Data Analytics Report – HW1

by Ugur Cetin / Due Date: 21OCT24

1. Introduction

Crowdfunding platforms such as Kickstarter and Indiegogo have become increasingly popular since the late 2000s, enabling creators, entrepreneurs, and even celebrities to launch new products and ideas. These platforms offer a unique opportunity for individuals to raise capital by directly engaging with their audience. However, not all projects achieve their funding goals, and understanding the factors that contribute to a project's success or failure is critical for future campaigns.

For this assignment, the focus is on analyzing a database of 1,000 sample crowdfunding projects. By exploring this data, the goal is to identify patterns, trends, and key factors that may influence a project's success. This analysis will provide insights into what differentiates successful campaigns from those that fall short, helping both creators and organizations make data-driven decisions in their future crowdfunding efforts.

2. Data Description

The dataset provided for this homework consists of 1,000 sample crowdfunding projects, supplied as an Excel file. It contains key variables relevant to project outcomes and financial performance, which are essential for understanding the trends in successful crowdfunding campaigns. The primary variables included are:

Goal: The target amount of funding (donation) the project aimed to raise.

Pledged: The total amount of money pledged by backers.

Date: The date when the campaign was launched or ended.

Outcome: The result of the campaign (e.g., successful, failed).

Category: The broad classification of the project (e.g., Technology, Art).

Sub-category: A more specific classification within the main category (e.g., Wearables under Technology).

Since this dataset is provided as part of a homework assignment, detailed information on the data collection method is not available and not critical for the scope of this task. However, the data offers a valuable opportunity to explore trends and patterns within a range of crowdfunding campaigns.

3. Data Cleaning and Preprocessing

- Separation of Category and Sub-category:
    - A new column was created for each, splitting the original data into two distinct columns to better analyze trends at both the category and sub-category levels.
- Date Transformation:
    - The date variable was reformatted to **standard calendar format**, ensuring consistency and enabling time-based analysis.
- Visual Enhancements with Conditional Formatting:
    - To make the raw data more visual and easier to interpret, **color conditioning** was applied to the **percent funded** and **outcome** columns. This highlights key metrics and outcomes at a glance.
- Creation of Additional Sheets and Analysis Tables:
    - As part of the assignment, the following sheets were added to the Excel workbook, each containing relevant data and visualizations:
        - PTable_1
        - PTable_2
        - Table_3
        - Crowdfunding Goal Analysis
        - Backers Data

- These sheets include **pivot tables, charts, and other analyses**, providing deeper insights into the data.

No further cleaning, such as handling **missing values** or **outliers**, was required, as the provided data was used as-is.

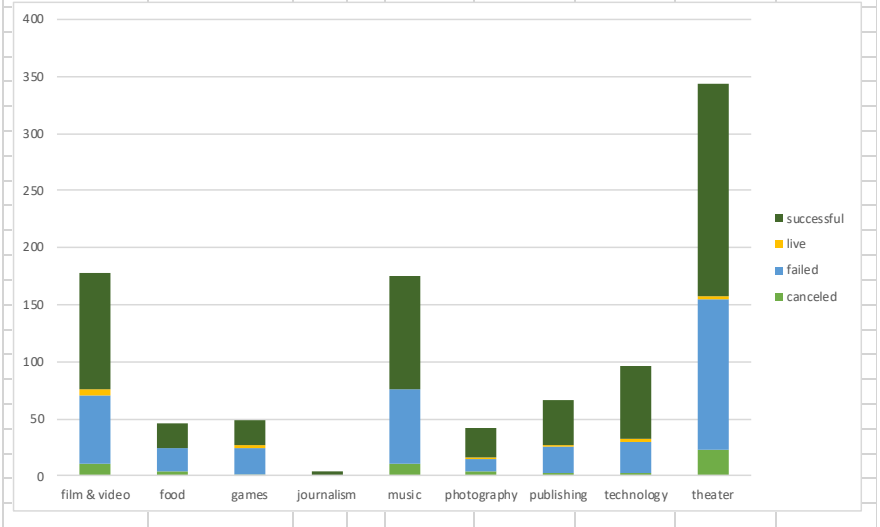
4. HW Questions

- **Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

The three key conclusions from the charts we created are as follows:

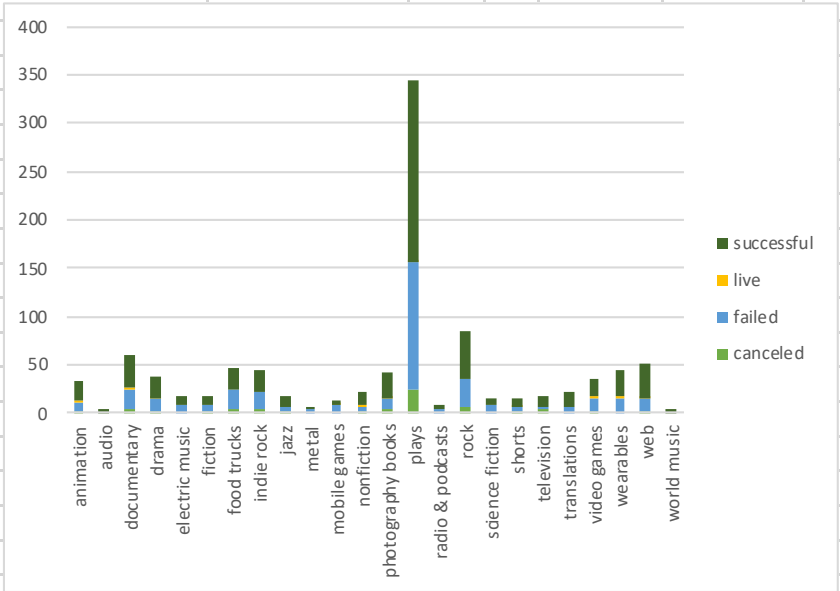
1. Based on **PTable_1**, the overall success rate of the projects is approximately **50%**. This success rate does not appear to be strongly influenced by project categories, or at least, the dataset does not provide sufficient evidence to establish a direct relationship between categories and success rates.

Count of outcome	Column Labels				
Row Labels	**canceled**	**failed**	**live**	**successful**	**Grand Total**
film & video	11	60	5	102	178
food	4	20		22	46
games	1	23	3	21	48
journalism				4	4
music	10	66		99	175
photography	4	11	1	26	42
publishing	2	24	1	40	67
technology	2	28	2	64	96
theater	23	132	2	187	344
Grand Total	**57**	**364**	**14**	**565**	**1000**



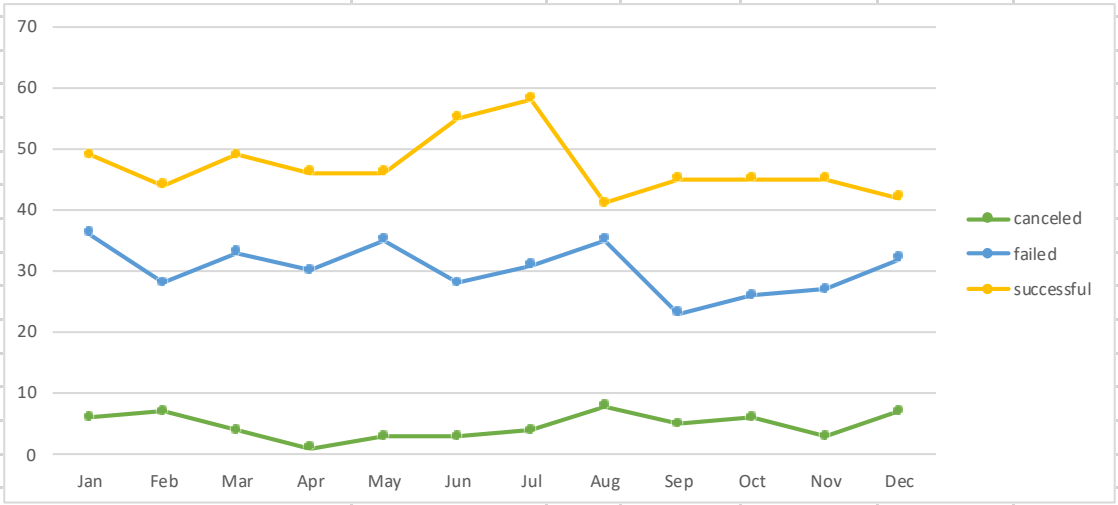
Based on **PTable_2**, like PTable_1, the success rate is approximately **50%**. However, the **PLAYS** sub-category has the highest number of projects.

Count of outcome	Column Labels				
Row Labels	**canceled**	**failed**	**live**	**successful**	**Grand Total**
animation	1	10	2	21	34
audio				4	4
documentary	4	21	1	34	60
drama	2	12	1	22	37
electric music		8		10	18
fiction	1	7		9	17
food trucks	4	20		22	46
indie rock	3	19		23	45
jazz	1	6		10	17
metal		3		4	7
mobile games		8	1	4	13
nonfiction	1	6	1	13	21
photography books	4	11	1	26	42
plays	23	132	2	187	344
radio & podcasts		4		4	8
rock	6	30		49	85
science fiction		9		5	14
shorts	1	5	1	9	16
television	3	3		11	17
translations		7		14	21
video games	1	15	2	17	35
wearables		16	1	28	45
web	2	12	1	36	51
world music				3	3
Grand Total	**57**	**364**	**14**	**565**	**1000**



2. The **PTable_3** indicates a direct relationship between the success rate and the month of the project. The success rate peaks in **July**, with a noticeable increase beginning in **May** and reaching its highest point in **July**.

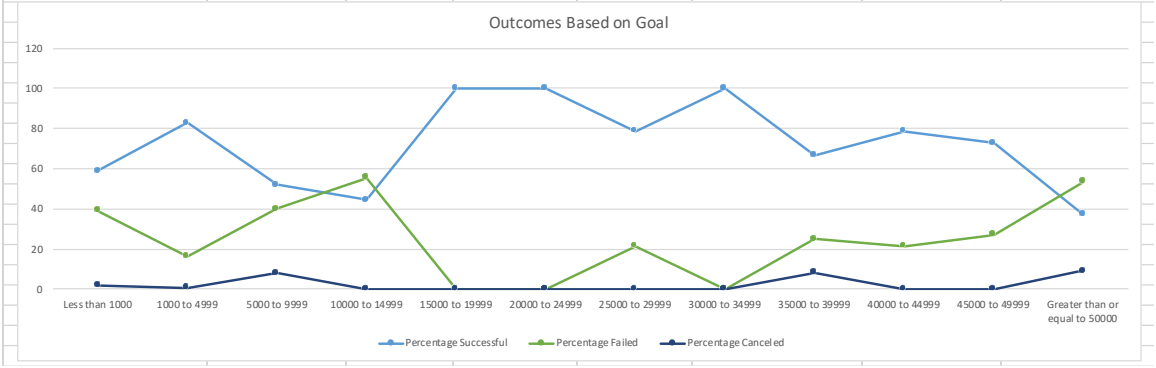
Count of outcome	Column Labels			
Row Labels	**canceled**	**failed**	**successful**	**Grand Total**
Jan	6	36	49	91
Feb	7	28	44	79
Mar	4	33	49	86
Apr	1	30	46	77
May	3	35	46	84
Jun	3	28	55	86
Jul	4	31	58	93
Aug	8	35	41	84
Sep	5	23	45	73
Oct	6	26	45	77
Nov	3	27	45	75
Dec	7	32	42	81
Grand Total	**57**	**364**	**565**	**986**



3. The Outcomes Based on Goal chart indicates that the success rate is higher for projects ranging from 15,000 to 35,000. However, the currency for these values is not always in USD. To properly analyze the data, we should add a column to convert all amounts to a common currency. **We cannot include this statement in our official report until we standardize the currency.**

4.

Goal	Number Successful	Number Failed	Number Canceled	Total Projects	Percentage Successful	Percentage Failed	Percentage Canceled
Less than 1000	30	20	1	51	59	39	2
1000 to 4999	191	38	2	231	83	16	1
5000 to 9999	164	126	25	315	52	40	8
10000 to 14999	4	5	0	9	44	56	0
15000 to 19999	10	0	0	10	100	0	0
20000 to 24999	7	0	0	7	100	0	0
25000 to 29999	11	3	0	14	79	21	0
30000 to 34999	7	0	0	7	100	0	0
35000 to 39999	8	3	1	12	67	25	8
40000 to 44999	11	3	0	14	79	21	0
45000 to 49999	8	3	0	11	73	27	0
Greater than or equal to 50000	114	163	28	305	37	53	9



- **What are some limitations of this dataset?**

This data set is not yet to analyze how can we increase the success rate. It is enough to see what happened or the outcome of the donations but doesn't give us hints on how to improve it or do it better next time.

- **What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

Geographic data is crucial for analyzing these factors. In my opinion, we need more information about the location, age groups, and ethnicities to gain a better understanding of what contributes to success rates. Additionally, it's important to know what types of marketing activities have been undertaken and what can be done to further enhance the success rate.

5. Statistical Analysis / Model Building

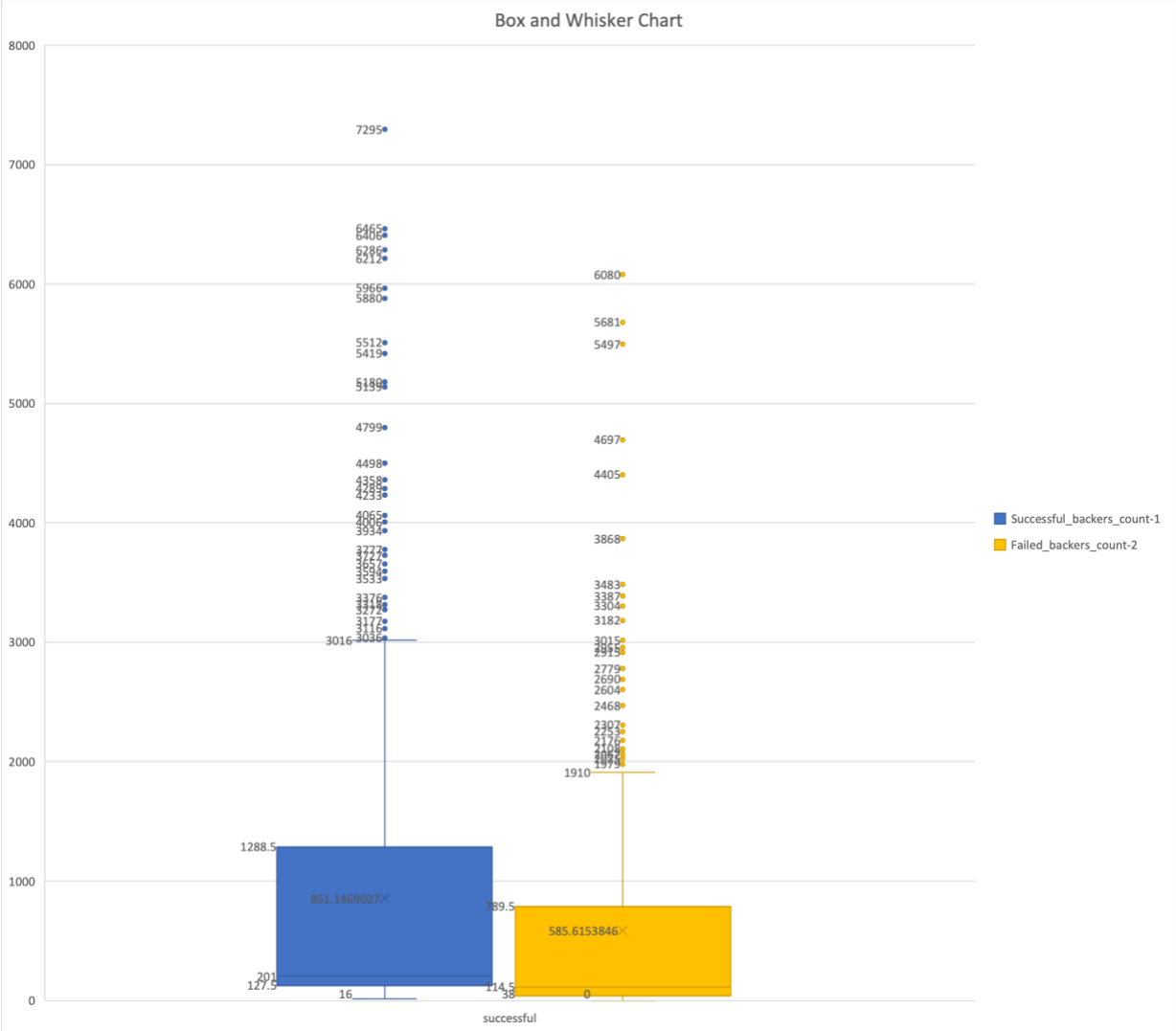
The **Box-and-Whisker chart** compares the distribution of the number of backers for **successful** and **failed** crowdfunding projects. Here are the key insights:

1. **Mean & Median Backers**:
    - **Successful projects** (in blue) have a higher mean & median backer count (~851 & 201) compared to **failed projects** (in yellow) with a mean of around 586, and a median of ~115.
    - This indicates that successful projects generally attract more backers than failed ones.
2. **Spread (Interquartile Range, IQR)**:
    - The range of backers (difference between Q1 and Q3) for successful projects is wider, showing more variability in the number of supporters.
    - Failed projects have a smaller IQR, suggesting less variation in their backer count.
3. **Outliers**:
    - Both successful and failed campaigns contain several **outliers**, indicating some projects attracted a significantly higher number of backers. For example, successful campaigns show extreme outliers with **7,259 backers** at the upper end.
4. **Range and Whiskers**:
    - Successful projects show a broader range, with some campaigns having more than 6,000 backers.
    - In contrast, failed projects typically attract fewer backers, with the whisker extending only up to around 1,910.

5. **Skewness**:
    - Both distributions appear **right-skewed**, meaning most projects attracted relatively fewer backers, but a few projects had an exceptionally high number of supporters.

Conclusion:

The chart highlights that **successful campaigns tend to have more backers** overall, with higher variability in backer counts. There are some exceptional outliers in both successful and failed projects, but successful campaigns generally perform much better in terms of backer engagement.



	Successful	Failed
Mean	851	586
Median	201	115
Min	16	0
Max	7295	6080
Variance	1603374	921575
SD	1266	960