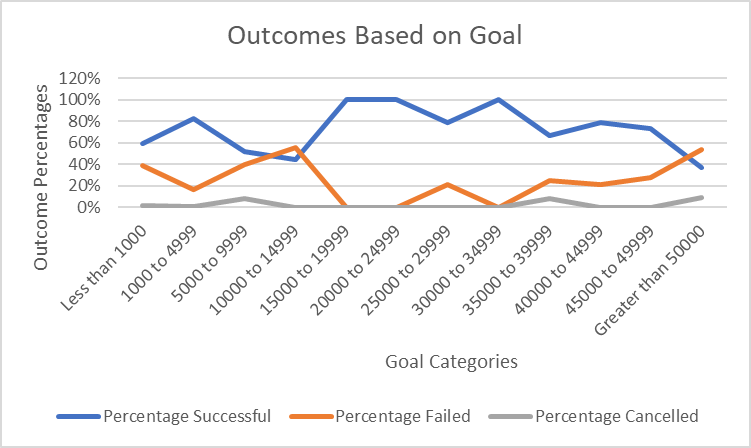
**Figure 2.** Sub-category outcome analysis.

**Temporal Patterns of Campaigns:** The temporal trend analysis indicates fluctuations in campaign outcomes throughout the year, with certain months showing higher numbers of failed campaigns. This pattern might be influenced by seasonal factors affecting backers' willingness to support projects or could be tied to the timing of campaign launches.



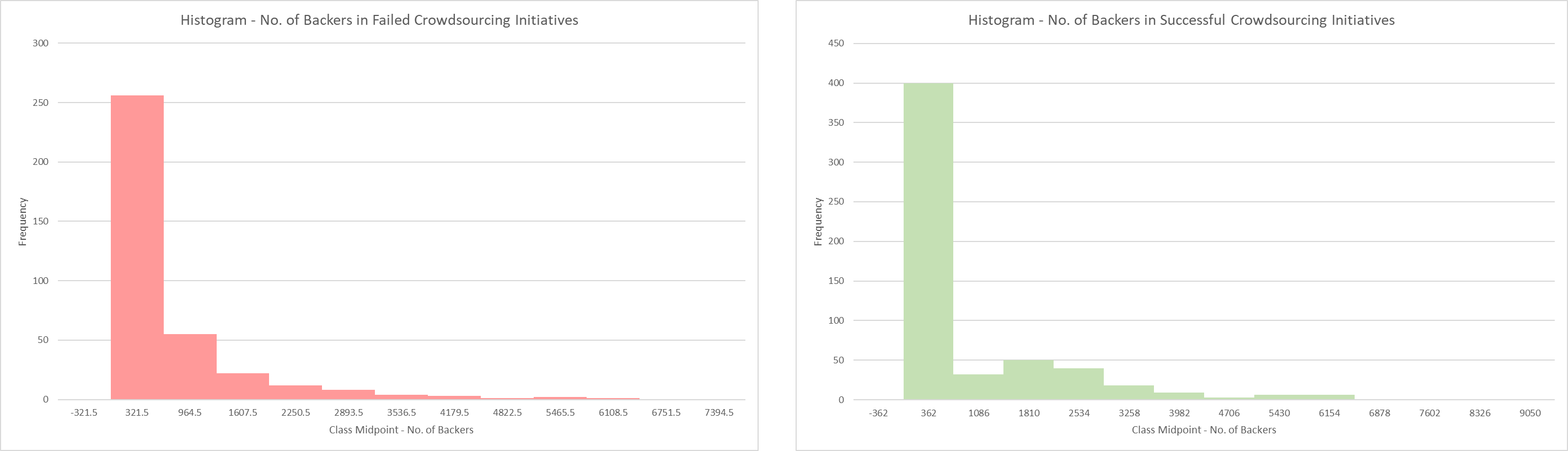
**Figure 3.** The temporal trend analysis by campaign outcome.

**Goal Setting and Outcomes:** The relationship between the set funding goals and outcomes indicates that campaigns with lower goals have a higher success rate, with diminishing success as goals increase. This trend underscores the importance of setting realistic funding goals that can be attractive to potential backers and achievable for the creators.



**Figure 4.** The relationship between the set funding goals and outcomes.

**Backer Engagement:** Histograms depicting the number of backers for failed and successful crowdfunding initiatives reveal that successful campaigns typically have a larger base of backers. This highlights the critical role of garnering substantial backer support in achieving crowdfunding success.



**Figure 5.** Histograms depicting the number of backers.

A statistical review of the number of backers in “successful” versus “unsuccessful” campaigns demonstrates that the group labelled as 'failed' has a variance of 924,113.45 and the group labelled as 'successful' has a variance of 1,606,216.59.

Variance is a measure of how spread-out numbers are. A higher variance indicates that the data points are more spread out from the mean, and from one another. Thus, with a higher variance value, the 'successful' campaigns exhibit more variability in the number of backers compared to the 'failed' campaigns.

This observation can make intuitive sense for several reasons:

* **Variety of Successful Campaigns:** Successful campaigns may vary widely in their appeal, goals, rewards, and promotion, which can attract a highly variable number of backers. Some successful campaigns may go viral and attract a very large number of backers, while others may just meet their goals, leading to a broader spread of data points.
* **Threshold Effect in Failed Campaigns:** There might be a threshold below which campaigns tend to fail, causing the number of backers to cluster within a smaller range below this threshold, thus resulting in less variability.
* **Resource Allocation:** Successful campaigns may have more resources to invest in marketing and promotion, potentially reaching a wider audience. This can lead to a larger and more variable number of backers as different marketing strategies and campaign qualities yield different results.
* **Engagement Levels:** Successful campaigns might engender differing levels of engagement from the backers. Some might incite a passionate response that results in high backer counts, while others may just have enough support to be successful.

In summary, the 'successful' campaigns showing more variability does make sense as success in crowdfunding is often influenced by a variety of factors that can lead to significant differences in the number of backers. This is indicative of the dynamic nature of crowdfunding, where different projects can have vastly different outcomes based on their content, promotion, and audience engagement.

1. **Proposed Additional Graphs**

Time limitations have restricted the ability to do an inclusive exploratory data analysis. For a more comprehensive understanding of the crowdfunding data, I would recommend adding the following graphs to enrich the insights, allowing for informed and value-based decision-making enabling both crowdfunding platforms and project creators to optimise their approach:

* **Cumulative Pledge Amount Over Time:** A line graph showing the cumulative pledge amount over the campaign duration to understand how funding momentum builds up for successful versus unsuccessful campaigns.
* **Backer Demographics Distribution:** If demographic data of backers is available, pie charts or bar graphs could display the distribution of backers by age, location, or other demographic factors.
* **Average Pledge by Category:** A bar graph showing the average amount pledged to campaigns in each category to see if certain types of campaigns tend to attract higher contributions.
* **Success Rate by Staff Pick and Spotlight:** A comparison chart showing the success rate of campaigns that were tagged as 'staff pick' or 'spotlight' versus those that were not to quantify the impact of these recognitions.
* **Correlation Heatmap:** A heatmap showing the correlation between variables like goal amount, number of backers, average donation, and percent funded. This could help identify which factors are most closely associated with campaign success.
* **Reward Levels Analysis:** If data on reward levels is available, a graph illustrating the most popular reward levels could provide insights into backer behavior and preferences.
* **Geographic Distribution of Campaigns:** A map visualization showing the geographic distribution of campaigns and their outcomes to identify potential regional market trends.
* **Time Series Analysis of Campaign Launches:** A time series graph showing the number of campaigns launched over time to spot trends and perhaps the best times of year to launch a campaign.
* **Word Cloud of Campaign Titles and Blurbs:** A word cloud generated from the titles and blurbs of successful versus unsuccessful campaigns to identify common keywords associated with each outcome.
* **Duration of Campaigns:** A histogram showing the length of campaign durations against their outcomes to determine if there's an optimal campaign length that correlates with success.

1. **Limitations of the data include:**

Herewith a non-exhaustive list of possible data limitations:

* **Temporal Coverage:** The dataset includes Year Created Conversion and Date Ended Conversion columns, but without context on the dataset's overall temporal range, it's challenging to assess trends over time or the impact of external factors (e.g. economic downturns) on crowdfunding success.
* **Lack of Detailed Project Description:** While there is a blurb column providing short project descriptions, the dataset lacks detailed information about the projects (e.g. full project descriptions, rewards offered, campaign duration). This limits a deeper understanding of what might contribute to a project's success or failure.
* **Categorical Broadness:** The categories (e.g. music, technology) may be too broad to draw specific conclusions about subsectors within these categories. Detailed analysis may require finer categorization or additional data on project specifics.
* **Missing Demographic and Geographic Data:** The dataset includes a country column but lacks more specific geographic or demographic data on project creators or backers. Such information could be valuable in analysing funding patterns and preferences among different groups.
* **Outcome Bias:** If the dataset primarily includes projects from a crowdfunding platform that publicly shares more successful projects than failed ones, it might bias analyses towards success factors and underrepresent the challenges faced by unsuccessful campaigns.

1. **Conclusions**

In conclusion, based on the analysis performed, these are the key observations:

**Category Resonance:** The 'theater' and 'music' categories are more successful in attracting crowdfunding support, whereas 'technology' projects face higher failure rates. This indicates that certain categories naturally elicit more enthusiasm and backing from the crowdfunding community.

**Goal Realism and Success:** There is a clear trend that lower funding goals are more often met, underscoring the need for realistic and attainable financial targets. Setting a goal too high can be a deterrent to potential backers, affecting the overall success of the campaign.

**Backer Engagement:** The success of crowdfunding campaigns is closely tied to the number of backers, with successful projects typically enjoying wider support. Effective engagement with the community is pivotal to securing the necessary backing to meet funding goals.

Based on these outcomes, the following is recommended:

* **Strategic Goal Setting:** Campaign creators should be equipped with data-driven insights to inform their goal-setting process, leveraging historical data on successful campaigns for comparable projects to inform their funding targets.
* **Focused Community Building:** Enhanced engagement strategies, particularly for underperforming categories like 'technology', should be developed. This could involve creating stronger narratives, showcasing potential impact, and fostering a sense of community around the project's goals.
* **Enhanced Campaign Support:** Crowdfunding platforms might provide additional resources and visibility options such as 'staff pick' and 'spotlight' to guide project creators towards crafting higher-quality and more appealing campaigns, especially in categories that struggle to attract funding.

Implementing these strategies will help both crowdfunding platforms and project creators optimise their approach, leading to improved outcomes and a more robust crowdfunding environment.

1. **References**

Howe, J. (2006). "The Rise of Crowdsourcing." Wired. This seminal article introduced the term crowdsourcing and discussed its potential to transform various industries by leveraging the power of the crowd.

Aitamurto, T. (2015). "Motivation Factors in Crowdsourced Journalism: Social Impact, Social Change, and Peer Learning." International Journal of Communication, 9, 3523–3543. This research explores the motivations behind participation in crowdsourced journalism, highlighting the diverse reasons individuals contribute to crowdsourced projects.

**Appendices**

**Appendix A.** Instructions provided for the University of Western Australia Bootcamp Challenge 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Module 1 Challenge  Start Assignment   * + **Due** Thursday by 23:59      * + **Points** 100      * + **Submitting** a text entry box or a website url   **Background**  Crowdfunding platforms like Kickstarter and Indiegogo have been growing in success and popularity since the late 2000s. From independent content creators to famous celebrities, more and more people are using crowdfunding to launch new products and generate buzz, but not every project has found success.  To receive funding, the project must meet or exceed an initial goal, so many organisations dedicate considerable resources looking through old projects in an attempt to discover “the trick” to finding success. For this week's Challenge, you will organise and analyse a database of 1,000 sample projects to uncover any hidden trends.  **Before You Begin**   * + Create a new space for this project called excel-challenge in either Dropbox or Google Drive. **Do not add this assignment to an existing repository**.   + Store your Excel workbooks here in this new space, and create a sharable link for submission.   **Files**  Download the following files to help you get started:  [Module 1 Challenge filesLinks to an external site.](https://static.bc-edx.com/data/dla-1-2/m1/lms/starter/Starter_Code.zip)  **Instructions**  A table contains a database of 1,000 sample crowdfunding projects.  Using the Excel workbook in your .zip file, modify and analyse the sample-project data and try to uncover market trends.   * + Use conditional formatting to fill each cell in the outcome column with a different colour, depending on whether the associated campaign was successful, failed, cancelled, or is currently live.   + Create a new column called Percent Funded that uses a formula to find how much money a campaign made relative to its initial funding goal.   + Use conditional formatting to fill each cell in the Percent Funded column according to a three-colour scale. The scale should start at 0 with a dark shade of red, and it should transition to green at 100 and blue at 200.   + Create a new column called Average Donation that uses a formula to find how much each project backer paid on average.   + Create two new columns, one called Parent Category and another called Sub-Category, that use formulas to split the Category and Sub-Category column into the two new, separate columns.   Category Stats   * + Create a new sheet with a pivot table that analyses your initial worksheet to count how many campaigns were successful, failed, cancelled, or are currently live per **category**.   + Create a stacked-column pivot chart that can be filtered by country based on the table that you created.   Subcategory Stats   * + Create a new sheet with a pivot table that analyses your initial sheet to count how many campaigns were successful, failed, or cancelled, or are currently live per **sub-category**.   + Create a stacked-column pivot chart that can be filtered by country and parent category based on the table that you created.   + The dates in the deadline and launched\_at columns use Unix timestamps. Fortunately for us, [this formulaLinks to an external site.](https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html) that can be used to convert these timestamps to a normal date.   + Create a new column named Date Created Conversion that will use [this formulaLinks to an external site.](https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html) to convert the data contained in launched\_at into Excel's date format.   + Create a new column named Date Ended Conversion that will use [this formulaLinks to an external site.](https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html) to convert the data contained in deadline into Excel's date format.   Outcomes Based on Launch Date   * + Create a new sheet with a pivot table that has a column of outcome, rows of Date Created Conversion, values based on the count of outcome, and filters based on parent category and Years.   + Now, create a pivot-chart line graph that visualises this new table.   + Create a report in Microsoft Word, and answer the following questions:   + Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?   + What are some limitations of this dataset?   + What are some other possible tables and/or graphs that we could create, and what additional value would they provide?   **Bonus**   * + Create a new sheet with 8 columns:   + Goal   + Number Successful   + Number Failed   + Number Cancelled   + Total Projects   + Percentage Successful   + Percentage Failed   + Percentage Cancelled   + In the Goal column, create 12 rows with the following headers:   + 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   + Greater than or equal to 50000   A table and corresponding graph showing the percentage of projects that are successful, failed, and cancelled based on their crowdfunding goal.   * + Using the COUNTIFS() formula, count how many successful, failed, and cancelled projects were created with goals within the ranges listed above. Populate the Number Successful, Number Failed, and Number Cancelled columns with these data points.   + Add up each of the values in the Number Successful, Number Failed, and Number Cancelled columns to populate the Total Projects column. Then, using a mathematical formula, find the percentage of projects that were successful, failed, or cancelled per goal range.   + Create a line chart that graphs the relationship between a goal amount and its chances of success, failure, or cancellation.   **Bonus Statistical Analysis**  Most people would use the number of campaign backers to assess the success of a crowdfunding campaign. Creating a summary statistics table is one of the most efficient ways that data scientists can characterise quantitative metrics, such as the number of campaign backers.  For an additional challenge, evaluate the number of backers of successful and unsuccessful campaigns by creating **your own** summary statistics table.   * + Create a new worksheet in your workbook, and create one column for the number of backers of successful campaigns and one column for unsuccessful campaigns.   A table containing a column for the number of backers of successful campaigns and a column for unsuccessful campaigns.   * + Use Excel to evaluate the following values for successful campaigns, and then do the same for unsuccessful campaigns:   + The mean number of backers   + The median number of backers   + The minimum number of backers   + The maximum number of backers   + The variance of the number of backers   + The standard deviation of the number of backers   + Use your data to determine whether the mean or the median better summarises the data.   + Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?   **Requirements**  **Conditional Formatting (10 points)**   * + Conditional formatting is applied appropriately to the outcome column (5 points)   + Conditional formatting is applied appropriately to the percent funded column (5 points)   **Column Creation (10 points)**   * + Six new columns were correctly created for:   + percent funded   + average donation   + category   + sub-category   + Date Created Conversion   + Date Ended Conversion   **Pivot Tables and Stacked Column Charts (15 points)**   * + Correctly created a pivot table that counts how many campaigns were "successful," "failed," "cancelled," or are currently "live" per category (7.5 points)   + Correctly created a stacked column pivot chart that can be filtered by country (7.5 points)   **Pivot Tables and Line Graphs (15 points)**   * + Correctly created a pivot table with a column of outcome, rows of Date Created Conversion, values based on the count of outcome, and filters based on parent category and Years (7.5 points)   + Correctly created a pivot chart line graph (7.5 points)   **Written Report (20 points)**   * + Presents a cohesive written analysis that:   + Draws three conclusions from the data (10 points)   + States limitations of the dataset and suggestions for additional tables of graph (10 points)   **Crowfunding Goal Analysis (10 points)**   * + Computed calculations of percentages for projects that were successful, failed, or were cancelled per goal range (5 points)   + Created a line chart showing the relationship between the goal’s amount and its chances at success, failure, or cancellation (5 points)   **Statistical Analysis (20 points)**   * + Computed calculations of the mean, median, min, max, variance, and stdev using Excel formulas (15 points)   + A brief and compelling justification of whether the mean or median better summarises the data (5 points)   **Grading**  This assignment will be evaluated against the requirements and assigned a grade according to the following table:   |  |  | | --- | --- | | **Grade** | **Points** | | A (+/-) | 90+ | | B (+/-) | 80–89 | | C (+/-) | 70–79 | | D (+/-) | 60–69 | | F (+/-) | < 60 |   **Submission**  To submit your Challenge assignment, click Submit, and then provide the URL to your Dropbox or Google Drive folder for grading.  **NOTE**  You are allowed to miss up to two Challenge assignments and still earn your certificate. If you complete all Challenge assignments, your lowest two grades will be dropped. If you wish to skip this assignment, click Next, and move on to the next Module.  Comments are disabled for graded submissions in BootCamp Spot. If you have questions about your feedback, please notify your instructional staff or your Student Success Advisor. If you would like to resubmit your work for an additional review, you can use the Resubmit Assignment button to upload new links. You may resubmit up to three times for a total of four submissions.  **IMPORTANT**  No matter how difficult the course becomes, you must always turn in original work. Plagiarism is not tolerated. If your instructional or support staff determine that you have plagiarized work, your Student Success Advisor will determine the appropriate course of action based on university policy. Such actions may include, but are not limited to, a documented plagiarism discussion, an incomplete or failing grade assignment, or ineligibility for graduation.  **It is your responsibility to include a note in the README section of your repo specifying code source and its location within your repo**. This applies if you have worked with a peer on an assignment, used code in which you did not author or create sourced from a forum such as Stack Overflow, or you received code outside curriculum content from support staff such as an Instructor, TA, Tutor, or Learning Assistant. This will provide visibility to grading staff of your circumstance in order to avoid flagging your work as plagiarized.  If you are struggling with a challenge assignment or any aspect of the academic curriculum, please remember that there are student support services available for you:   * + Ask the class Slack channel/peer support.   + AskBCS Learning Assistants exists in your class Slack application.   + Office hours facilitated by your instructional staff before and after each class session.   + [Tutoring GuidelinesLinks to an external site.](https://docs.google.com/document/d/1hTldEfWhX21B_Vz9ZentkPeziu4pPfnwiZbwQB27E90/edit?usp=sharing) - schedule a tutor session in the Tutor Sessions section of Bootcampspot - Canvas   + If the above resources are not applicable and you have a need, please reach out to a member of your instructional team, your Student Success Advisor, or submit a support ticket in the Student Support section of your BCS application.   **References**  Data for this dataset was generated by edX Boot Camps LLC, and is intended for educational purposes only.  [Previous](https://bootcampspot.instructure.com/courses/5205/modules/items/1224114)[Next](https://bootcampspot.instructure.com/courses/5205/modules/items/1224117)  © 2024 edX Boot Camps LLC    *From <*[*https://bootcampspot.instructure.com/courses/5205/assignments/76292?module\_item\_id=1224116*](https://bootcampspot.instructure.com/courses/5205/assignments/76292?module_item_id=1224116)*>* |