COMP-590 Project 2 Final Report

Wanyi Chen, Mila Korobkina, Caroline Lu, and Saumya Ray

Kickstarter Data Analysis

We used a dataset created by Mickaël Mouillé that presents data collected from the Kickstarter platform, which is a website that allows users to crowdfund their creative projects. We found this dataset using Kaggle.com and it is available here: http://bit.ly/590ProjectData. The dataset examines 14 attributes of Kickstarter campaigns. These attributes include: campaign name, campaign category, monetary campaign goal, campaign deadline, amount of money pledged campaign state (success or failure), and number of backers.

We hope to examine the following research questions:

1. Does the campaign length impact the success rate?
2. What is the relationship between the goal ($) and the success of the campaign?
    a. Distribution of (goal ($) - raised amount ($))
3. What is the relationship between a single pledge amount and goal (%)?
    a. Do people pay more in expensive campaigns?
    b. Do people pay differently in different countries?
4. How does the data vary across different categories of campaigns?
    a. Which category has a higher success rate? (break down by countries) ?
    b. Which categories have higher pledges/goals? Lower pledges/goals?

We assigned one research question to each team member, and the following sections describe our team's progress in answering the research questions.

The first two questions look at effects of different fields on success rates, while the second and third questions also look at campaign goal relationships. The third and fourth questions overlap in their analysis of countries and categories. Overall, kickstarters are not that successful and there is a fair amount of variation with what types of campaigns succeed geographically as well as fiscally. There are many possible relationships to explore within this dataset and this analysis could help kickstarter become a more effective fundraising platform.

Campaign Length and Success Rate

Caroline Lu

I will be analyzing the relationship between campaign length and the success of the campaign. The first task I needed to complete was moving the dataset to moxie. Once I had done that, I needed to decide what sort of MapReduce programs I need to write. I determined that I need the following fields from the dataset: ID, Deadline, Launched, State, Currency, Goal, and Pledged. I modified a MapReduce program to output those six fields. This program will be fairly similar to the HostCount example. I will use the command line, Matlab, and Spark to calculate the duration of the campaigns and to complete further analysis.

We ran into many issues while attempting to convert the dataset into a usable format. Once I had finally converted my data to a usable format, I wanted to compare campaign length to the state of the campaign (success, fail, or canceled). One annoying characteristic of the dataset is that there were 10-15 records scattered throughout which have 1/1/70 as a launch date, which is obviously incorrect. I have no idea what caused this anomaly, but deleted all of those records.

Category	Campaign Length
1	≤ 7 days
2	8-22 days
3	23-37 days
4	38-52 days
5	53-67 days
6	68-82 days
7	83-92 days

I divided my data based on these seven categories. I wanted the first category to encompass only seven days so that I would be able to draw conclusions about short campaigns. Similarly, the last category encompasses only 9 days. All other categories cover 14 days.

There are many conclusions that can be drawn from the following graphs, descriptive statistics, and CDF plots. Below is a selection of interesting deductions.

- The majority of all campaigns lasted 23-37 days.

- 89.7% of campaigns that lasted 83-92 days and 86% of campaigns that lasted 68-82 days failed, suggesting that campaigns that succeed do not last as long as campaigns that fail. 0.7% of successful campaigns lasted 83-92 days, while 1.1% of failed campaigns lasted 83-92 days.

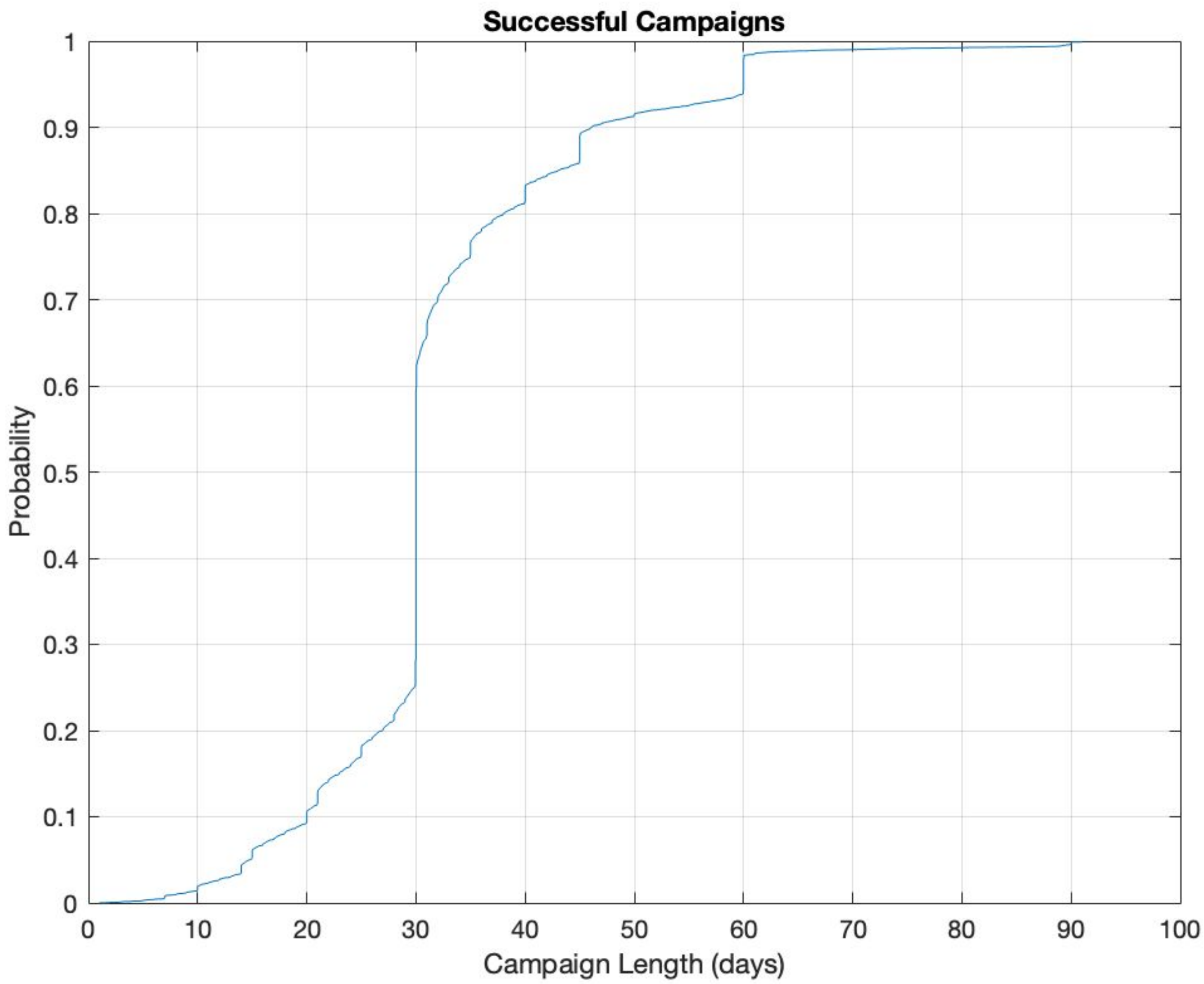
- In comparison, only 42.3% of campaigns that lasted 8-22 days failed, and only 51.5% of campaigns that lasted 23-37 days failed.

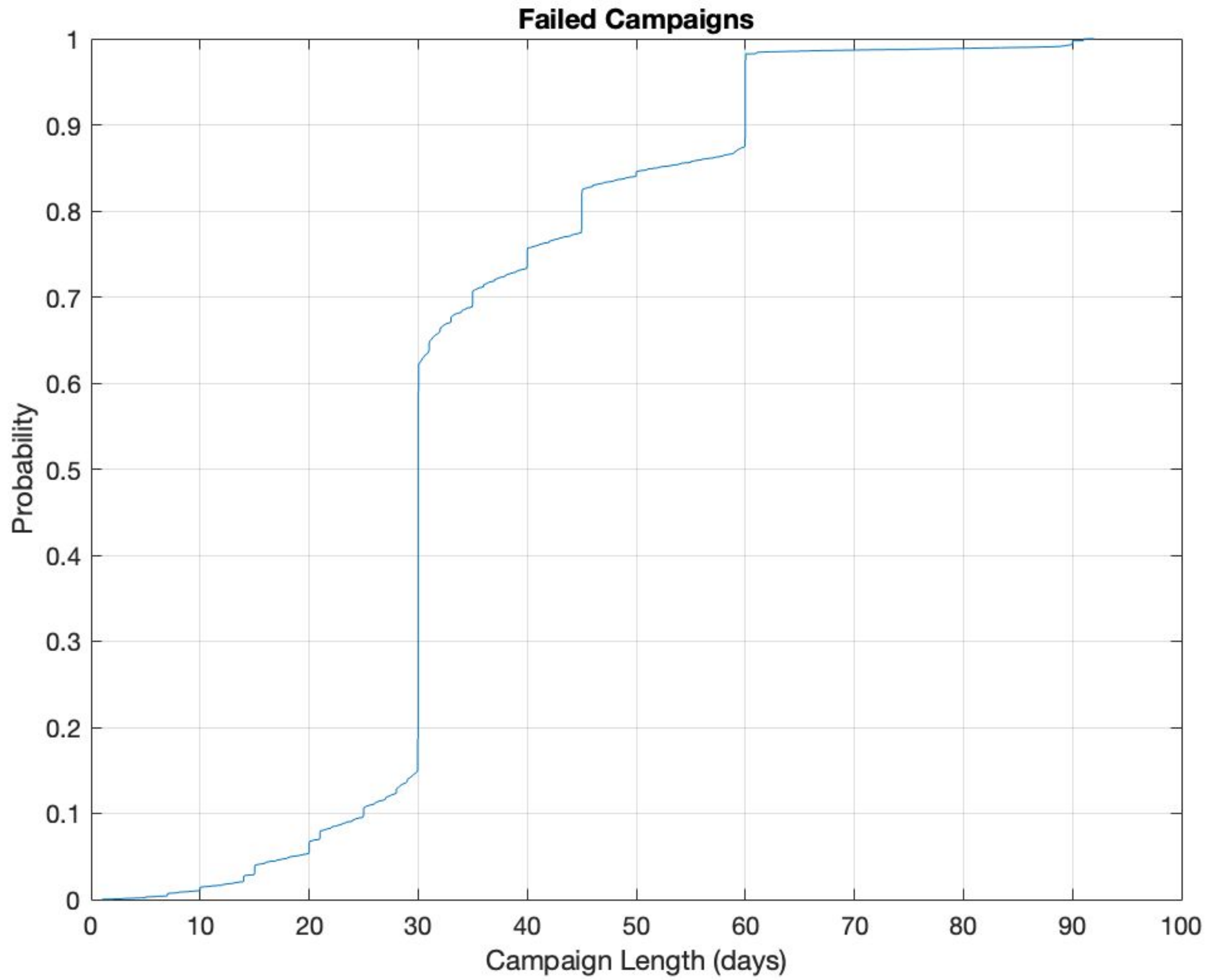
- Only 3.1% of campaigns that lasted 83-92 days and 3.2% of campaigns that lasted 68-82 were canceled, suggesting that even though the campaigns did not succeed, the organizers chose to continue the campaign until its deadline, rather than cancel.

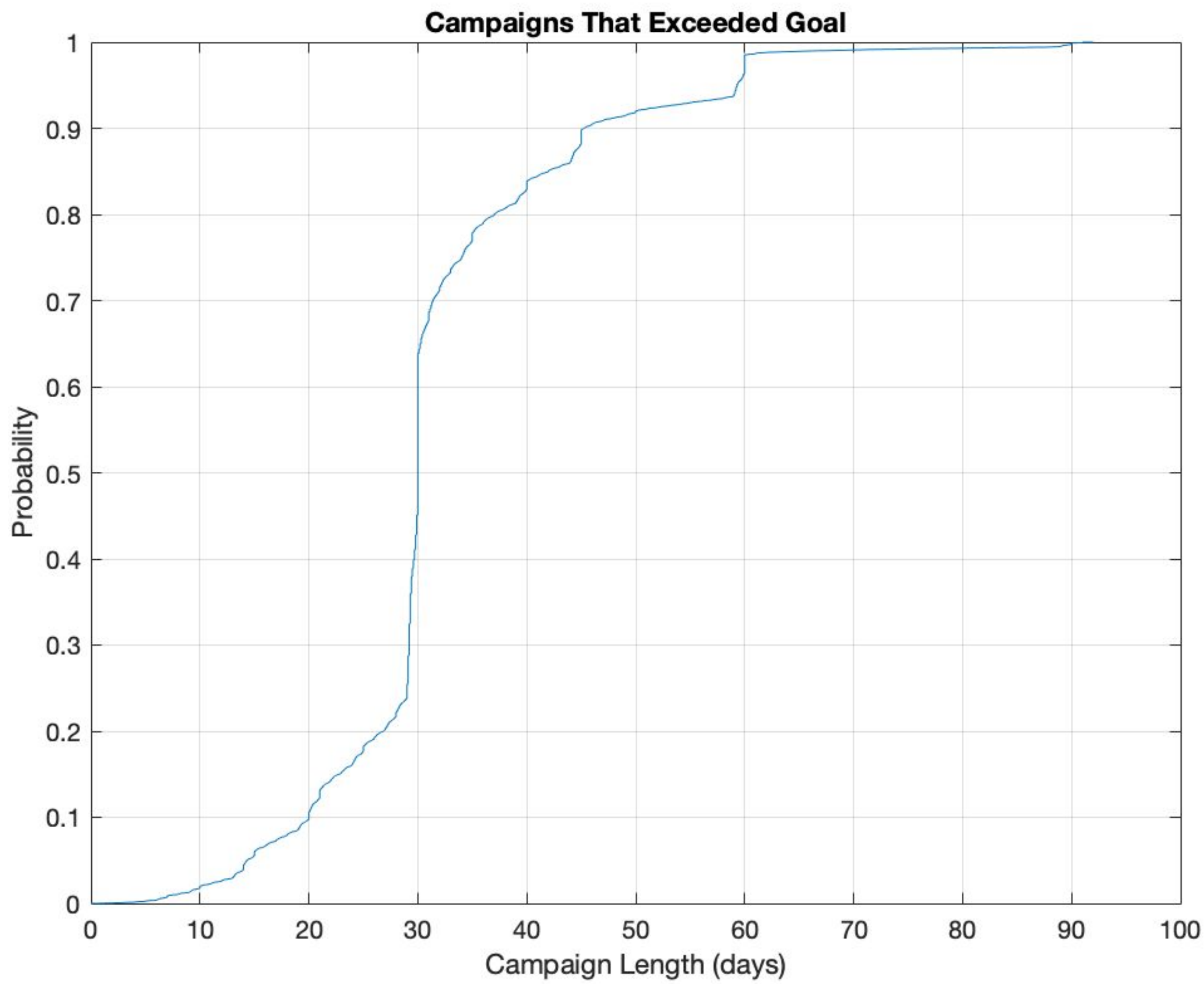
- In comparison, 12.8% of campaigns that lasted 53-67 days, 10.6% of campaigns that lasted 38-52 days, and 9.6% of campaigns that lasted 23-37 days were canceled.

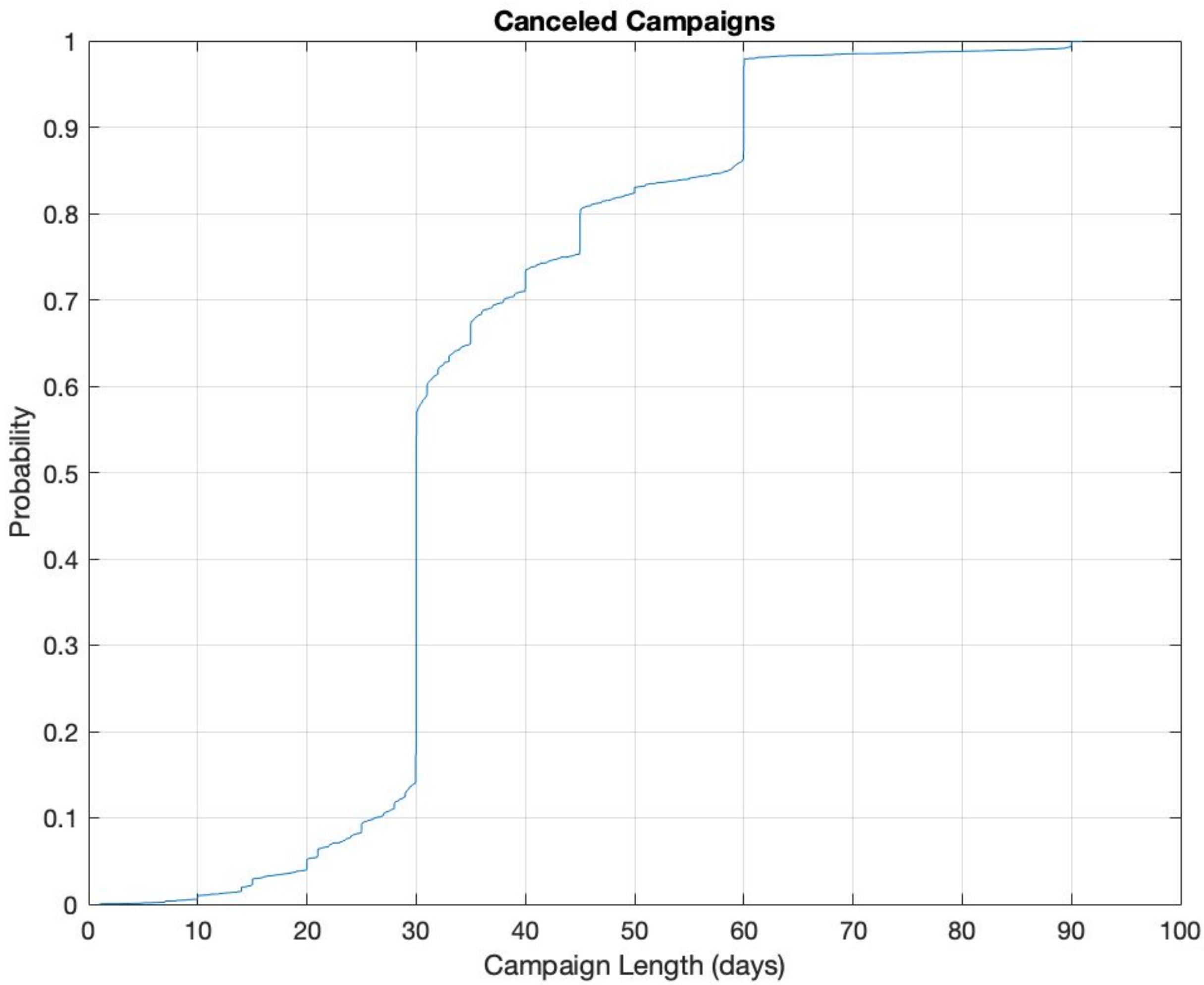
- The histograms for lengths of successful, failed, and canceled campaigns are surprisingly similar. The counts steadily increase until around 30 days, where there is the largest spike. After approximately 30 days, the counts drops down around 40 days, then drops down even lower around 50 days, and then increases around 60 days. On all three histograms, we can also see an increase around 90 days.

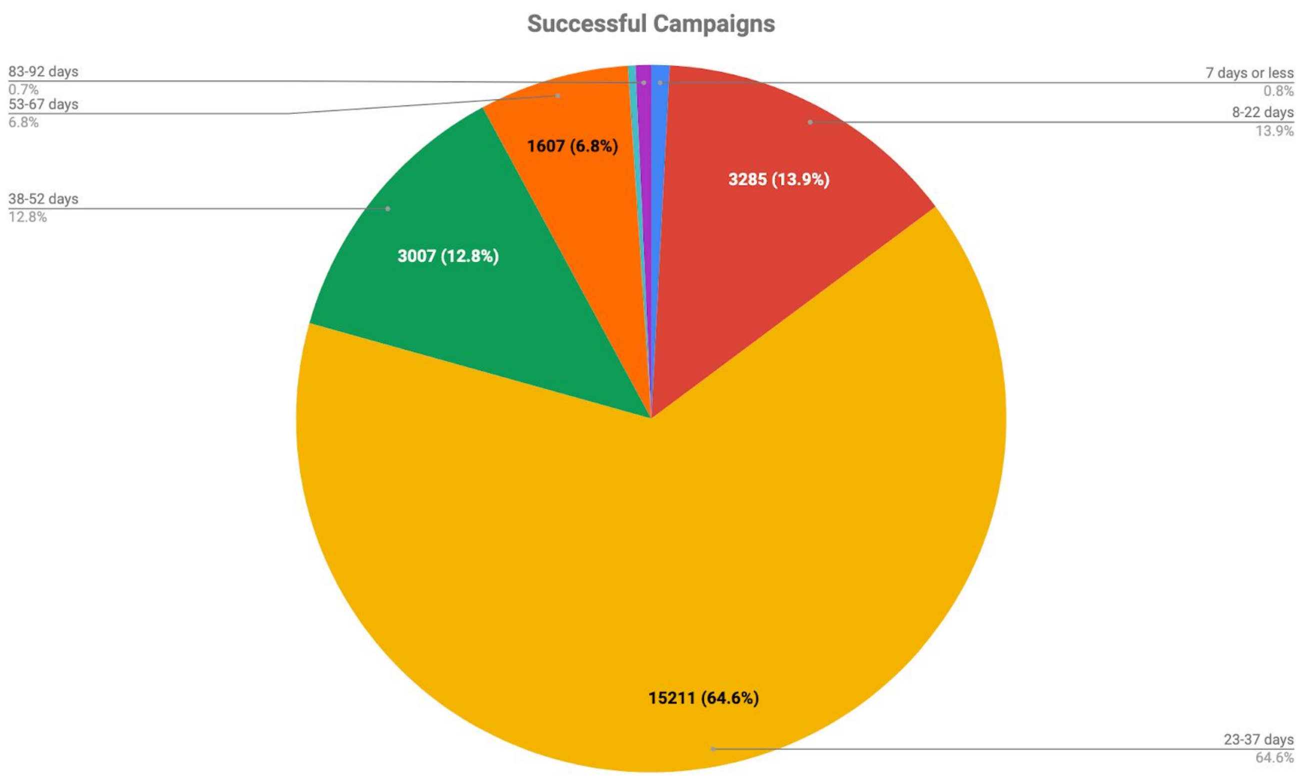
Measure	Value
Median	30 days
Mean	34 days
Mode	30 days
Variance	169.79 days
Standard Deviation	13.03 days
Coefficient of Variation	38.32
Minimum	0.01 days (14.4 minutes)
Maximum	92 days
First quartile (0.25)	30 days
Second quartile (0.5)	30 days
Third quartile (0.75)	38.24 days
Fourth quartile (1)	92 days

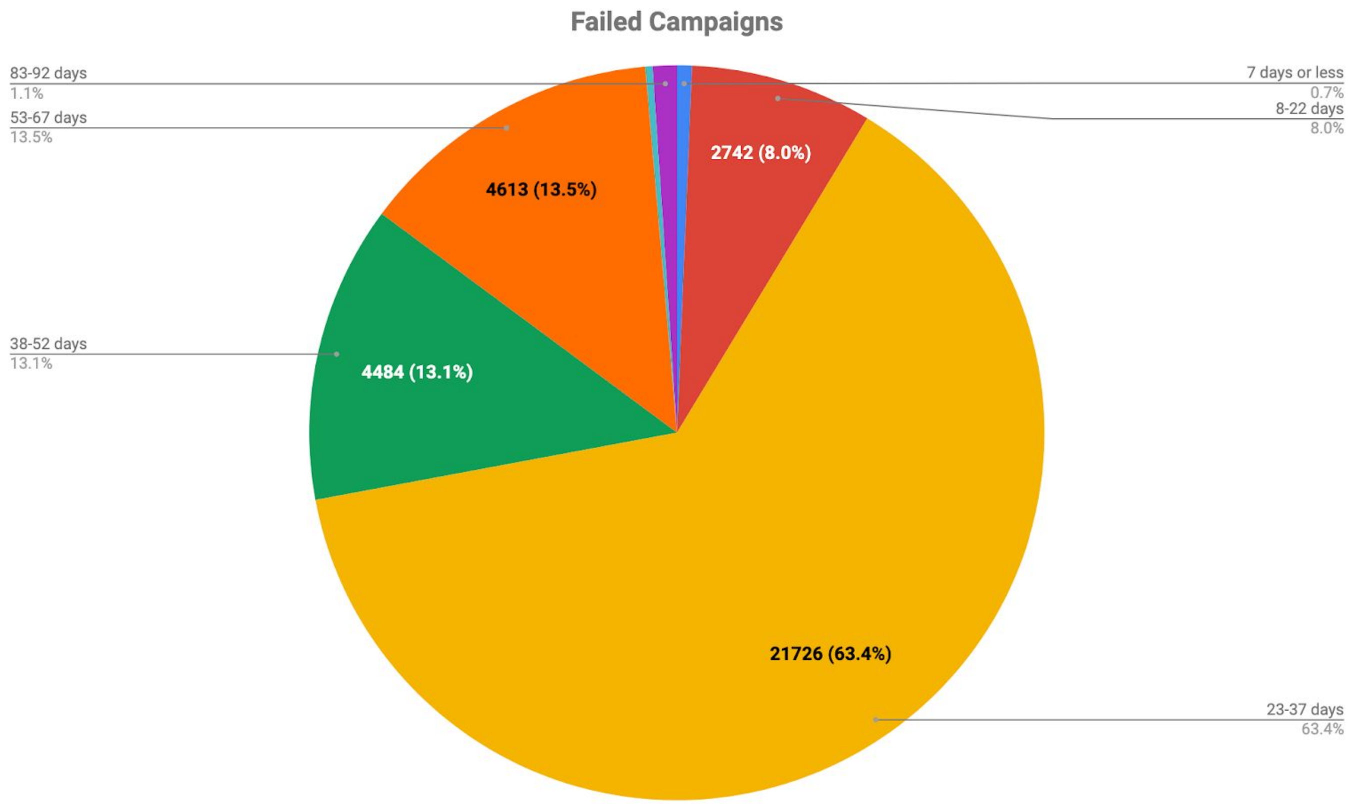

Successful Campaigns
Probability
Campaign Length (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

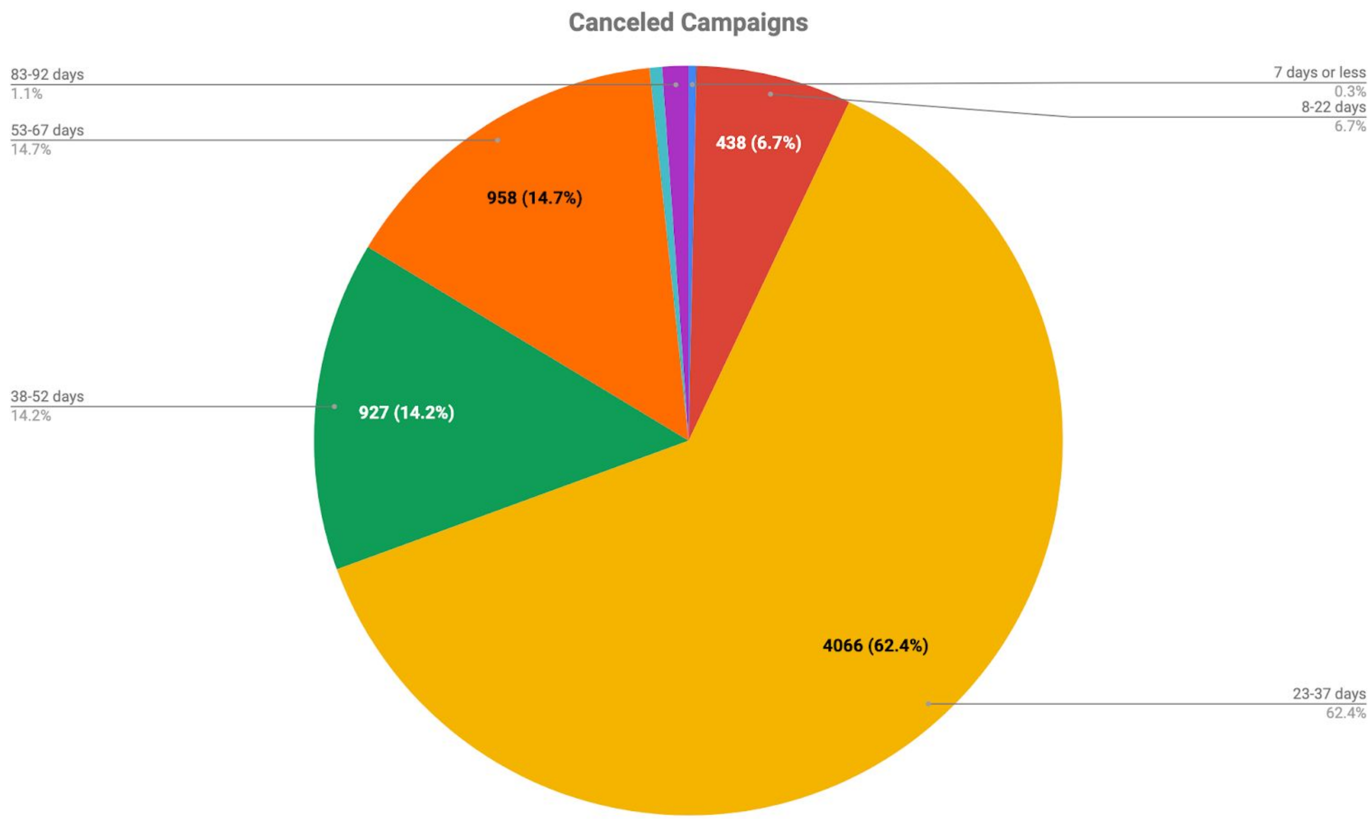

Failed Campaigns
Probability
Campaign Length (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

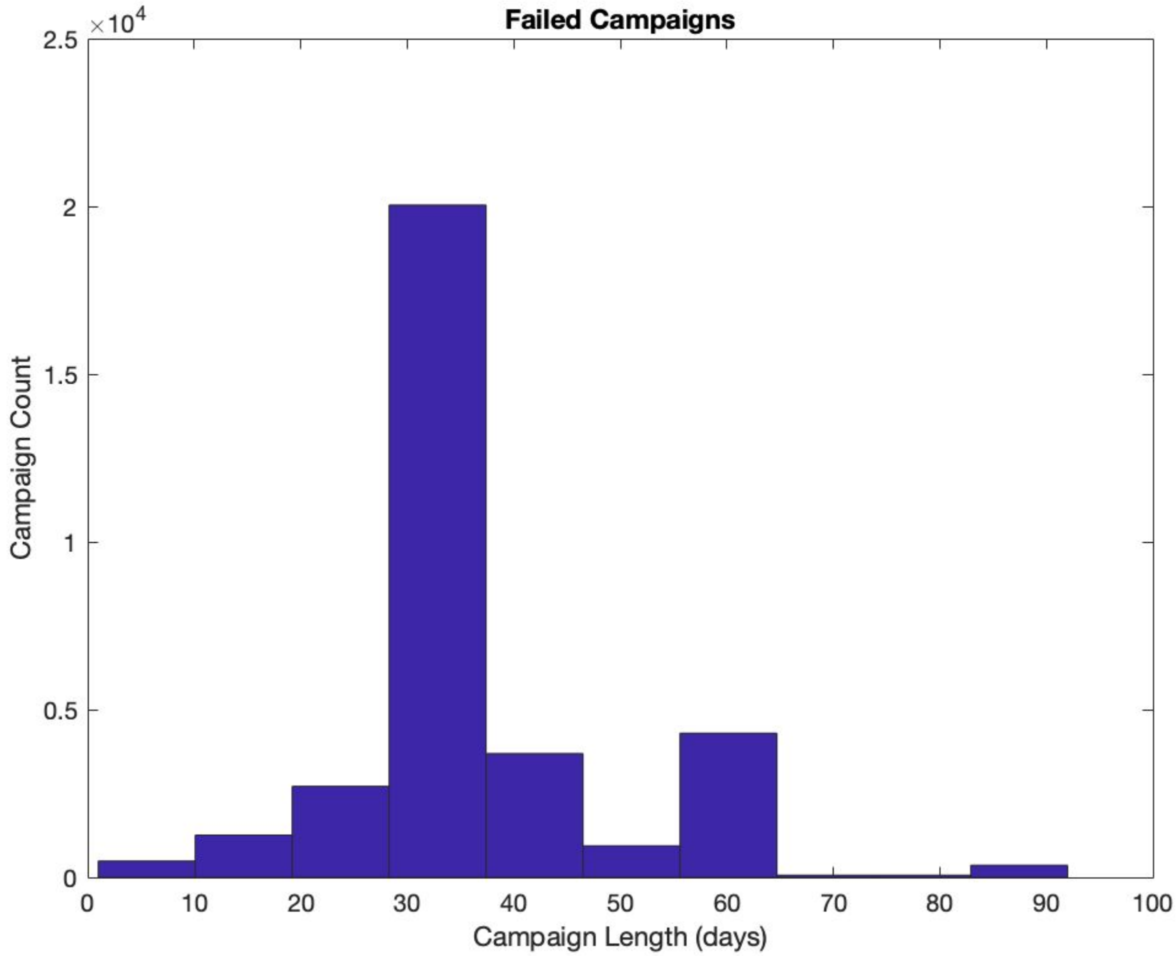

Campaigns That Exceeded Goal
Probability
Campaign Length (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

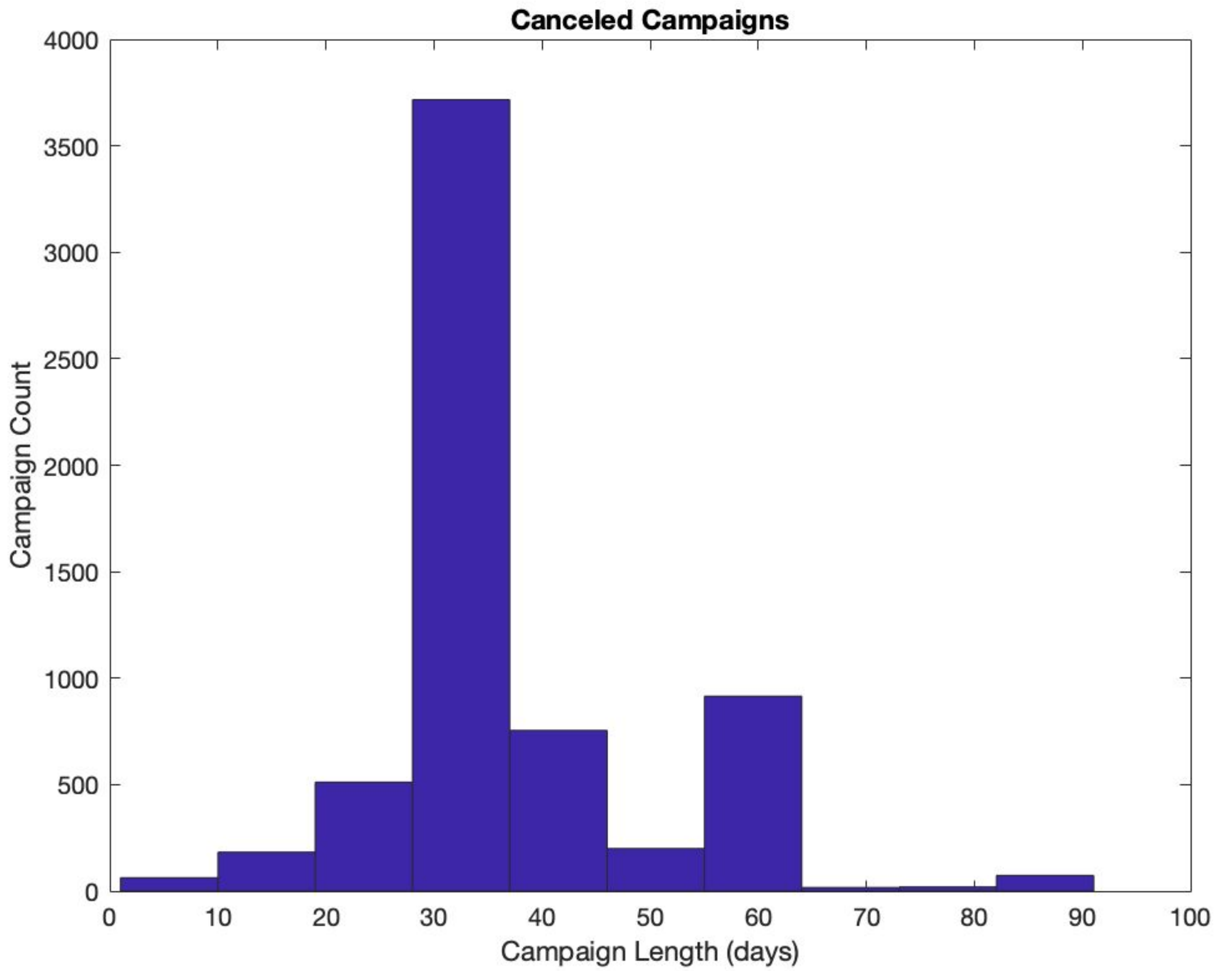

Canceled Campaigns
Probability
Campaign Length (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

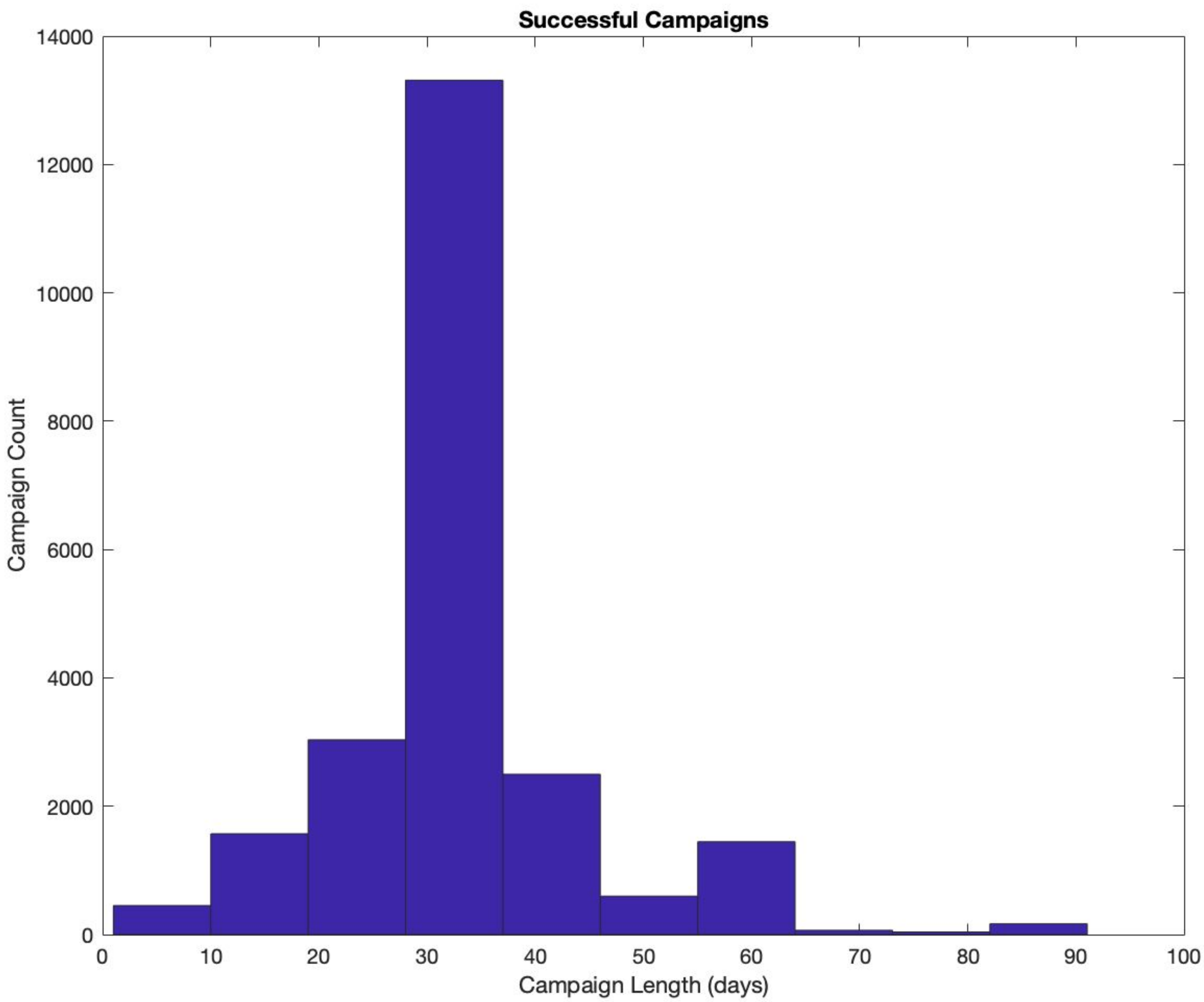

Successful Campaigns
83-92 days
0.7%
53-67 days
6.8%
38-52 days
12.8%
7 days or less
0.8%
8-22 days
13.9%
1607 (6.8%)
3285 (13.9%)
3007 (12.8%)
15211 (64.6%)
23-37 days
64.6%


Failed Campaigns
83-92 days
1.1%
53-67 days
13.5%
38-52 days
13.1%
7 days or less
0.7%
8-22 days
8.0%
23-37 days
63.4%
2742 (8.0%)
4613 (13.5%)
4484 (13.1%)
21726 (63.4%)

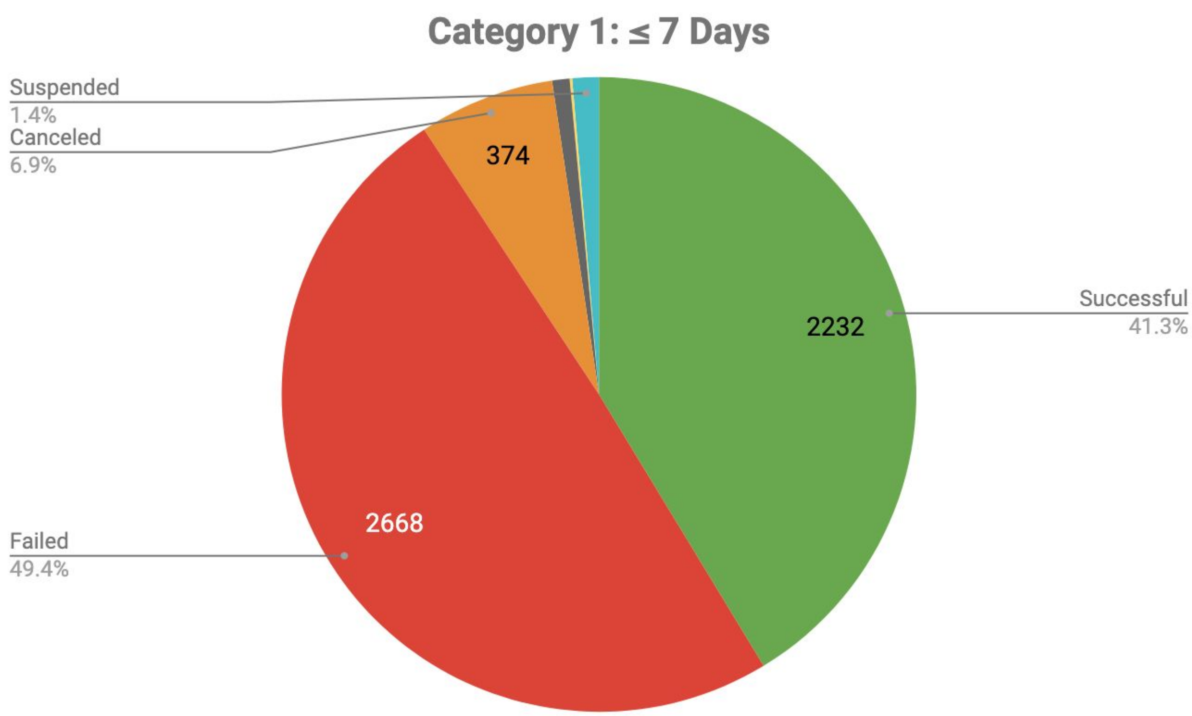

Canceled Campaigns
83-92 days
1.1%
53-67 days
14.7%
38-52 days
14.2%
7 days or less
0.3%
8-22 days
6.7%
23-37 days
62.4%
438 (6.7%)
958 (14.7%)
927 (14.2%)
4066 (62.4%)

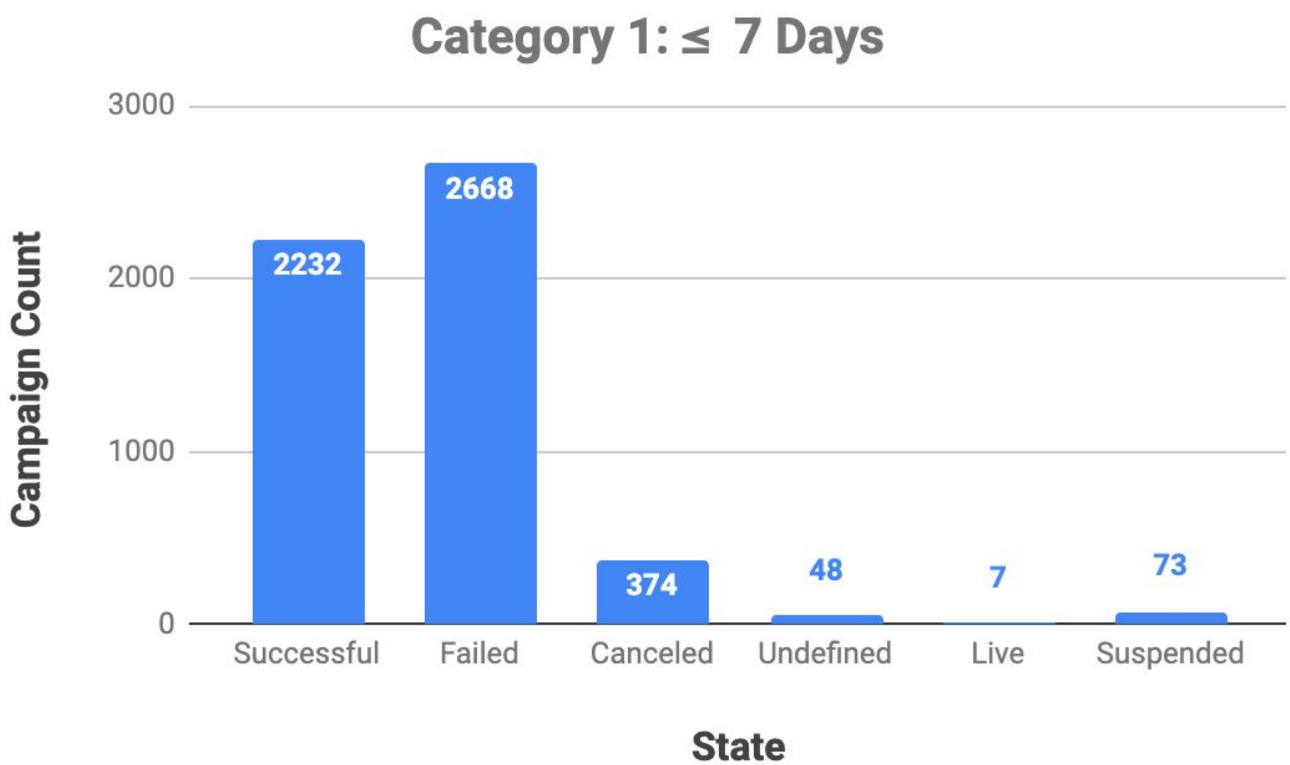

Failed Campaigns
×10^4
2.5
2
1.5
1
0.5
0
Campaign Count
0
10
20
30
40
50
60
70
80
90
100
Campaign Length (days)


Canceled Campaigns
Campaign Count
Campaign Length (days)
0
500
1000
1500
2000
2500
3000
3500
4000
0
10
20
30
40
50
60
70
80
90
100



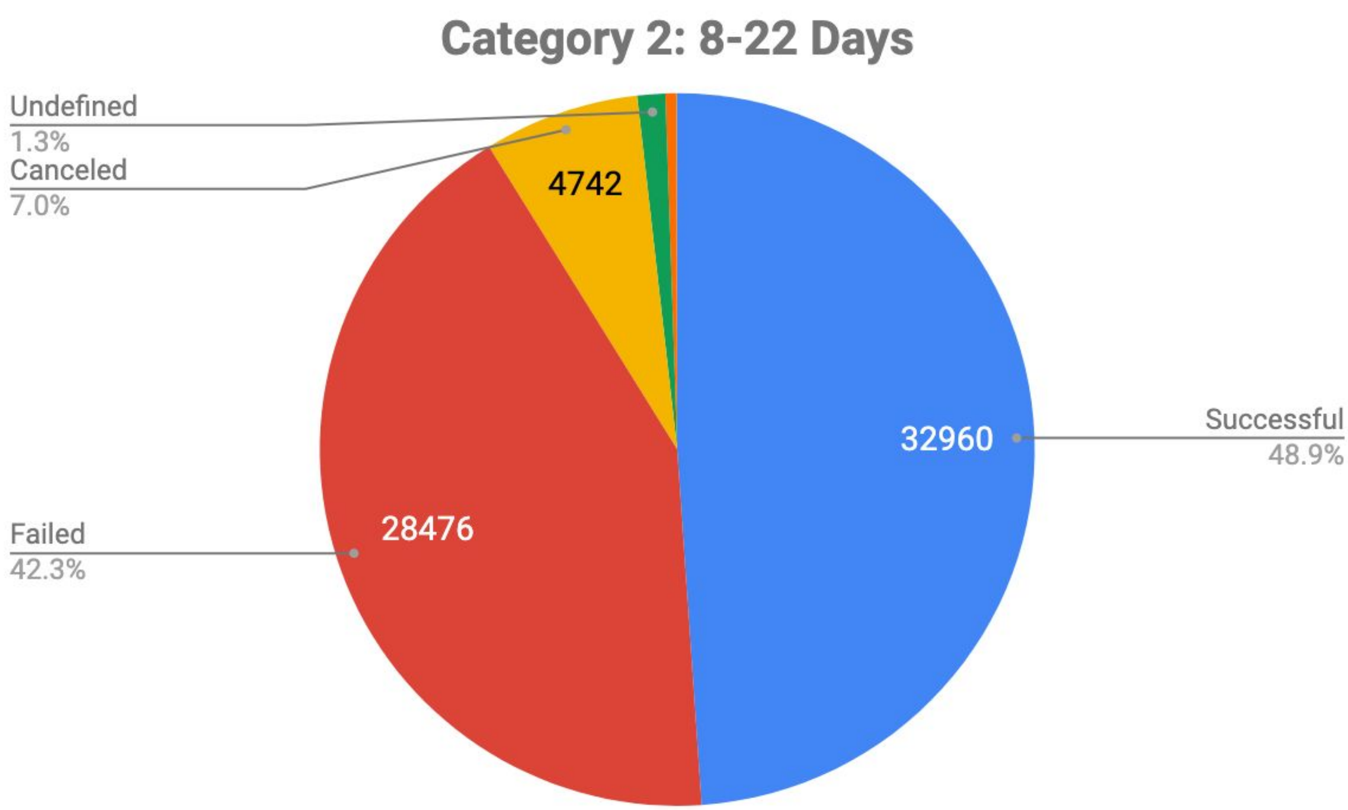
Successful Campaigns
Campaign Count
14000
12000
10000
8000
6000
4000
2000
0
0
10
20
30
40
50
60
70
80
90
100
Campaign Length (days)

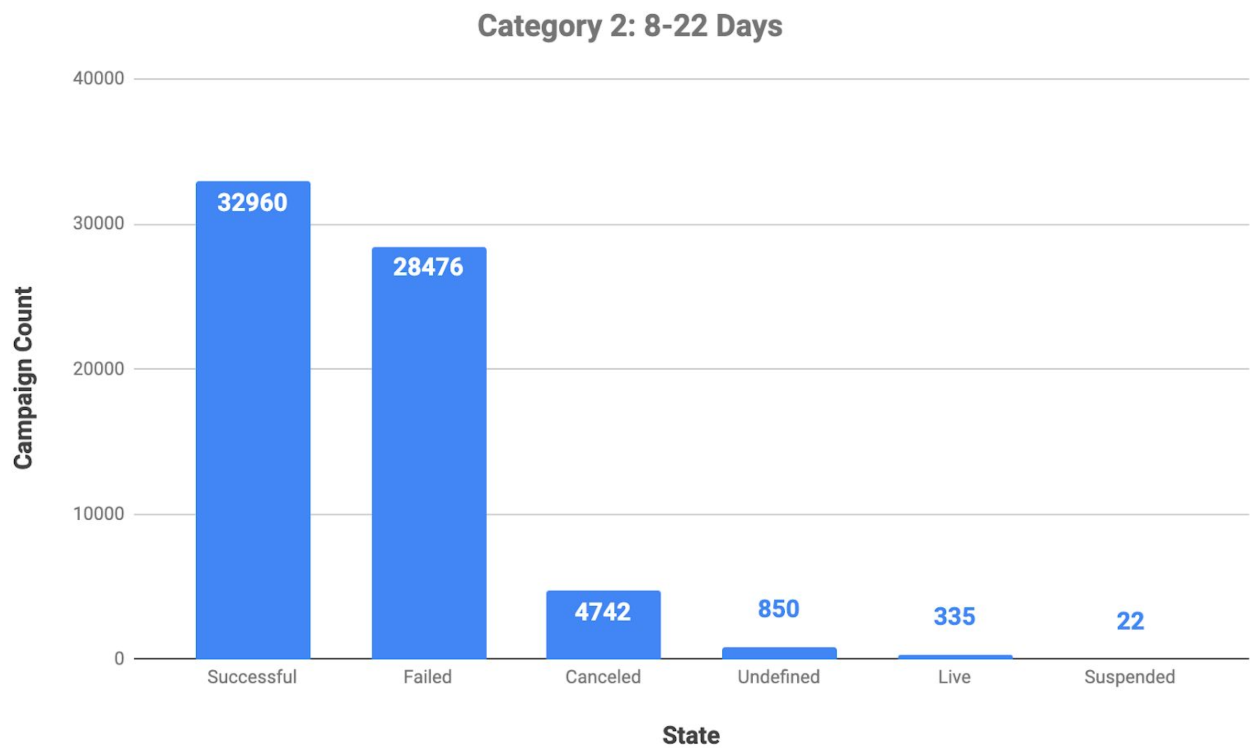

Category 1: ≤ 7 Days
Suspended
1.4%
Canceled
6.9%
374
2232
Successful
41.3%
2668
Failed
49.4%


Category 1: ≤ 7 Days
Campaign Count
3000
2000
1000
0
2232
2668
374
48
7
73
Successful
Failed
Canceled
Undefined
Live
Suspended
State

CATEGORY 1

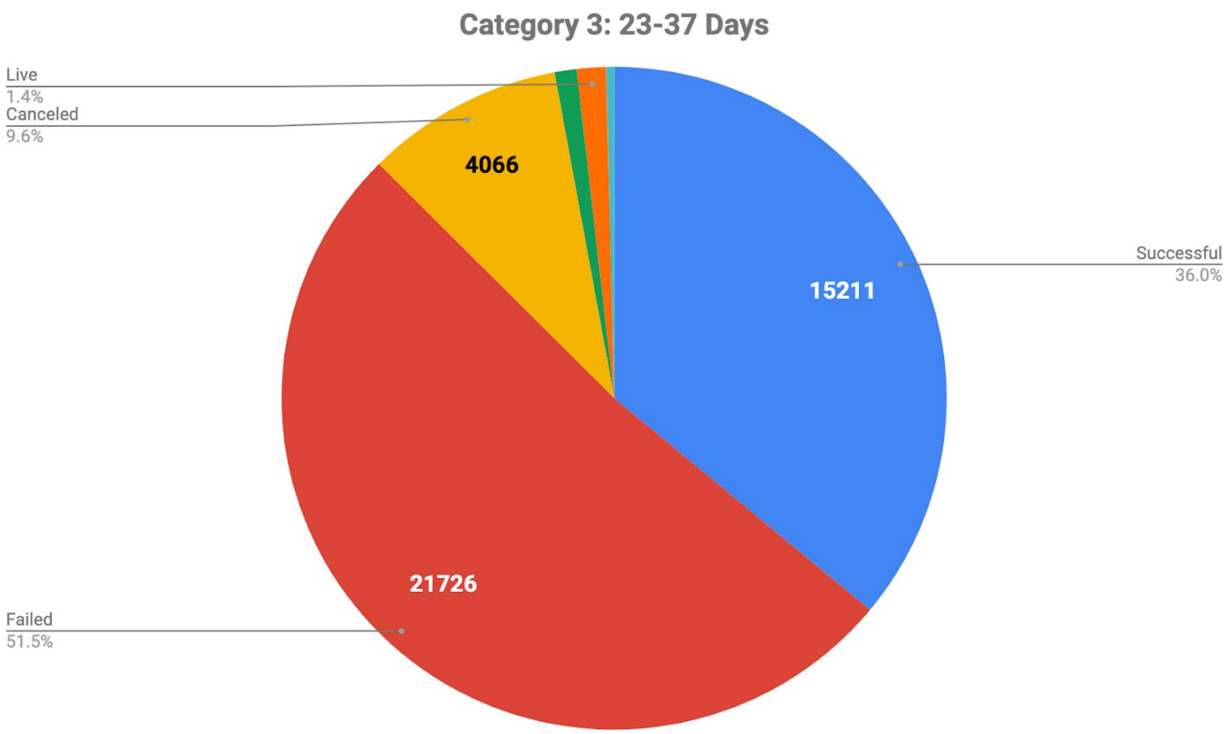
State	Count
Successful	2232
Failed	2668
Canceled	374
Undefined	48
Live	7
Suspended	73

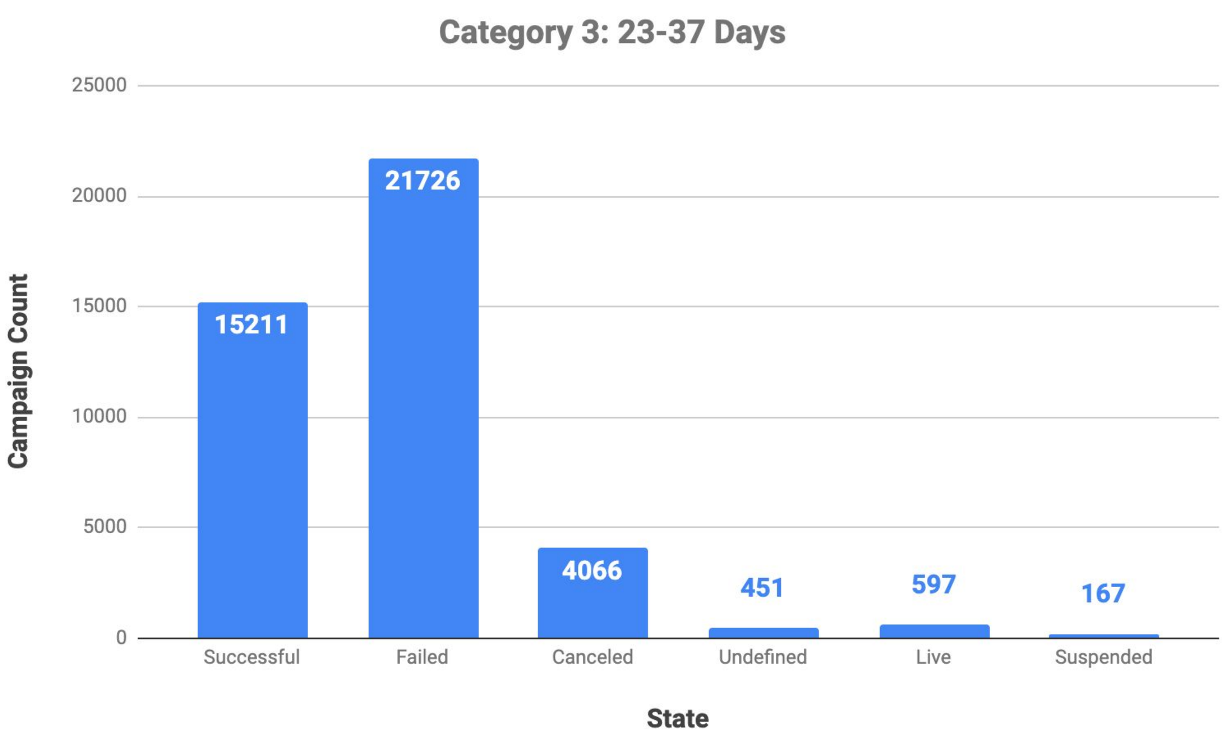




CATEGORY 2

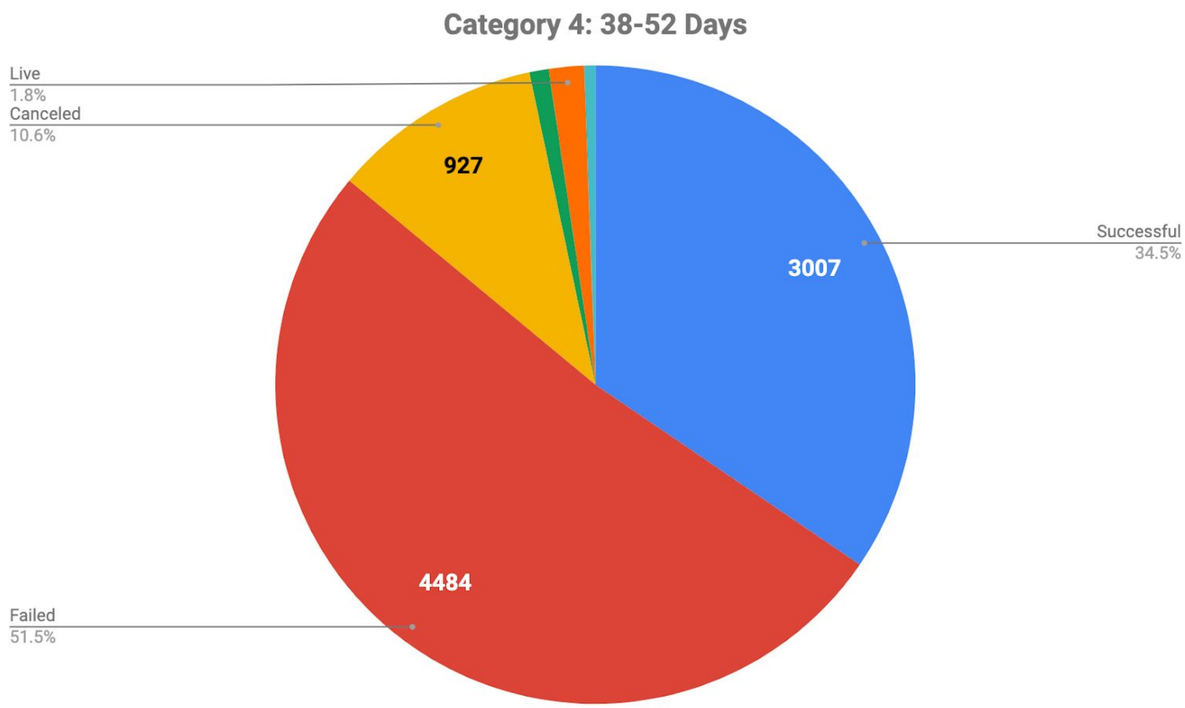
State	Count
Successful	32960
Failed	28476
Canceled	4742
Undefined	850
Live	335
Suspended	22

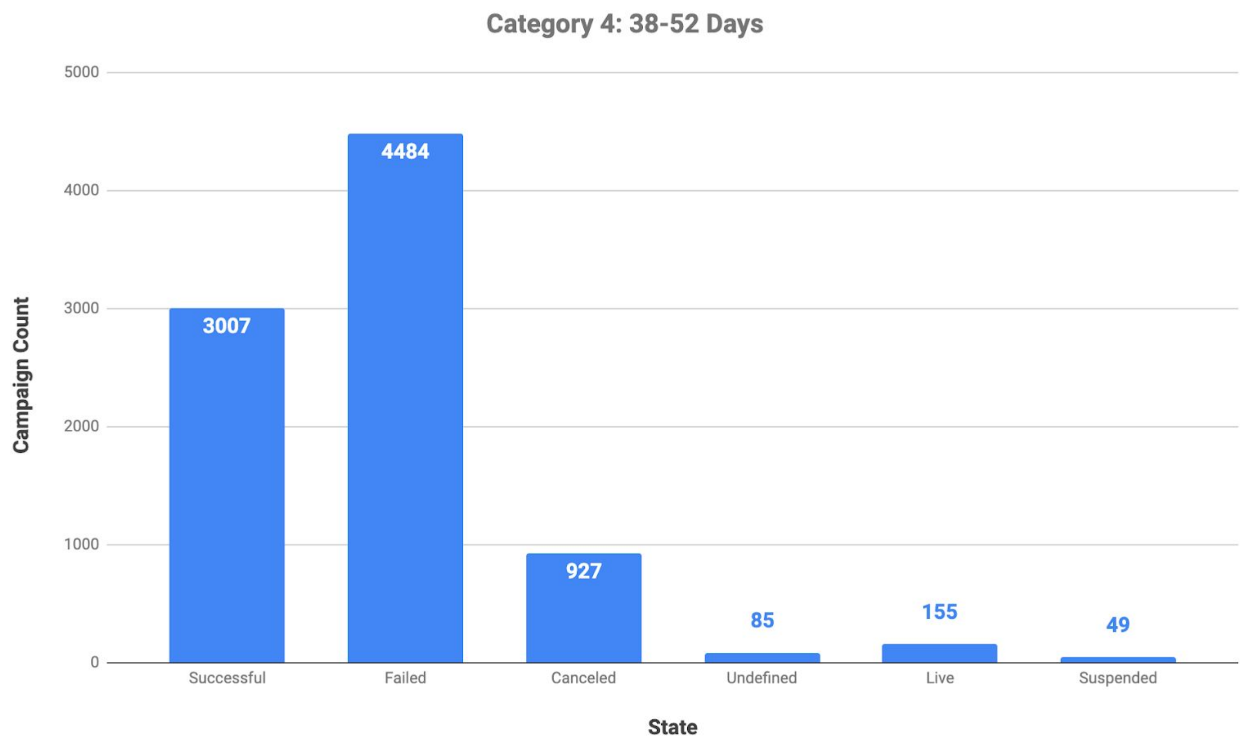

Category 3: 23-37 Days
Live
1.4%
Canceled
9.6%
4066
Successful
36.0%
15211
21726
Failed
51.5%


Category 3: 23-37 Days
25000
20000
15000
10000
5000
0
Campaign Count
15211
21726
4066
451
597
167
Successful
Failed
Canceled
Undefined
Live
Suspended
State

CATEGORY 3

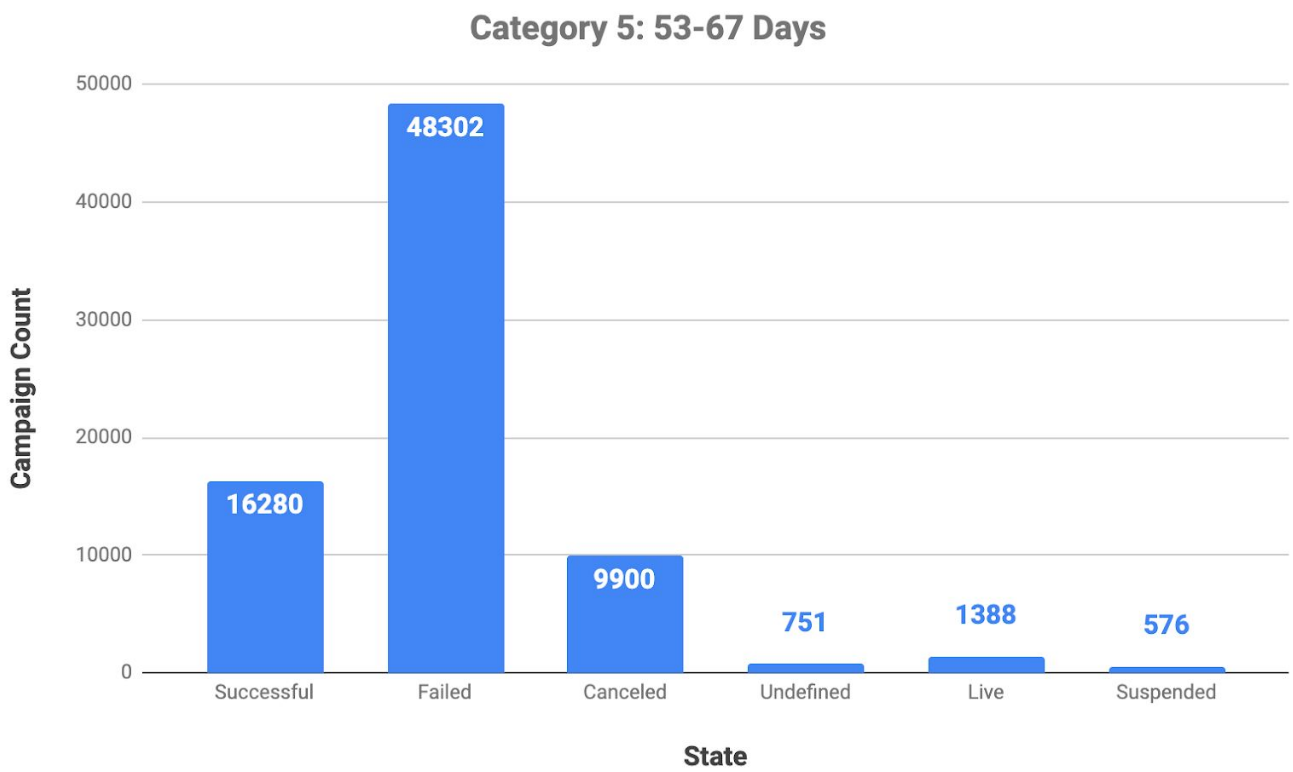
State	Count
Successful	15211
Failed	21726
Canceled	4066
Undefined	451
Live	597
Suspended	167

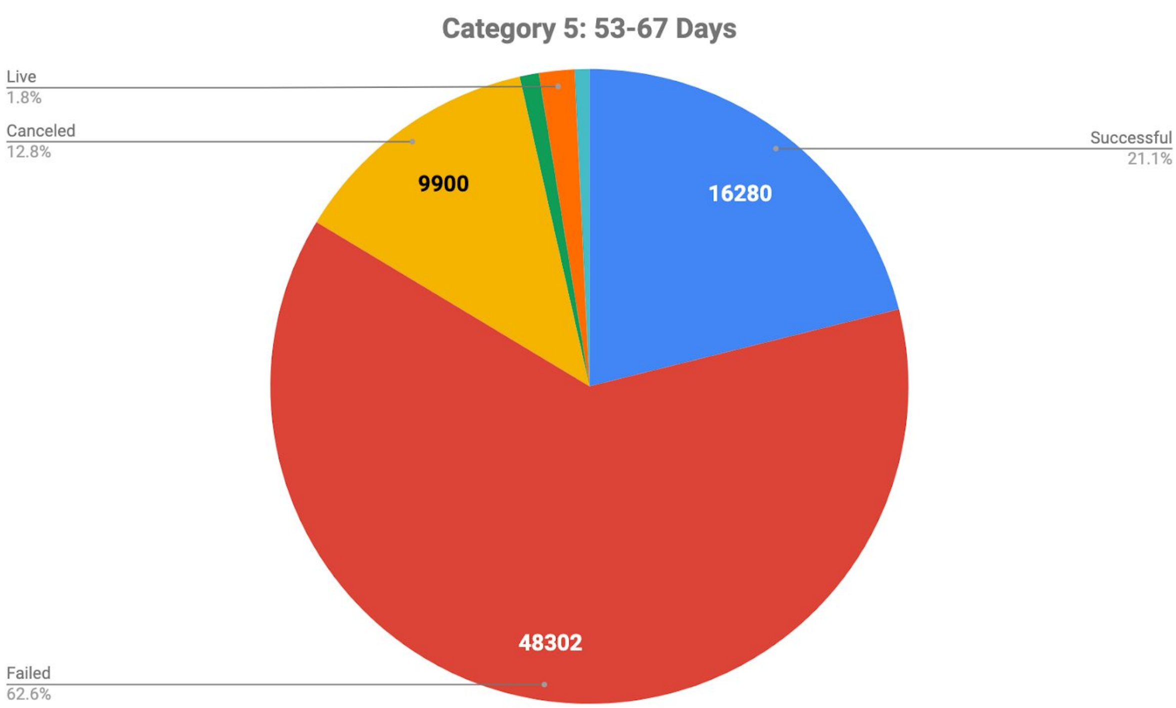




CATEGORY 4

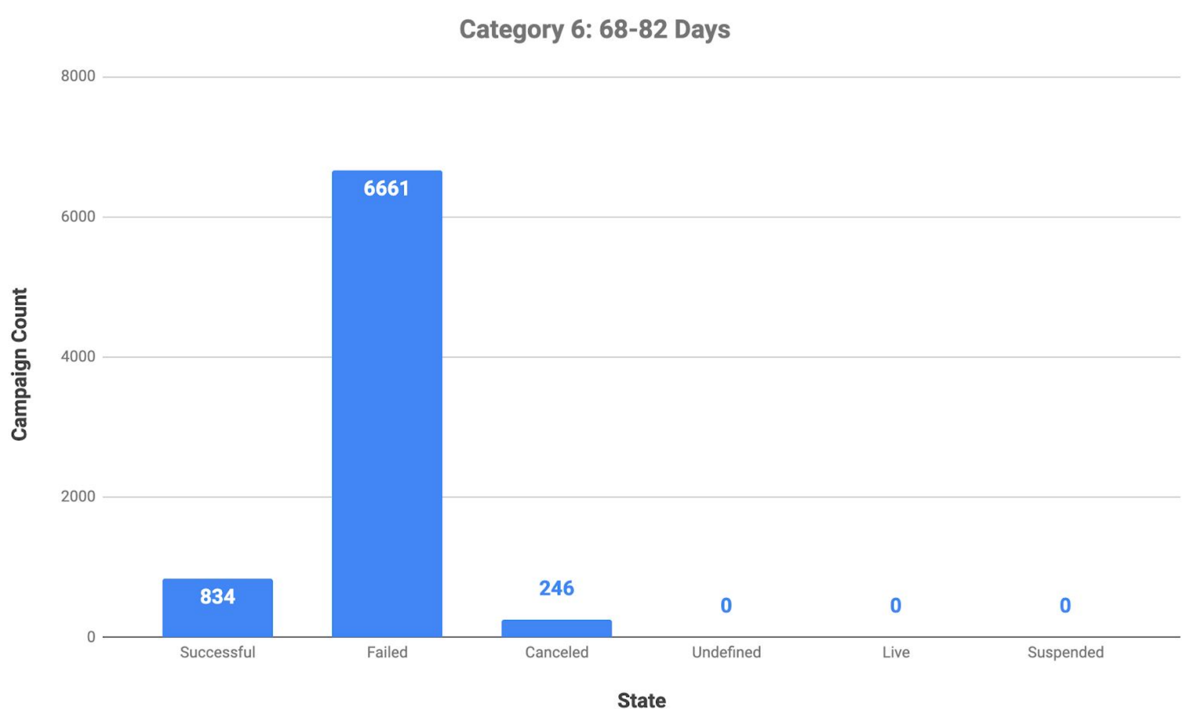
State	Count
Successful	3007
Failed	4484
Canceled	927
Undefined	85
Live	155
Suspended	49

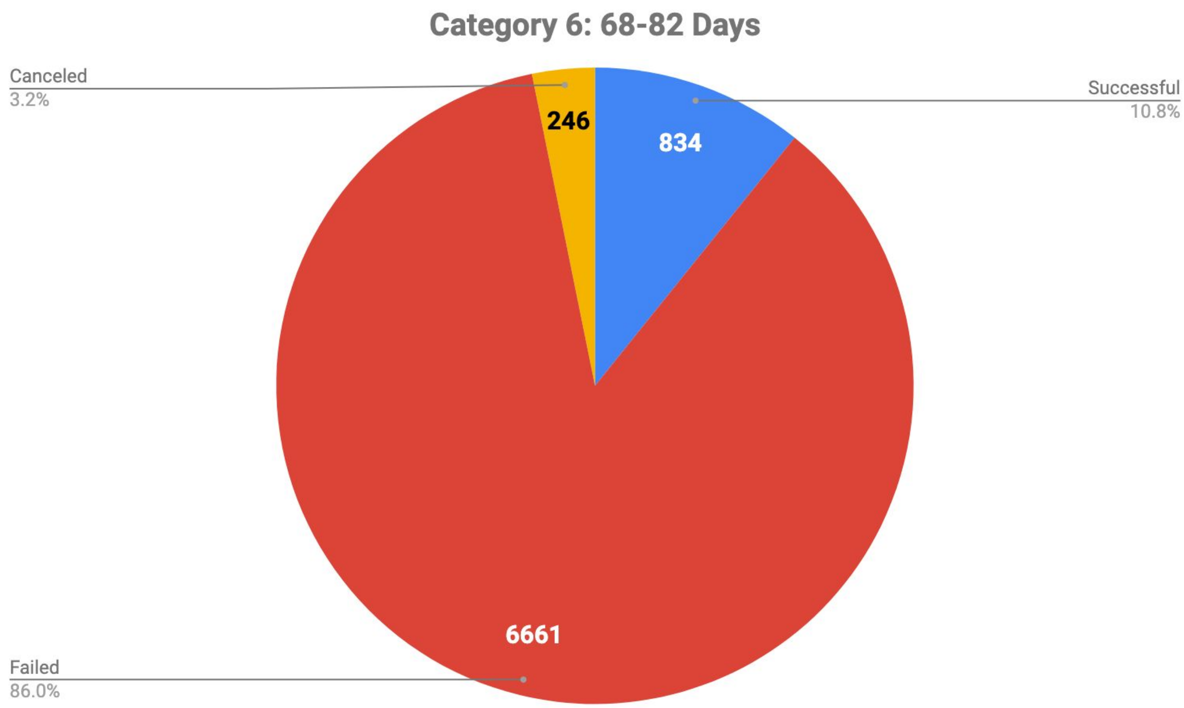

Category 5: 53-67 Days
Campaign Count
50000
40000
30000
20000
10000
0
48302
16280
9900
751
1388
576
Successful
Failed
Canceled
Undefined
Live
Suspended
State


Category 5: 53-67 Days
Live
1.8%
Canceled
12.8%
Successful
21.1%
9900
16280
48302
Failed
62.6%

CATEGORY 5

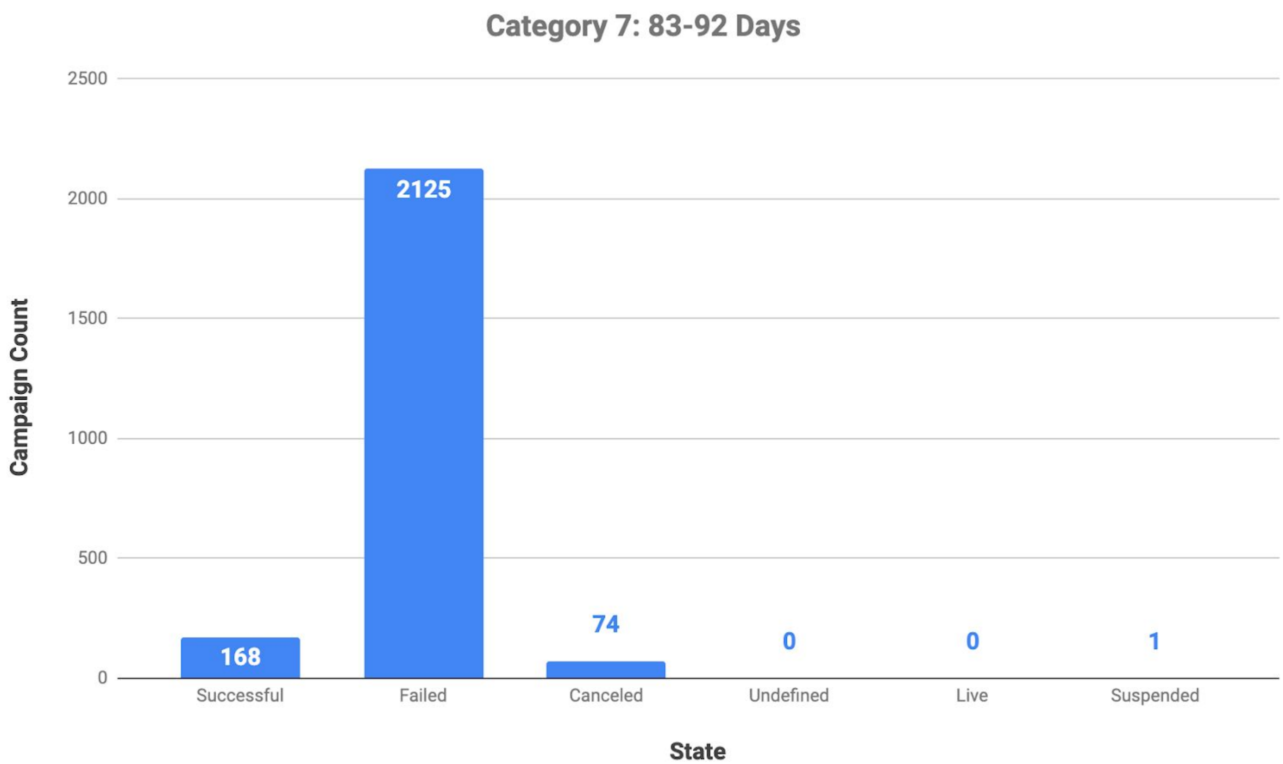
State	Count
Successful	16280
Failed	48302
Canceled	9900
Undefined	751
Live	1388
Suspended	576

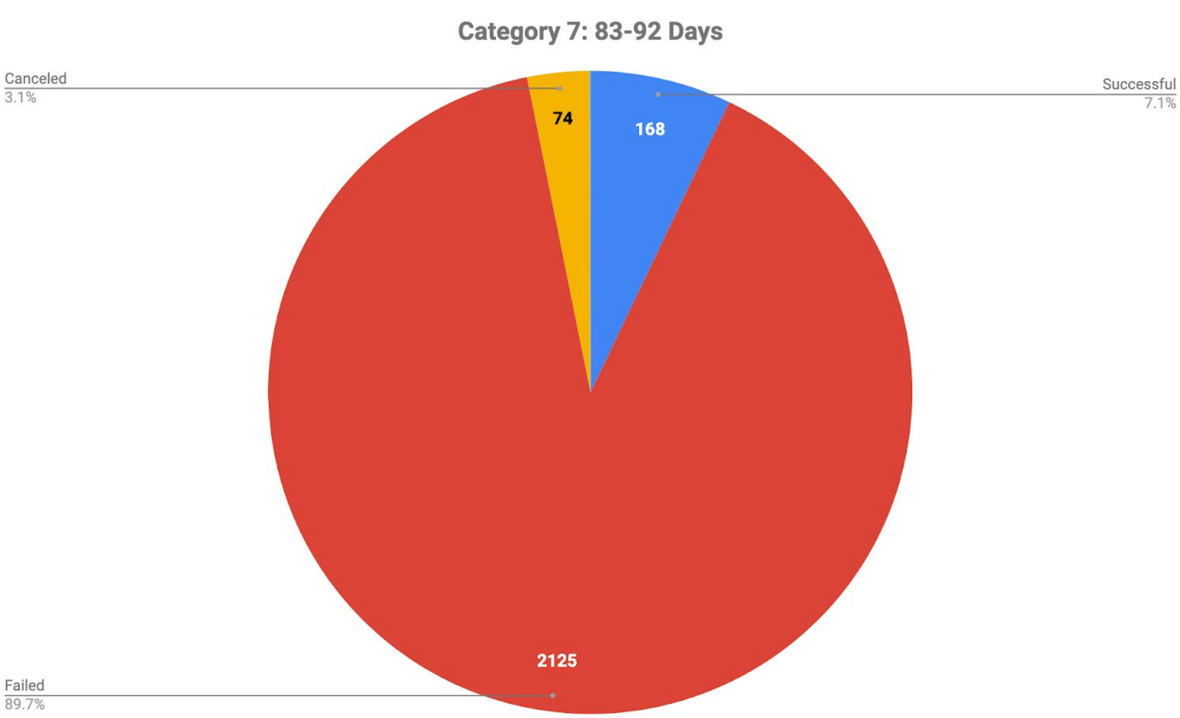




CATEGORY 6

State	Count
Successful	834
Failed	6661
Canceled	246
Undefined	0
Live	0
Suspended	0


Category 7: 83-92 Days
Campaign Count
2500
2000
1500
1000
500
0
2125
168
74
0
0
1
Successful
Failed
Canceled
Undefined
Live
Suspended
State


Category 7: 83-92 Days
Canceled
3.1%
Successful
7.1%
74
168
2125
Failed
89.7%

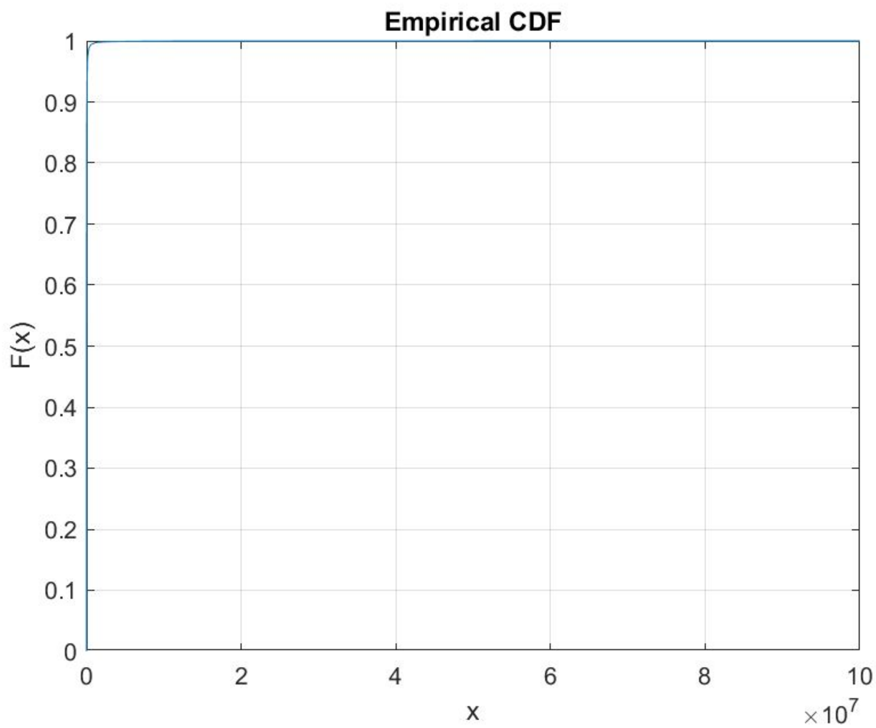
CATEGORY 7

State	Count
Successful	168
Failed	2125
Canceled	74
Undefined	0
Live	0
Suspended	1

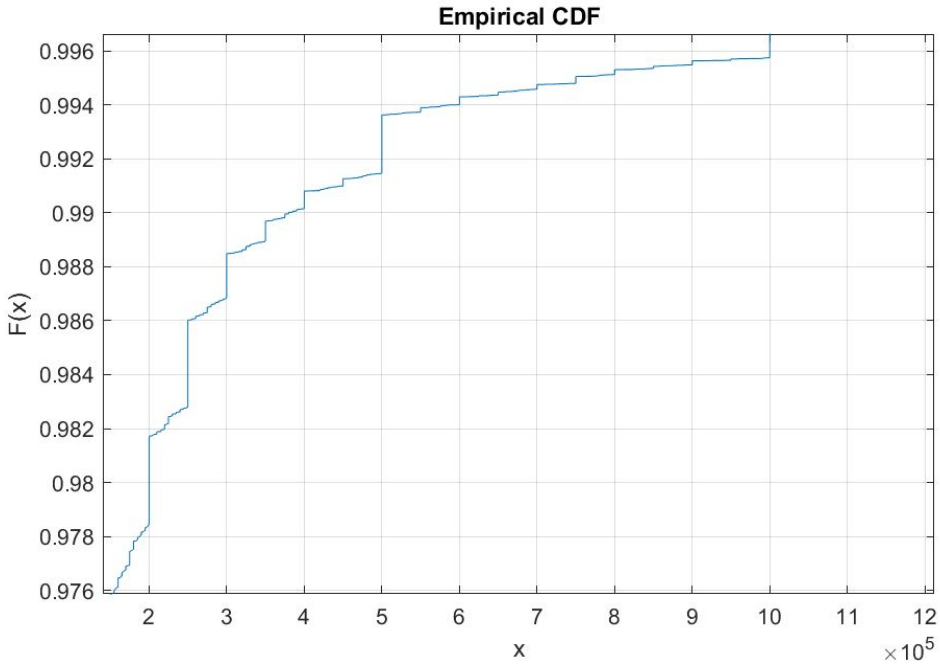
Relationship Between Goal and Success of Campaign & Goal and Amount Pledged

Mila Korobkina

In this section, I analyzed the relationship between fiscal goal amounts and the success of campaigns, or states. I also looked at the differences between amounts pledged and goals for these campaigns. I looked at overall trends as well as trends within groups based on different ranges of goal amounts. I used MapReduce and command line to get the categories of goal, state and pledged and also used them to break down the data into the different comparison buckets. I used Excel and Matlab to plot the data. Below Are the CDF plots of all of the goals, the first is the overall view while the next figure is a closer view at the actual curve in the data.



All Kickstarter Goals CDF



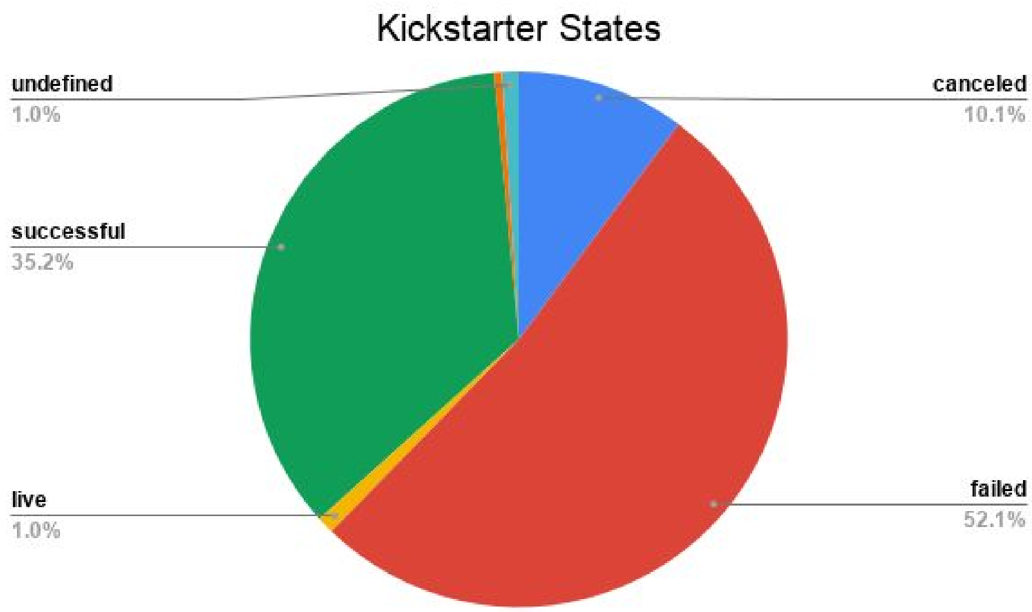
All Kickstarter Goals (Zoomed In)

Kickstarter Goal Measurements (USD)

Measure	Value
Median	$5,000
Mean	$48,193
Mode	$5,000
Variance	$1.3536e+12
Standard Deviation	$1,163,400
Coefficient of Variation	$24.1409
Minimum	$0.01
Maximum	$100,000,000
First quartile (0.25)	$1,500
Second quartile (0.5)	$5,000
Third quartile (0.75)	$15,550
Fourth quartile (1)	$100,000,000

Overall, kickstarter goals were incremented in tens, halves or quarter values. This makes raising certain amounts easier because pre-set donation amounts allow for campaigns to actually meet their goals by making sure that adding up to the exact goal is possible. Within the curve, campaign goals tended to stay towards the lower price end. This could be explained by the type of things kickstarters are usually used for as well as cultural associations with asking for money and crowdfunding.

I also plotted all of the states (successful, failed, canceled, undefined, live, suspended) within the entire dataset. Below are the results:

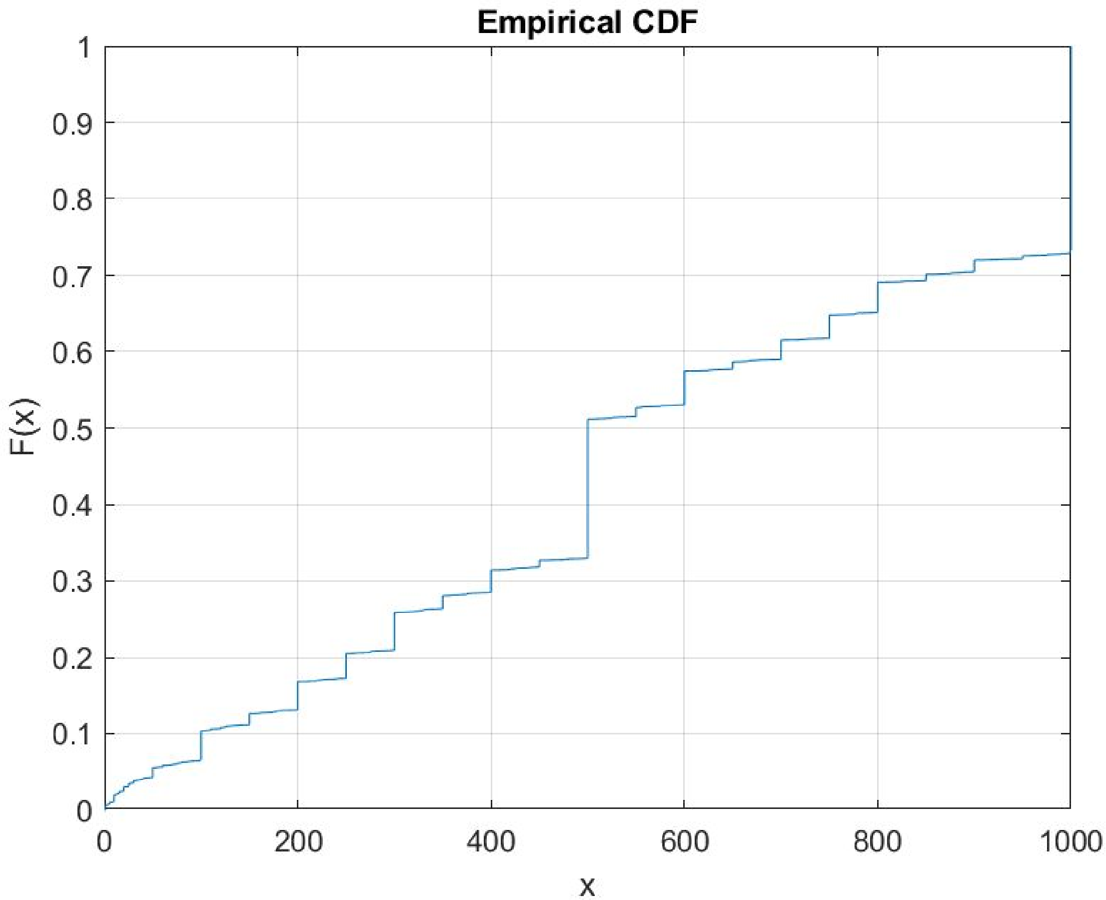


Approximately 52% of all of the kickstarter campaigns failed. However, 35.2% did succeed. This indicates that kickstarter campaigns aren't as useful as they could be. A potential solution could be more customized recommendations for some of the metrics we analyze in this report, such as goal, duration or category, to help campaigners get the most of this platform.

I decided to further break down the data by goal buckets based on different price margins. I did this based on what seemed to be socially acceptable price margins instead of buckets with even ranges. I wanted to see if there were any trends within certain goal amounts successes and whether and extremely low, or feasible goal, was more successful than something more expensive. Granted, this does not take into consideration the category, or the "story" behind a campaign so a more circulated or compelling campaign may raise more money than a less "popular campaign", regardless of the goal. Below are the goal groups as well as the CDF goal plots and state plots for each goal group.

Kickstarter Goal Amount Groups (USD)

Group 1	Group 2	Group 3	Group 4
x <= $1,000	$1,000 < x <= $10,000	$10,000 < x <= $100,000	$100,000 < x

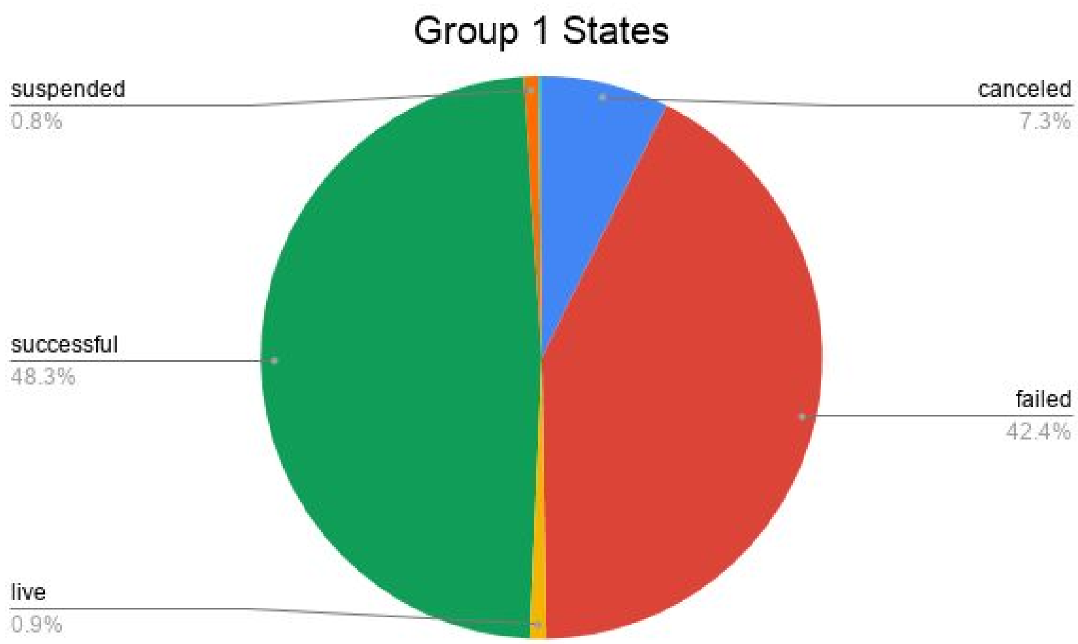


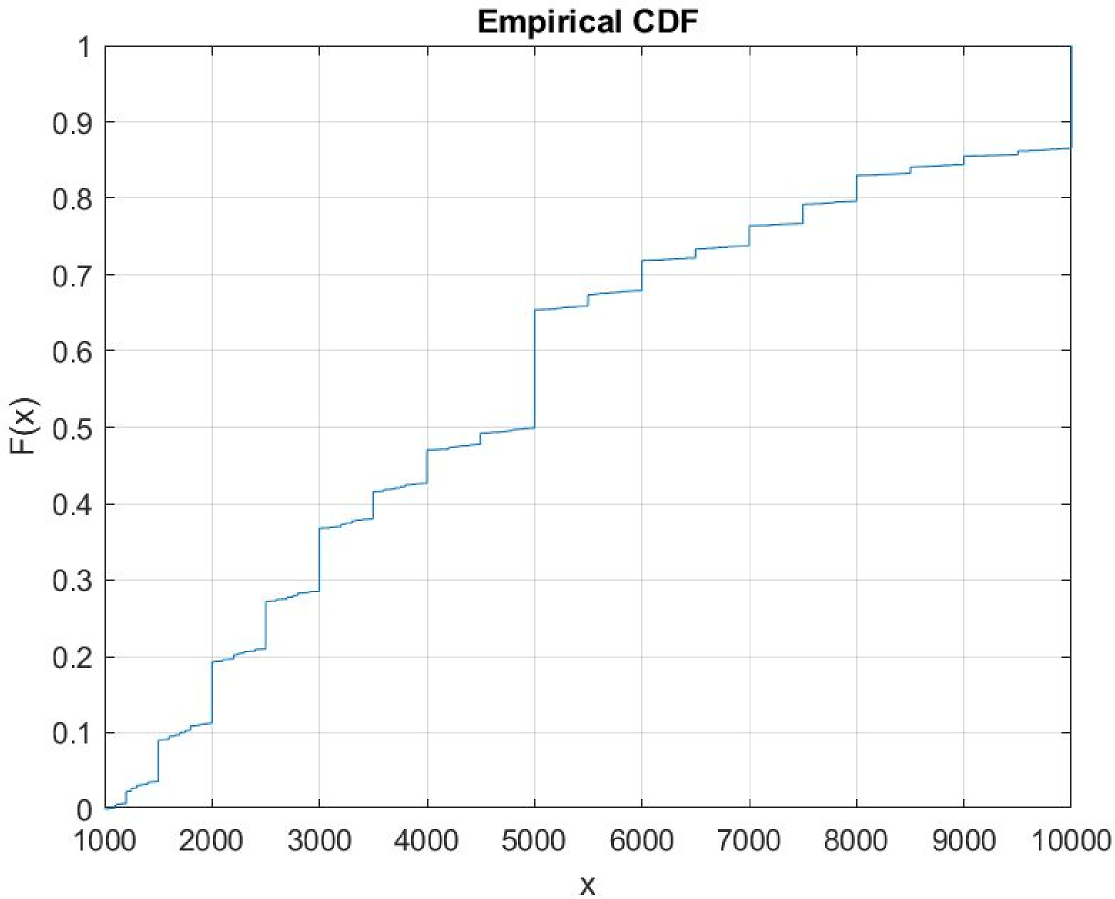
Goal Group 1 (x <= $1,000) CDF

Goal Group 1 Measurements (USD)

Measure	Value
Median	$500
Mean	$589.6988
Mode	$1,000
Variance	$1.0629e+05
Standard Deviation	$326.0151

Coefficient of Variation	$ 0.5529
Minimum	$0.01
Maximum	$1,000
First quartile (0.25)	$200
Second quartile (0.5)	$500
Third quartile (0.75)	$1,000
Fourth quartile (1)	$1,000



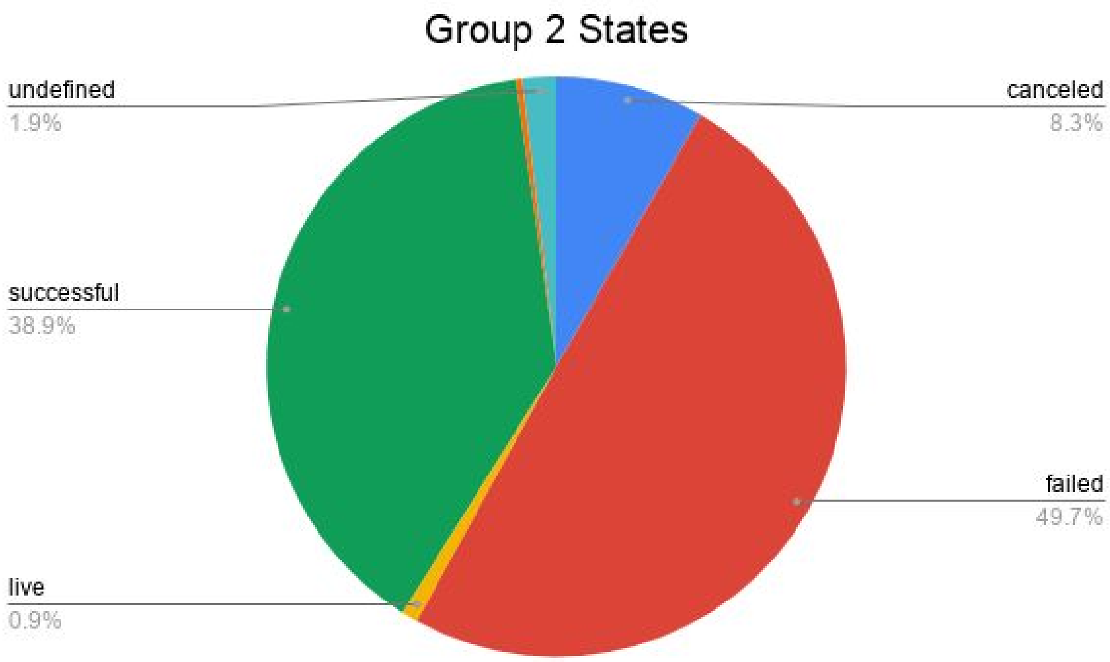


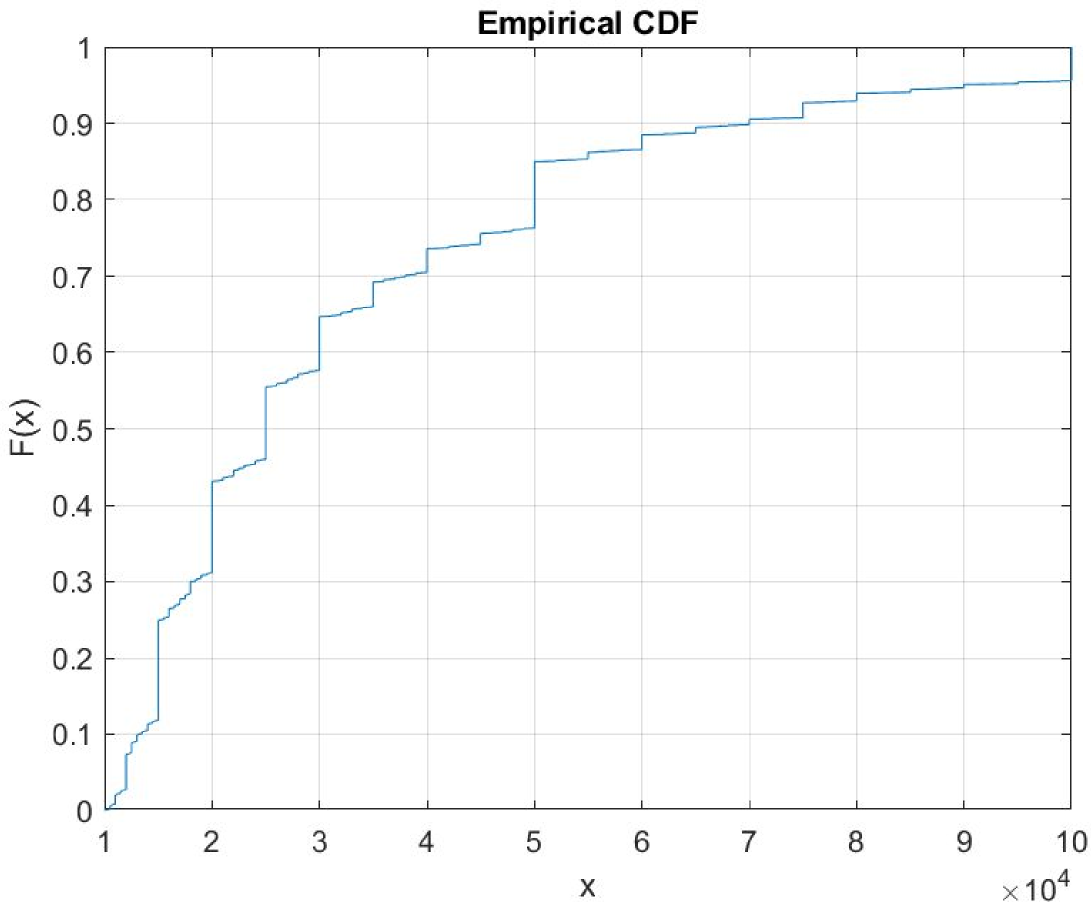
Goal Group 2 ($1,000 < x <= $10,000) CDF

Goal Group 2 Measurements (USD)

Measure	Value
Median	$5,000
Mean	$4,945.90
Mode	$5,000
Variance	$7.8757e+06
Standard Deviation	$2,806.40
Coefficient of Variation	$0.5674
Minimum	$1,001
Maximum	$10,000

First quartile (0.25)	$200
Second quartile (0.5)	$500
Third quartile (0.75)	$1,000
Fourth quartile (1)	$10,000



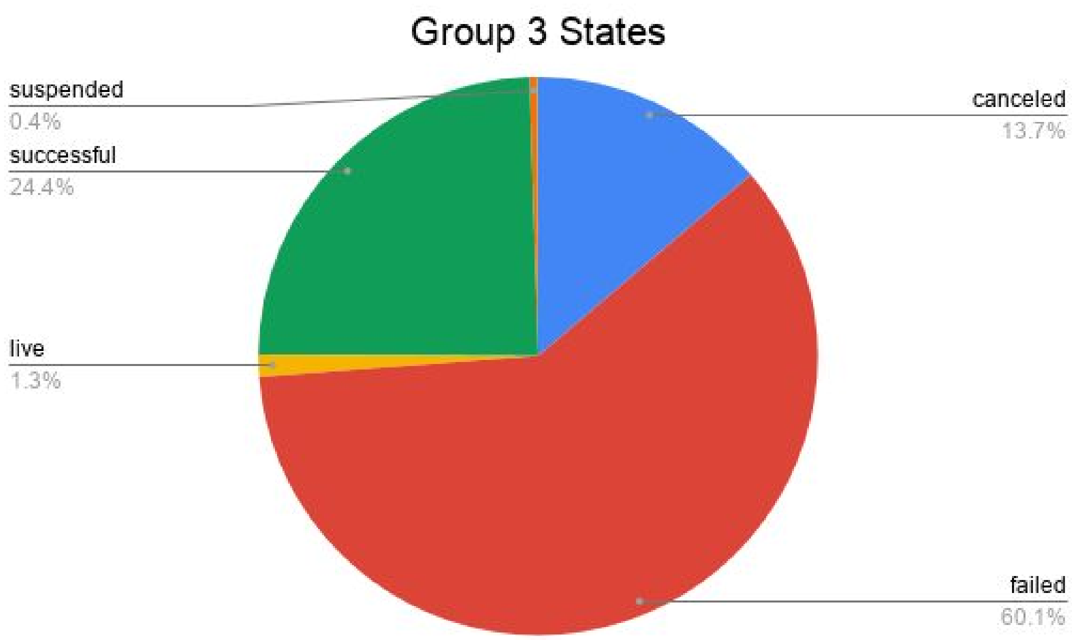


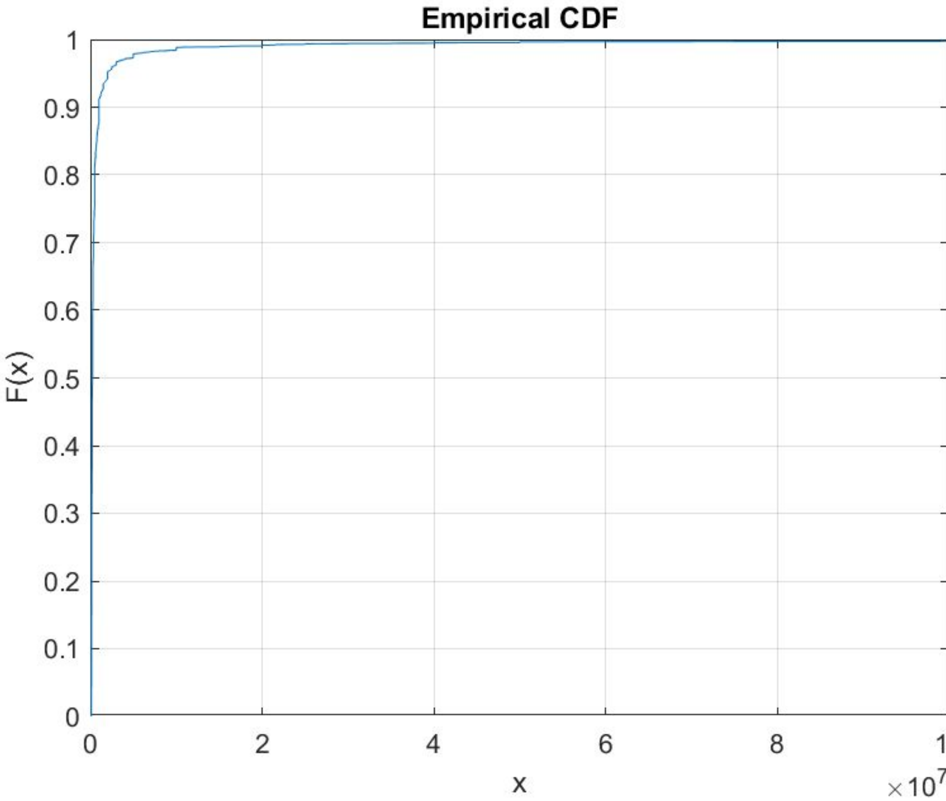
Goal Group 3 ($10,000 < x <= $100,000) CDF

Goal Group 3 Measurements (USD)

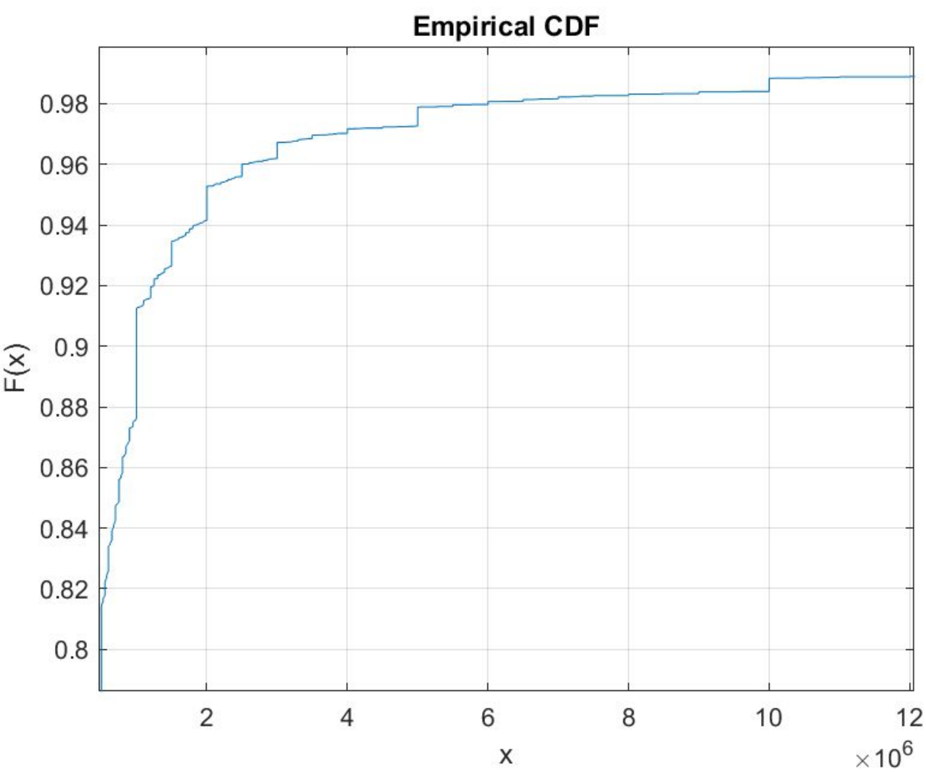
Measure	Value
Median	$25,000
Mean	$33,342
Mode	$15,000
Variance	$5.3252e+08
Standard Deviation	$23,076
Coefficient of Variation	$0.6921
Minimum	$10,001
Maximum	$100,000

First quartile (0.25)	$15,000
Second quartile (0.5)	$25,000
Third quartile (0.75)	$50,000
Fourth quartile (1)	$100,000

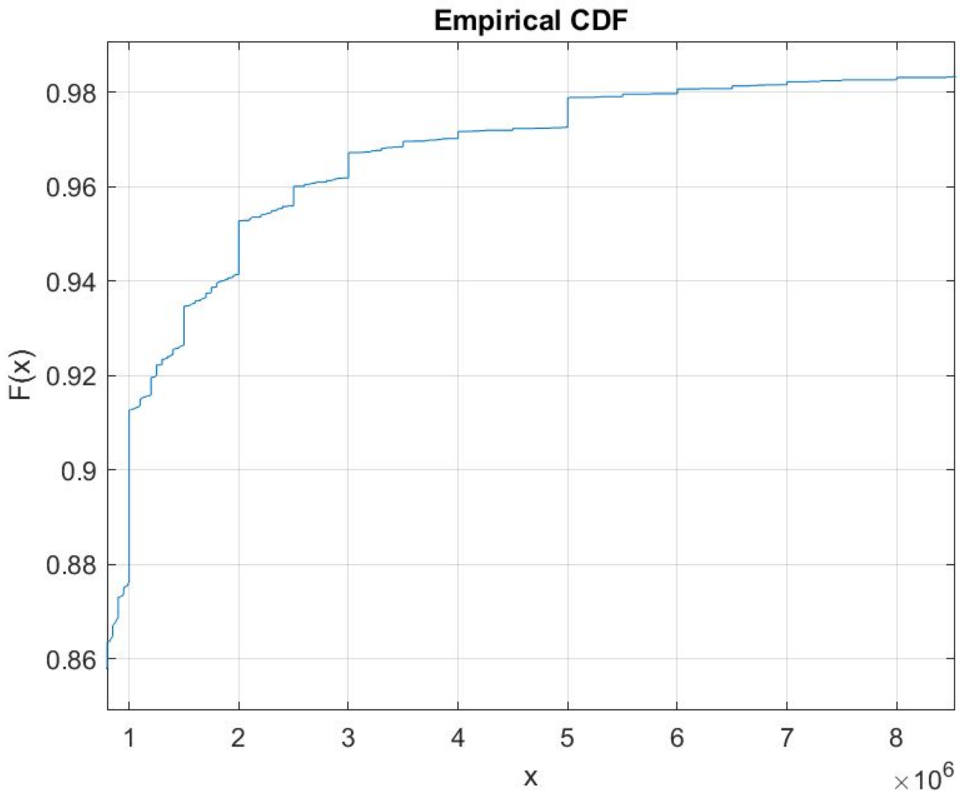




Goal Group 4 (x > $100,000) CDF



Goal Group 4 Zoom 1 (x > $100,000) CDF

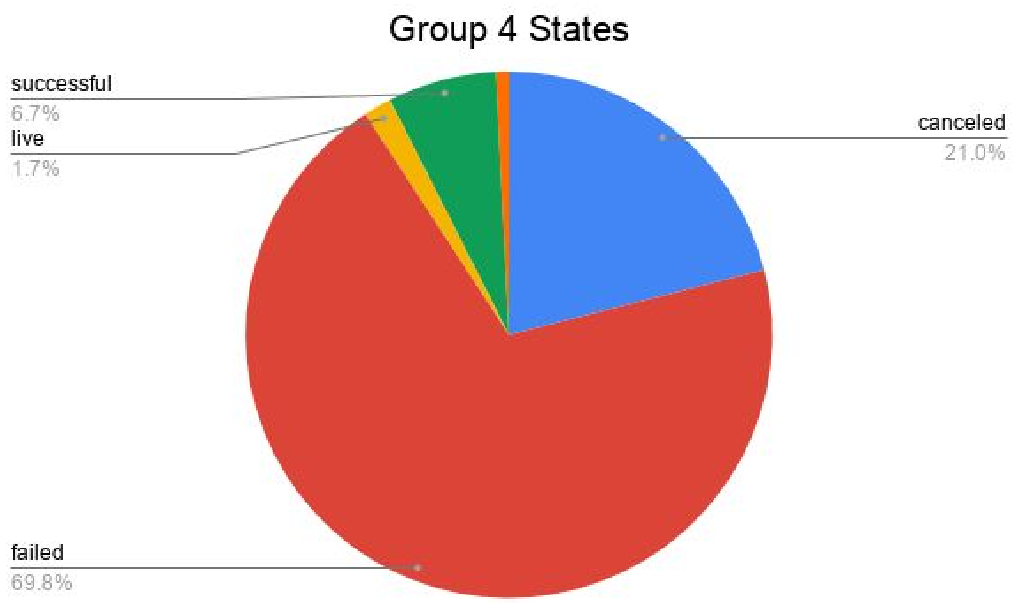


Goal Group 4 Zoom 2 (x > $100,000) CDF

Goal Group 4 Measurements (USD)

Measure	Value
Median	$250,000
Mean	$1,046,900
Mode	$150,000
Variance	$3.8333e+13
Standard Deviation	$6.1913e+06
Coefficient of Variation	$24.7654
Minimum	$100,001
Maximum	$100,000,000
First quartile (0.25)	$150,000
Second quartile (0.5)	$250,000

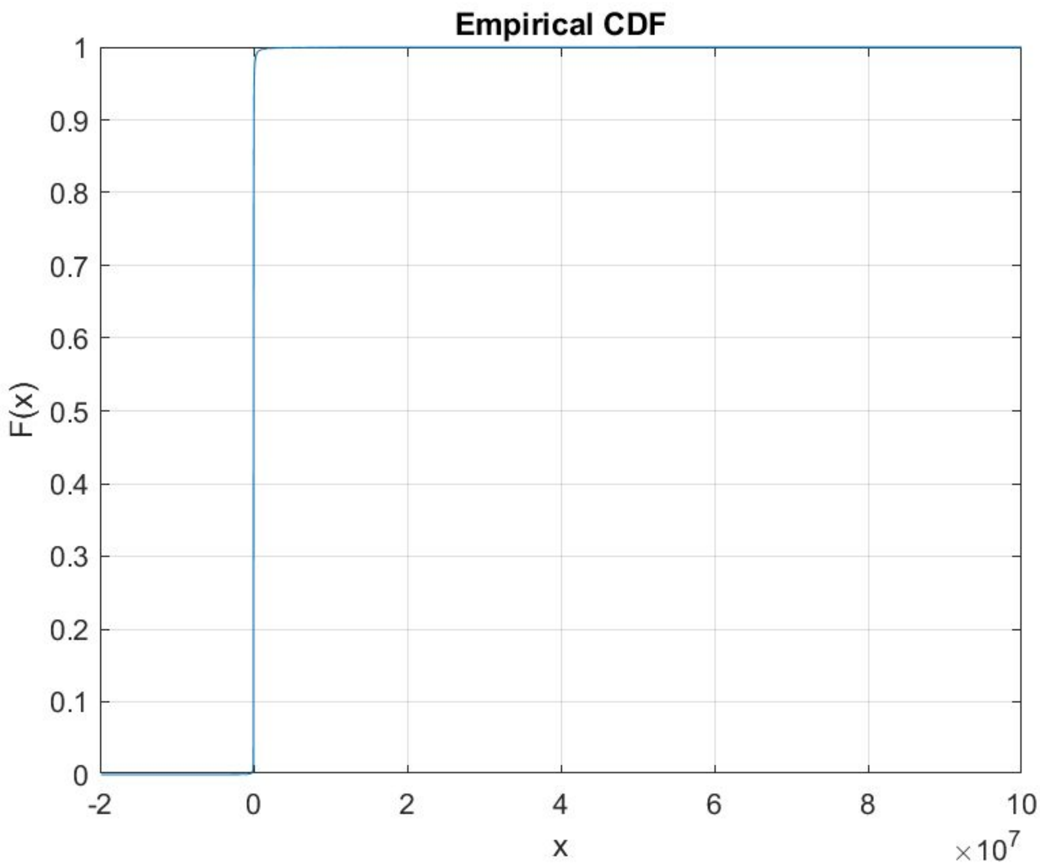
Third quartile (0.75)	$500,000
Fourth quartile (1)	$100,000,000



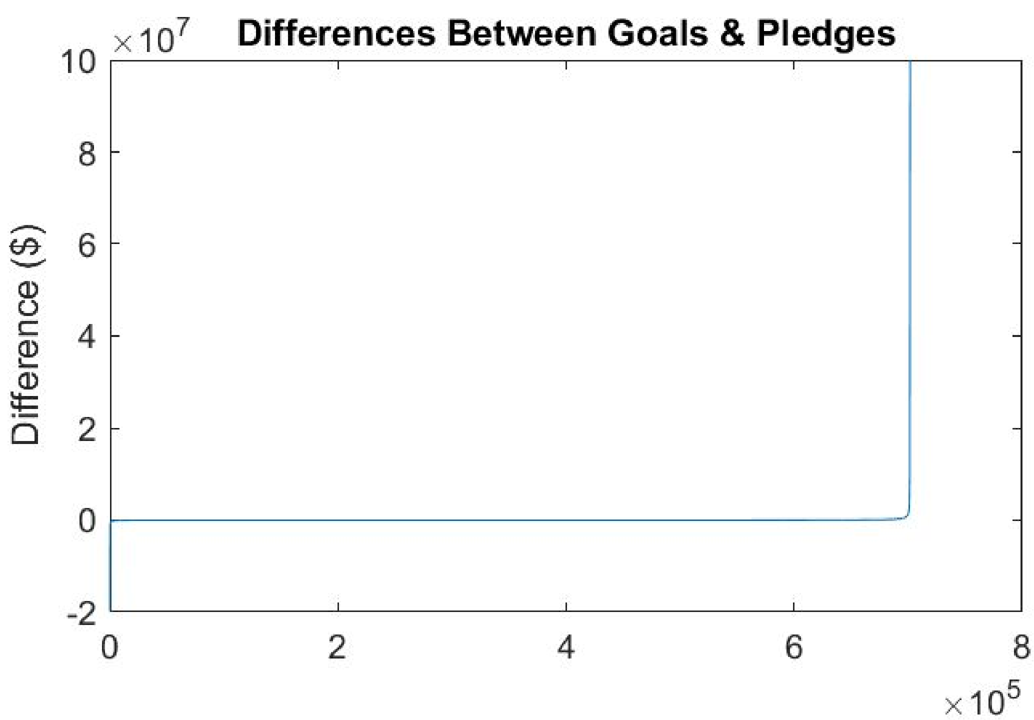
As mentioned earlier, goals tended to be set at certain even increments. This can be seen more in the first 3 groups, which have smaller ranges, and in those groups, the most common goal amounts were around 50% of the total amount. The fourth group did not have a set maximum and contained some extremely high values, which most likely skewed the goal set to lower values.

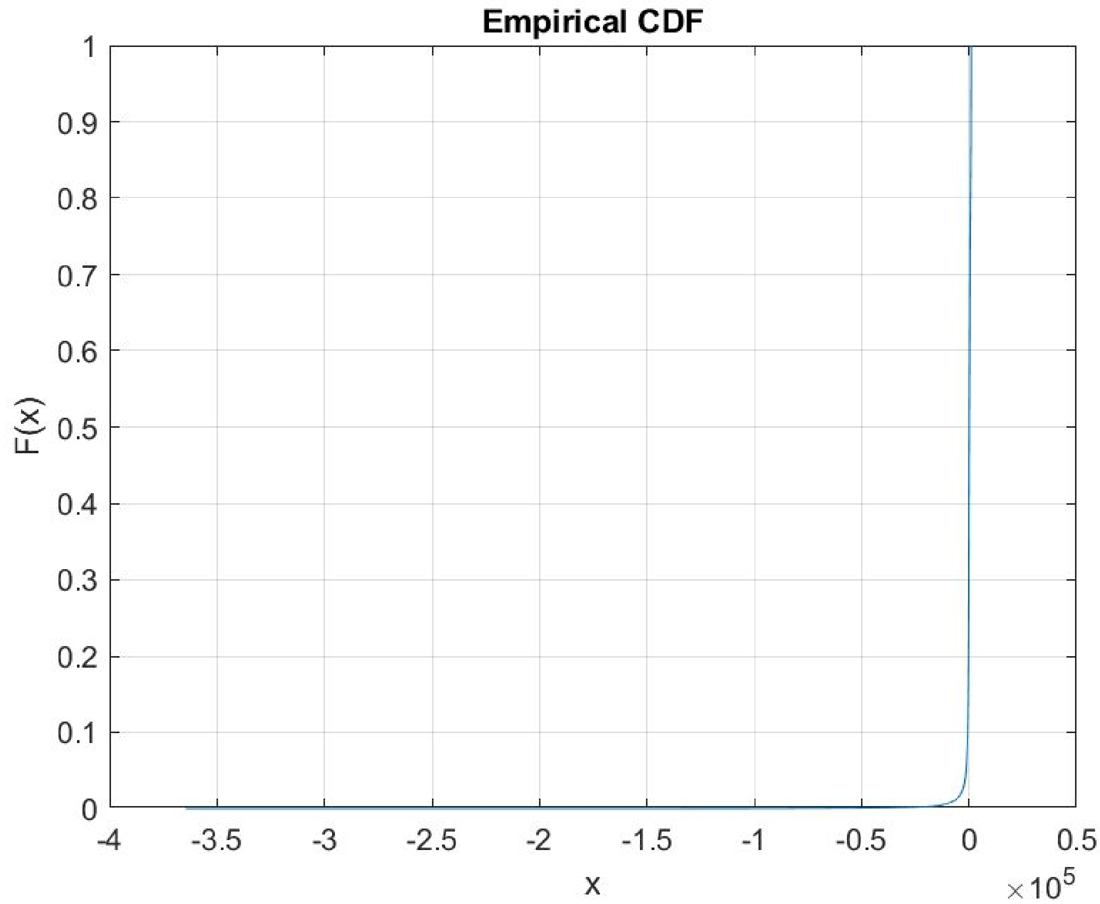
As the goal increased, both failure and cancellation rates also increased, while success rates decreased. This could be due to campaigns running out of time and not reaching their goal or that they were overpriced and dissuaded viewers from donating. It seems that lower goals tend to succeed more often than higher goals.

I also wanted to look at the disparity between goals and amonuts pledged overall, as well as in the 4 goal groups. In order to find this, I subtracted the amounts pledged from their respective goals and then plotted them both completely and in each bucket. The resulting plots are below, including a CDF where the x-values are the differences and a plot where the y-values are the differences. In the first plot for each section, x-values <= 0 mean that the campaign either met its goal or raised more than the goal amount and therefore was successful. Any positive x-values indicate a positive difference and therefore a campaign that did not reach its goal. The same rules apply for all of the second plots, but instead for the y-values.

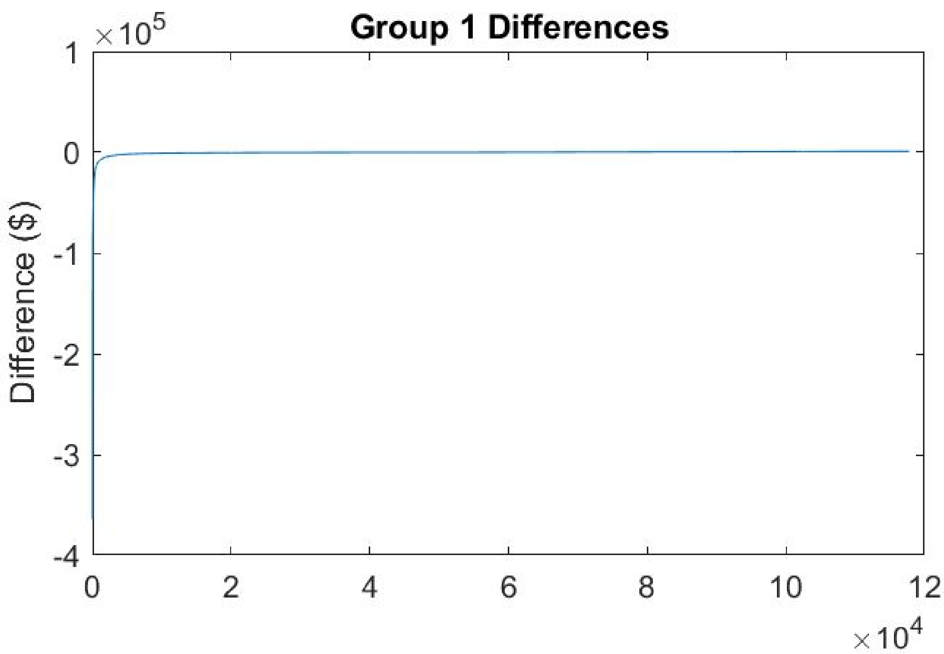


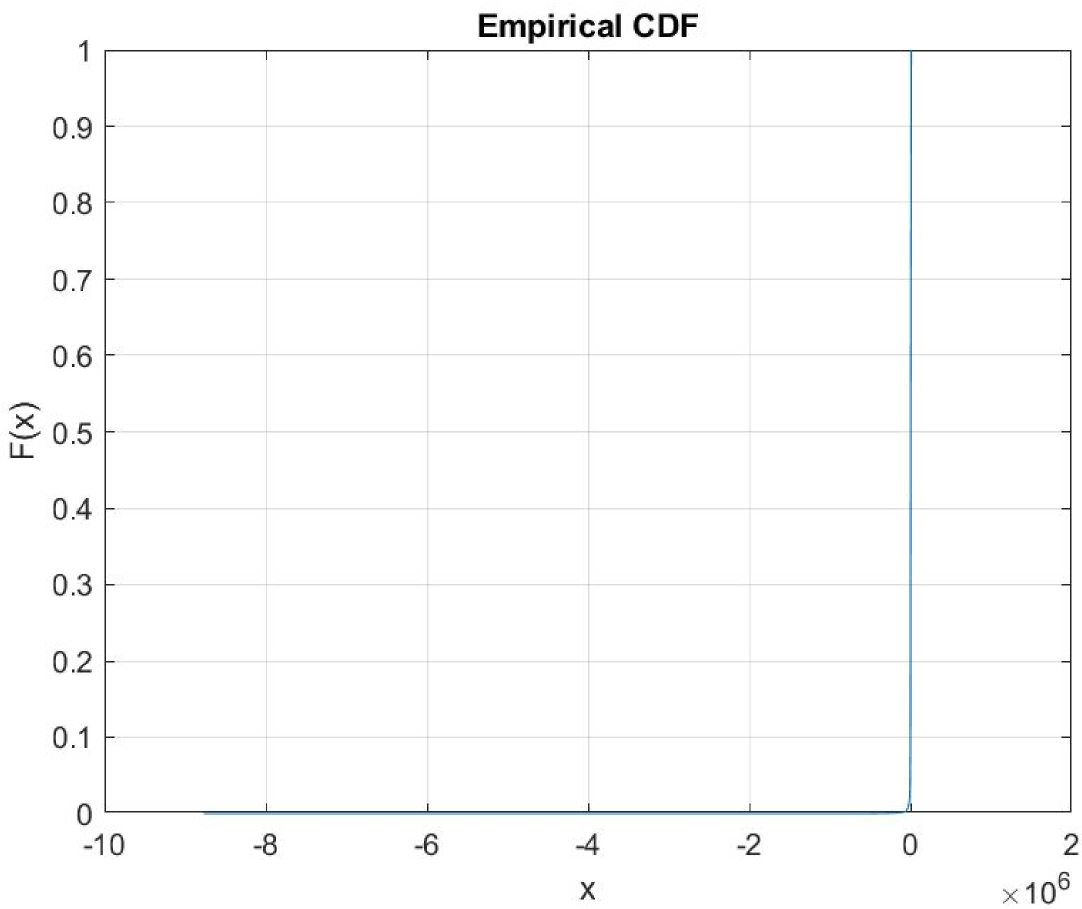
Difference Between All Goals and Amounts Pledged (USD)



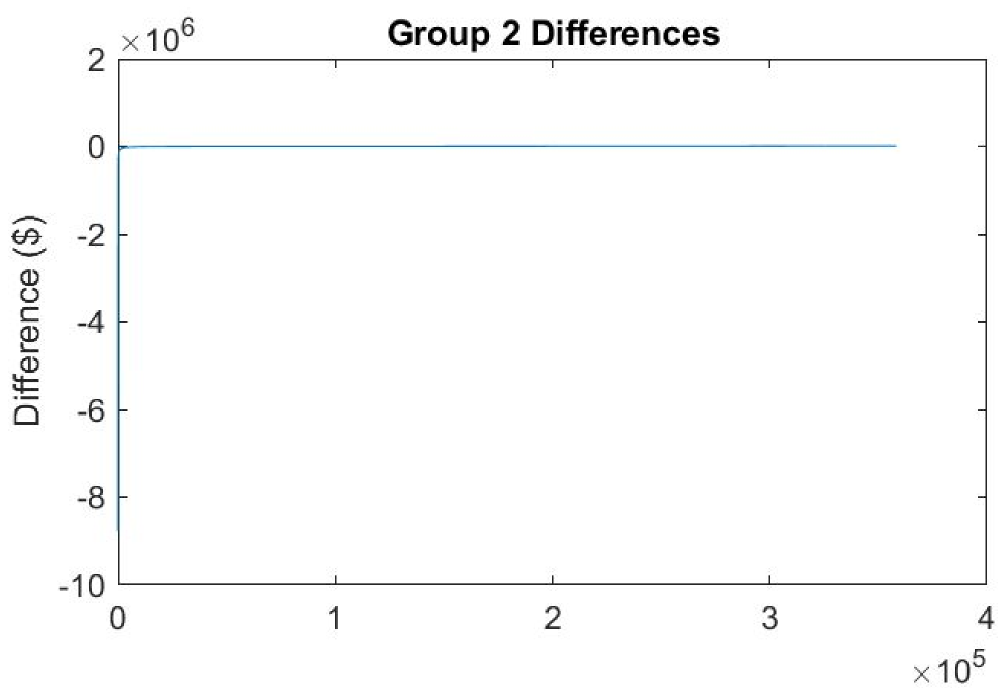


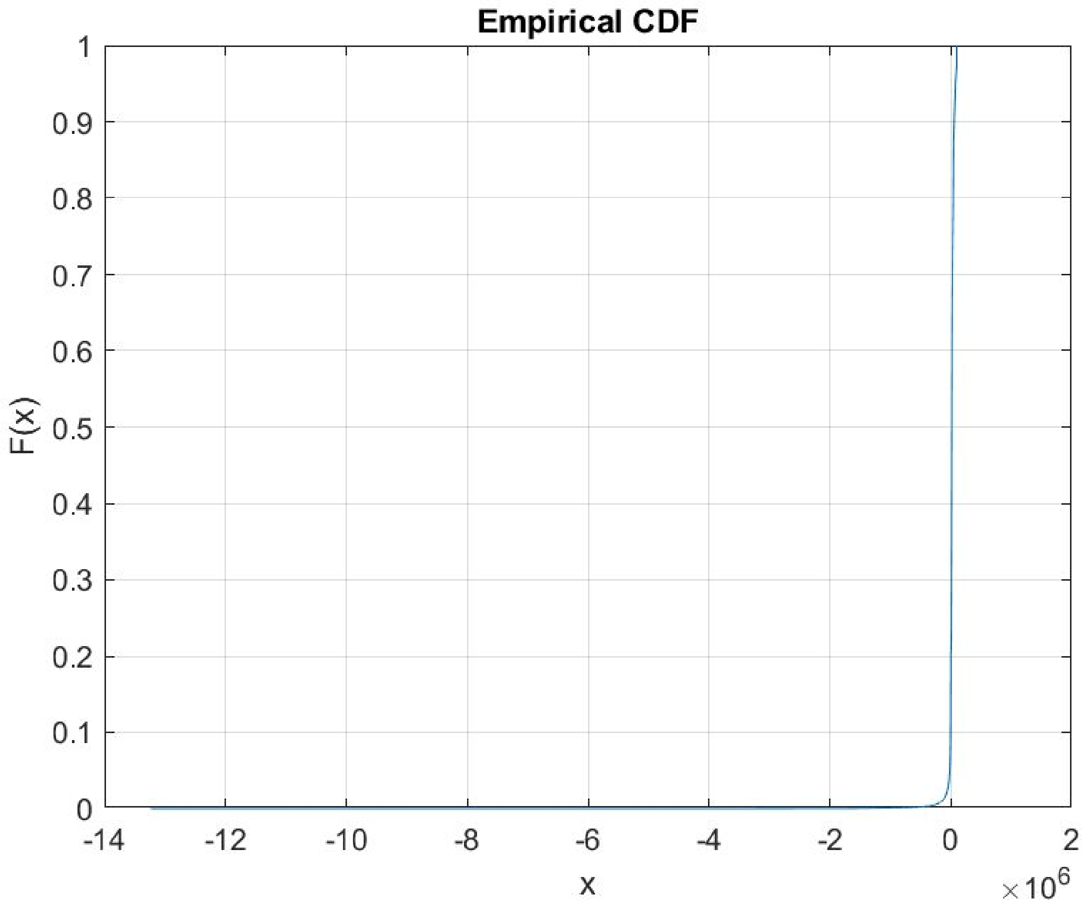
Difference Between Group 1 Goals and Amounts Pledged (USD)



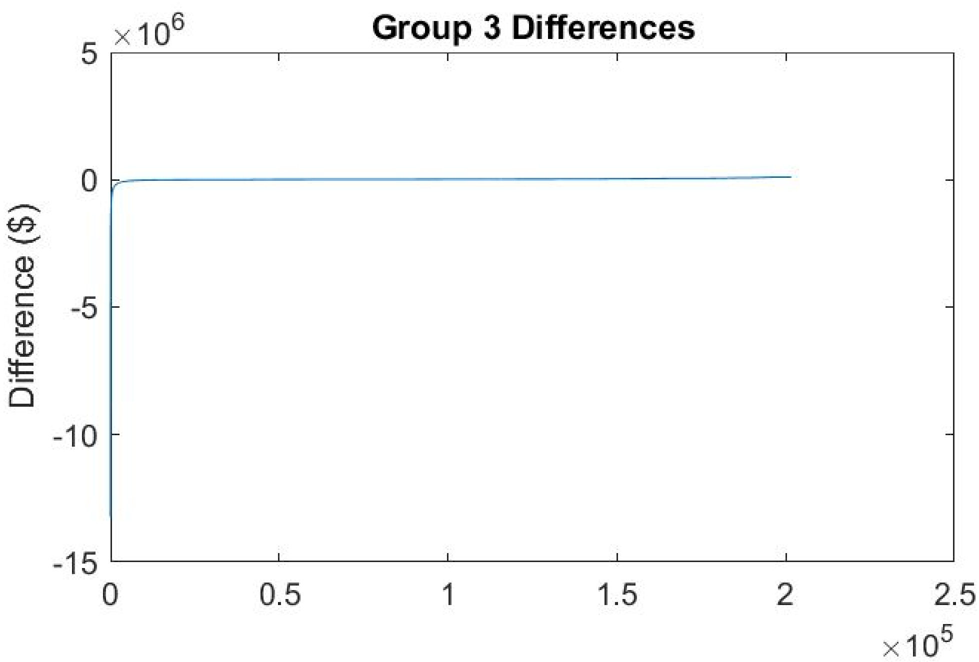


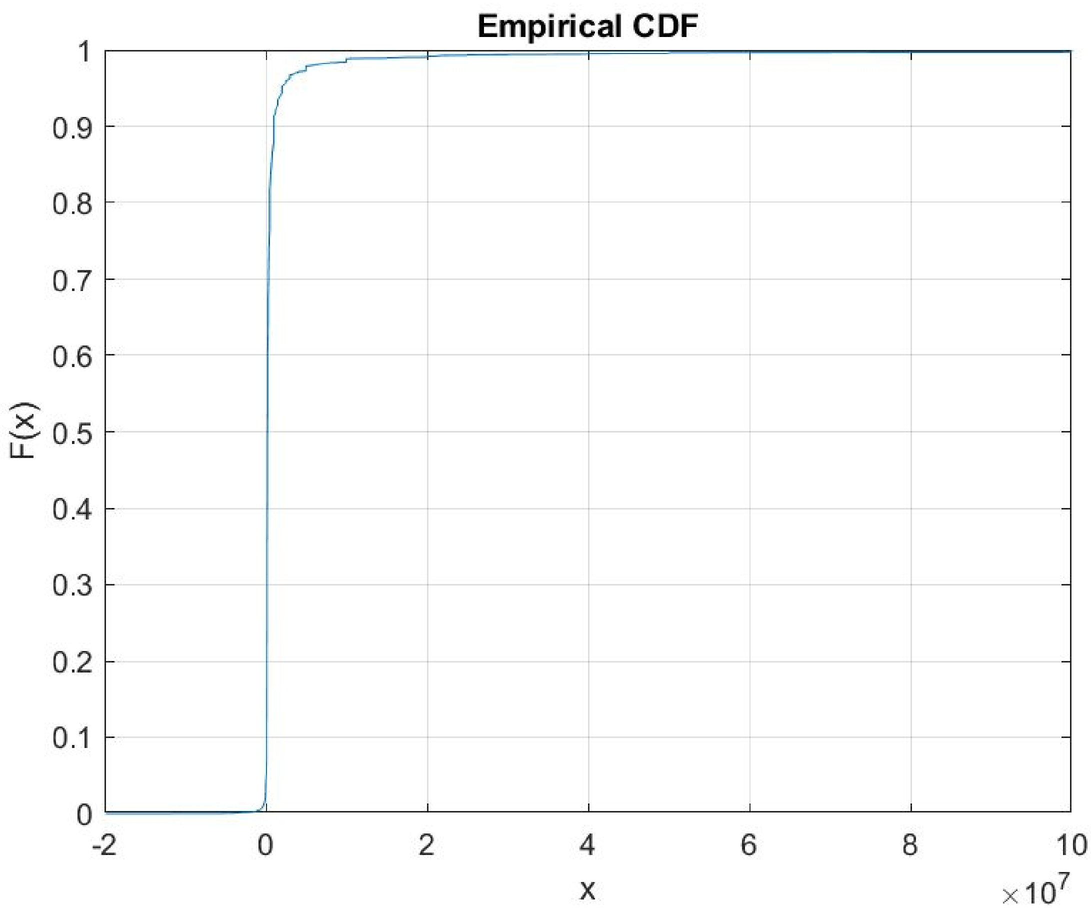
Difference Between Group 2 Goals and Amounts Pledged (USD)



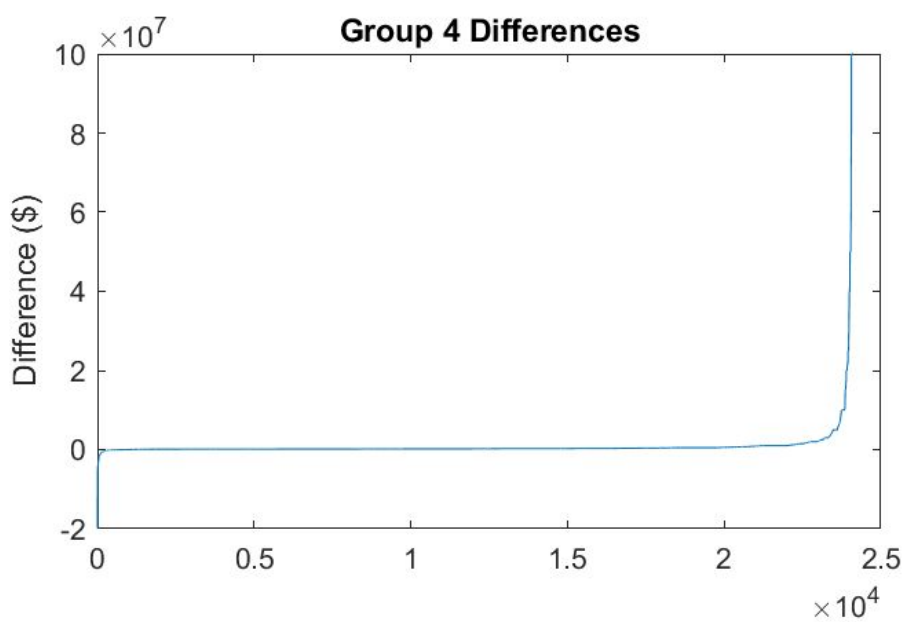


Difference Between Group 3 Goals and Amounts Pledged (USD)





Difference Between Group 4 Goals and Amounts Pledged (USD)



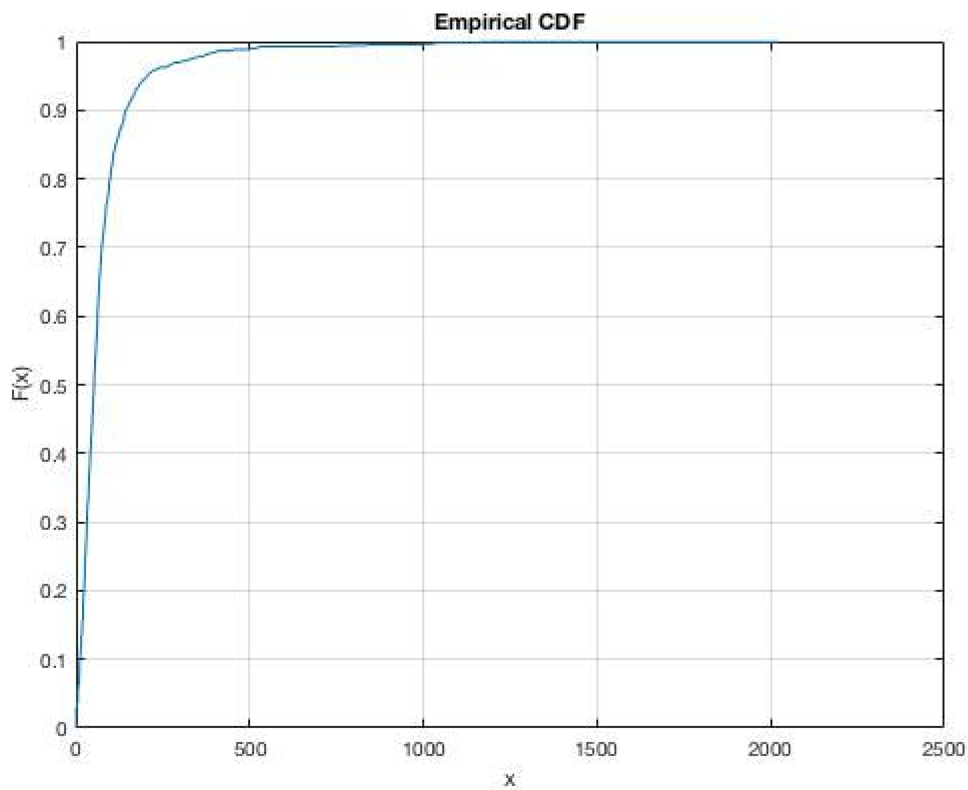
Overall, campaigns tended to have positive differences, which correlates to the fact that those include failed, cancelled, suspended and potentially live and undefined campaigns. As seen in the previous visuals about the states of each group, the lower goals tended to have more 0 or negative differences, while higher goals had more positive differences. Group 1 has a longer negative tail, while Group 4 has a longer positive tail. The difference trends follow the state distributions of each group.

Based on the analysis of these relationships, it does not seem like kickstarter is as effective or efficient of a platform as it could be. More analysis on duration and category along with this could help kickstarter recommend smarter campaign strategies to its users and improve the percentage of successful campaigns.

Relationship Between Single Pledge Amount and Goal

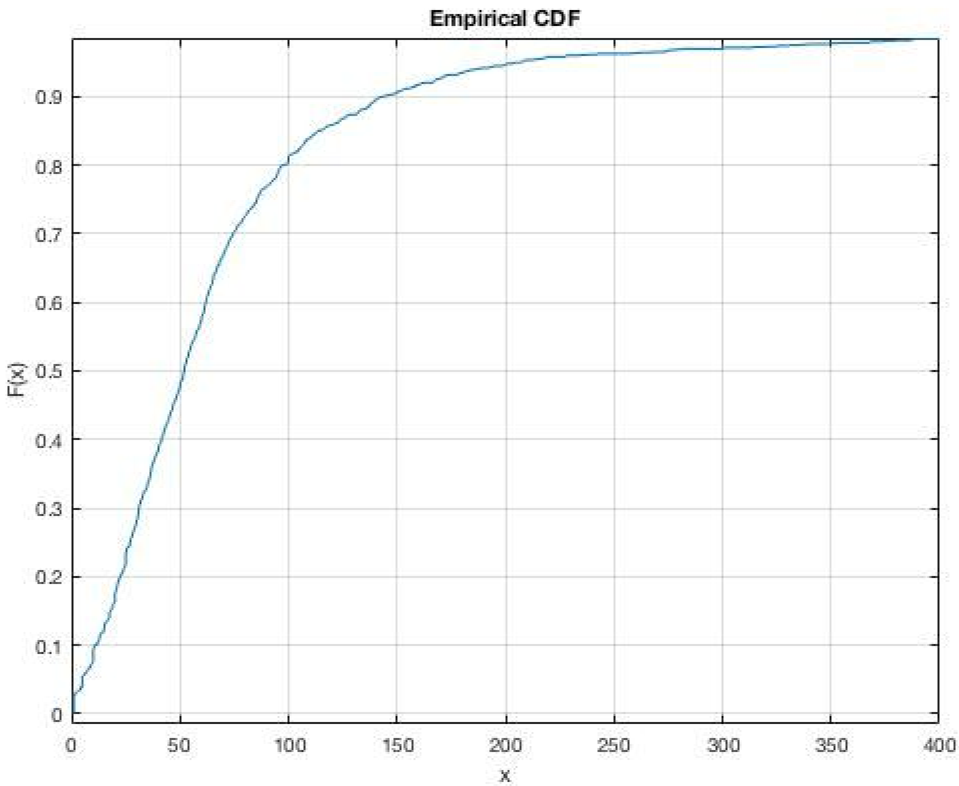
Wanyi Chen

This section explores the relationship between the average pledge amount and the fiscal goal amount. Do people give more money to projects that have higher goals, or do people give similar amounts regardless of how high the goal is? How does the trend various across different areas? To find out, I investigated the following data fields: backers, pledged, goal, and currency. The average pledge amount can be calculated by dividing the pledged amount by the number of backers. Investigate the data separately based on different currencies revealed trends across different areas. I first plotted CDF graphs of average pledge amounts. The figure below shows the average pledge amounts of all campaigns done in US dollars.



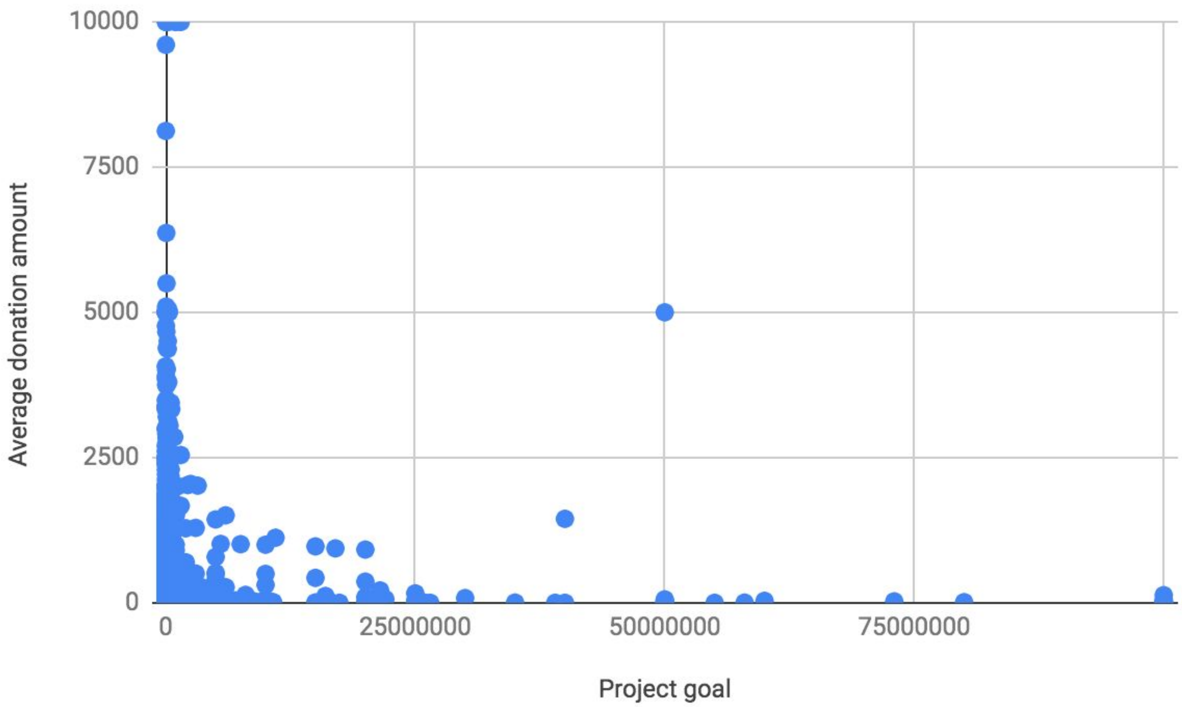
USD - average pledge amount

As shown in the graph above, most campaigns have similar individual pledge amounts, except for a few outliers. Zooming in the graph reveals clearer trend. About 90% of individual pledge amounts are less than $150, and about 50% are less than $50.



USD - average pledge amount (zoomed in)

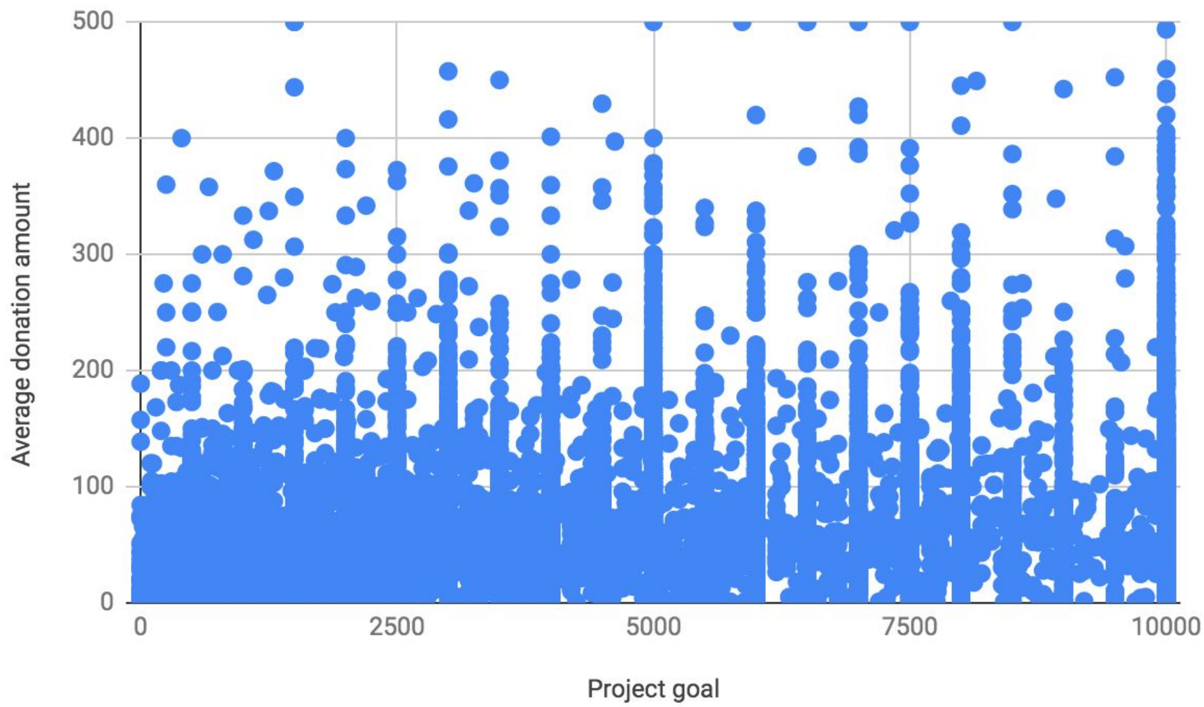
To further investigate, I then plotted graphs using goals as the x-axis and average pledge amounts as the y-axis to see their relationships. The figure below shows the graph for all campaigns done in US dollars.



USD goal-average pledged amount graph

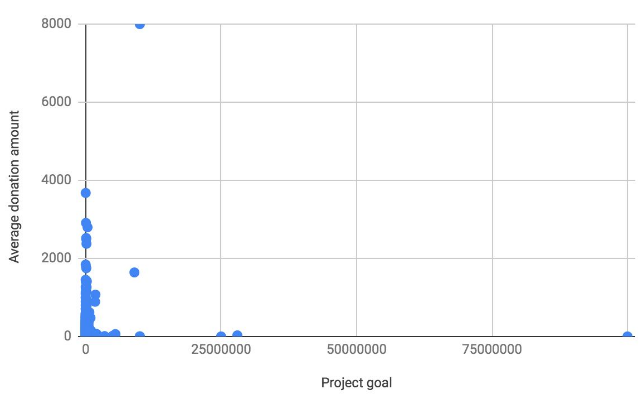
There is a strong inverse relationship between project goal and average donation amount. Most campaigns' goals are less than $10,000, as discussed in the previous section. However, for the few outliers whose goals are far more ambitious (> $25,000,000), backer tend to pay very few amounts. On the other hand, as discussed above, an average backer pay less than $150 in most campaigns. Yet the outliers (campaigns whose average pledged amount are more than $2500) tend to be campaigns with lower goals. One possible explanation is that people may find campaigns with lower goals more realistic, so they may contribute more to help achieve the goal faster. Meanwhile, campaigns with extremely high goals may seem unrealistic, so they attract fewer backers and people pay less.

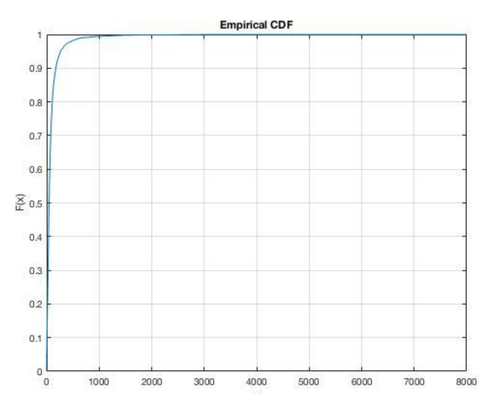
However, for the vast majority of the campaigns, campaign goals cannot predict average pledge amounts. Zooming in the previous graph shows that there is no identifiable relationship between campaign goals and average pledge amounts for campaigns with goals less than $10,000.



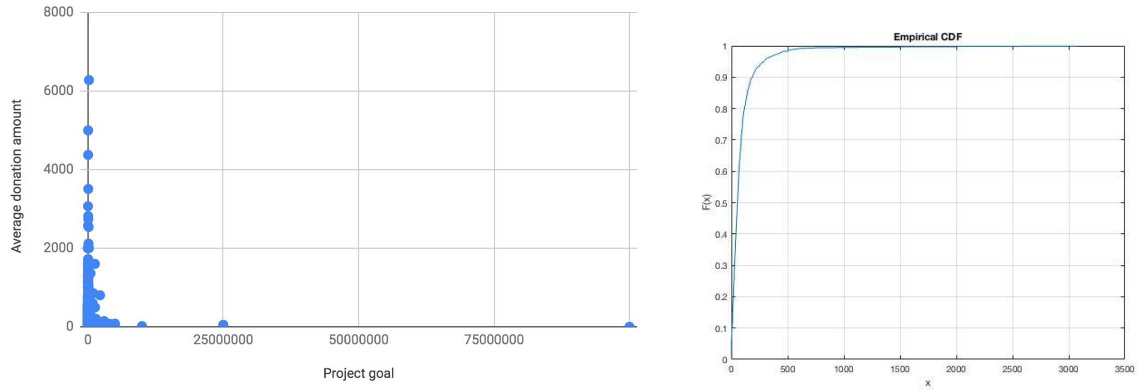
USD goal-average pledged amount graph (zoomed in)

Western countries follow this trend. Graphs for campaigns done in western (North American, European, and Australian) currencies are listed below. The graphs all look very similar to USD's graph.

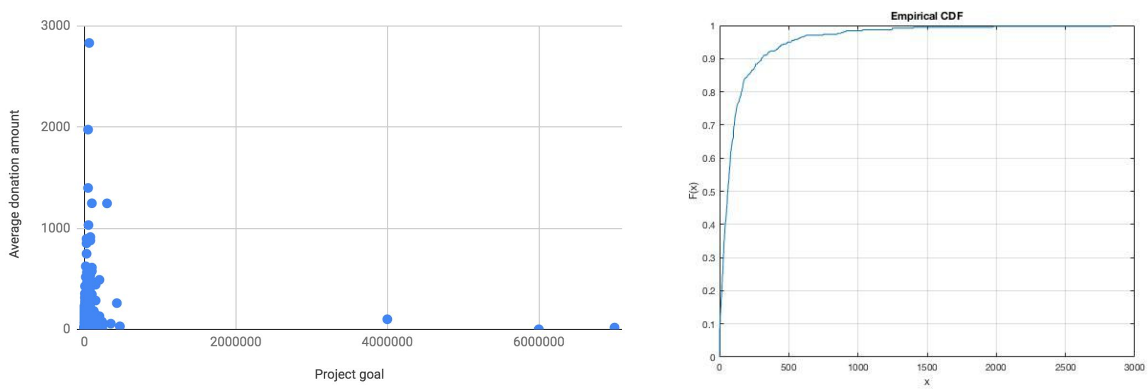




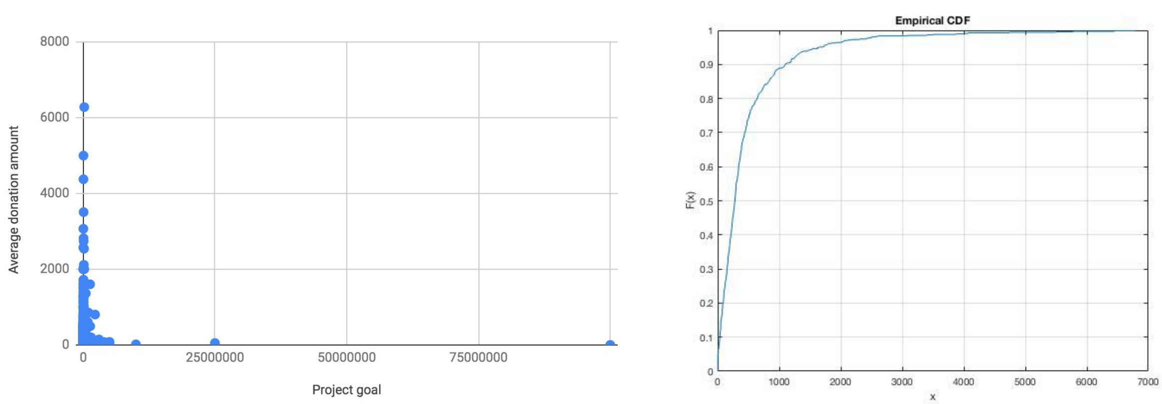
AUD (Australian Dollar)



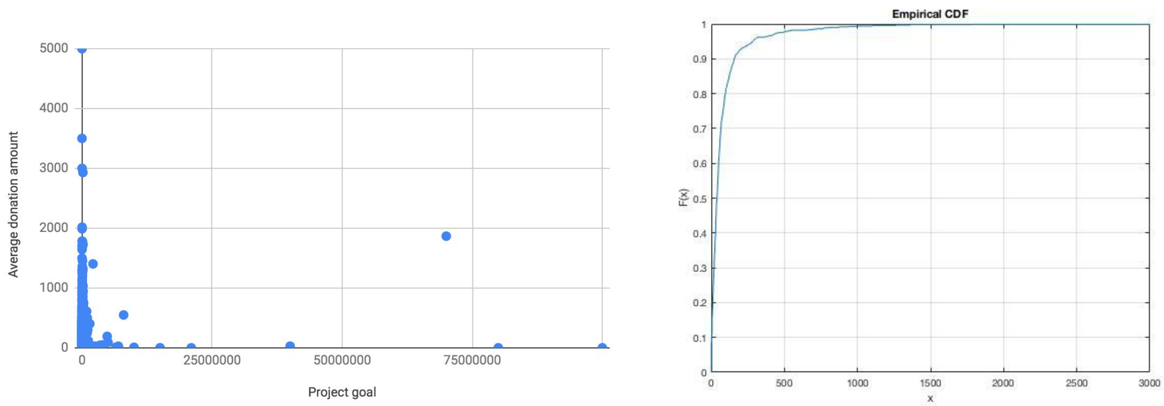
CAD (Canadian dollar)



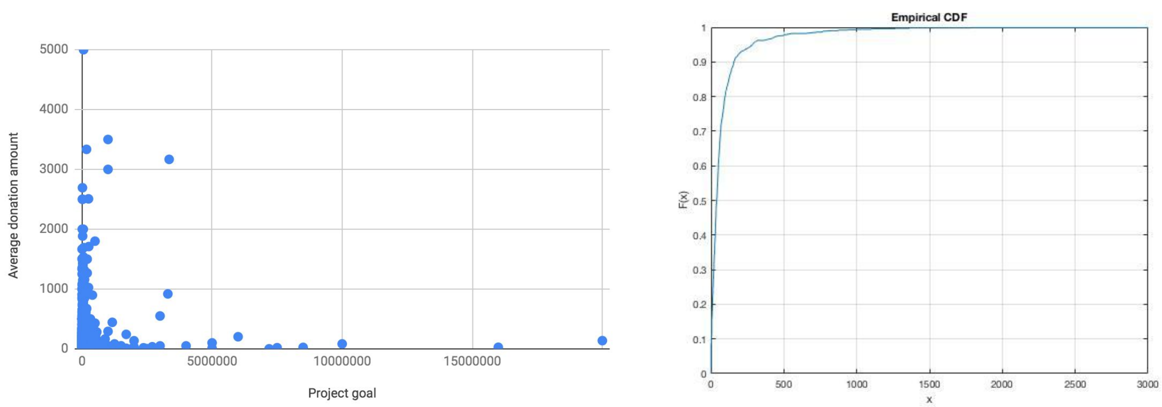
CHF (Swiss Franc)



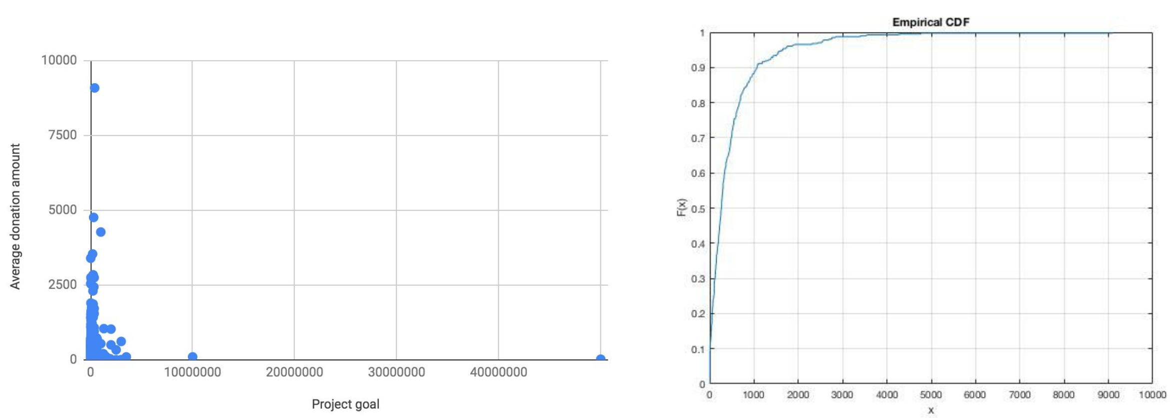
DKK (Danish krone)



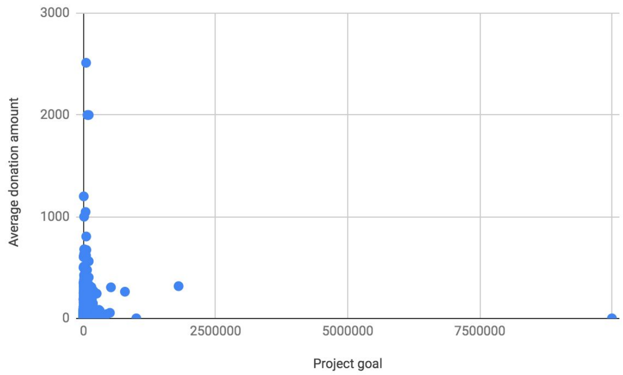
EUR (Euro)

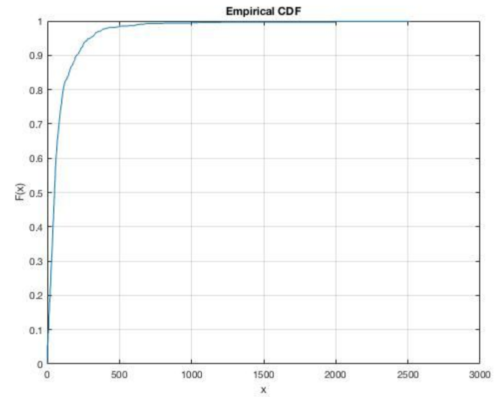


GBP (Great British pound)

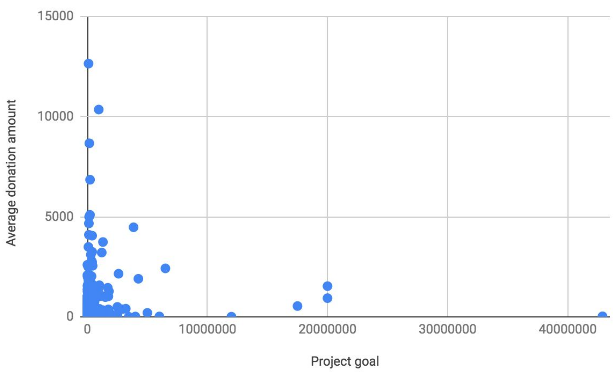


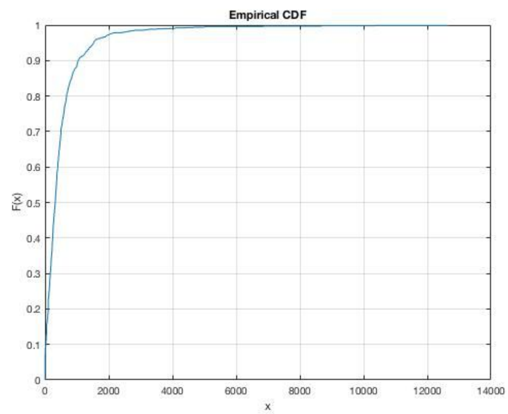
NOK (Norwegian krone)





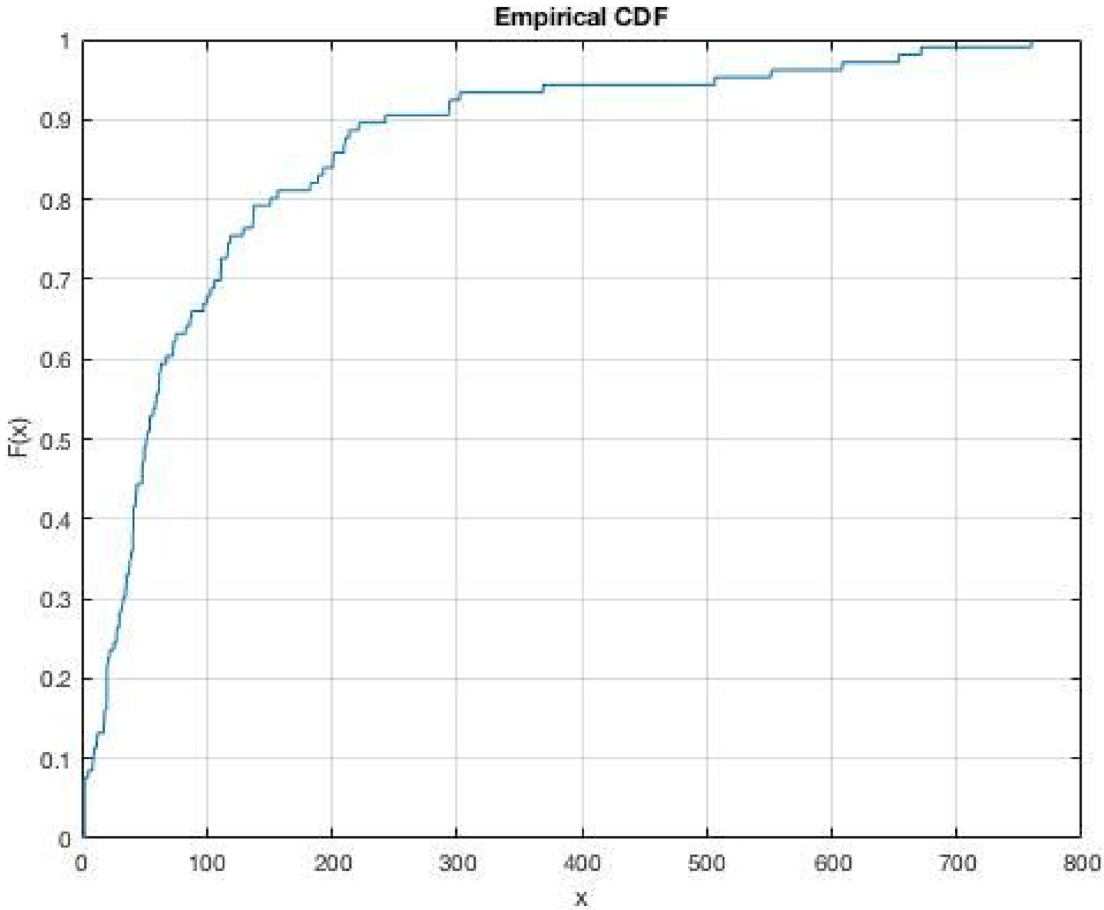
NZD (New Zealand dollar)





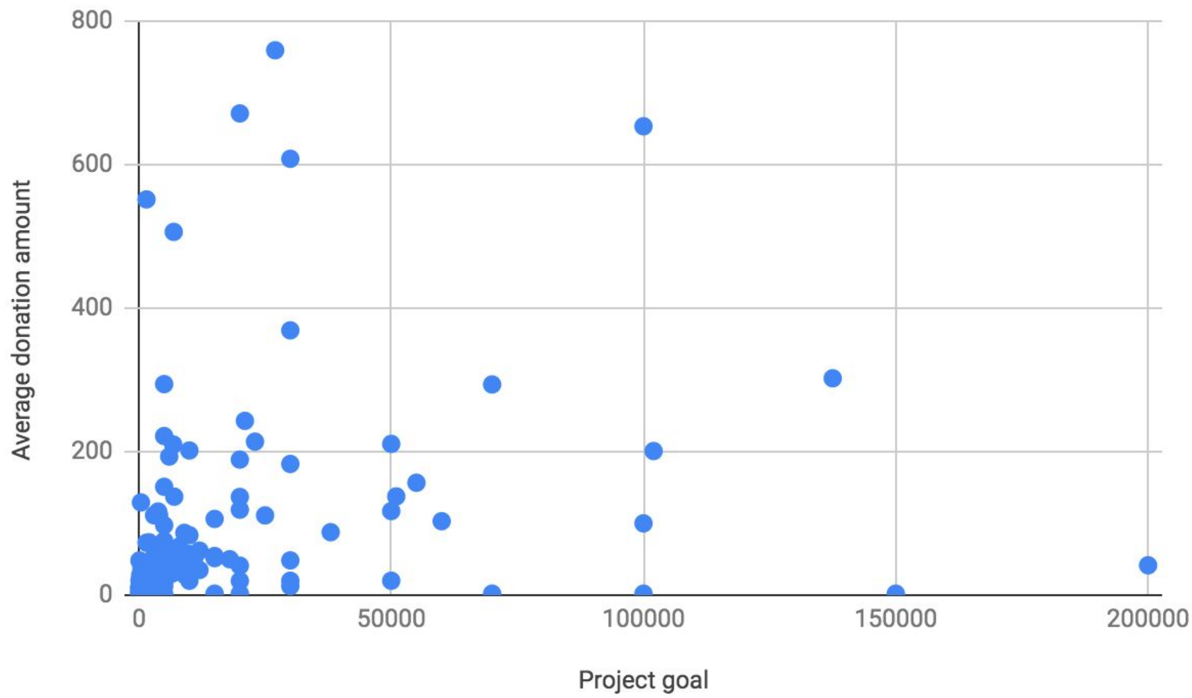
SEK (Swedish krona)

However, non-western countries may not follow this trend. Take Singapore for example. As the following CDF shows, Singapore has less outliers than the U.S. In the U.S., 90% of average pledged amounts are less than $150 USD, but the highest average pledged amount can be as high as $2000 USD. Yet in Singapore, 90% of average pledged amounts are less than $250 SGD (about $184 USD), and the highest average pledged amount is less than $800 SGD (about $590 USD). Although the distribution of Singapore's average pledged amounts still has a long tail, it is more even than USA's distribution.

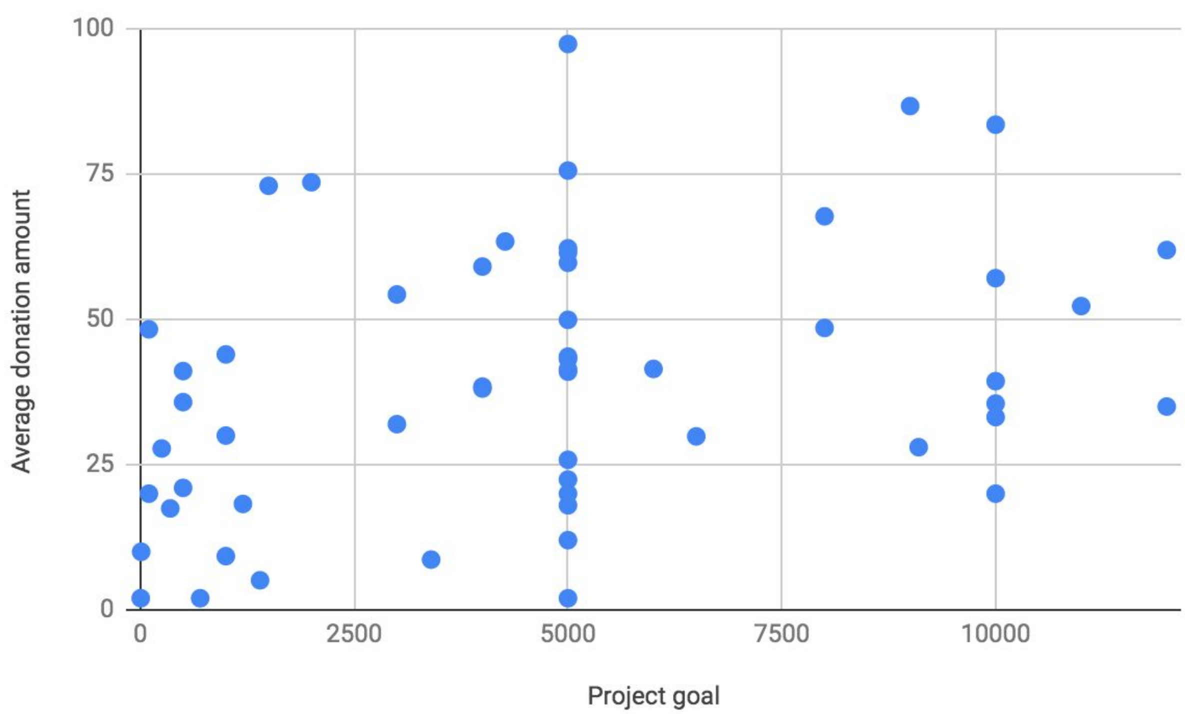


SGD (Singapore dollar) - average pledged amount

As for the relationship between average pledged amount and project goal, Singapore also has fewer outliers, as shown in the graph below. Comparing to USA's plot, Singapore's plot is more scattered. Zooming in, the graph still shows no predictable relationship between average pledged amount and project goal.

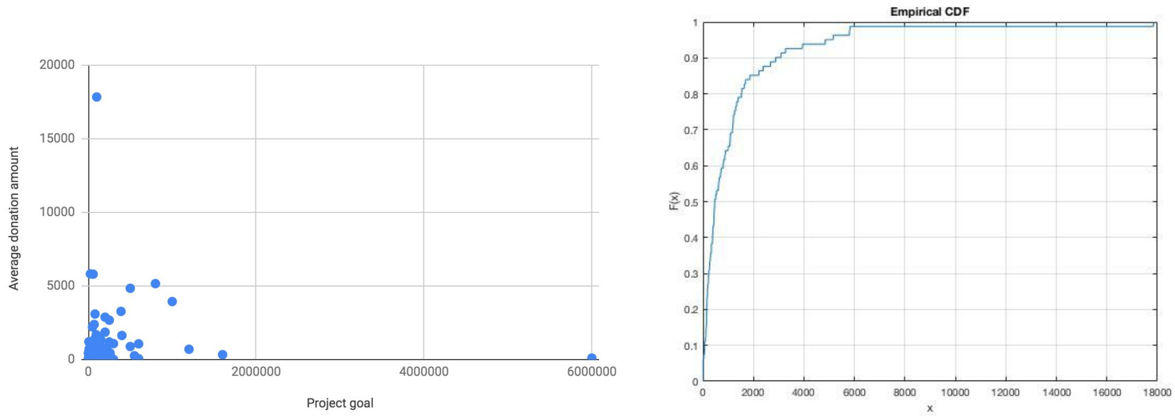


SGD goal-average pledged amount graph

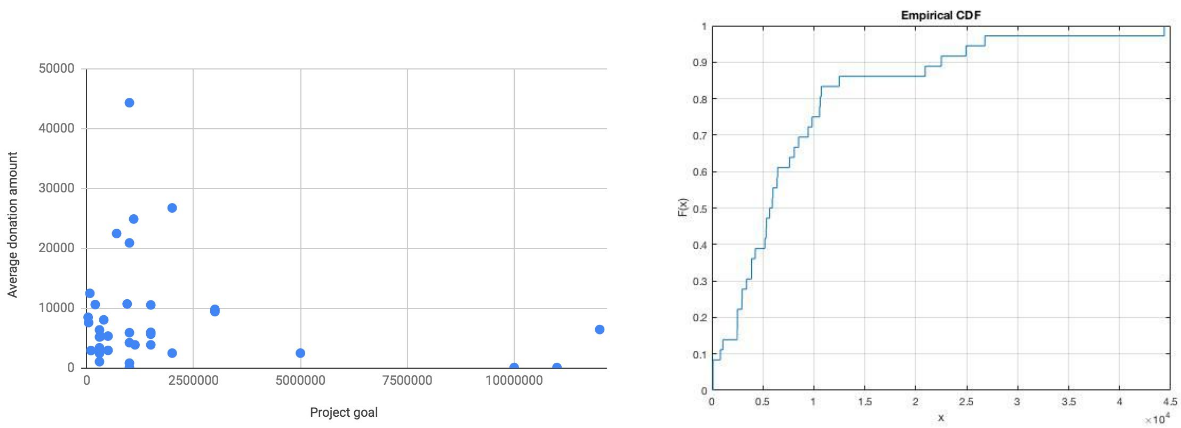


SGD goal-average pledged amount graph (zoomed in)

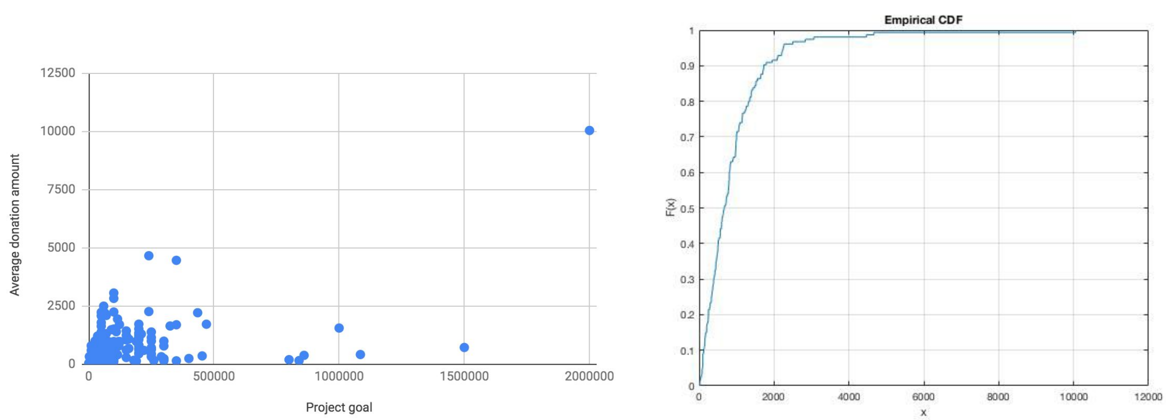
Non-western countries/areas (Hong Kong, Japan, and Mexico) follow a similar trend. Their graphs are listed below. However, it remains unclear whether the difference between western and non-western countries is truly due to differences in culture or is simply due to a lack of data for non-western countries. Non-western countries have much fewer entries in the dataset, and therefore they may also have fewer outliers.



HKD (Hong Kong dollar)



JPY (Japanese yen)



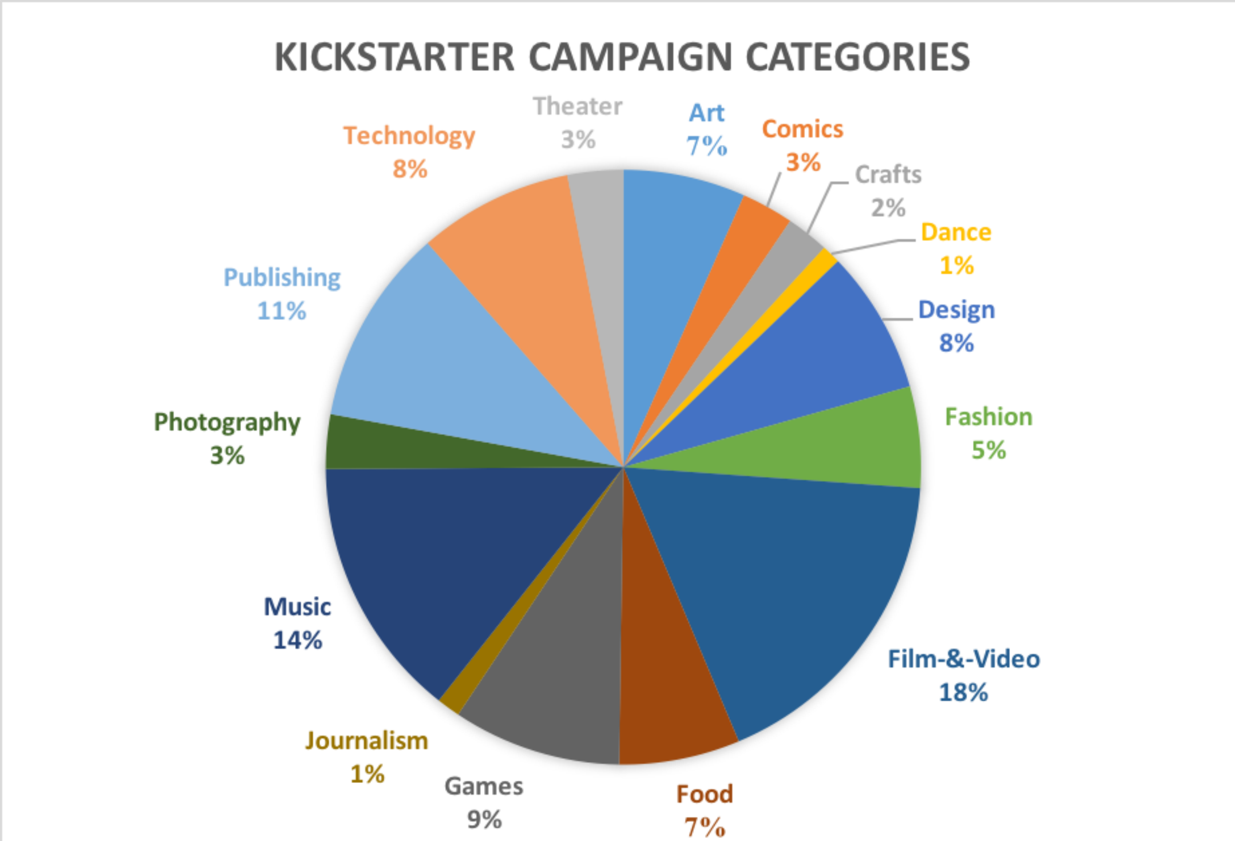
MXN (Mexican Peso)

Data Variation Across Categories

Saumya Ray

In this section, the relationship between campaign categories, country of origin, and outcome is examined. There are a variety of categories that can be chosen for a Kickstarter project. Each main category can have multiple subcategories. In all, there are fifteen main categories and 120 subcategories. I decided to specifically look at main categories due to the large number of subcategories. The following table shows how many campaigns were found for each main category in this dataset. The subsequent graph shows the prevalence of each campaign category in comparison to the whole dataset.

Category	Campaign Count
Art	45948
Comics	19562
Crafts	15991
Dance	7140
Design	53935
Fashion	38059
Film-&-Video	121247
Food	45374
Games	63235
Journalism	8824
Music	98567
Photography	20443
Publishing	74095
Technology	58103
Theater	20875

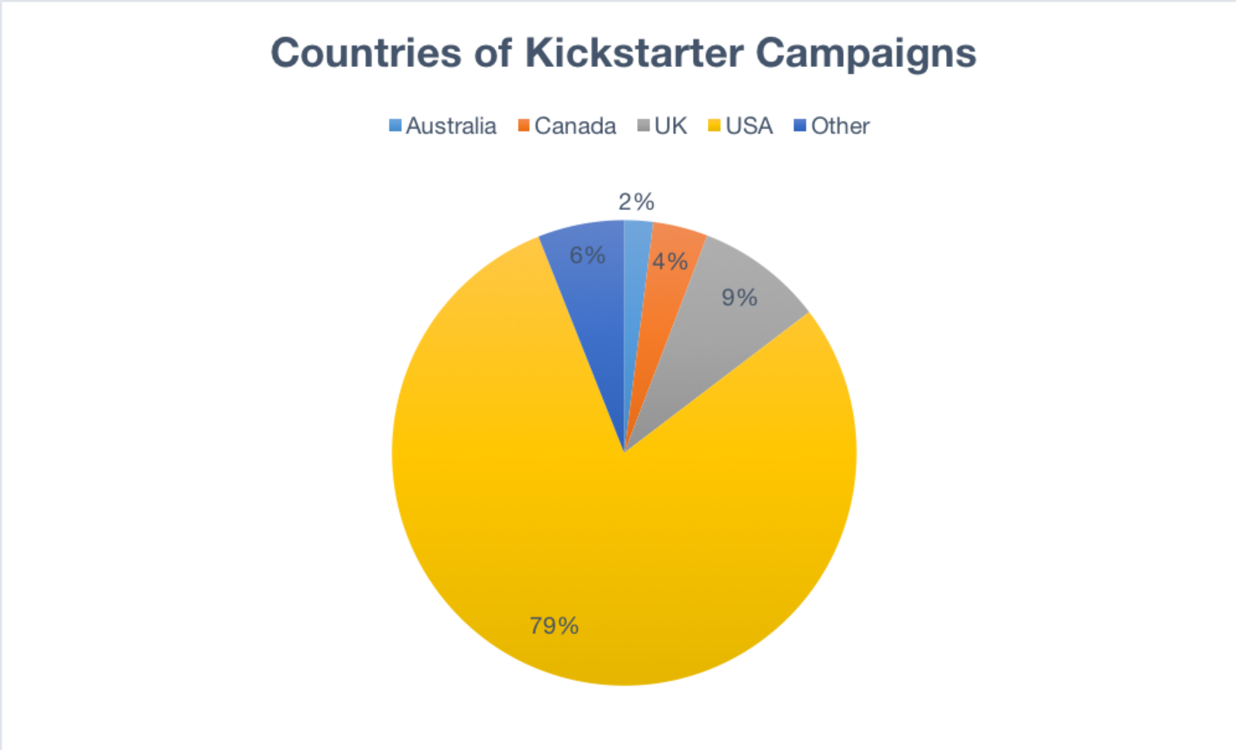


The most common categories used in this dataset are Film & Video, Music, Publishing, Games, and Technology. Later in this report, I will be comparing the prevalence of categories to the overall success rate of these categories.

Kickstarters campaigns in this dataset originated from 22 different countries that were located in North America, Europe, Asia and Oceania. The United States had the most Kickstarters, with over 540,000 campaigns. Other highly represented countries in this dataset were Australia, Canada, and the United Kingdom. Countries were labeled in the dataset with two letter country codes. The following table shows the number of campaigns in each country. The graph after that shows the prevalence of each country in the dataset.

Country Code	Country	Campaign Count	Percent of Total
AT	Austria	962	0.140682353
AU	Australia	13767	2.013278542
BE	Belgium	989	0.144630819
CA	Canada	26242	3.837615712

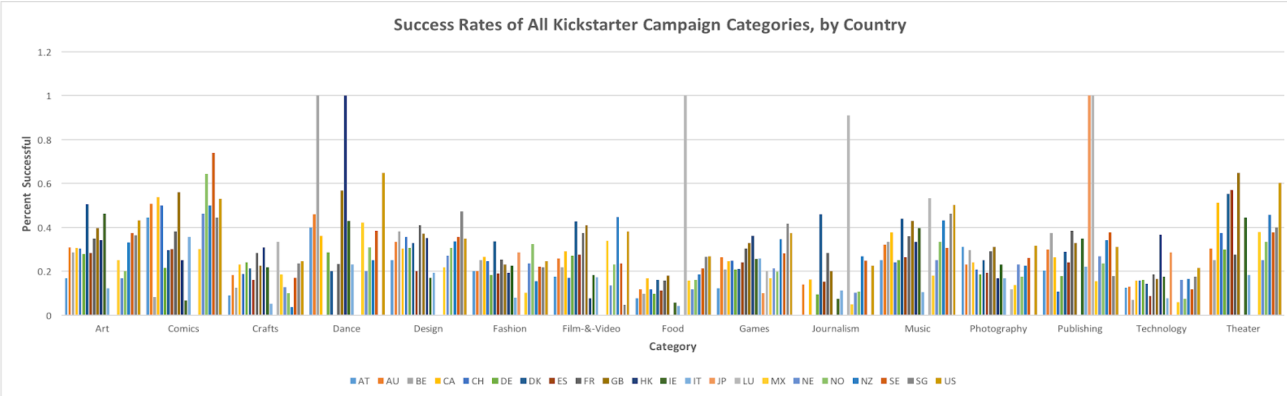
CH	Switzerland	1191	0.174171188
DE	Germany	6722	0.9830216
DK	Denmark	1851	0.270689227
ES	Spain	3535	0.516956465
FR	France	4764	0.696684752
GB	UK	60172	8.799520335
HK	Hong Kong	710	0.103830011
IE	Ireland	1367	0.199909332
IT	Italy	4512	0.65983241
JP	Japan	39	0.005703339
LU	Luxembourg	100	0.014623945
MX	Mexico	1935	0.282973341
NE	Netherlands	4973	0.727248797
NO	Norway	1219	0.178265893
NZ	New Zealand	2525	0.369254618
SE	Sweden	2967	0.433892456
SG	Singapore	662	0.096810518
US	USA	542606	79.35040435
Total		683810	



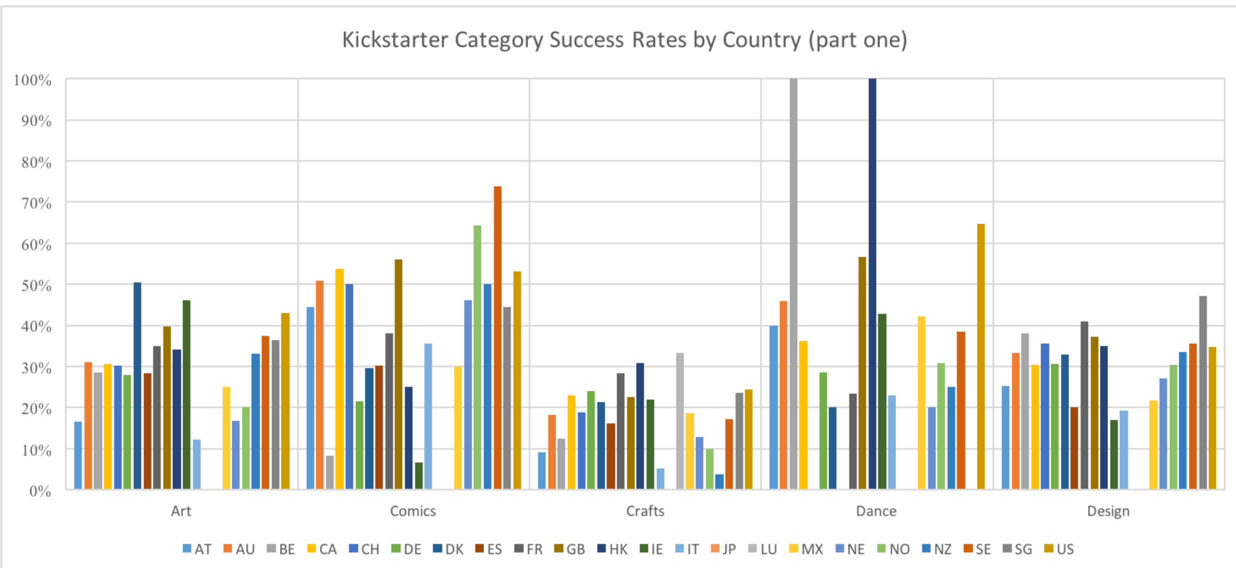
With this overall base data, I will be analyzing which categories have a higher success rate than others, and which categories are not as successful. I will also look at which categories are the most successful for each country that has Kickstarters. In addition to success metrics, I will also be exploring which categories of campaigns can receive more than 100% of their initial goal.

To analyze this data, I wrote a few MapReduce and Spark programs using the following fields: category, main_category, goal, pledges, backers, and country. This resulting data will be used to see which categories are more successful in online crowdfunding through Kickstarter in a specific county.

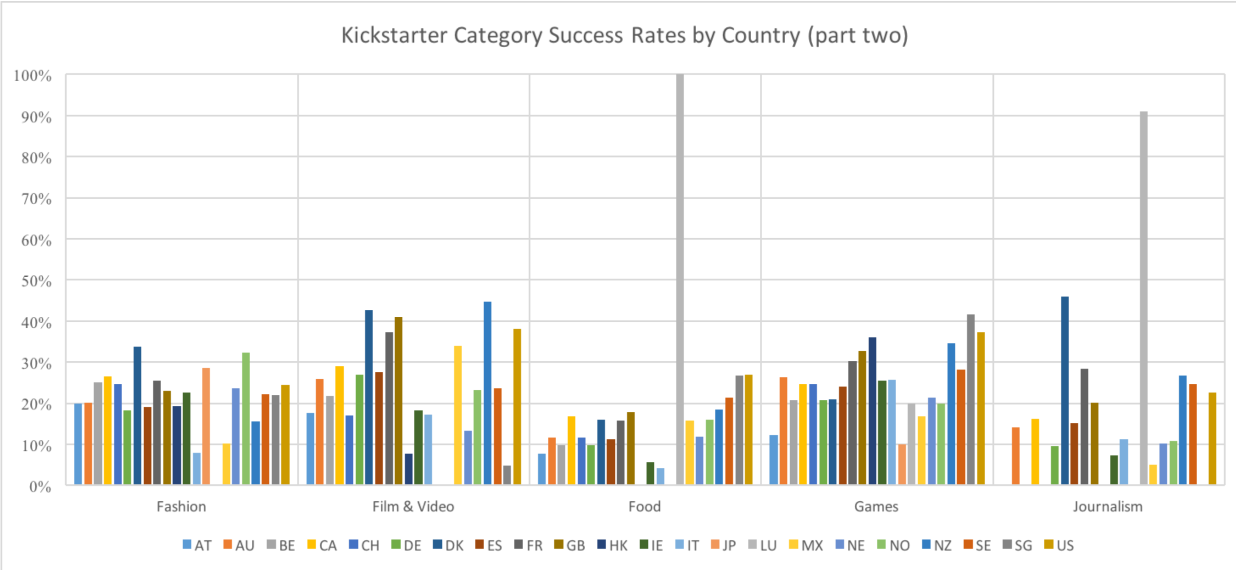
The following graph shows the success rates of all Kickstarter campaigns for each country. This is intended to be only used for visualization purposes.



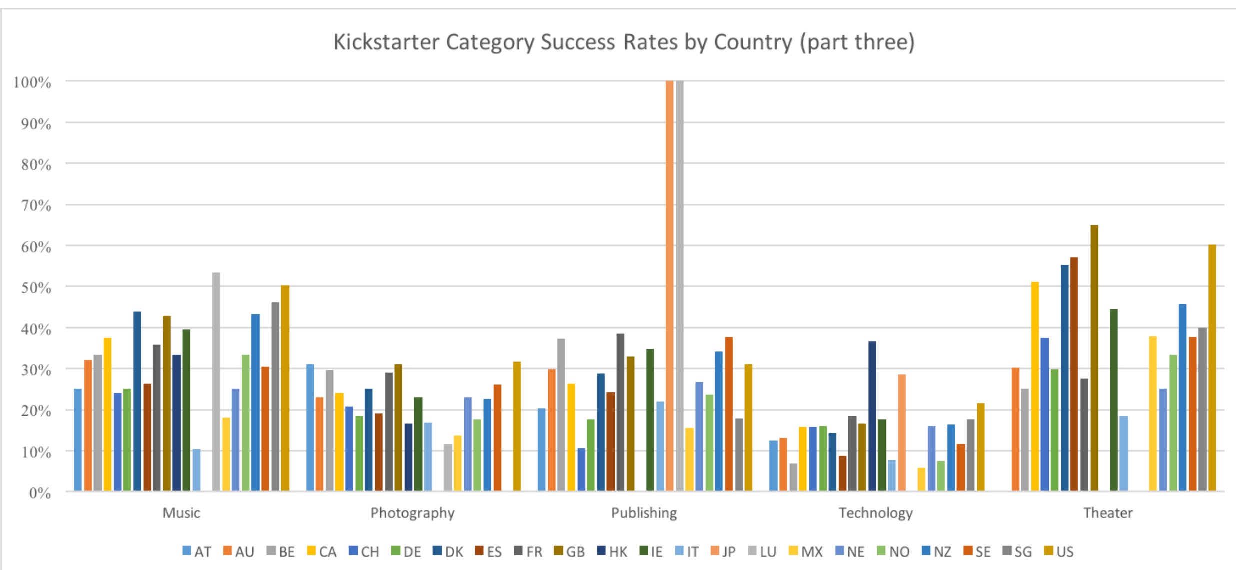
To better see the data, I split the previous graph into three graphs with five categories each.



In the above graph, success rates were the highest for Dance and Comics. Most success rates were under 50% and typically ranged between 30-50%. Craft-based Kickstarters were the least successful.

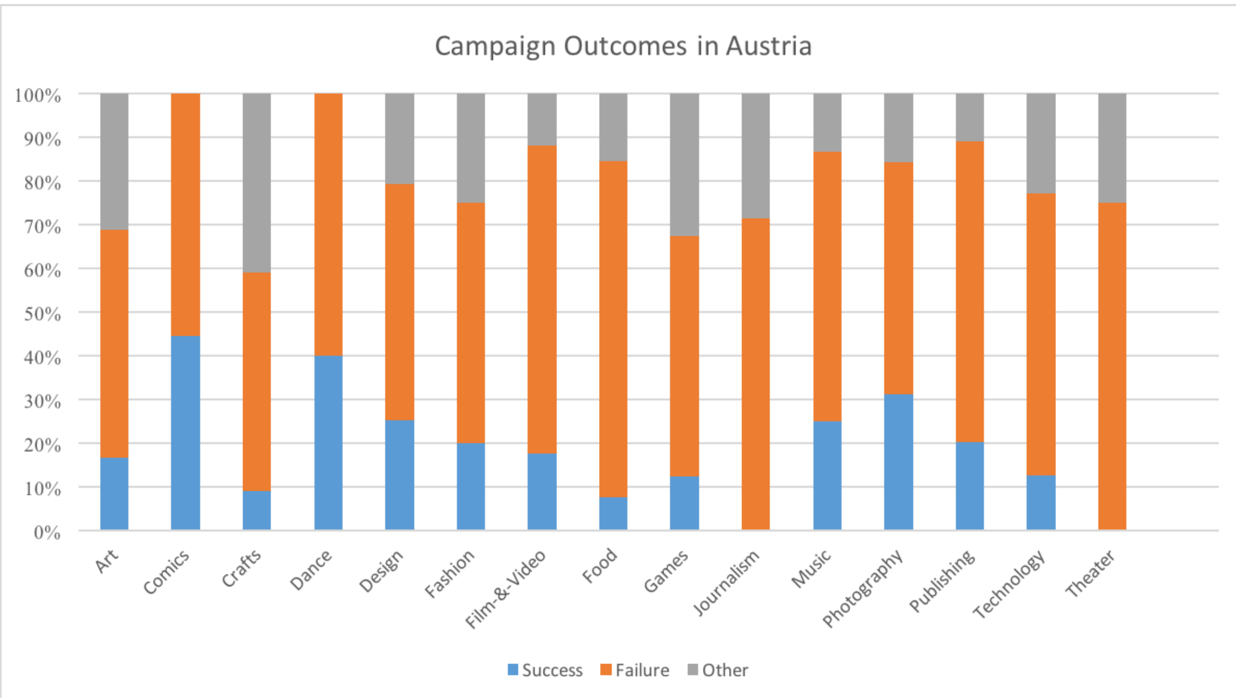


In the above cohort of categories, success rates were mostly under 40%. There are two outliers in this graph, both with the country Luxembourg. Since this country has one of the fewest campaigns in the dataset, a successful outcome to the campaign would lead to a skewed average. Film & Video and Games were the most successful categories.

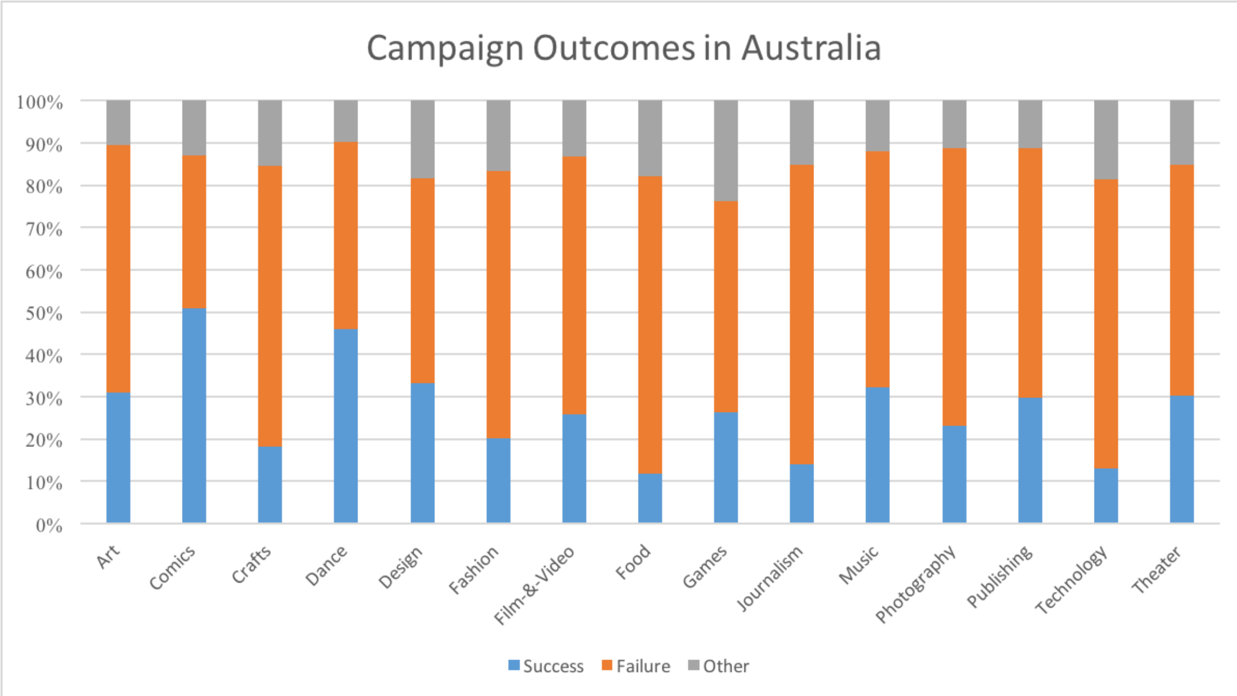


For the last group, success rates for each country in each category were mostly under 50%. The most successful categories were Music, Publishing, and Theater. There were also two outliers in this graph, Japan and Luxembourg in Publishing. The 100% success rate these two countries have in this category is likely because both countries have very few Kickstarter campaigns.

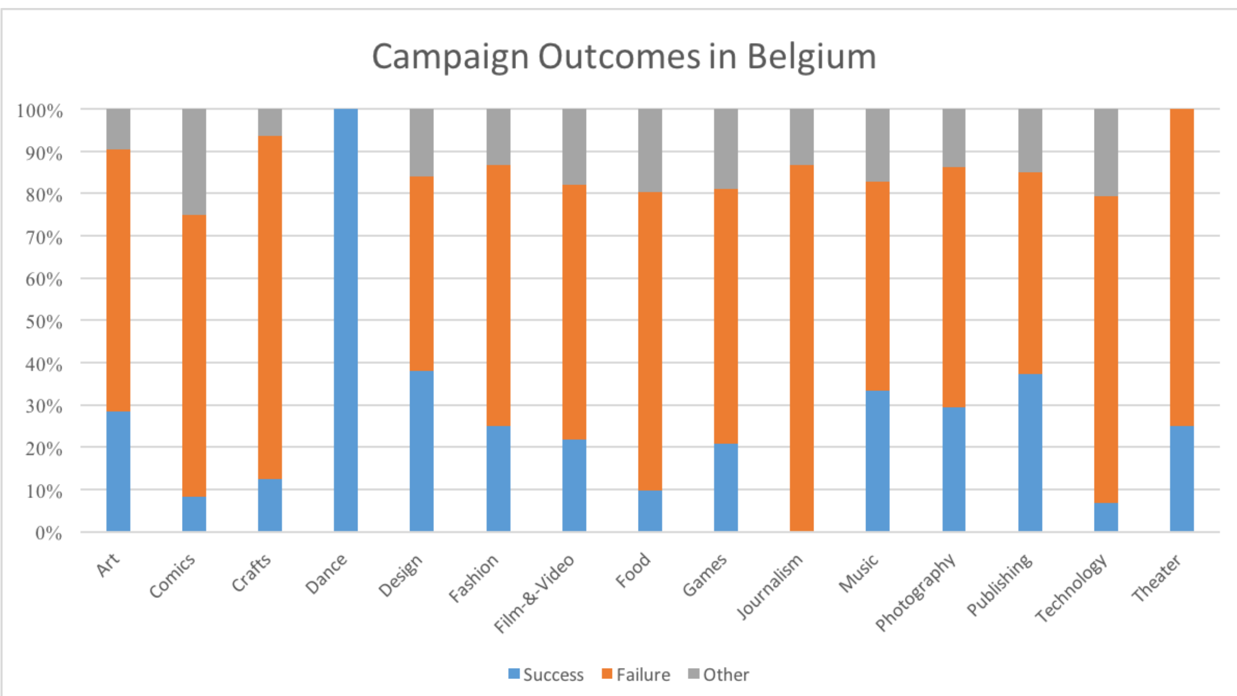
Next, I will be looking specifically at each of the 22 countries that have Kickstarter campaigns and their outcomes. There are several different outcomes a campaign can have: successful, failed, canceled, live, and suspended. Since most campaigns either succeeded or failed, I grouped canceled, live, and suspended into the "other" category. The following graphs detail these three outcomes for each country.



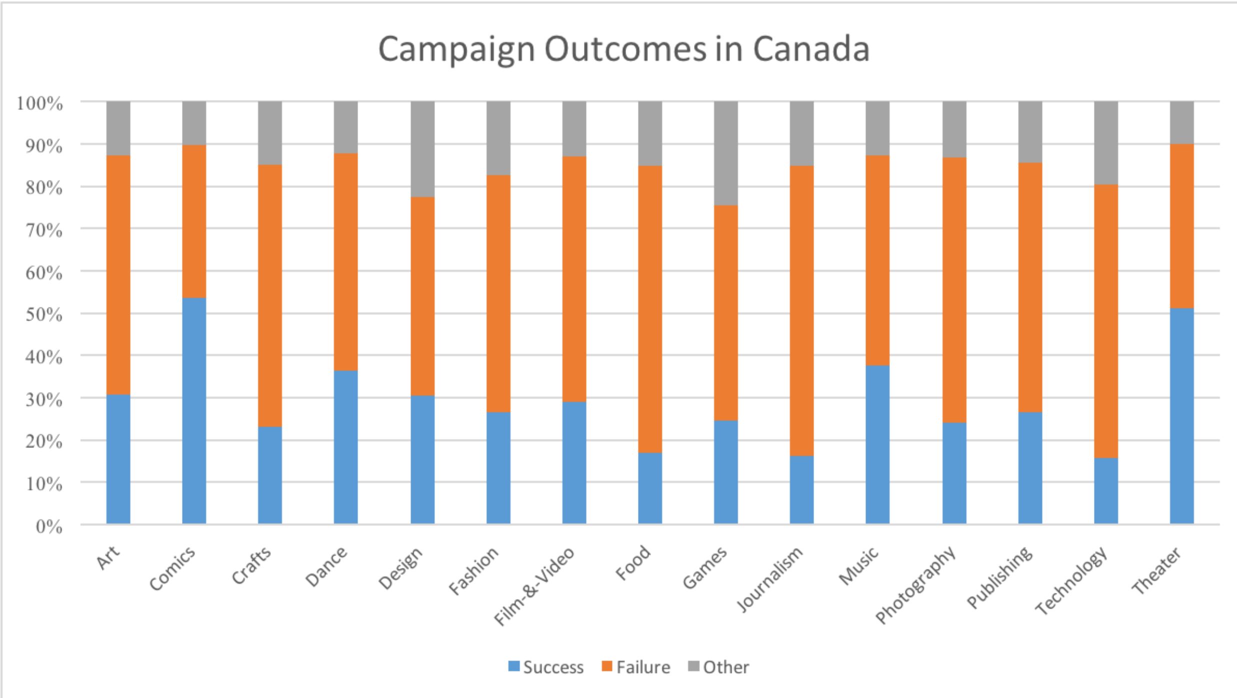
The most successful categories in Austria are Comics and Dance. The least successful categories are Journalism and Theater.



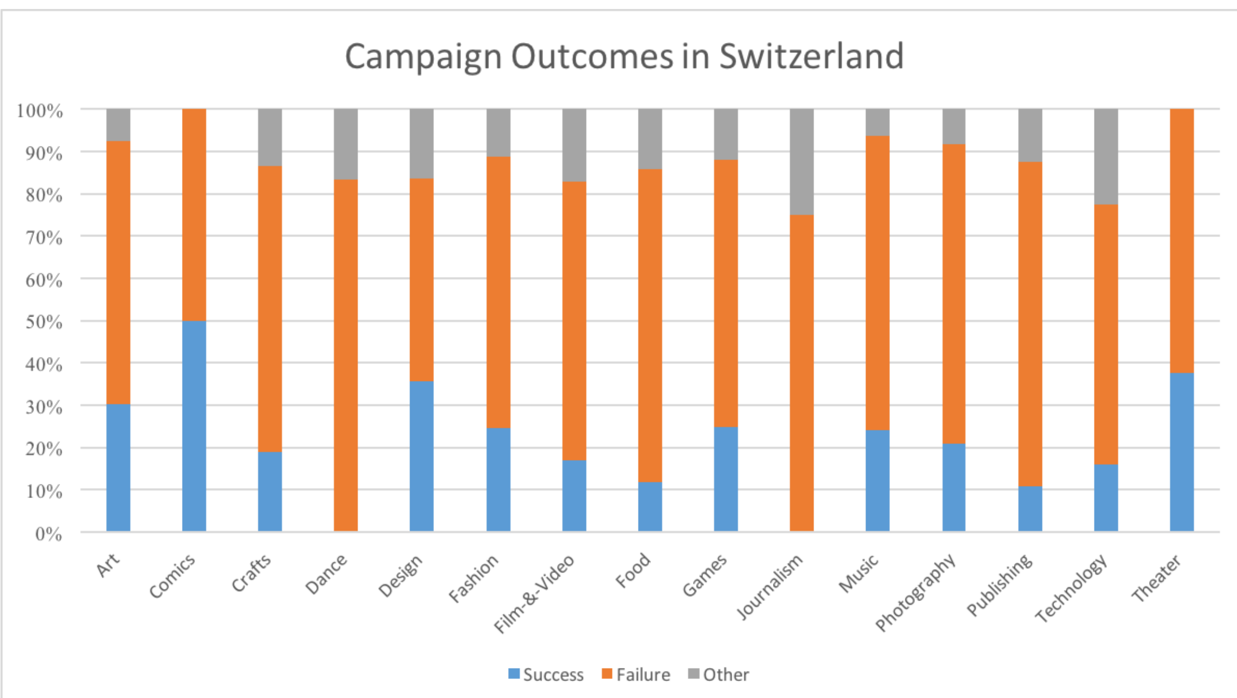
The most successful categories in Australia are Comics and Dance.
The least successful categories are Food and Technology.



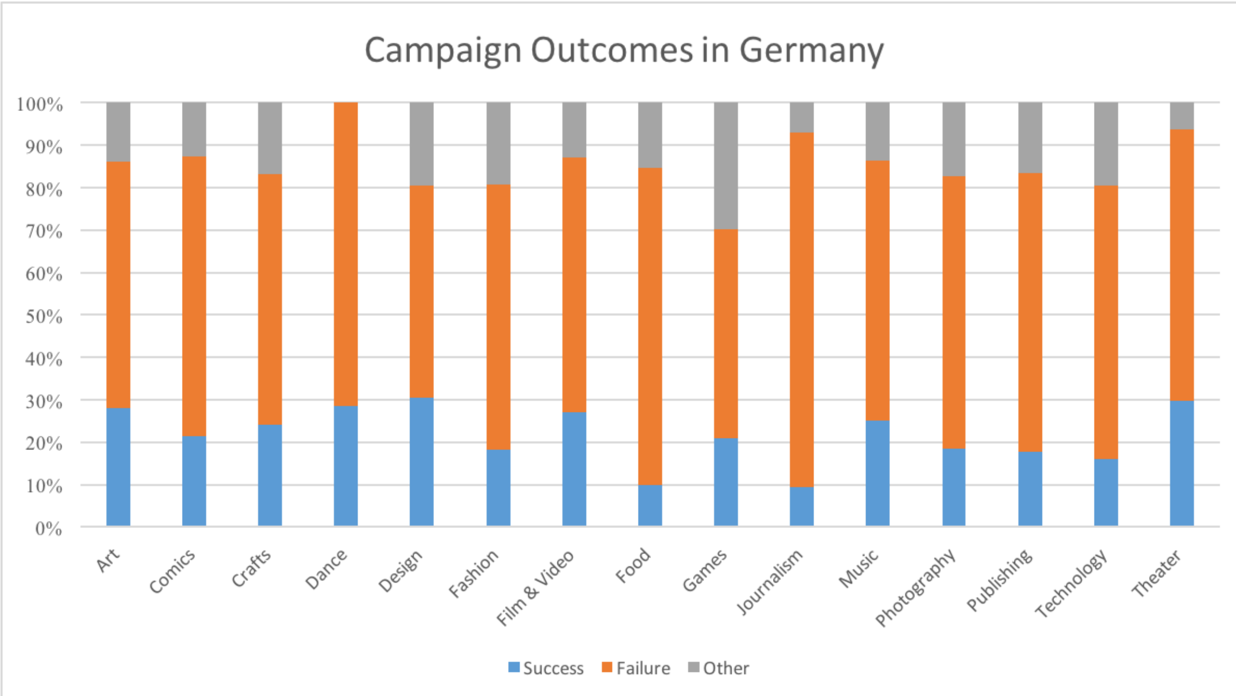
The most successful categories in Belgium are Dance and Design.
The least successful categories are Comics and Journalism.



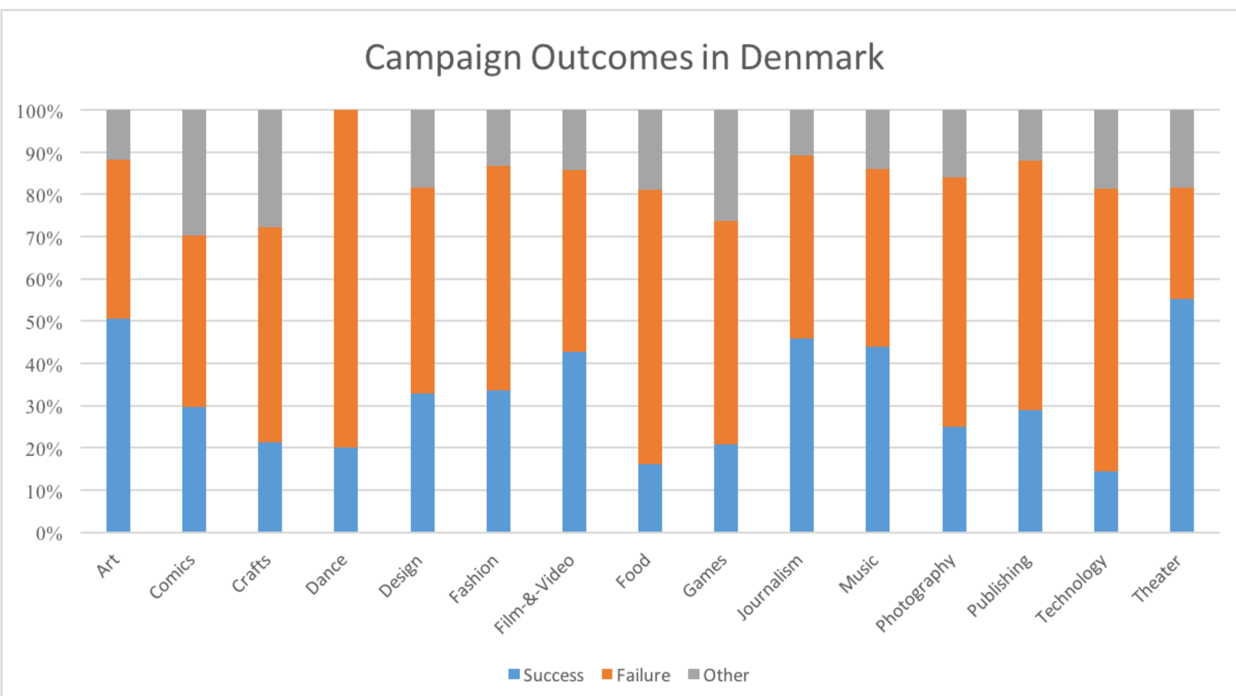
The most successful categories in Canada are Comics and Theater.
The least successful categories are Food, Journalism, and Technology.



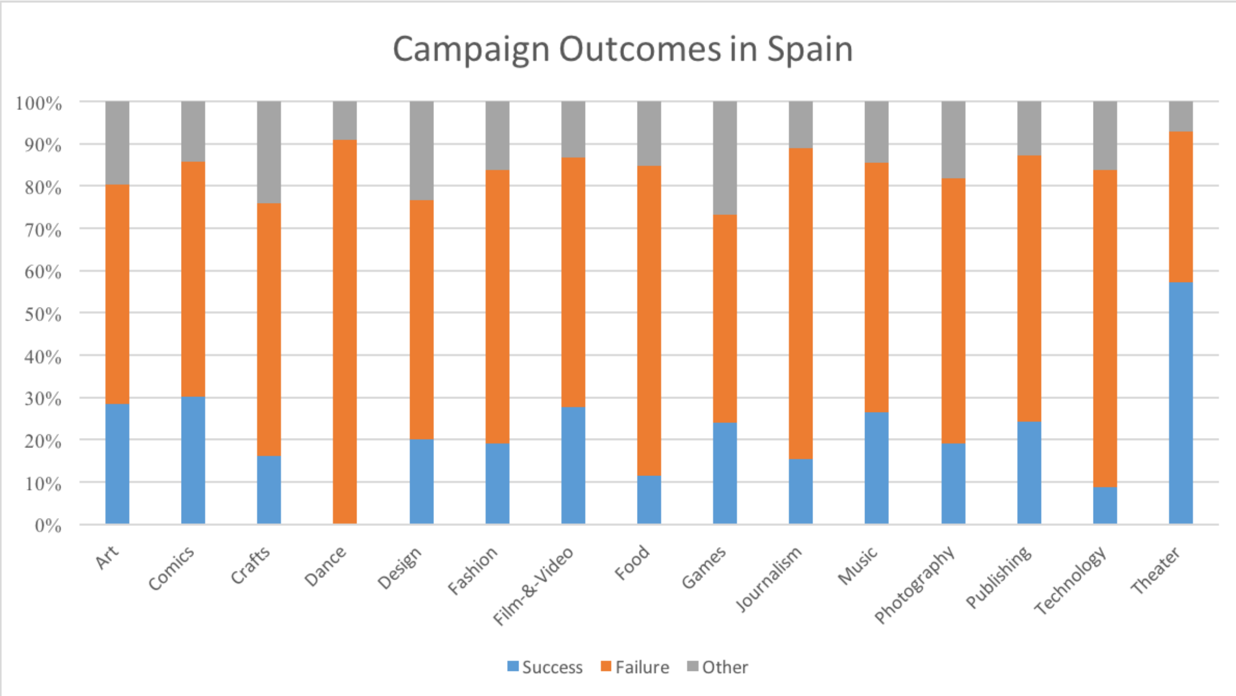
The most successful categories in Switzerland are Comics and Theater.
The least successful categories are Dance and Journalism.



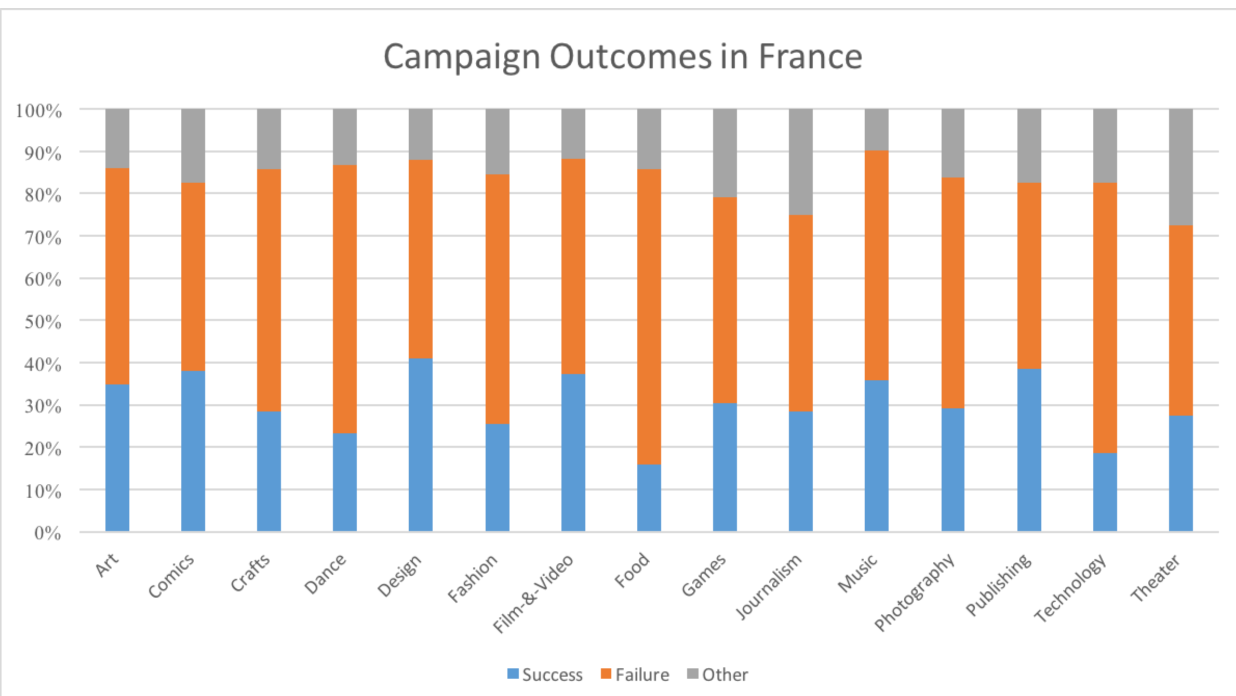
The most successful categories in Germany are Design and Theater.
The least successful categories are Food and Journalism.



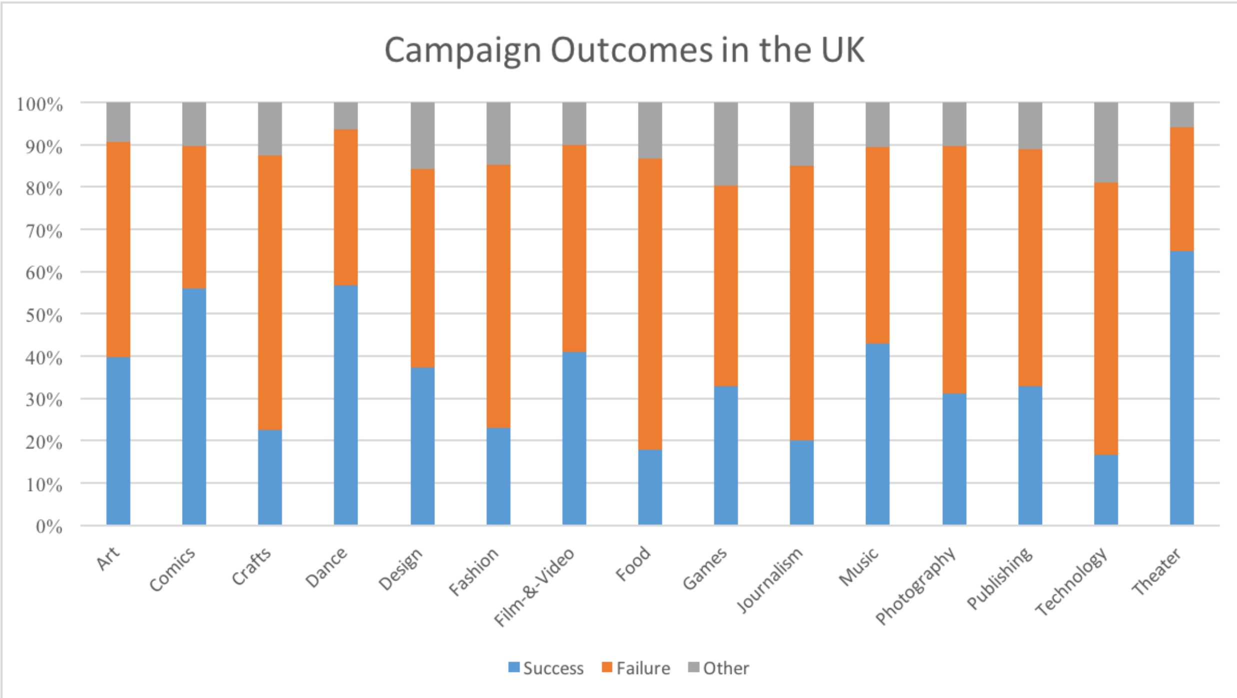
The most successful categories in Denmark are Art and Theater.
The least successful categories are Food and Technology.



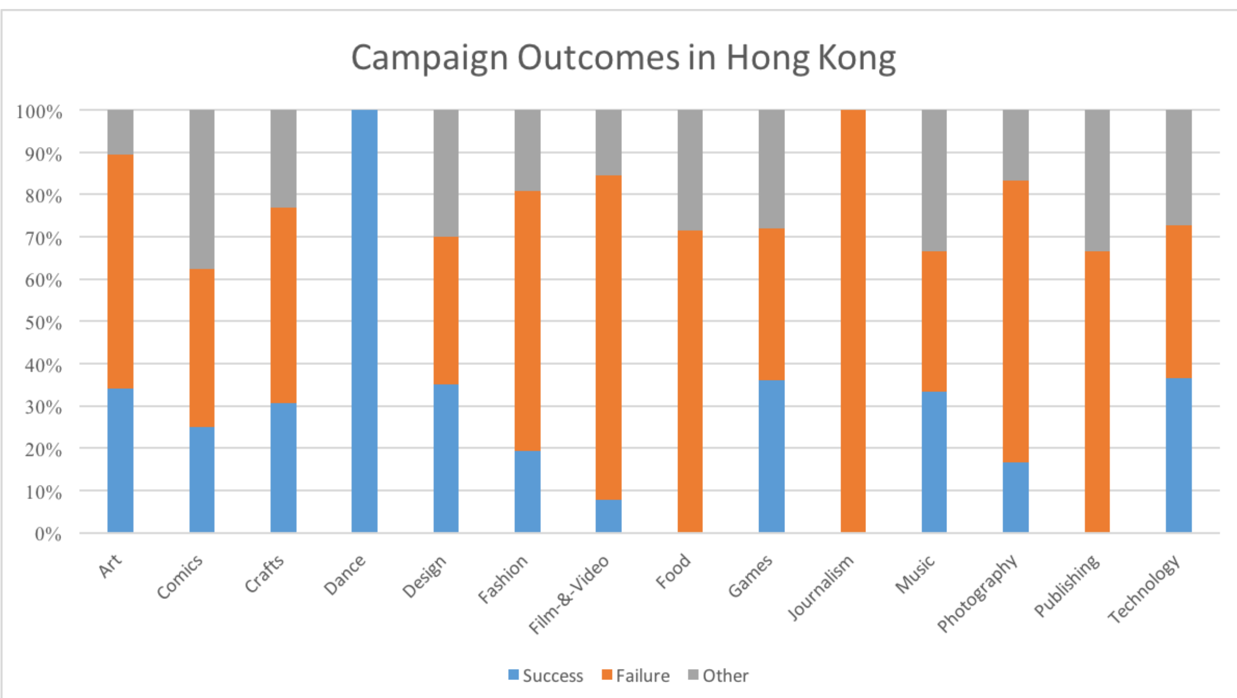
The most successful categories in Spain are Comics and Theater.
The least successful categories are Dance and Technology.



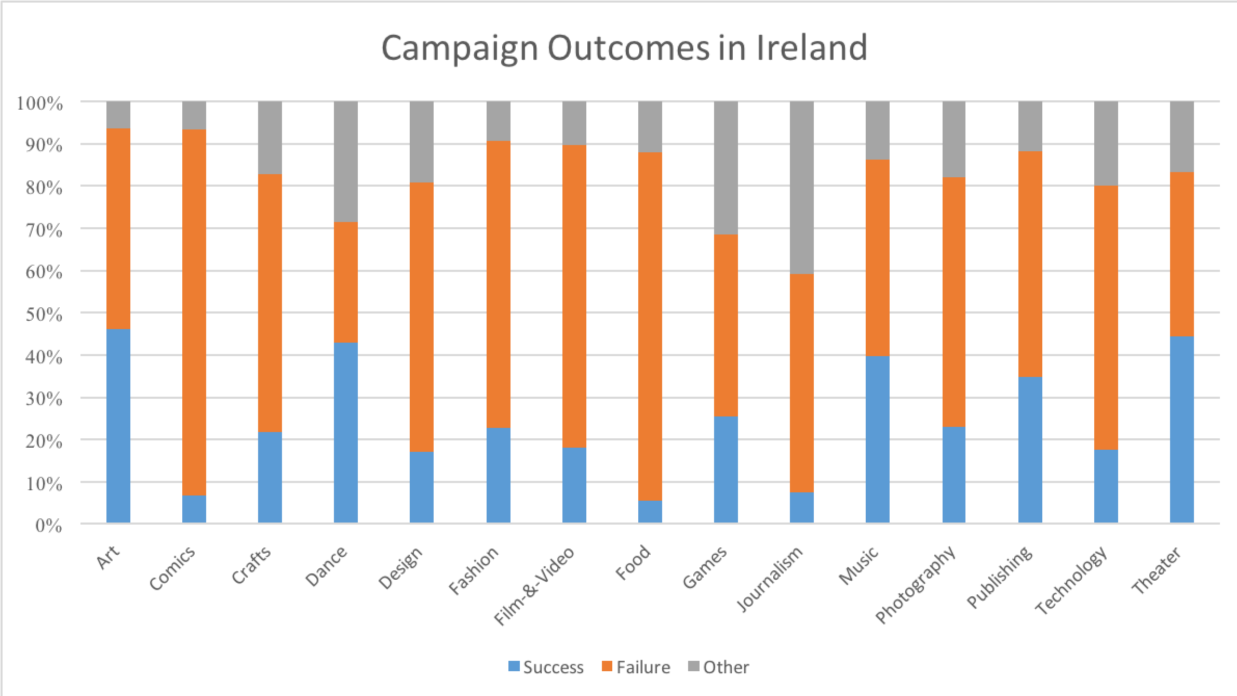
The most successful categories in France are Design and Publishing.
The least successful categories are Food and Technology.



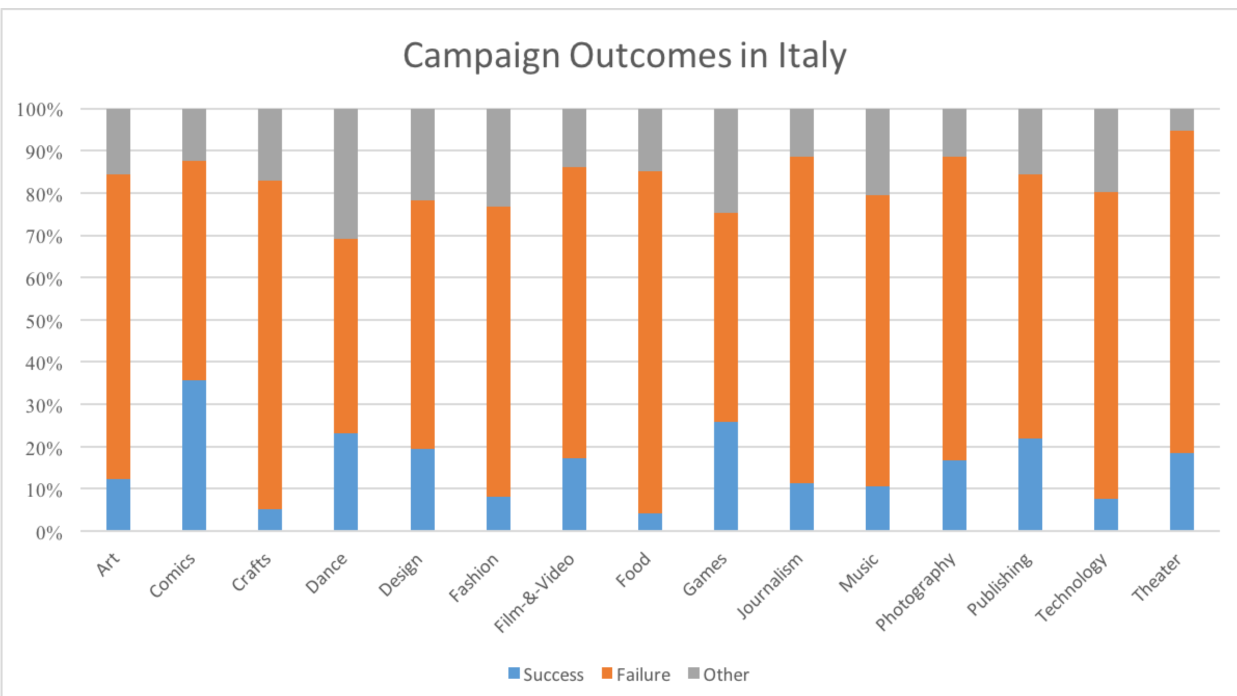
The most successful categories in the UK are Comics, Dance, and Theater.
The least successful categories are Food and Technology.



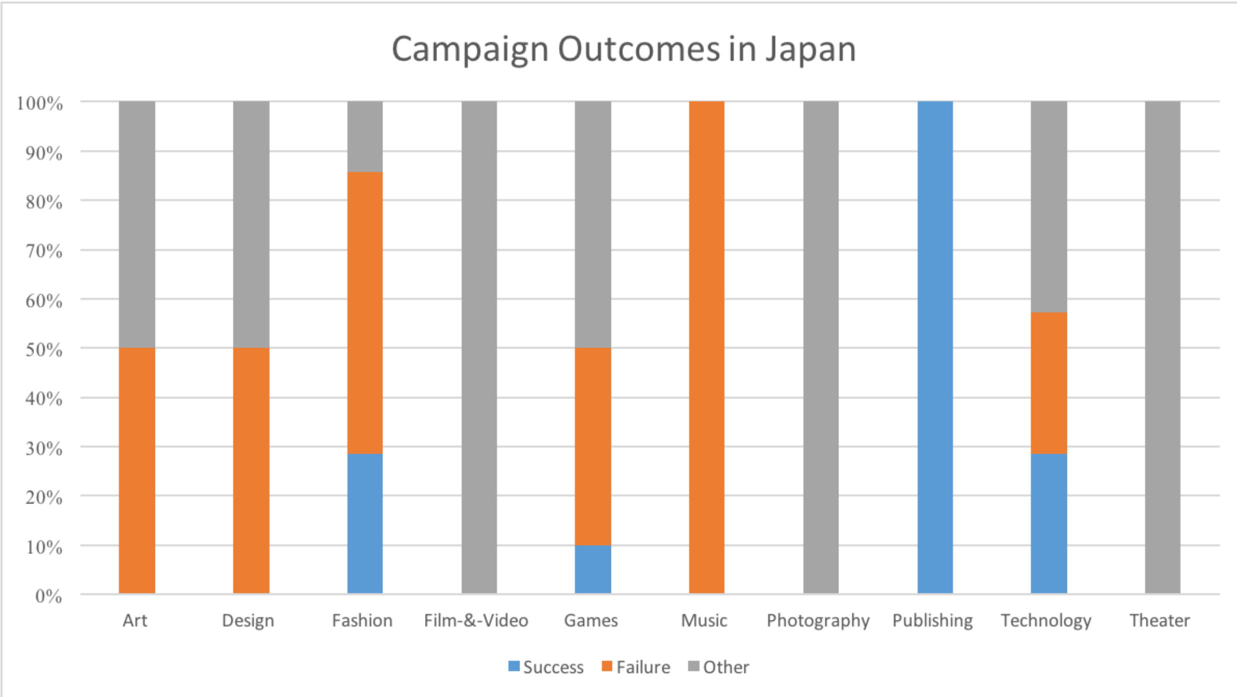
The most successful categories in Hong Kong are Technology and Dance.
The least successful categories are Journalism and Publishing.



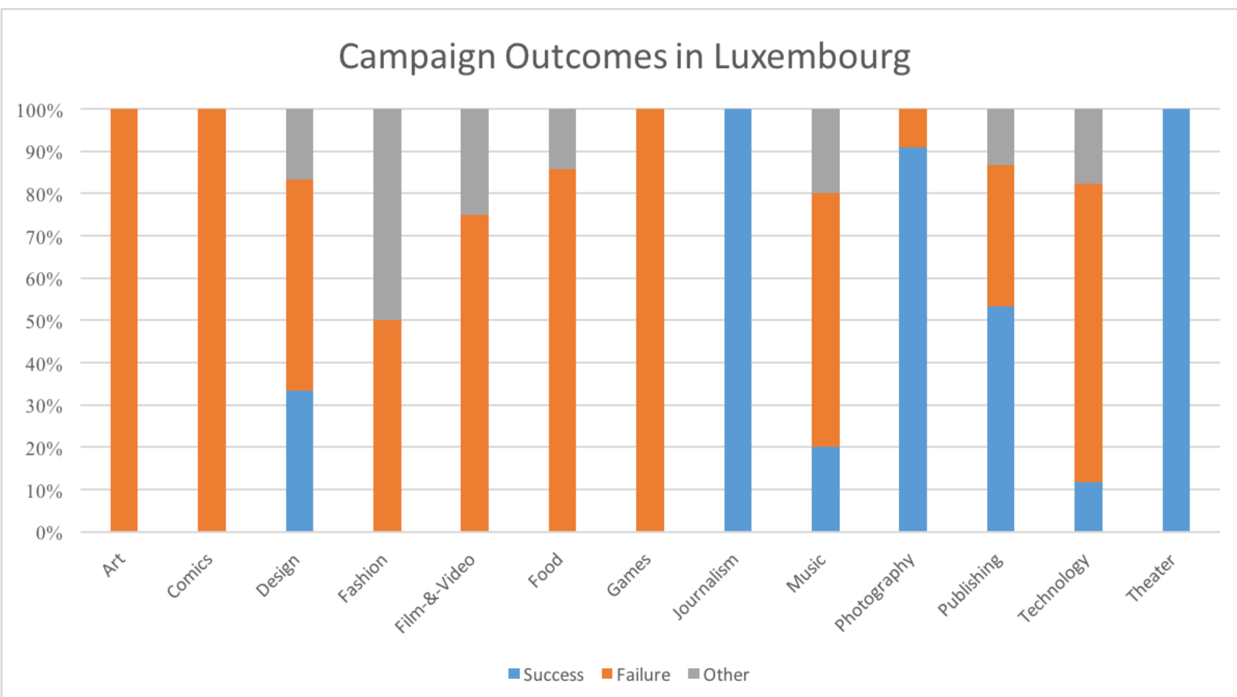
The most successful categories in Ireland are Art, Theater, and Dance.
The least successful categories are Comics, Food, and Journalism.



The most successful categories in Italy are Comics, Games, and Dance.
The least successful categories are Food and Crafts.

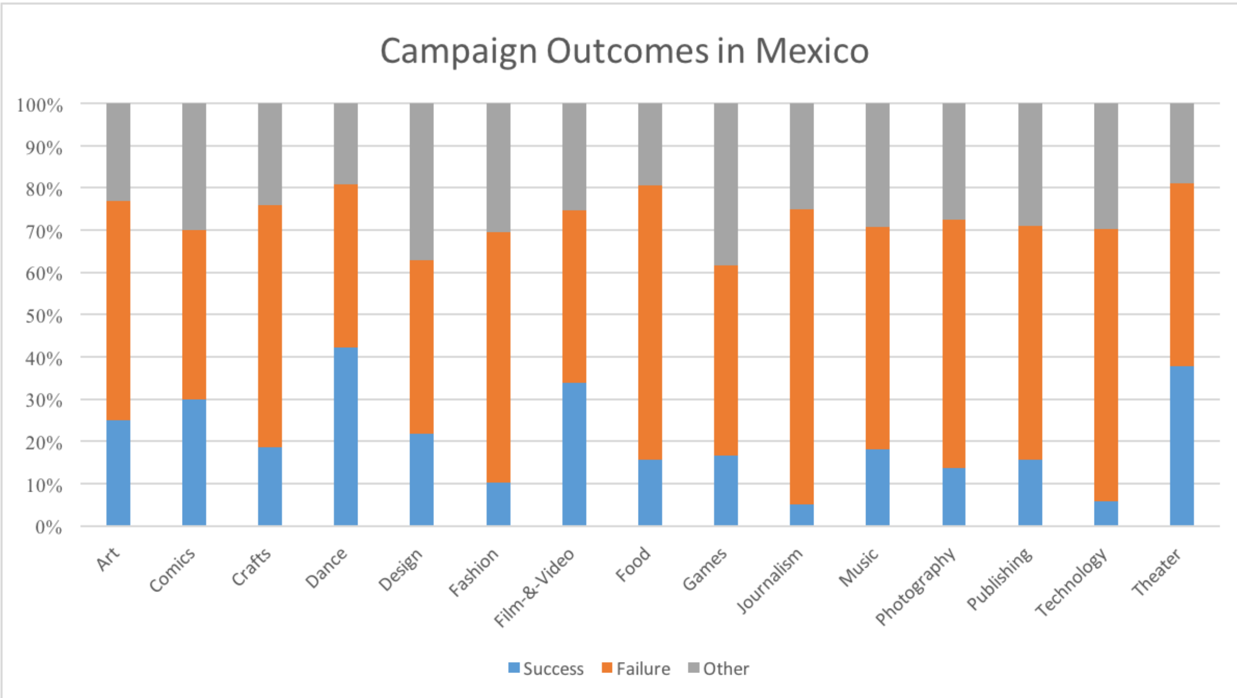


The most successful categories in Japan are Publishing, Fashion, and Technology.
The least successful categories are Music, Art, and Design.

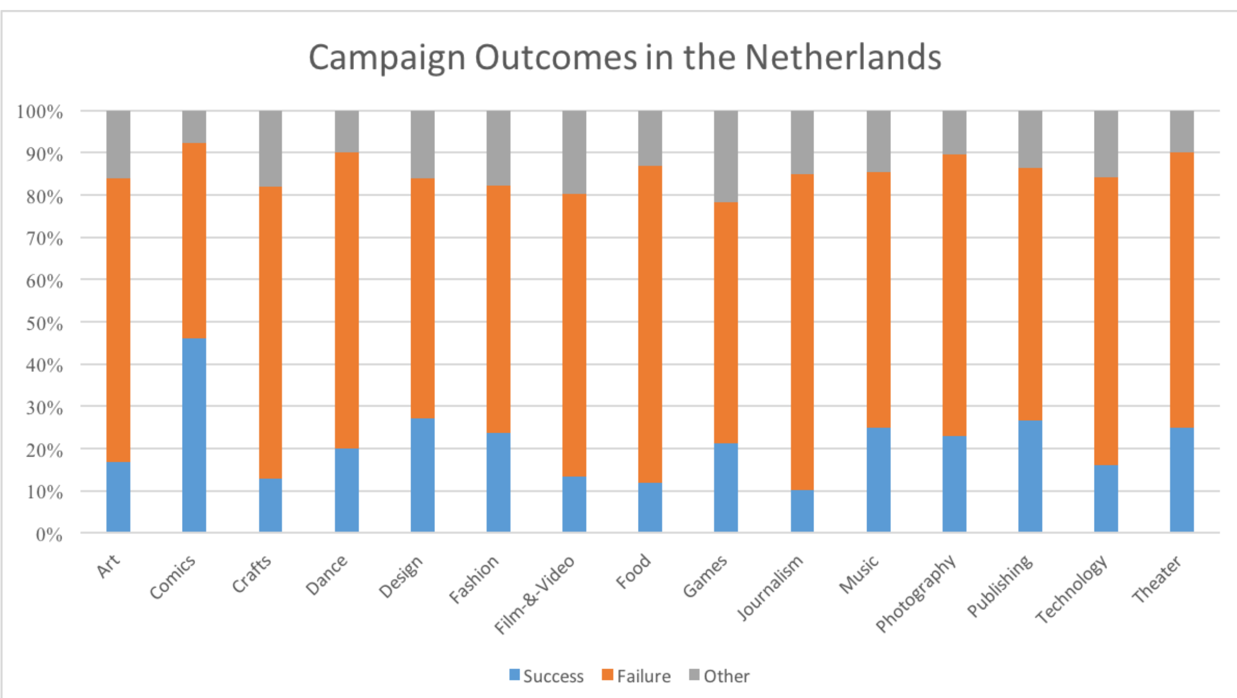


The most successful categories in Luxembourg are Journalism, Theater, and Photography.

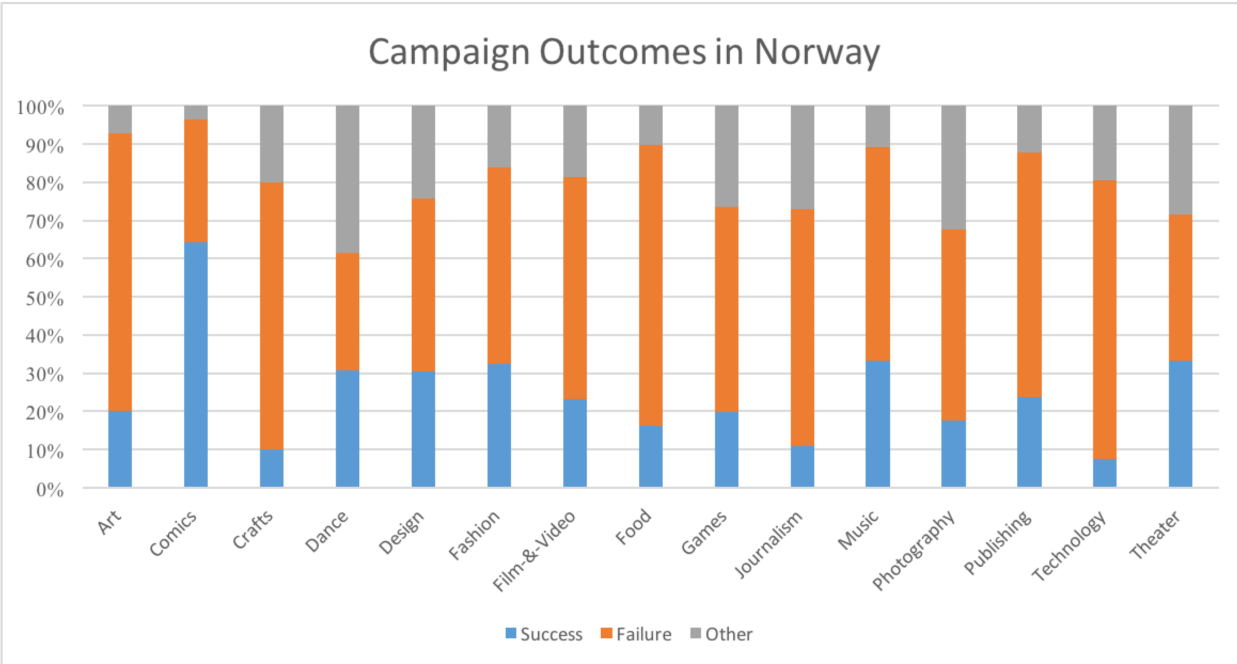
The least successful categories are Art, Comics, Fashion, Film & Video, Food, and Games, all of which had no successful campaigns.



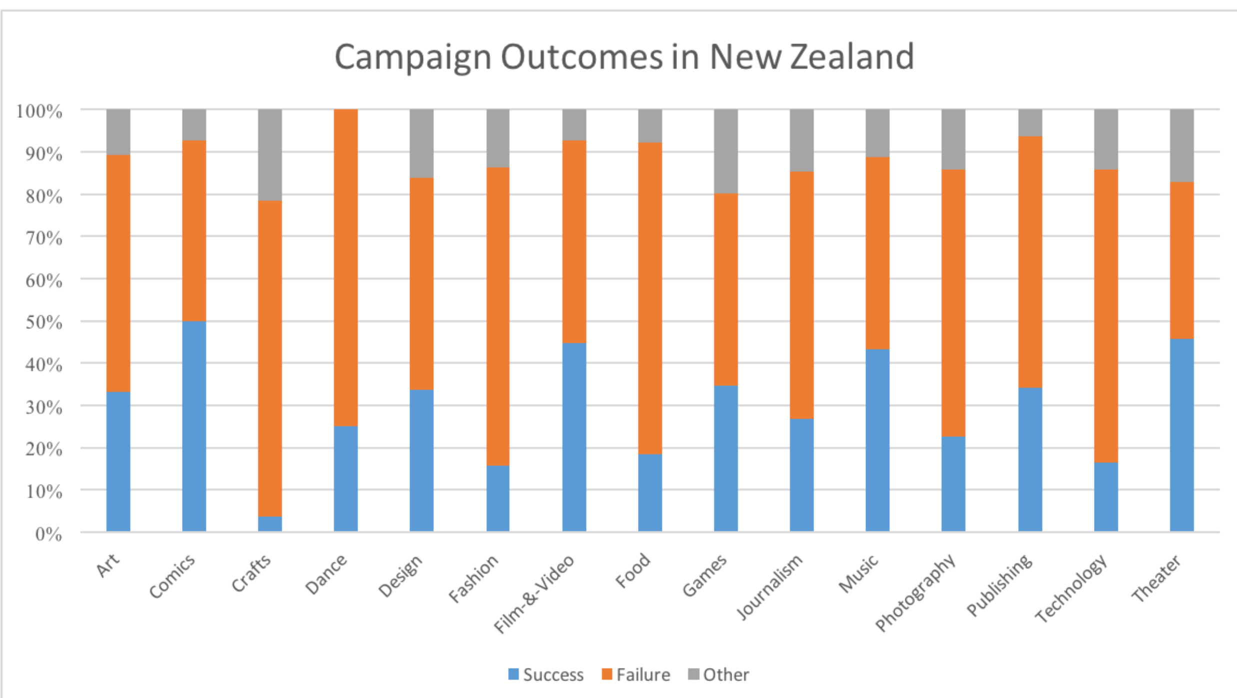
The most successful categories in Mexico are Theater and Dance.
The least successful categories are Journalism and Technology.



The most successful categories in the Netherlands are Comics, Publishing, and Design.
The least successful categories are Crafts and Journalism.

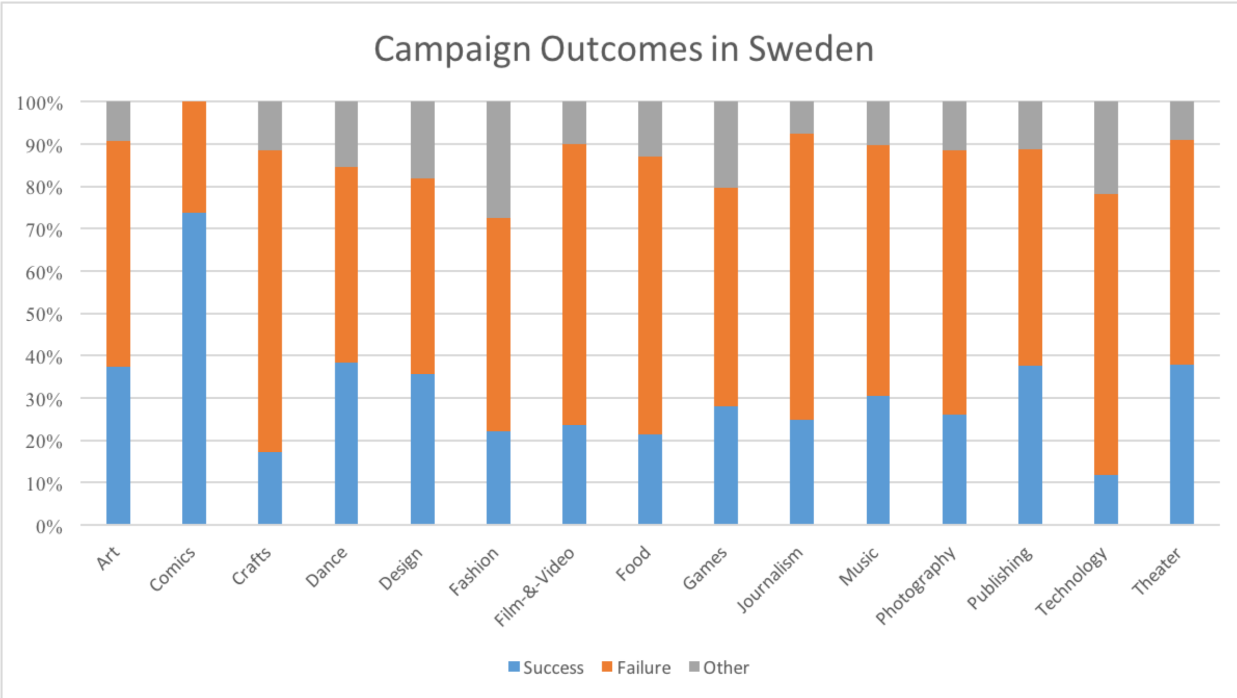


The most successful categories in Norway are Comics, Fashion, Music, and Theater.
The least successful categories are Crafts and Technology.

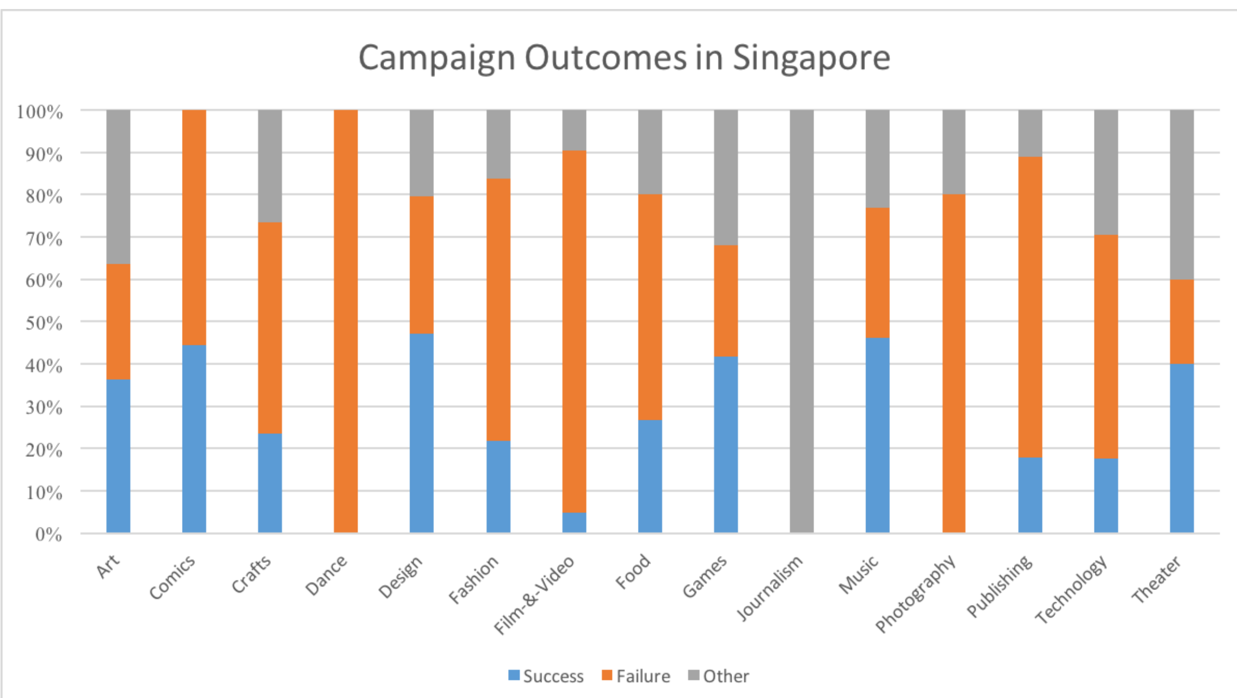


The most successful categories in New Zealand are Comics, Film & Video, and Theater.

The least successful categories are Crafts and Fashion.

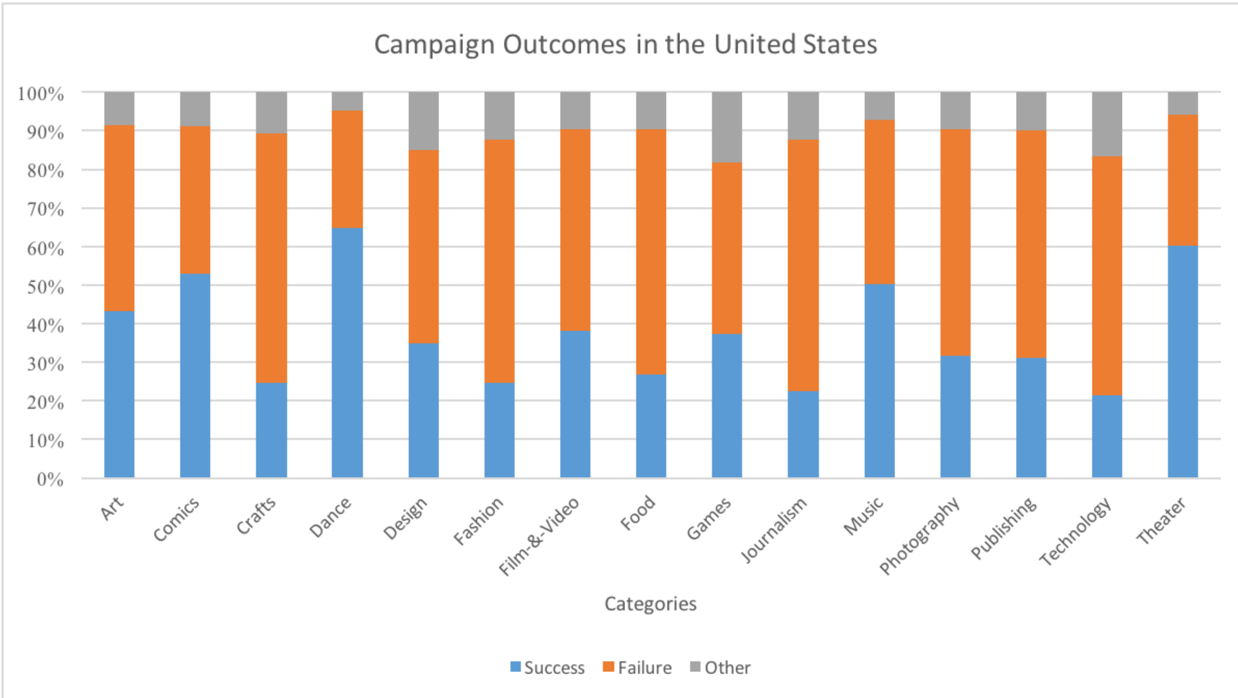


The most successful categories in Sweden are Comics, Dance, and Theater.
The least successful categories are Crafts and Technology.



The most successful categories in Singapore are Design and Music.

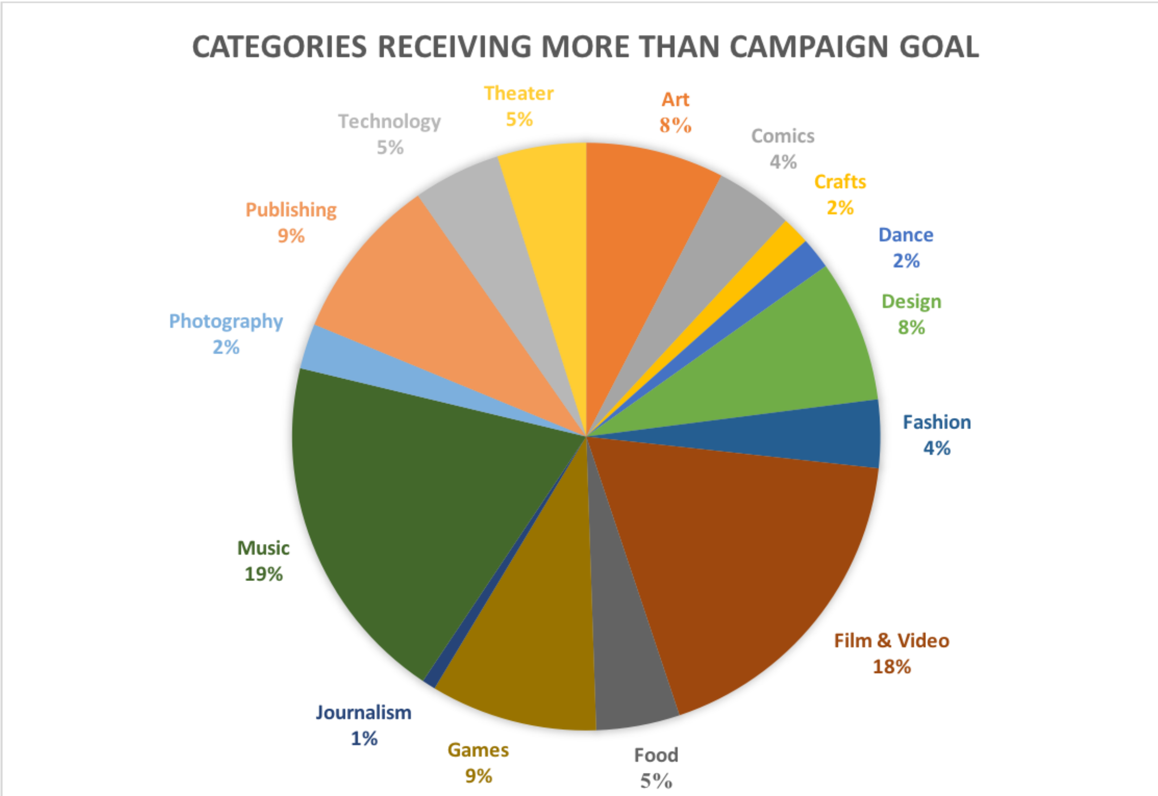
The least successful categories are Journalism and Photography.



The most successful categories in the United States are Theater and Dance.
The least successful categories are Fashion and Technology.

Some categories tend to be successful in a majority of countries. Common successful categories were Comics, Dance, and Theater. The most common categories with the least successful rates were Food, Journalism, and Technology.

Lastly, I analyzed which categories received more pledge funding from backers than the official goal stated in the campaign. The number of campaigns that surpassed their goal were counted for each category. The following graph shows the prevalence of each category in a group of campaigns receiving more than 100% funding. Music, Film & Video, Publishing, and Games were the categories with the most campaigns that exceeded their goal. This is probably because these are the four most common categories on Kickstarter.



Looking at categories, countries, and campaign success rates, the category of a campaign definitely plays a role in its success. What was most surprising was that Technology was not a very successful category, especially since it is a field that typically gets invested in. Some categories are consistently more successful than others. It is interesting how some of the most successful categories are some of the least represented in the dataset. This applies to Comics which only appears in 3% of the data, Dance which was 1%, and Theater which was 3%. It could be that campaigns in these categories fit a particular niche and were "competing" against less campaigns. However, the Games, Publishing, and Music categories were largely successful and accounted between 9-14% each in the dataset.

Breaking down the success rates of campaigns in each category for each country showed how funding trends varied. Categories successful in a country could indicate which industries are growing, valued more in society, or may not be able to gather funding from traditional sources. Many countries had the same pairing for successful categories - Comics and Dance. The least successful categories also appeared in similar pairs together. Certain categories are more conducive to success on the Kickstarter platform for funding projects.

In this report, our team analyzed various factors of the Kickstarter dataset. We used almost all of the metrics that describe Kickstarter projects, from the dates of a campaign to the amounts pledged, to the countries and categories of the campaign. All of these parts of the dataset are useful in determining what potentially contributes to the success of a campaign.