2018

Kickstarter Project Proposal

TEAM – SKINNY SNAKES

HARSHA GUPTA , MOHAMMED ARBAZ PATEL , REEYA PATIL

04/12/2018 | Prof Yegin Genc

TABLE OF CONTENTS

1: Introduction:

Kickstarter is an American public benefit corporation [1] based in Brooklyn, New York, that maintains a global crowd-funding platform focused on creativity.[2] The company's stated mission is to "help bring creative projects to life".[3] Kickstarter has reportedly received more than $1.9 billion in pledges from 9.4 million backers to fund 257,000 creative projects, such as films, music, stage shows, comics, journalism, video games, technology and food related projects.[4]

Everything on Kickstarter must be a project with a clear goal, like making an album, a book, or a work of art. A project will eventually be completed, and something will be produced by it.

Our report analyses the dataset from kickstarter and outlines whether a project would be successful or not. As a creator starts a project, it asks for funding amount (goal) and a deadline by which the project should be launched. The crowd may give financial support to the project if they are interested. If the project gathers enough funding, the project begins development. If the project does not meet the sufficient goal no money is charged to the user.

There are a lot of budding startup projects that has been happening in today's world and the reason for us choosing a kickstarter dataset is that whenever a new budding entrepreneur wants to start a project he can analyze what project he should go for on the basis of the project that has the most success rate. Our project tries to analyze which parameters makes a project successful. Furthermore, we estimate/predict the success or failure of the project.

The report would help creators analyze parameters that effect the success rate of a project, help them estimate their backers for a particular project, the amount of funding a particular category project is likely to get, and also help them realize effects of novelty in a particular category and hence foresee their success.

2: Data:

2.1: Data Collection

For this analysis, we will be analyzing a dataset named as "Kickstarter". Kickstarter is a community that has more than 10 million people involved who are creative tech enthusiast people that try bringing creative projects to life. The projects range from publishing, movies, food, fashion, games, apps. Kickstarter is basically an all or nothing project. Either the project gets successful or not.

The dataset was obtained from the Kaggle platform. We downloaded one csv file. The datasets consists of 378661 observations (rows) and 17 variables (columns) of whom 10 are categorical and 7 are numeric. Column metadata looks as follows:

- ID – internal kickstarter ID
- Name- name of the project
- Main category- category of the campaign
- Category- sub category of the campaign
- Currency- currency used to support
- Deadline- deadline for crowdfunding
- Goal- amount required for the completion of the project
- Launched- date and time on which the project was launched
- Pledged- amount pledged by the crowd
- State – State of the project. (eg: successful/failed)
- Backers – number of people funding the project
- Country- country listed for the project
- USD Pledged- Converted amount to US Dollars
- Duration – Difference between the deadline date and the launch date
- Rank – Novelty of the project in each category

The dataset was loaded and processed using the R programming language. The data has about 3797 missing values(N.A) which we handle in data cleaning and profiling.

3: Methodology:

The project tries to infer as much as possible through exploratory analysis (covered in this report). The exploratory analysis uses statistical and mathematical techniques to plot graphs, visualize data and answer a few questions related to our research goal.

Furthermore, we delve into techniques of machine learning as we progress further in our coursework. We try to predict the success/failure of our project using exploratory analysis. Next, we move on to prepare Machine Learning models using linear regression, logistic regression and other popular machine learning models. Also we will try to analyze and compare our predictions through different loss functions and fine tune our parameters to improve accuracy.

4: Exploratory Data Analysis:

For our exploratory analysis, we have decided to focus on a certain number of question which are as follows:

1. What is the percentage state of project on Kickstarter?
2. What category of projects gets the maximum number of backers?
3. What is the number of projects in each category?
4. Does keeping the project open for a longer duration get more funding?
5. What projects are affordable by the crowd?
6. Does project novelty determine success?

4.1: What is percentage state of the projects on Kickstarter?

Output:

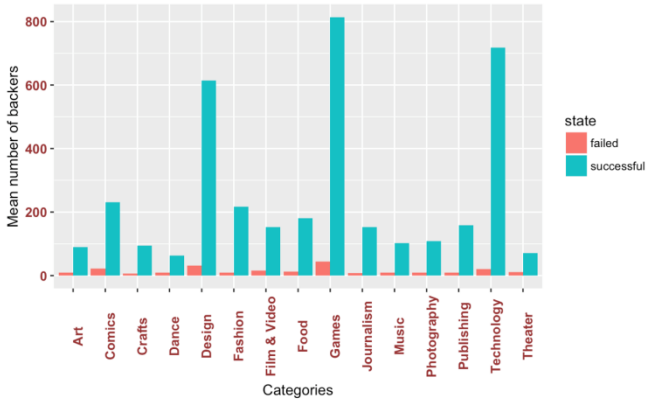
```
  canceled     failed       live successful  suspended  undefined
10.2410864 52.2153060  0.7391836 35.3762336  0.4875073  0.9406831
```

Inference:

The inference that we derive from the code above is that approximately half of the projects on kickstarter have failed. The projects that have been successful is approximately 35%. The number of projects that are live, suspended or undefined are negligible.

4.2: Which category of the projects get the maximum number of backers?

Output:

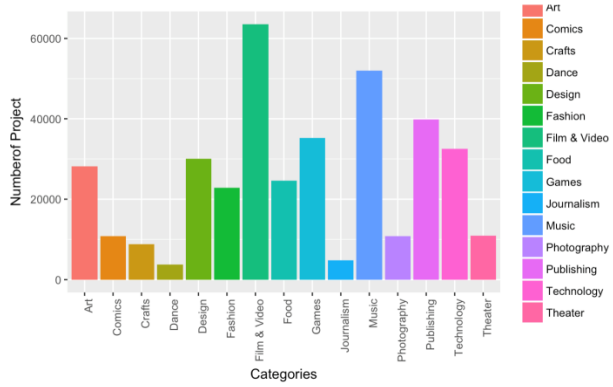


Inference:

From this output, we can conclude that the "games" category of the projects has the most number of backers. Hence, we can make an inference on the fact that the project goal amount can be set higher as compared to other projects since there is a probability of getting more number of backers.

4.3:What is the number of projects in each category?

Output:

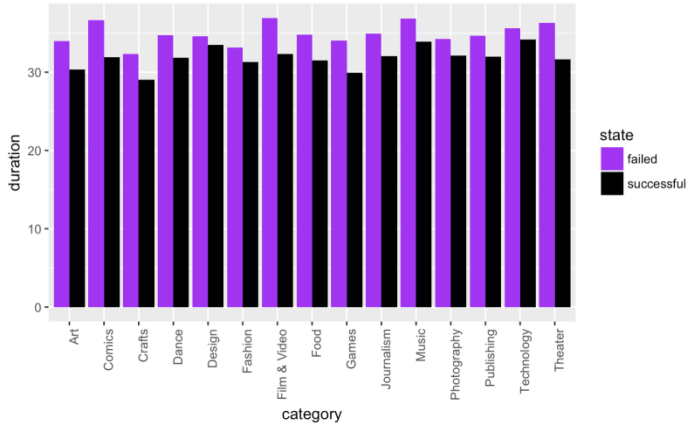


Inference:

From the output above, we can conclude that the "Film and Video" category has the maximum number of projects in Kickstarter.

4.4: Does keeping the project open for a longer duration get more funding ?

Output:

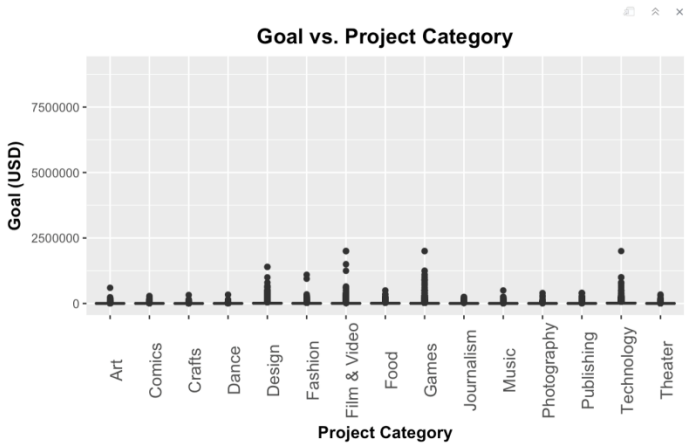


Inference:

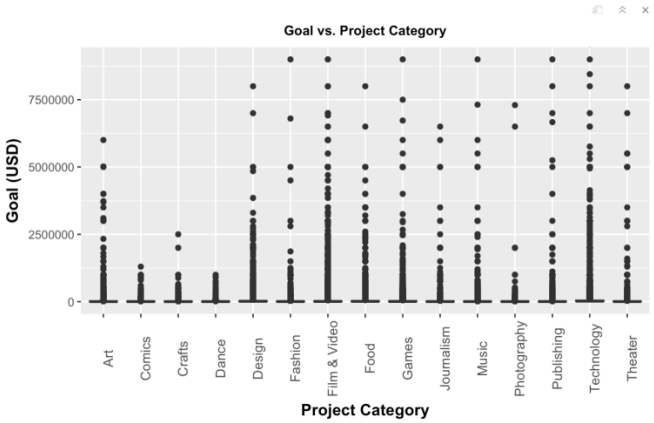
Every project has a launch date which is when the creator launches it for funding. It also has a deadline by which it is supposed to get funding by the crowd. We try to analyze whether keeping extended deadlines guarantee that the project would eventually be funded by the crowd? As we see from the graph above, the average duration/project (deadline - launch date) for failed projects is higher than that of successful projects. The data is thus counter-intuitive. We can however say that keeping the project open for a longer duration does not guarantee success of the project.

4.5:What Projects are affordable by the crowd?

4.5.1:Output for successful projects:



4.5.2:Output for unsuccessful projects:

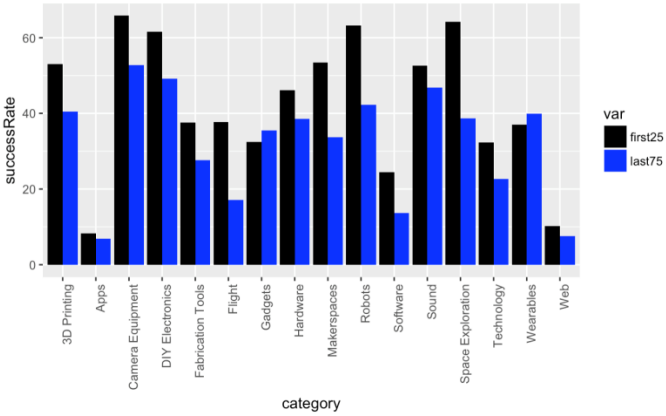


Inference:

We plotted graphs for the funding goal for every project in every category. We do this for both successful and failed projects. We observe that there are no successful projects having a funding goal above 2.5million us dollars. However, there are a lot of failed projects having such high funding goals. We can clearly say that as there are no successful projects with a required funding of 2.5 million us dollars, such high amounts may not be affordable by the crowd. The thresholds however, differ for different categories.

4.6: Does project novelty determine success

Output:



Inference:

We ranked the projects in every sub-category in Technology field in ascending order of their launch dates. We then compared the success rates for the top 25% ranked projects to those which arrived later. As we see that earlier projects included in the same sub category especially in the technology field have higher chances of success than the ones which arrive later. Thus, novelty of the project makes it more likely to be successful.

5: Conclusion:

We tried to answer some basic questions by analyzing the dataset of Kickstarter project. We analyzed what is the percentage state of projects on Kickstarter? And found out that approximately half of the projects on Kickstarter have failed. We even explored the projects that has the maximum number of backers on Kickstarter and found out that people are more interested in backing the "games" project. We also found out about what is the number of projects in each category? and came to know that "Film and Video" has the most number of ongoing projects. We determined the duration of failed and successful projects in each category as well as the goal amount set by the successful and unsuccessful projects. Later on we evaluated if the novelty of the project determines success and concluded that for each subcategory in technology the success rate does depend on novelty of the project.

References :

1.Isaac, Mike (September 21, 2015). "Kickstarter Focuses Its Mission on Altruism Over Profit". The New York Times.

2."Kickstarter crowdfunding site officially launches in Canada". The Canadian Press. 10 September 2013. Retrieved 14 October 2013.

3.Gannes, Liz (May 29, 2010). "Kickstarter: We Don't Have Anything Against Celebrity Projects". All Things D.

4."OMG". Kickstarter. Retrieved 23 September 2015.