

# *Data Analysis Report*

**By Ansh Anand, Chi-Cheong Lui, Sahand Saremi, Jason Tran, Mahumd Iqbal Halim**

School of Continuing Studies, University of Toronto

## Introduction

Kickstarter is a global crowdfunding platform based out of Brooklyn, New York that helps companies raise money for their creative ideas. Every project on Kickstarter is started with an idea which can be anything like making an album or designing a piece of technology. This idea then becomes the basis of the project. The project is then released to a pool of 15.5 million users worldwide with a clear quantitative goal known as the pledge. Pledge can be any sum of money that is contributed by the pool of users on the Kickstarter platform. Now another thing to keep in mind is that though the project is released to so many users there are only a few that back the project by pledging. These users are called backers, and it all depends upon the pledge and backers if a project is going to be successful or not. Kickstarter Projects have a lot of attributes to take into consideration like pledged amount, goal amount etc. For this report we will see how attributes can be used to predict the success rate of the project and to check whether it depends only on pledge and backers for a project to become successful or there are other attributes like category that also play a role in the success of a project.

In this project we hope to find answers to questions like which Kickstarter project category is most successful, how much money to ask for and if Kickstarter does help new companies to grow.

→ *Keywords: Kickstarter,*

# KICKSTARTER

*Kickstarter a crowdfunding company with over 15 million users.*

## Data Preparation

Data Collection is process of gathering Data and aligning information on attributes of interest. For this report Data collected was originally taken from the Kickstarter Platform by an enthusiast Mickael Mouille. In the dataset the Kickstarter projects are differentiated by various columns (*Figure 1*).

Columns	Signification
<b>Id</b>	Number given to Kickstarter project internally by the platform.
<b>Name</b>	Name of the project
<b>Category</b>	Sub-Category given to the project. (*Unique Values)
<b>Main_Category</b>	Main Category given to the project. (*Unique Values)
<b>Currency</b>	Currency used to support the project.
<b>Deadline</b>	Last date to support the project.
<b>Goal</b>	The amount of money asked by the creator of the project.
<b>Launched</b>	Date of launch of the project.
<b>Pledged</b>	The amount of money given to the project by the Kickstarter backers.
<b>State</b>	The condition of the project (*Unique Values)
<b>Backers</b>	Number of people that support the project by given it money(pledge).
<b>Country</b>	Most money pledged from which country.
<b>USD pledged</b>	Amount of money pledged in usd.

*Figure 1: Table illustrating different columns in the dataset and brief description of their significance.*

With over Thirty-Seven Thousand Kickstarter projects collected the main categories were narrowed out to be 15, some of them being Film & Video, Music, Technology etc.

Keeping this in mind some of our key points to research were:

1. Analyzing the Kickstarter Projects to find most popular categories, how much money each of the categories generates, success rates and time it usually takes to trend. This would give a pretty good estimate to predict future projects success.
2. Analyzing Kickstarter as a crowdfunding company to get a better understanding if this is the way to go for entrepreneurs.

### Analyzing Kickstarter Projects

On initial inspection of data, we find that the Film & Video as a main category have the maximum number of projects with approximately Sixty-Four Thousand Projects with Music following very closely with almost Fifty-Two Thousand Projects. To our surprise this popularity doesn't hold up when we compare the main categories to its success rate.

```
df.main_category.value_counts().sort_values(ascending=False)
```

Film & Video	63585
Music	51918
Publishing	39874
Games	35231
Technology	32569
Design	30070
Art	28153
Food	24602
Fashion	22816
Theater	10913
Comics	10819
Photography	10779
Crafts	8809
Journalism	4755
Dance	3768

Name: main\_category, dtype: int64

*Figure 2:  
Illustrating the  
number of  
projects in each  
main category.*

State of projects when the data was recorded can be divided into Successful, Failed, Canceled, Live, Suspended and Undefined. So, when we compare projects with their success rate, we find that the most successful categories are Dance with 62 % success rate and Theater with around 59 % success rate.

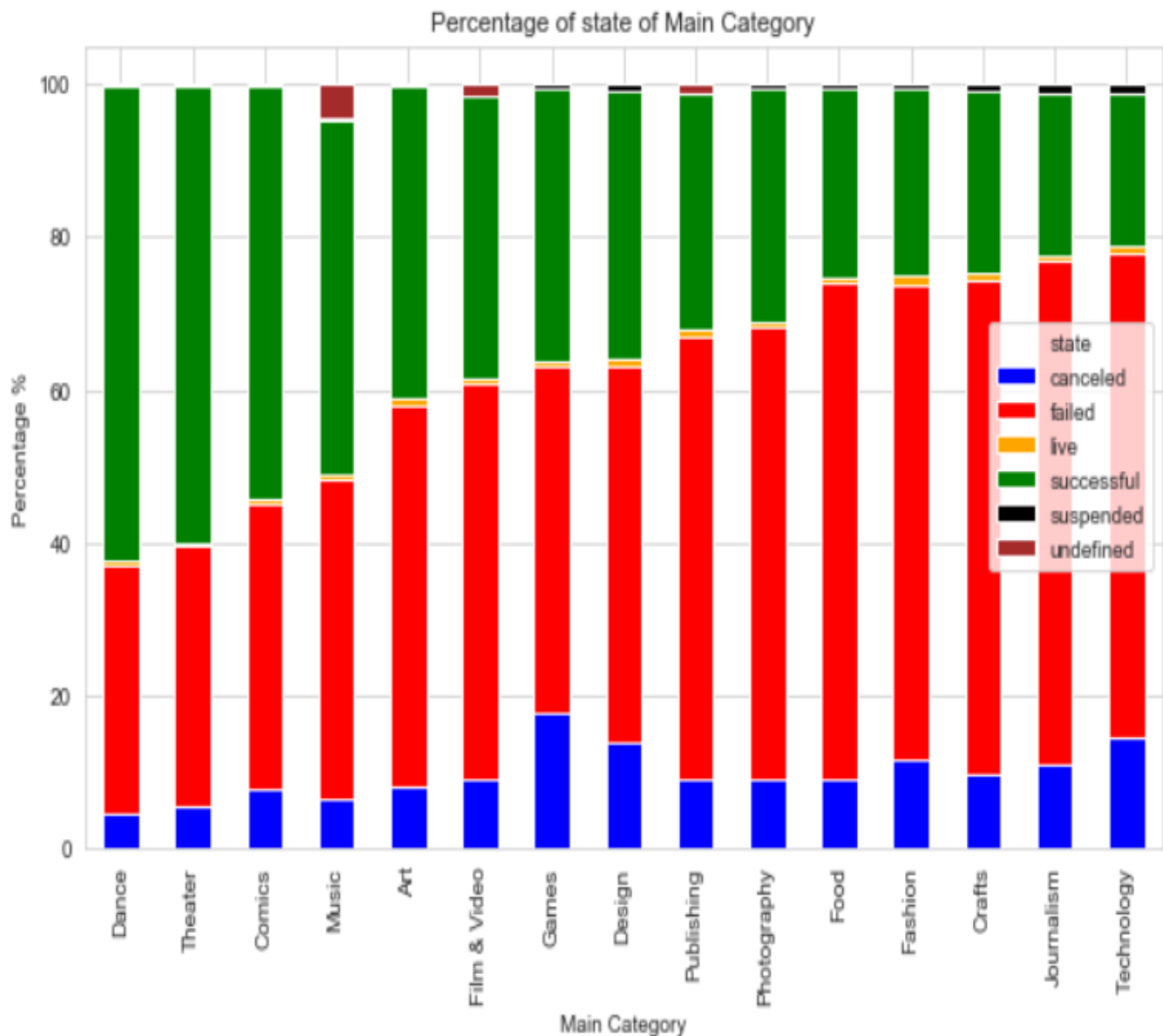
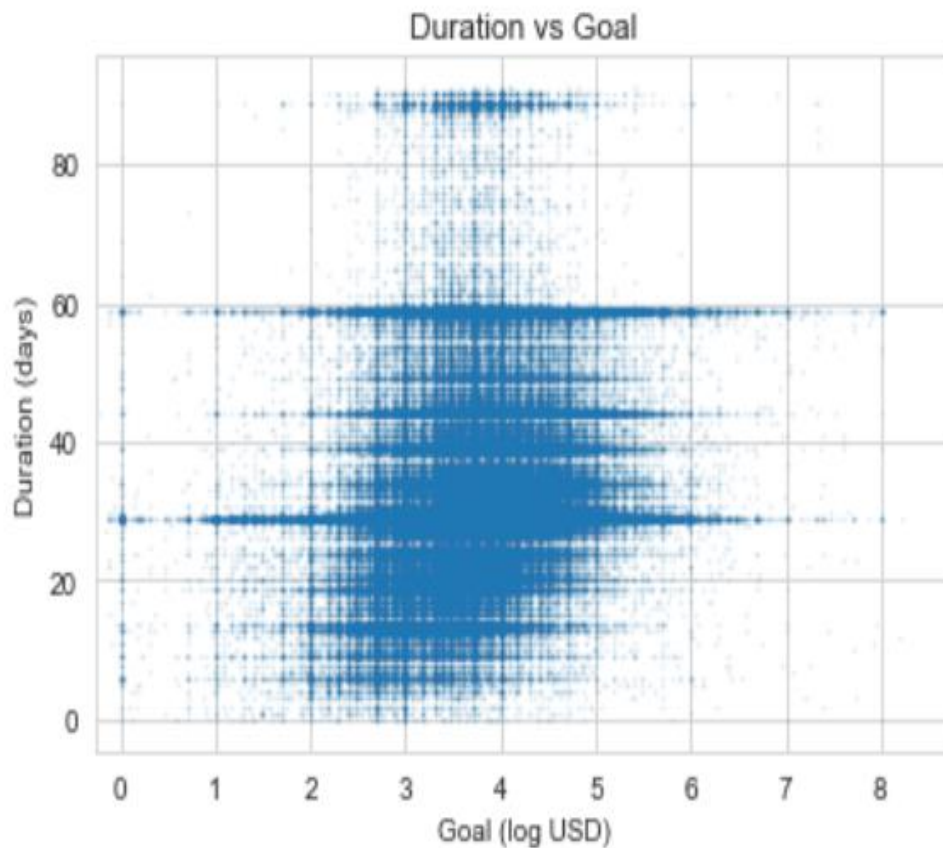


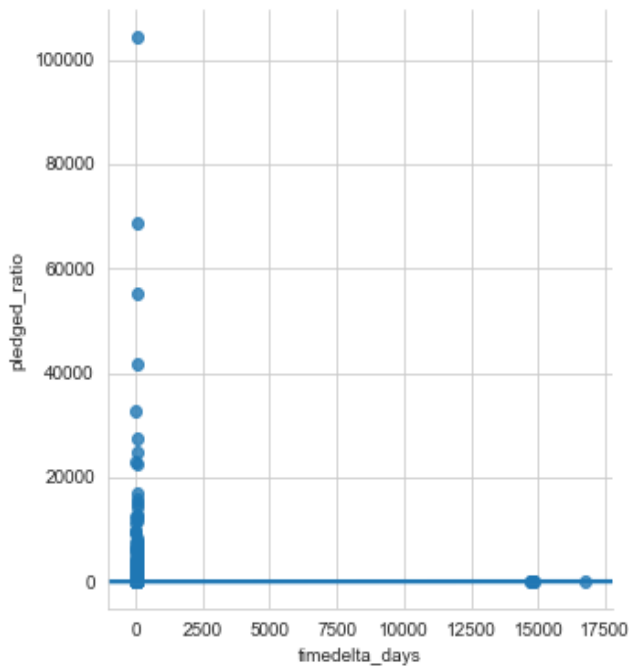
Figure 3: Graph between main categories and state of the project expressing the percentage of success.

To get an idea of how much time does it take to reach the goal amount of a project on Kickstarter a plot was made with taking into an account a new attribute created called duration. Duration is the time difference between Launch date and the deadline date. Though this doesn't give us an exact estimate in what time the goal was reach as some projects gain more popularity than other, but it gives us a pretty close approximation. Closing looking into Figure 4, we see that most projects that are successful take about 5 – 60 days to achieve their goal with high accents on 30-day and 60-day marks.



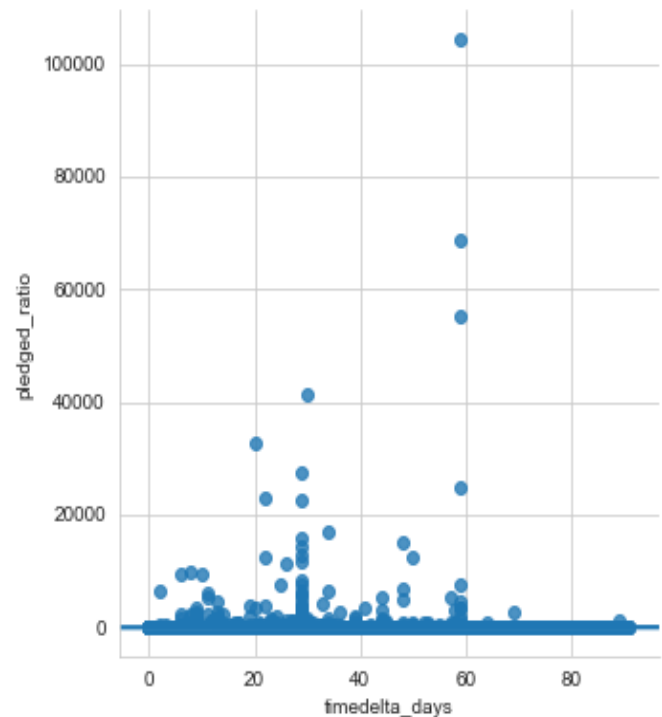
*Figure 4: Scatterplot between Duration and Log10 of Goal.*

Taking the same duration attribute, we can make a correlation plot with pledged-ratio (amount pledged divided by the goal amount). We find that the as it is the data shows an outlier where the duration is very high compared to the other values given us correlation value of 0.0004 which is quite low suggests no correlation (Figure 5). On moving out the outliers by refining the data we find that the we get a much better correlation value of 0.01 but still low to say that they are correlated in any way (Figure 6).



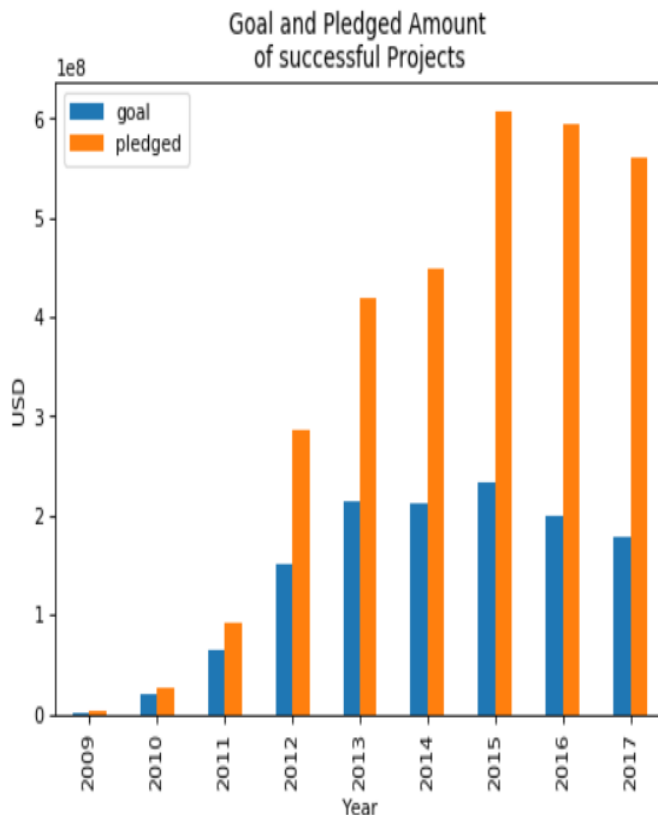
*Figure 5: Correlation Scatterplot between pledged-ratio and duration with outliers.*

*Figure 6: Correlation Plot between pledged-ratio and duration without outliers.*

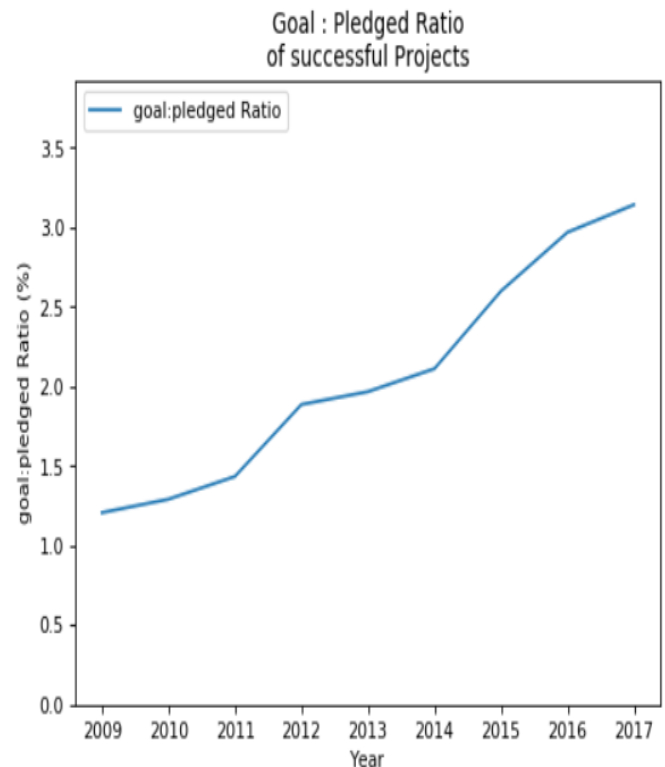


## Analyzing Kickstarter Platform

When looking into growth of the platform over the years we see that the amount of goal and amount pledged has increased drastically. This may suggest that due to increase in the user base of the platform there were more pledges which attracted more projects to the platform for the following year. *Figure 8* shows that the amount money received from the backers increased to almost 3 times what the goal was at end of 2017 from around 1.1 times in 2008. This gradual increase in the ratio is a healthy sign that the Kickstarter Platform is a viable option for entrepreneurs to get funding for their ideas. for the following year.



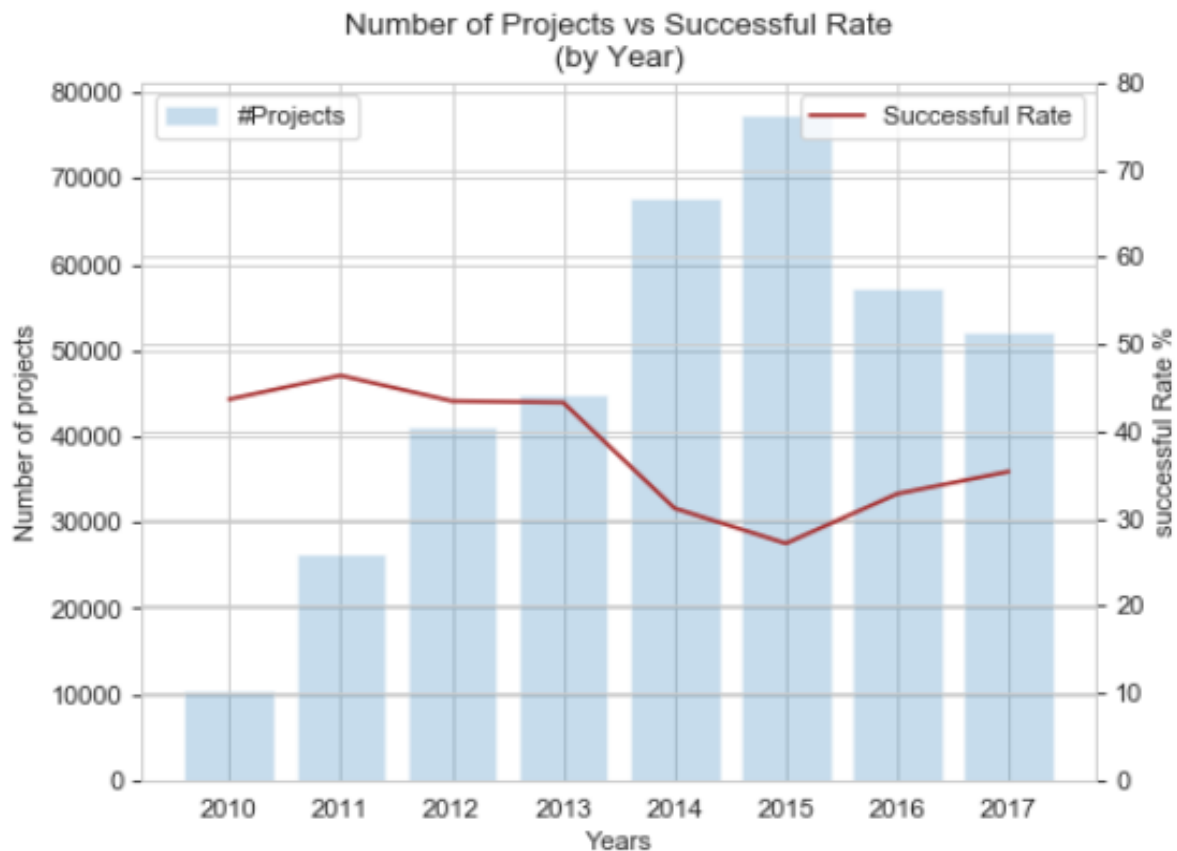
*Figure 7: Graph showing a comparison of goal vs pledged from 2008 to 2017.*



*Figure 8: Graph depicting a new variable goal-pledged ratio from 2008 to 2017.*



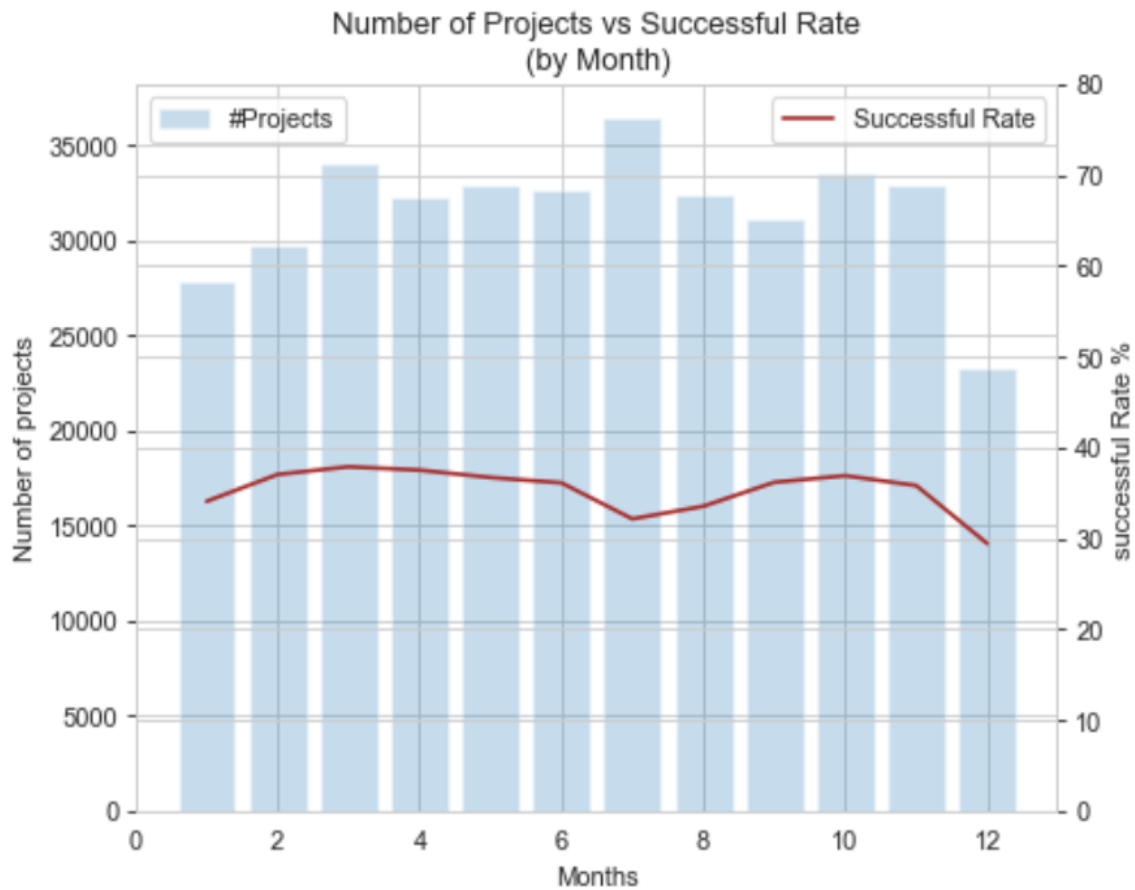
Though the goal-pledged ratio in *Figure 8* shows such a clear indication as to how the overall census is, it doesn't give an exact idea if entrepreneurs would be able to raise the goal amount if they launch soon. So, to comprehensively understand that issue we analyzed the Number of projects launched in a particular year vs the success rate of that year (*Figure 9*).



*Figure 9: Illustration denoting the number of projects in each year vs the success rate of that year*

Now we see a more realistic view as to how projects were successful in each year. The figure depicts a major dip from 2013 to 2014 and continues to dip going into 2015, after which it

relatively starts to increase again. Which suggests that success rate there is approximately a 40 percent chance the projects in 2018 would be successful. To narrow down more into a year, do a similar graph with different months as shown in *Figure 10* we see that the results somewhat remain consistent with major lows in January, July and December. We can then deduce that maybe the best time to launch projects is in months other than the ones listed above. Another distinct trait shown in the graph is that when no. of projects increases in a month the success rate relatively tends to go down.



*Figure 10: Illustration denoting the number of projects in each month over the years vs the success rate of each month.*

## **Conclusion**

Kickstarter has seen an increase in the user base since its initial launch. With more and more projects launching every year it has quickly become a renowned crowdfunding company where ideas do come to life. Before conducting the analysis, we were hoping to find more correlations between different attributes to make a predictive model but that was not the case. Though plotting and comparing attributes does tell us about some details it doesn't really help to predict if future projects can become successful or not. Another thing worth mentioning is that the success has a different meaning when compared to another attribute. We quickly found out that the no. of projects in category doesn't really correspond to having more success rate. Nor does increasing the number of days to complete the goal increase the chance of success. In conclusion to the questions asked earlier we would say the best way to achieve success in terms of reaching the goal the new entrepreneurs should maybe set a realistic goal amount keep the duration of the project to be between 30 to 60 days and to find best success rate avoid starting the projects in the months of January, July and December.