What Makes A Successful Startup Company?

Lab Group - Tessa Grabowski, Himanshu Jain, Paul Noujaim, Tri Truong

Introduction

Our general research question to find out what are the most important factors in contributing to the success of a startup project. Our dataset is comprised of the statistics from over 300,000 Kickstarter proposals, collected directly from the Kickstarter Platform (found on Kaggle). It includes variables that could be essential to determining the success of a startup such as, the amount of money pledged to a startup, the number of backers the project has, or the industry the company is in. The key variable we are looking at is the binary variable called "state", which shows which startups were successful and which were not. This will be our response variable in the analyses. We think this analysis of this data would be particularly useful if any member of our group wanted to start our own creative project by giving us an idea about which factors are key in indicating the future success of a startup.

Kickstarter is a global crowdfunding platform where different products can be listed in different categories like music, arts, technology etc. Till date, the company has received over \$4.6 billion in funding from almost 17.2 million backers. We believe that it would be interesting to analyze the data from this company to recognize the reason behind its success and how useful it might be for upcoming projects. Many new products are launched everyday so it would also provide an insight to the developer as to what products would have a higher rate of success.

The different variables in this dataset are- ID, name (name of the project), category(category of the project), main category, currency, deadline, goal (amount of money required), launched, pledged (amount of money the project got), state, backers, country, and usd pledged. We believe that some of these variables would be really important in providing us an insight about the data/company.

Data Analysis Plan

```
Observations: 323,750
Variables: 17
$ ID             <dbl> 1000002330, 1000004038, 1000007540, 1000011046, 10000...
$ name           <chr> "The Songs of Adelaide & Abullah", "Where is Hank?", ...
$ category       <chr> "Poetry", "Narrative Film", "Music", "Film & Video", ...
$ main_category  <chr> "Publishing", "Film & Video", "Music", "Film & Video"...
$ currency       <chr> "GBP", "USD", "USD", "USD", "USD", "USD", "USD", "USD...
$ deadline       <dttm> 2015-10-09 11:36:00, 2013-02-26 00:20:50, 2012-04-16...
$ goal           <dbl> 1000, 45000, 5000, 19500, 50000, 1000, 25000, 125000,...
$ launched       <dttm> 2015-08-11 12:12:28, 2013-01-12 00:20:50, 2012-03-17...
$ pledged        <dbl> 0.00, 220.00, 1.00, 1283.00, 52375.00, 1205.00, 453.0...
$ state          <chr> "failed", "failed", "failed", "canceled", "successful...
$ backers        <dbl> 0, 3, 1, 14, 224, 16, 40, 58, 43, 0, 100, 0, 0, 11, 1...
$ country        <chr> "GB", "US", "US", "US", "US", "US", "US", "US", "US",...
$ `usd pledged`  <dbl> 0.000000, 220.000000, 1.000000, 1283.000000, 52375.00...
$ X14            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ X15            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
$ X16            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ X17            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...

# A tibble: 6 x 2
# Groups:   state [6]
  state           n
  <chr>       <int>
1 canceled    32354
2 failed     168221
3 live         4428
4 successful 113081
5 suspended    1479
6 undefined    3555
```

From this table, we see that most of the companies have a state of either failed or successful. Thus when we are analyzing our data, we will filter it to only include these companies. We are interested in what makes a project successful versus unsuccessful, so any entry with a different state will be irrelevant to our analysis.

```
# A tibble: 149 x 2
# Groups:   category [149]
   category         n
   <chr>        <int>
 1 3D Printing    556
 2 Academic       706
 3 Accessories   2053
 4 Action         606
 5 Animals        218
 6 Animation     2264
 7 Anthologies    549
 8 Apparel       5478
 9 Apps          4860
10 Architecture   622
# ... with 139 more rows

# A tibble: 15 x 2
# Groups:   main_category [15]
   main_category     n
   <chr>         <int>
 1 Art           23975
 2 Comics         8753
 3 Crafts         7187
 4 Dance          3375
 5 Design        23872
 6 Fashion       18398
 7 Film & Video  57679
 8 Food          21229
 9 Games         28008
10 Journalism     4073
11 Music         46744
12 Photography    9680
13 Publishing    34233
14 Technology    26128
15 Theater        9972
```

Looking at the numbers in the categories and main categories, we can get a better idea at what kind of projects are more successful than others. Knowing this will help us narrow down what the crowd is interested

in investing in, and what fields people are more likely to succeed in if they were to create their own project.

The variable `pledged` tells the amount of money pledged by the crowd, or users of the site.

```
# A tibble: 2 x 5
  state        mean   min        max      range
  <chr>       <dbl> <dbl>      <dbl>      <dbl>
1 failed      1369.     0     768087     768087
2 successful 21490.     1 20338986. 20338985.
```

Successful Projects Generally Receive More Pledged Money
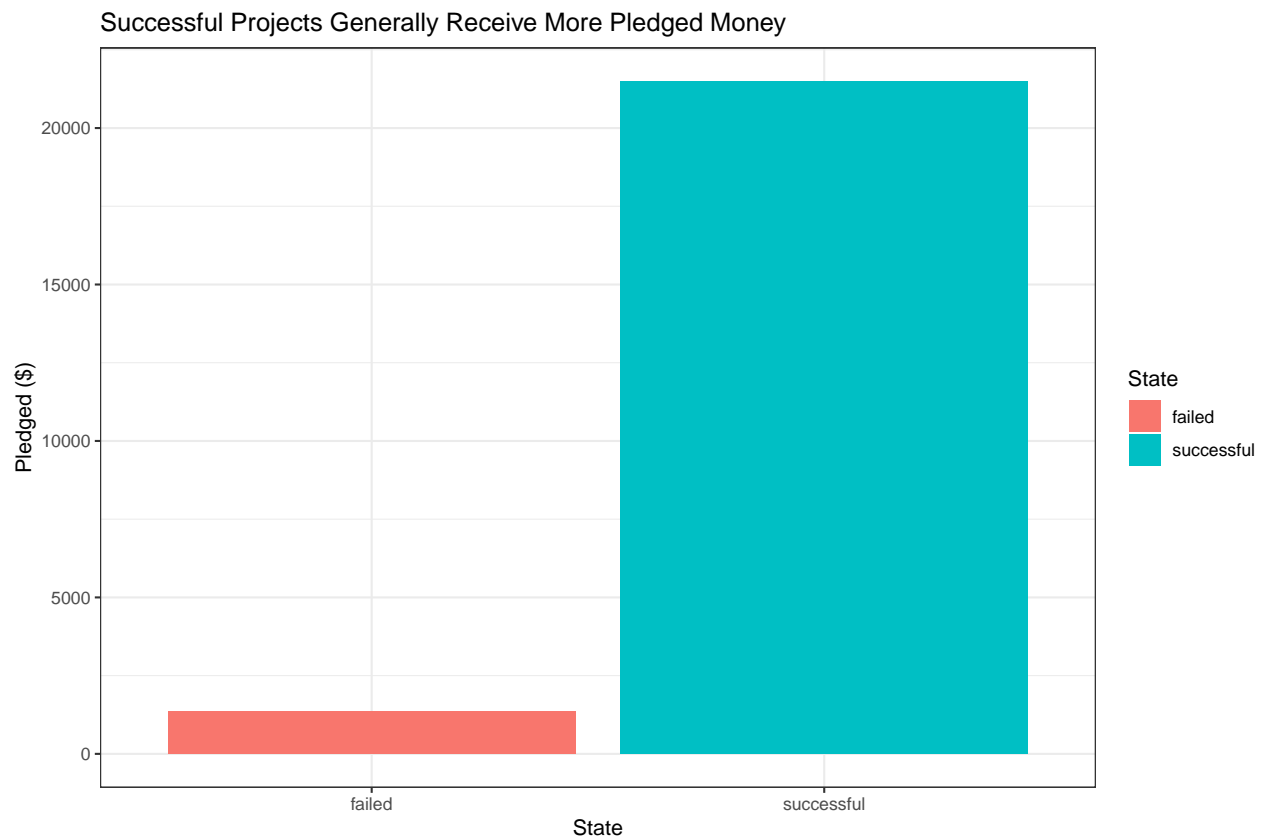
This table shows that there is a large gap between the amount of money pledged to projects that failed as opposed to those that were successful. The average amount of money pledged to successful projects is more than ten times greater than that pledged to failed projects. We can also see this in the bar plot above, as the bar for successful projects is significantly higher than the bar for failed. In addition, the range of money pledged to successful projects is significantly greater than the range of failed projects. These statistics suggest that the variable `pledged`has an effect on determining the state of a project.

Another variable we are interested in is `goal`. `Goal`gives the amount of money needed by a creator to fund and finish their project.

Failed Projects Generally Require More Money

This visualization shows that failed projects generally were much more expensive to create than successful projects. This is very useful in our analysis, since it suggests that there is possibly a threshold to how much money can be required by the creator before the project becomes unreasonable and fails.

In our analysis it could be interesting to compare the goal of a creator to how much they received from pledged money as well. Looking at these two variables and how they interact could be very telling of the outcome of the project.

Statistical Methods

Further, in our analysis, we will use other statistical methods to answer our question of what makes a project successful. We will create our own hypothetical projects and determine statistics for the needed variables. Then we will use the knn method to predict if this project will be successful or not based on the data we have.

We also will use confidence intervals for the mean amount of money required, number of backers necessary, amount pledged, etc. that we expect a successful company to have.

Finally, we will use a linear regression model to figure out which variables are the most significant in determining the eventual 'state' of the startup, whether they ended up being successful or unsuccessful.

Using these methods, we will hopefully be able to determine why certain projects are successful while others are not and use this information to help us in the future if we ever want to start our own companies.