

Predictive Analytics : Project 1



UNDER THE GUIDANCE OF dr. upender Subramanian

BY: KHUSHBOO RATHORE & KAMINI RAVICHANDRAN

kxr190005 & kxr180006

**Executive Summary:**

Kickstarter is an American public-benefit corporation, that maintains a global crowdfunding platform focused on creativity and merchandising. Our objective is to explore data and present the most interesting findings. To understand this data we conducted descriptive analysis using Python and SAS along with visualizations using Tableau Dashboards. While performing descriptive analysis, we checked the independence of the variables using statistical tests and noted the correlation present among variables. We have also tried scaling and normalizing few data columns in order to gain better understanding of the data.

We started our analysis by observing that major contribution in project proposals were coming predominantly from the United states. This is understandable since it is a US based platform and thus saw the highest level of participation over the years. We can also observe a time trend in number of project proposals submitted which are gradually increasing from 2009 till 2015 after which there is a decline in the number of projects received. There is also decrease in success rate during these years. This might be due to an exponential increase in the number of projects submitted which might have caused increase in scrutiny and expectations from proposal by the backers. Furthermore, we saw an increase in the number of projects submitted and backer funds for games, design and technology over the years due to high return of investment and demand for these categories. The optimum duration for each project completion was found to be 32.16 days. This might be due to shorter durations creating a sense of urgency for the backers who can look into the project proposal and will make the decision without delay. We can also see that the categories having the most backers might still not have the highest success rate. This can be due to higher competition in that category, thus increasing the difficulty level for success.

Since Kickstarter is an innovative platform, creativity and applicability of the idea of the project is the most important factors which can drive success rate in different categories. We can see that projects from Games category has the highest number of submissions and maximum backers, still the success rate is low while categories such as Dance which has lowest number of submissions and minimum backers, still the success rate is high. We need to gather more data to know the reason for sure such as proposal’s creativity content, innovative ideas and what’s new, and whether it has practical application or not, to be able to cater wider audience. Also how does the investor test genuinity – i.e. if the proposal is for a new product or a modified version of a pre-existing product. Hence, data is insufficient to perform predictive analytics.

**Introduction:**

**What is Kickstarter?** It is an online crowdfunding platform for creative projects. The company helps bring creative projects to life.

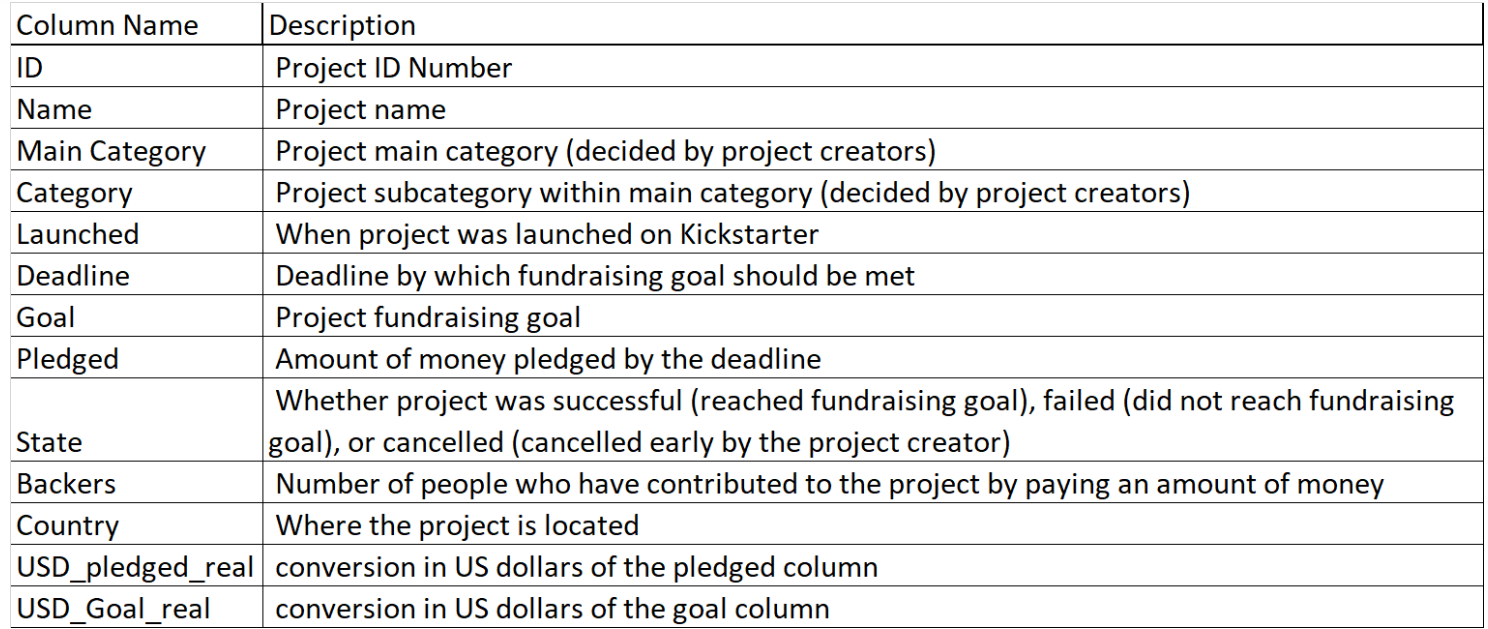
**Steps to follow**: First step to raise funds, is to describe the project, set a goal for the amount you will raise, and ask “backers” to invest in the project. If the goal is reached, then all the funds can be received. Otherwise, the project will not be funded, and all the money is returned back to the backers.

**Interesting Facts About Kickstarter** : It has reportedly received a commendable $1.9 billion in pledges from 9.4 million backers to fund 257,000 creative projects, ranging from films, technology and food-related projects. It is a great initiative to enable innovative ideas to surface to life, for instance, a game called exploding kittens raised $8.7 million and another popular project raised $9,192,055 for an efficient Travel Jacket. Kickstarter makes money by taking 5% of the total amount of money that is funded on the site. Those who want to be funded get a group of backers to donate to their project and then promise the backers with an incentive or reward.

**What we want to find** : What is the optimum duration set for each project deadline? Which locations around the world are most of the projects successful? What are the project categories that are receiving the most funds and succeeding? Is there any inter-dependency between categories?

**What we are going to do**: In this project, we conduct exploratory data analysis with 375,093 past Kickstarter projects across 14 variables and visualize them to obtain profitable insights and patterns between the variables that could be managerially relevant.

**Variables we are working with**:



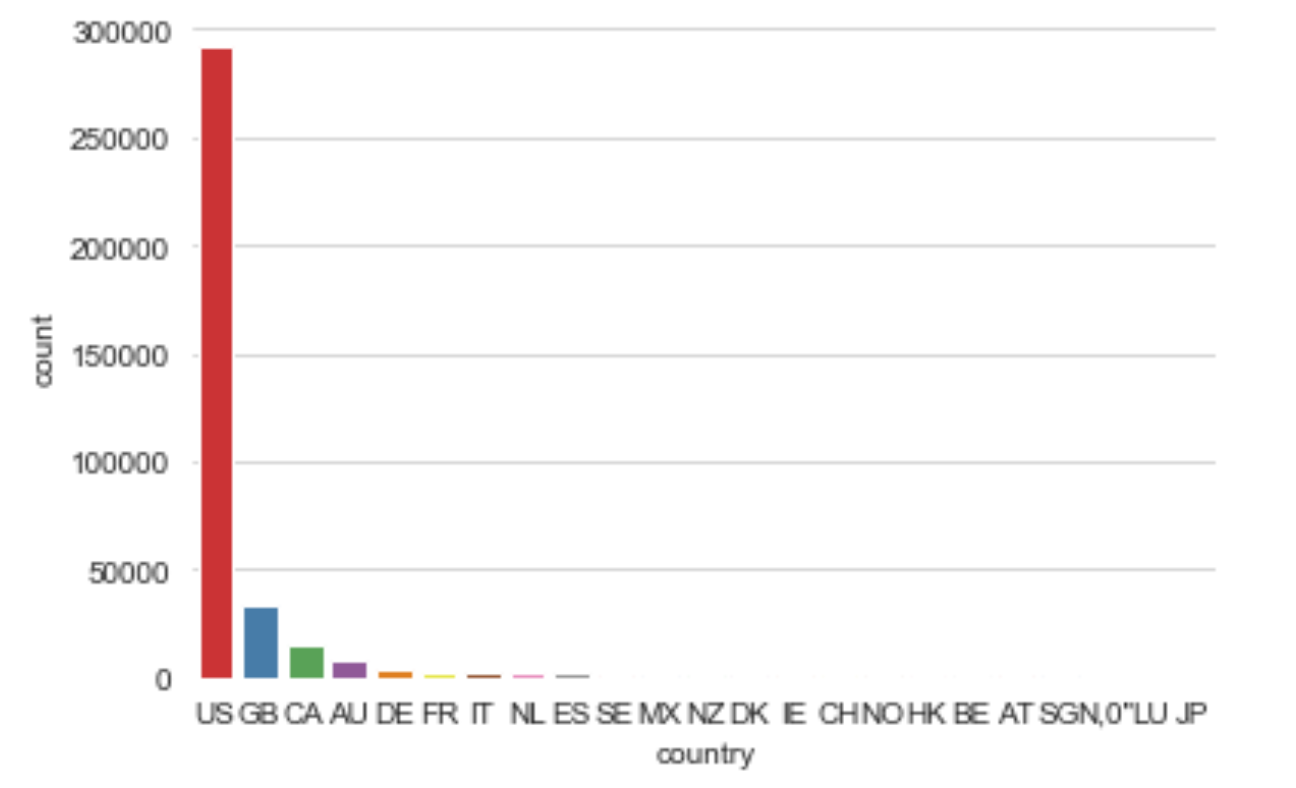
**Insights:**

1. Analysis across Countries
2. Analysis across the years
3. Analysis across average length of campaigns
4. Analysis across Categories

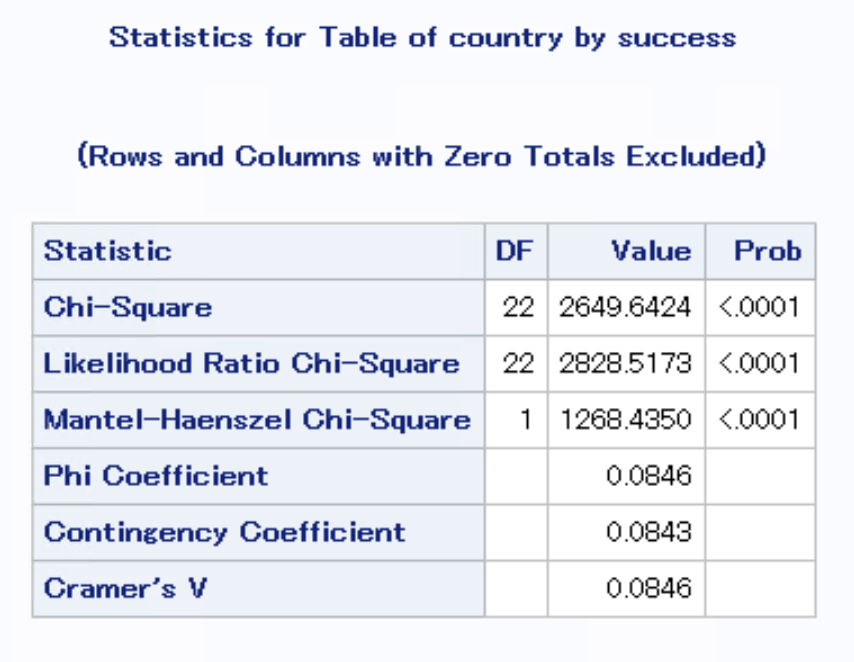
**Insight 1: Analysis across Countries**

Here, we can see that majority of the projects seeking funds in Kickstarter are from the United states with highest success rate in raising funds.

Over 78 percent of the overall projects proposed are from United States. It is obvious from the distribution and count plot of projects across countries that the distribution of the data is not normal. It seems that the data is skewed.

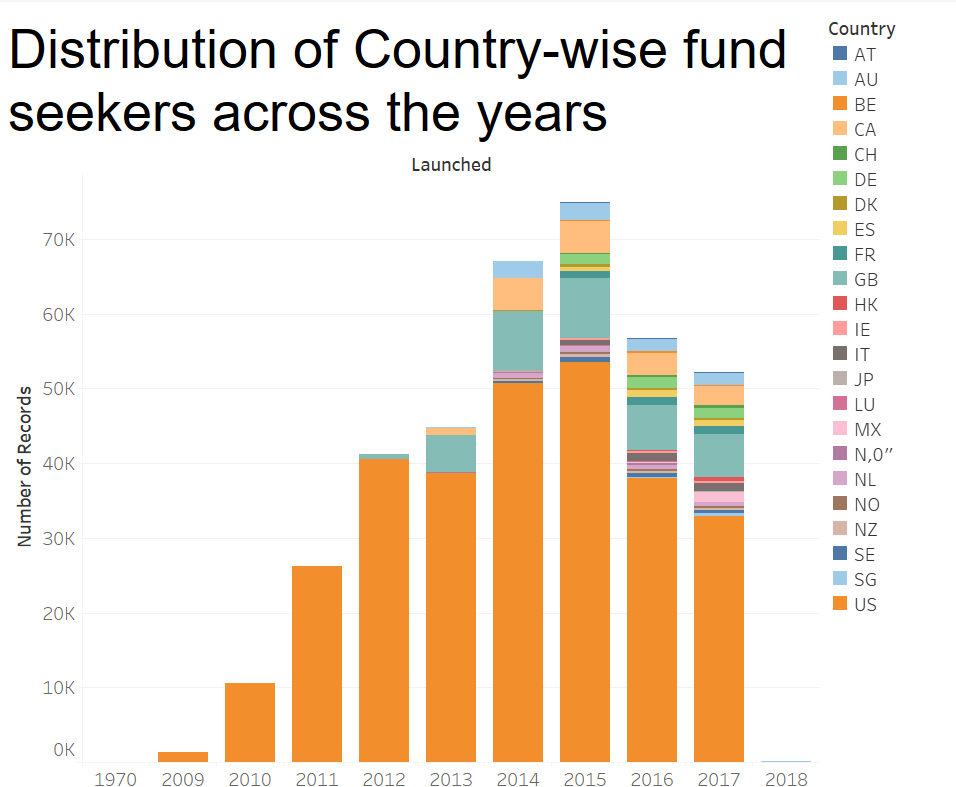


We have also executed a chi square test which indicates that at least one of the Country has higher rate of success compared to others since P value is < 0.0001.



The top 5 countries which have the highest number of projects proposed are shown below. We can see that there is a huge difference in the numbers between US and GB (UK). The reason for this could be that Kickstarter was launched in USA and from the data we can see that for 3 years (2009, 2010 and 2011) all the projects submitted were from US only and thus platform was yet to gain global recognition.

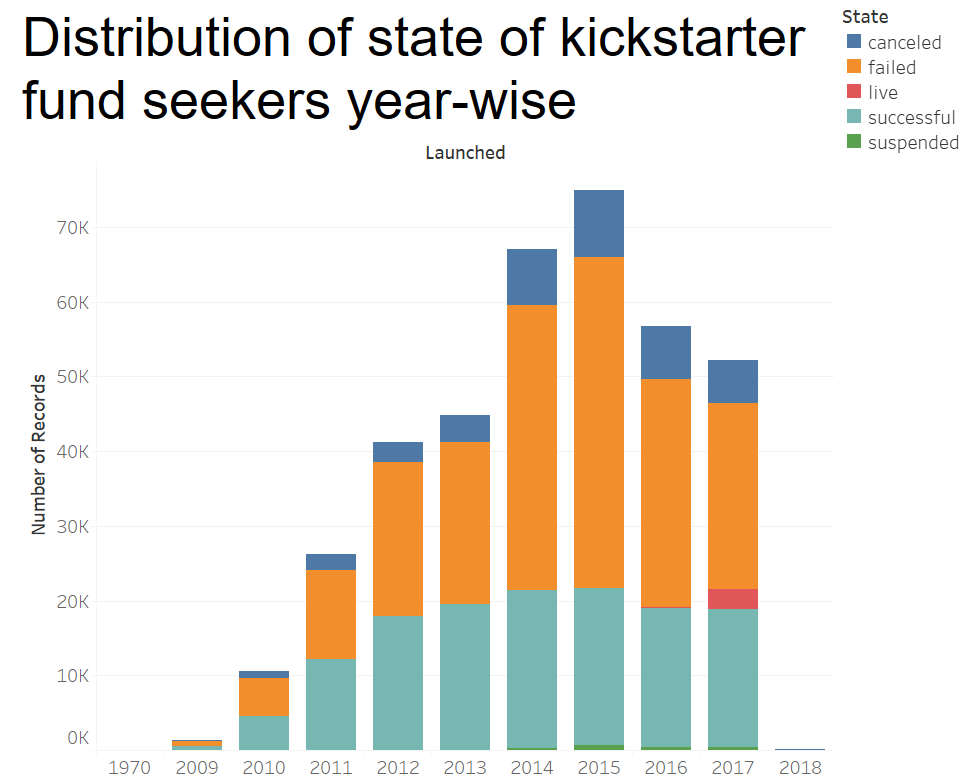
After having a huge surge in project proposals (260,237) in 2011, that’s when Kickstarter gained attention from people outside US and we can see a small portion of submission from UK (617). In the upcoming years we can see increasing submissions from other countries as well.



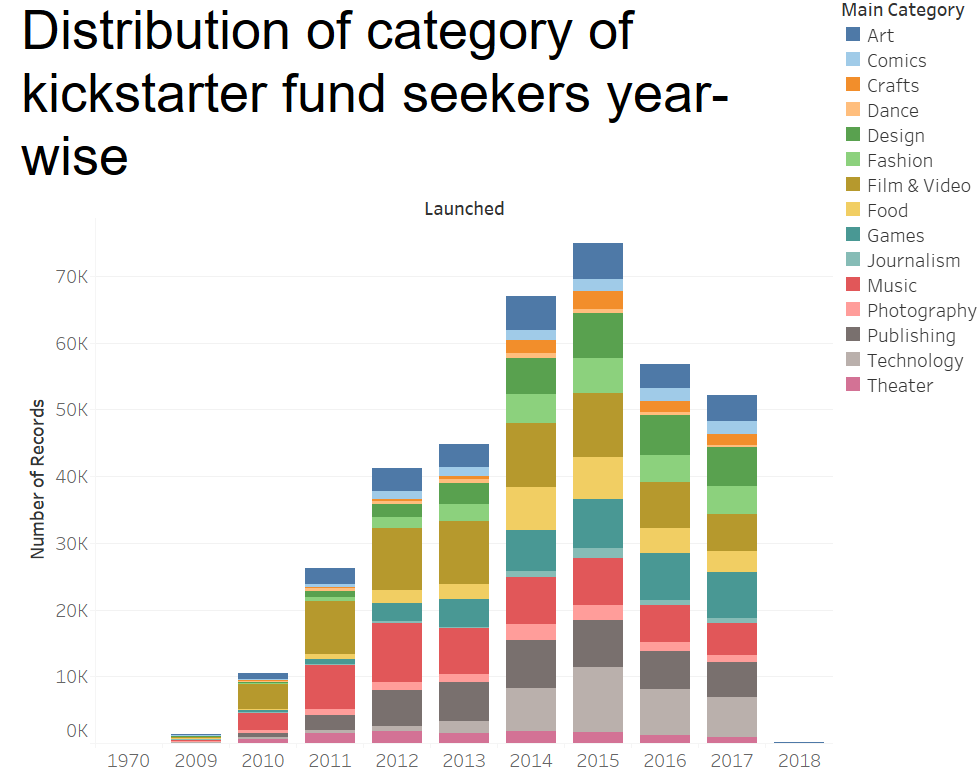
Given below we have the top 5 countries who have successfully submitted the projects.



**Insight 2: Analysis across the years**



1. We can observe that the successful projects have been rising till 2015 and reduced from then on. This can be attributed to the fact that the level of competition for these projects have gone up with time and backers have become more careful in investing bigger amounts into these projects as they weren’t getting much returns and hence reduced their investments in new projects due to which projects weren’t able to meet the pledged amount which led to the declining number of successes.
2. In the years 2014 and 2015 even though we saw a good surge in the number of Startups seeking funding from Kickstarter, there are more number of canceled and failed projects. This is due to the fact that the platform became popular after being in the market for 4 years and it started gaining international acclaim, which caused more people to submit project proposals trying to follow the trend . But they might not have been very innovative and upto the mark which led to more projects in failing and getting cancelled.



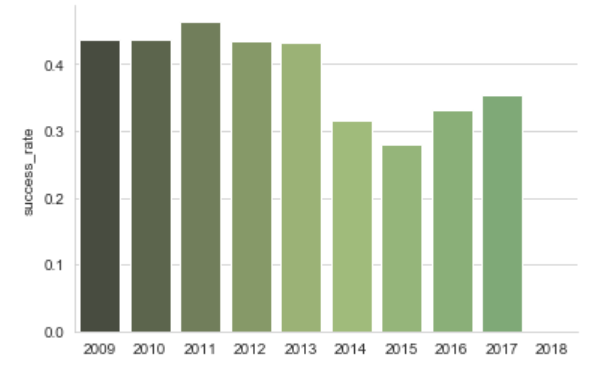
Some of the prominent observations are as below.

1. Years 2014 and 2015 have seen a good surge in the number of Startups that seek funding from Kickstarter.
2. Crowdfunding for video games on Kickstarter is down sharply in 2016 compared with the same period in 2015, suggesting that roughly the same number of video games projects are successfully backed, but at much lower amounts.

This lack of big-earning projects could also point to a general public skepticism for ambitious crowdfunding ventures in video games, or ones that are direct sequels to games postponed long ago by mainstream publishers. Kickstarter donors may also now realize that the higher the funding target, the longer it will take to see results, even past the two-year goal many video game Kickstarters set between funding and delivery.

1. The number of startups as a percent of the number seeking funds is steadily increasing over years for design and fashion categories while increasing at a much higher rate for Technology and gaming spaces. However, tech has seen a small dip in the last couple of years.

**success\_rate per year**

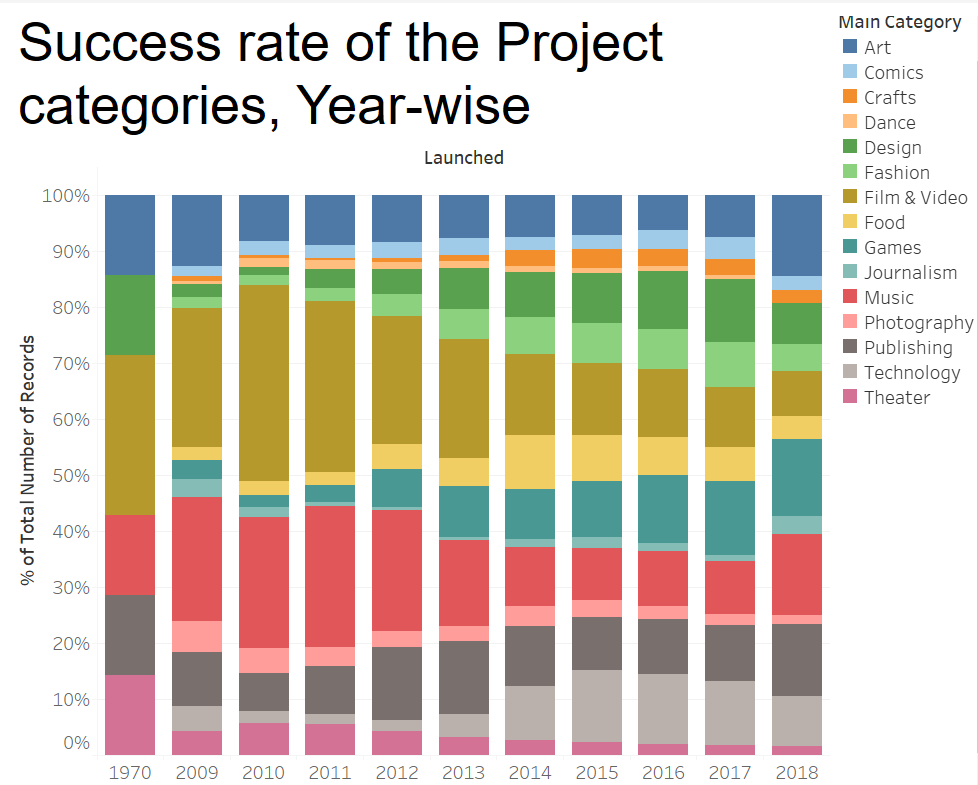




Here we can see that the highest success rate is 46.38% in 2011 with just 26237 projects. This is probably why the Kickstarter initiative started becoming more popular from 2011. To justify this we can observe the number of backers increase more than 3 times and number of projects also increased to double during 2011 to 2012.

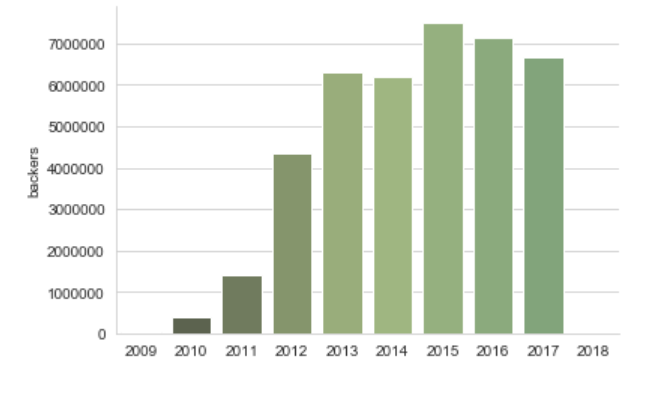
From the years 2011 to 2013 we observe that the success rate is constant as both backers and number of projects rose steadily.

We can also see a decline in success rate from 2013 to 2015 which might have been due to more number of submissions which provided more choices for Backers to invest in, so they probably very selective about their investment.



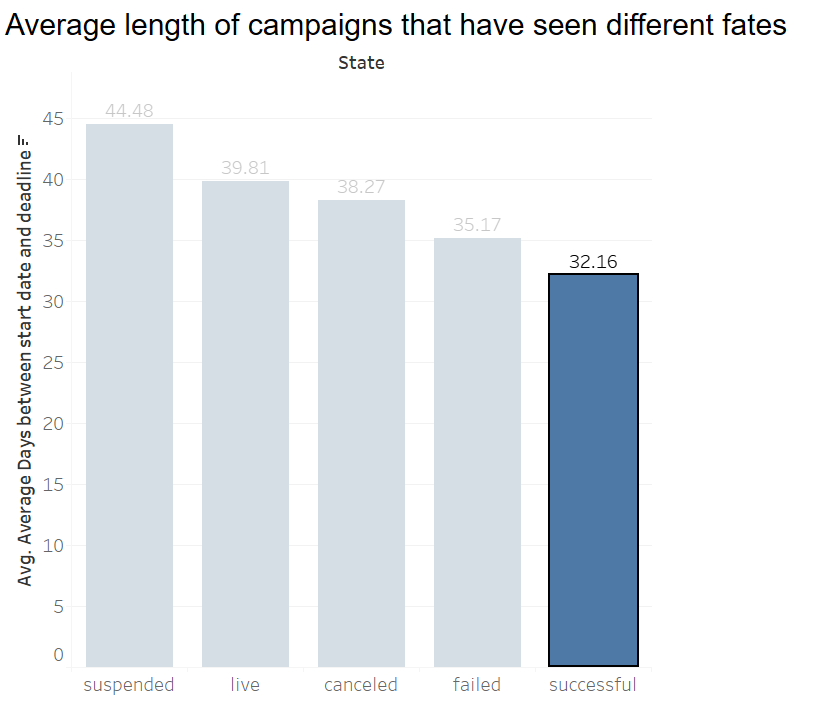
1. A major chunk of Kickstarter projects (around 40%) are films, music projects, and videos. This makes sense: Kickstarter was [founded](https://en.wikipedia.org/wiki/Kickstarter#History) by a trio of Burning Man enthusiasts in 2009 as a way for creative types and artists to raise money.
2. But over time Music, Film & videos are seeing a decline in being successful in campaign while technology, Games and Design are increasingly successful in raising funds because they have lesser investment and more returns along with gaining popularity in usage over years.
3. Technology and design (which frequently overlap, as many tech projects are categorized under “product design”) make up 17%, and games clock in at 10%. For, instance Exploding Kittens was a game with little investment which became one of the biggest hits of kickstarter due to the large revenue it created.
4. This could possibly explain the increased interest in the backers for Games, Technology and Design based startups.

**Distribution of Backers Year-wise**



In graph, we can observe that the number of backers is increasing over the years till 2015 then there is a decline in number of backers. This might be due to repeated experience in the failure in projects at this time, hence backers started backing out of pledging funds.

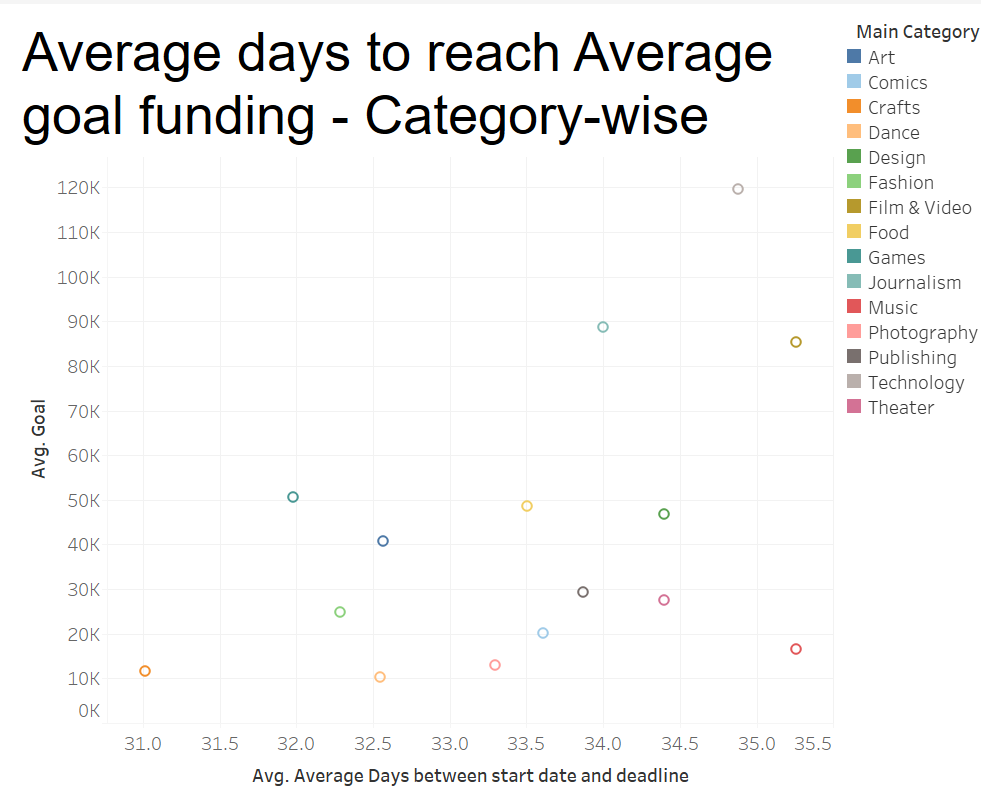
**Insight 3: Analysis across average length of campaigns**



From this graph, it appears that, on an average a campaign that has been successful has about **32.16 days** to achieve the feat.

The ones that are suspended, cancelled and failed as the average campaign length is, more time does not create more urgency. Instead it makes it easier for backers to procrastinate, and sometimes they forget to come back at all.

More time doesn’t help the project creator either. Though the funding deadline may seem to be a project’s enemy, it’s actually its ally. Even far shorter durations can be fatiguing for a creator.

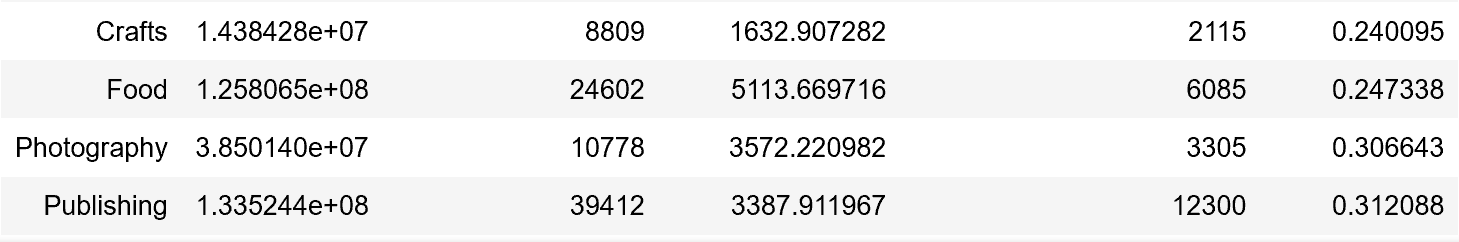
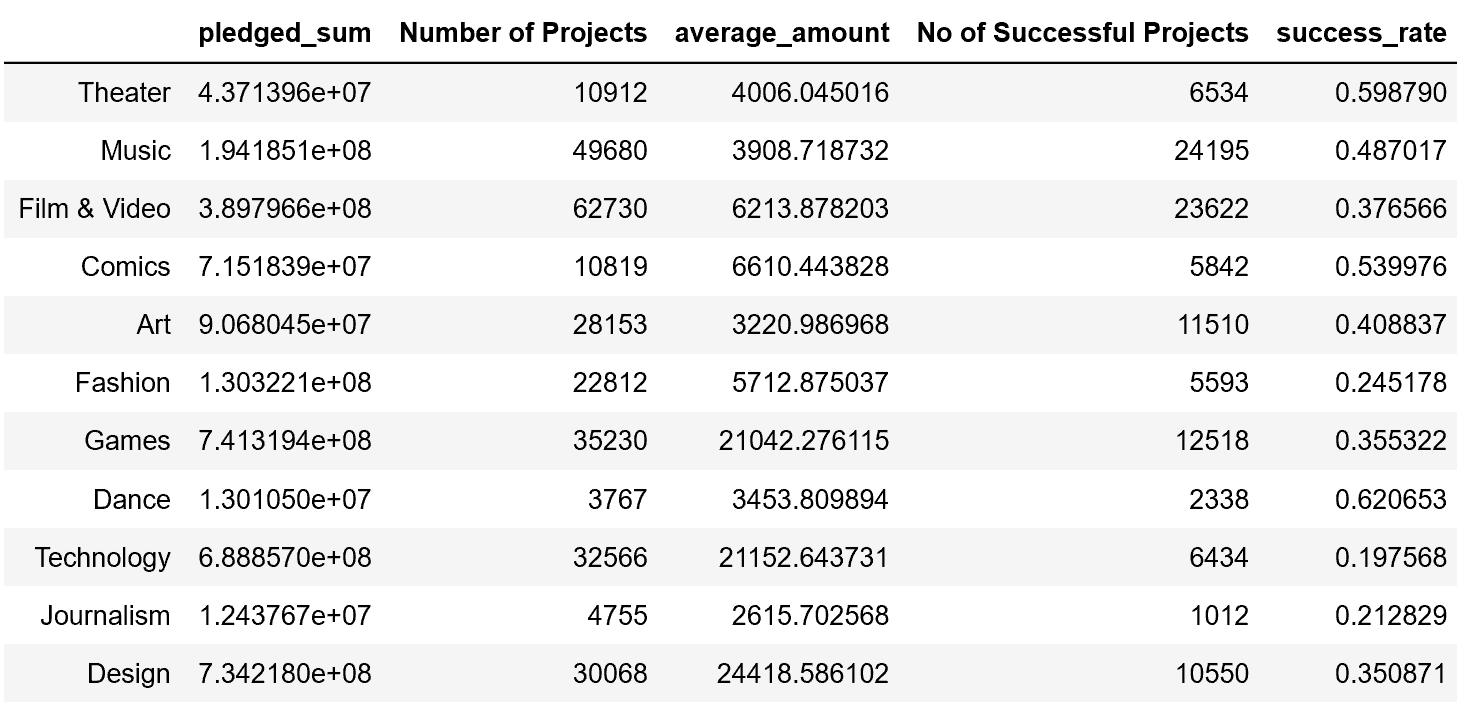


1. On an average crafts related start-ups seems to reach the average Goal faster than any other category. Probably because the average USD required by this community is also less than all other categories.
2. Start ups in Gaming space seem to be reaching the average goal faster than most other categories.

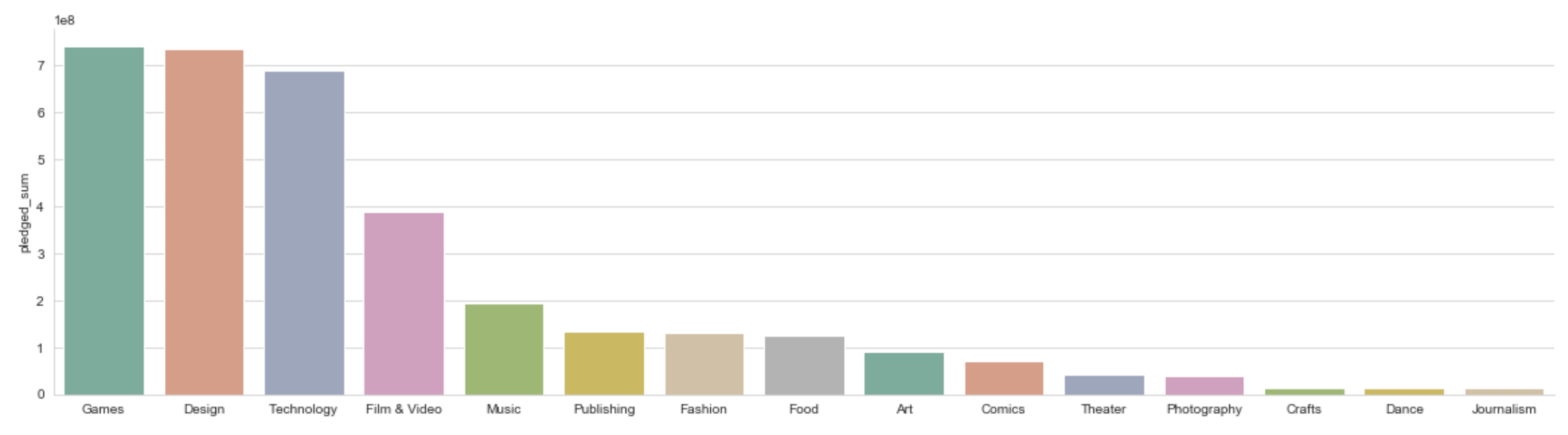
Due to the latest development of GPU, most devices are getting cheaper. Also, the number of customers got increased which caused an increase in the demand of games and hence a need to attain this gaming goal as soon as possible.

**Insight 4: Analysis of Categories**

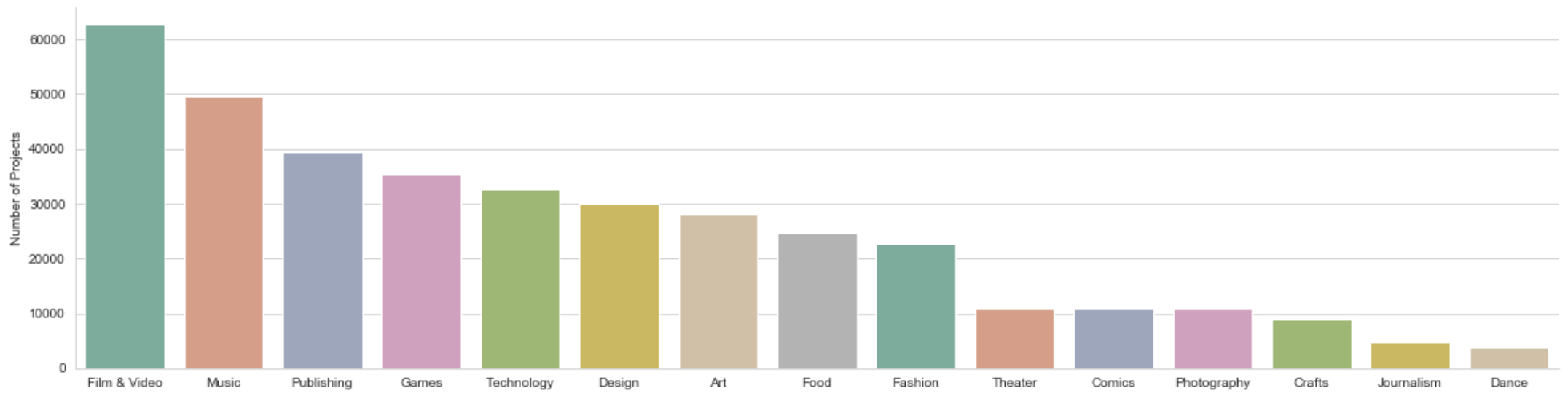
We have 15 major categories in which the projects are submitted for funding.



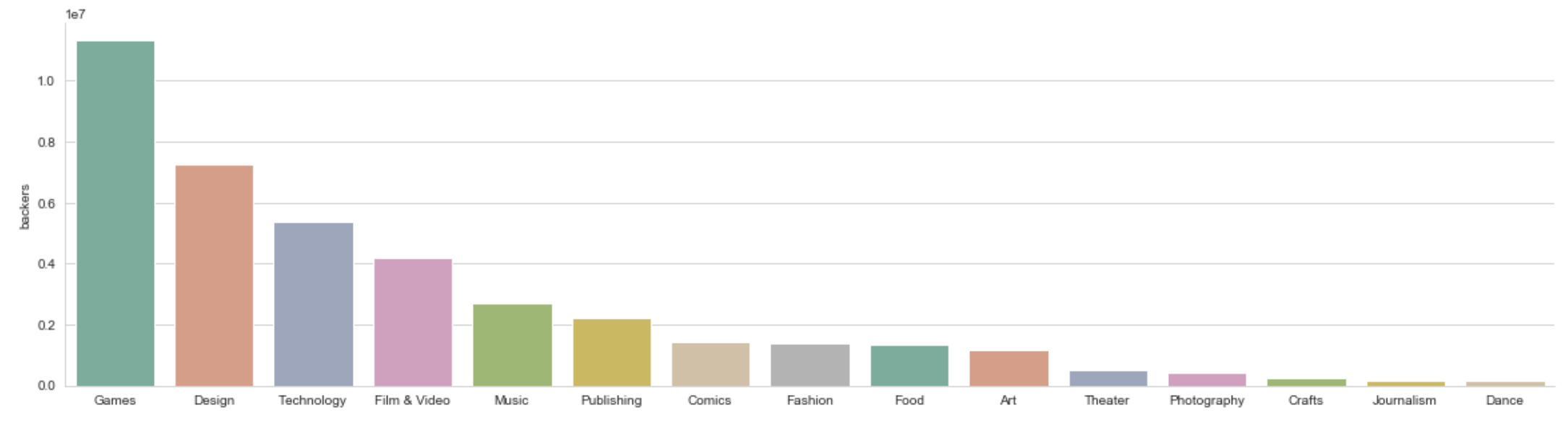
On analyzing the pledged amount, we can see that ‘Games’ category has brought the highest pledged amount of about $741319400.00 followed by ‘Design’ and ‘Technology’. We can infer from this that backers are more invested in these categories.



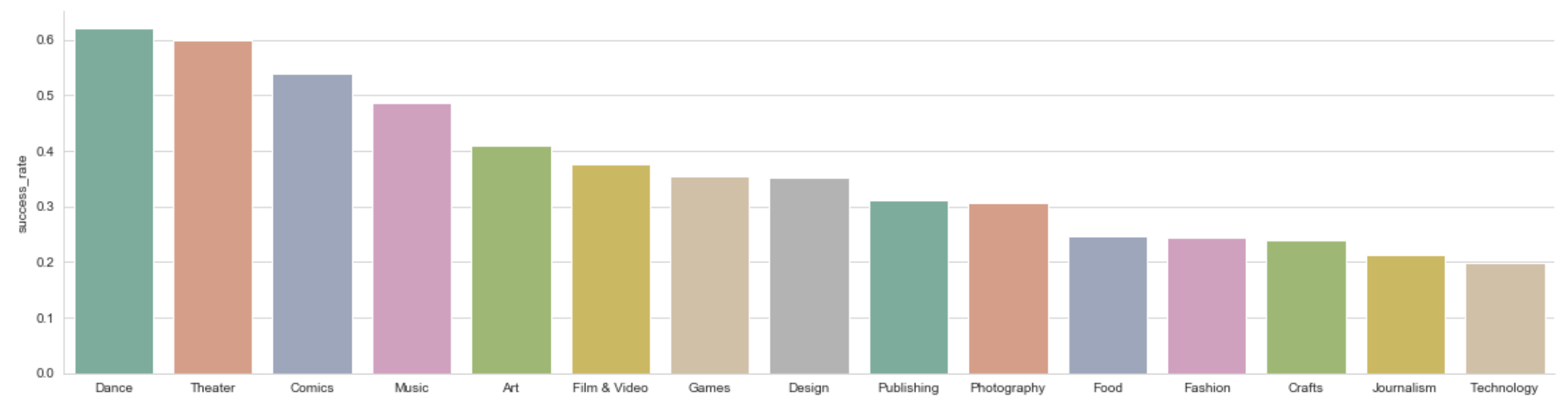
Even though ‘Journalism’ has raised lowest pledged amount, it might be due to less submission of project ideas in this category. Similarly, for categories such as ‘Dance’ and ‘Crafts’, there are lesser number of proposed projects. Here, the maximum number of projects are submitted in ‘Film & Video’ category (62730) which is approximately double the number of projects raised in ‘Games’, ‘Design’ and ‘Technology’.



In order to analyze which category is more successful, we have also investigated the number of backers available in each category. We can observe that Games has the maximum number of backers which explains why it raises the most money as compared to other categories. Games attracts more backers due to its popularity, interests and wide applicability.



Finally we have verified the success rate for projects in categories and surprisingly the success rate is highest in ‘Dance’ followed by ‘Theater’ and ‘Comics’. Even though these categories have less number of projects, since the success rate is high, we might be due to the fact that in spite of low budget of Dance projects and fewer backers who are interested in them, they are innovative enough to raise sufficient funds.



So we can say that it might be beneficial to propose a project in ‘Games’, ‘Design’ and ‘Technology’ as they have maximum backers investing into these projects but at the same time, there is higher competition which is leading to higher failures. In categories such as ‘Dance’, ‘Theater’ and ‘Comics’, there is less competition thus having higher chance to succeed.

**Appendix**

**SAS Code**

*ods html; ods graphics on; ods listing close;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*LIBNAME project 'E:\Users\kxr190005\Documents\My SAS Files\HWs\Project';*

*run;*

*proc import datafile='E:\Users\kxr190005\Documents\My SAS Files\HWs\Project\DATA.csv'*

*out=kickstart*

*DBMS=CSV;*

*GETNAMES=YES;*

*Guessingrows=max;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*proc format noprint;*

*value $main\_category\_num\_var*

*Food = 1*

*Games = 2*

*Photography = 3*

*Art = 4*

*Fashion = 5*

*'Film & Video' = 6*

*Music = 7*

*Technology = 8*

*Publishing = 9*

*Crafts = 10*

*Comics = 11*

*Design = 12*

*Dance = 13*

*Journalism = 14*

*Theater = 15*

*;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*data kickstart\_sf;*

*set kickstart;*

*if state in ("live", "suspended") then delete;*

*if state in ("successful") then success = 1;*

*if state in ("failed", "canceled") then success=0;*

*informat LDate mmddyy10. DDate mmddyy10.;*

*LDate = input (launched, MMDDYY10.);*

*DDate = input (deadline, MMDDYY10.);*

*LaunchMonth = month(LDate);*

*LaunchYear = year(LDate);*

*EndMonth = month(DDate);*

*EndYear = year(DDate);*

*Duration = (DDate - LDate);*

*numgoal = input(goal,8.);*

*numpledged = input(pledged,8.);*

*numbackers = input(backers,8.);*

*if duration <=30 then InMon = 1;*

*if 30 < duration <=60 then InMon = 2;*

*if 60 < duration <=90 then InMon = 3;*

*if duration > 90 then InMon = 4;*

*if LaunchYear in (1970, 2018) then delete;*

*format LDate mmddyy10. DDate mmddyy10.;*

*d\_main\_category = input(put(main\_category, $main\_category\_num\_var.), 3.);*

*if main\_category = 'Food' then d\_food = 1; else d\_food = 0;*

*if main\_category = 'Games' then d\_games = 1; else d\_games = 0;*

*if main\_category = 'Photography' then d\_photography = 1; else d\_photography = 0;*

*if main\_category = 'Art' then d\_art = 1; else d\_art = 0;*

*if main\_category = 'Fashion' then d\_fashion = 1; else d\_fashion = 0;*

*if main\_category = 'Film & Video' then d\_film\_video = 1; else d\_film\_video = 0;*

*if main\_category = 'Music' then d\_music = 1; else d\_music = 0;*

*if main\_category = 'Technology' then d\_technology = 1; else d\_technology = 0;*

*if main\_category = 'Publishing' then d\_publishing = 1; else d\_publishing = 0;*

*if main\_category = 'Crafts' then d\_crafts = 1; else d\_crafts = 0;*

*if main\_category = 'Comics' then d\_comics = 1; else d\_comics = 0;*

*if main\_category = 'Design' then d\_design = 1; else d\_design = 0;*

*if main\_category = 'Dance' then d\_dance = 1; else d\_dance = 0;*

*if main\_category = 'Journalism' then d\_journalism = 1; else d\_journalism = 0;*

*if main\_category = 'Theater' then d\_theater = 1; else d\_theater = 0;*

*log\_goal = log(goal);*

*log\_usd\_goal\_real = log(usd\_goal\_real);*

*log\_timeline = log(Duration);*

*\*Transform sqrt;*

*sqrt\_backers = sqrt(backers);*

*sqrt\_pledged = sqrt(pledged);*

*sqrt\_usd\_pledge\_real = sqrt(usd\_pledged\_real);*

*\*Transform sq;*

*sq\_backers = backers\*\*2;*

*sq\_pledged = pledged\*\*2;*

*sq\_usd\_pledge\_real = usd\_pledged\_real\*\*2;*

*\*interaction variable;*

*pledged\_backers = log(pledged\*backers);*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*/\*Looking at the frequencies by country, state and category\*/*

*proc freq data=kickstart order = freq;*

*tables category main\_category country success;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*/\*chi square test for independence - are variables independent or not\*/*

*/\*at least one of the categories has higher rate of success compared to others since P value is < 0.0001 \*/*

*/\*at least one of the MAIN categories has higher rate of success compared to others since P value is < 0.0001 \*/*

*/\*at least one of the Country has higher rate of success compared to others since P value is < 0.0001 \*/*

*proc freq data=kickstart\_sf order = freq;*

*tables (category main\_category country) \* success/ chisq;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*/\*Analysing further to look at the differences in success and failure categories within each category group\*/*

*proc sort data=kickstart\_sf;*

*by main\_category;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*proc freq data=kickstart\_sf order = freq;*

*tables (category) \* success/ chisq;*

*by main\_category;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*/\*Analysing further to look at the differences within Countries\*/*

*proc sort data=kickstart\_sf;*

*by country;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*proc freq data=kickstart\_sf order = freq;*

*tables (country) \* success/ chisq;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*/\*Analysing further to look at the differences within Countries\*/*

*proc sort data=kickstart\_sf;*

*by duration;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*proc freq data=kickstart\_sf order = freq;*

*tables (duration) \* success/ chisq;*

*run;*

*proc sgplot data=kickstart\_sf;*

*title "Histogram";*

*histogram duration;*

*yaxis values=(0 to 100 by 10);*

*xaxis values=(0 to 200 by 30);*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*\* Create random sample of 80% records;*

*PROC SURVEYSELECT DATA = kickstart\_sf OUT = kickstart\_data\_sample outall samprate=0.8 seed=1234567;*

*RUN ;*

*\* Create training set and test set;*

*data kickstart\_train kickstart\_test;*

*set kickstart\_data\_sample;*

*if selected then output kickstart\_train;*

*else output kickstart\_test;*

*run;*

*\* Create test set;*

*data kickstart\_train;*

*set kickstart\_train;*

*if selected then train\_y=success;*

*run;*

*proc contents data=kickstart\_train;*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*\* Histograms;*

*PROC means data=kickstart\_train min max median p25 p75;*

*var Numgoal;*

*run;*

*proc univariate normal data=kickstart\_train;*

*title "Histogram";*

*var d\_main\_category Numgoal log\_goal numbackers usd\_pledged\_real usd\_goal\_real duration log\_timeline;*

*histogram / normal (mu = est sigma = est);*

*run;*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*\* Distributions;*

*proc freq data=kickstart\_train;*

*tables success;*

*run;*

*proc freq data=kickstart\_train;*

*tables d\_main\_category;*

*run;*

*/\*Correlations\*/*

*proc corr data=kickstart\_train;*

*var success pledged\_backers usd\_pledged\_real usd\_goal\_real duration d\_food d\_games d\_photography d\_art d\_fashion d\_film\_video d\_music d\_technology d\_publishing d\_crafts d\_comics d\_design d\_dance d\_journalism d\_theater;*

*title 'Correlations';*

*run;*

*/\*there is high positive correlation between pledged\_backers and success which is understandable as for projects having higher pledged amount and more backers*

*will have more chance to succeed. Also for projects which tends to succeed attract more backers\*/*

*/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/*

*\* Model Selection;*

*/\*Running stepwise selection to filter out significant variables to the model\*/*

*proc logistic data=kickstart\_train;*

*title "Stepwise";*

*model train\_y(event='1') = pledged\_backers log\_goal sqrt\_pledged sqrt\_backers sqrt\_usd\_pledge\_real log\_usd\_goal\_real duration d\_food d\_games d\_photography d\_art d\_fashion d\_film\_video d\_music d\_technology d\_publishing d\_crafts d\_comics d\_design d\_dance d\_journalism d\_theater / selection= stepwise stb rsquare;*

*run;*

*/\*All the model parameters are significant and fits the criteria except sqrt\_pledged column\*/*

*/\*The Forward elimination analysis (SELECTION=FORWARD) starts Model with no predictors, to eliminate insignificant variables\*/*

*proc logistic data=kickstart\_train;*

*title "Forward";*

*model train\_y(event='1') = pledged\_backers log\_goal sqrt\_pledged sqrt\_backers sqrt\_usd\_pledge\_real log\_usd\_goal\_real duration d\_food d\_games d\_photography d\_art d\_fashion d\_film\_video d\_music d\_technology d\_publishing d\_crafts d\_comics d\_design d\_dance d\_journalism d\_theater / selection= forward stb rsquare;*

*run;*

*/\*All the model parameters are significant and fits the criteria\*/*

*/\*The backward elimination analysis (SELECTION=BACKWARD) starts Model with all predictors, to eliminate insignificant variables\*/*

*proc logistic data=kickstart\_train;*

*title "Backward";*

*model train\_y(event='1') = pledged\_backers log\_goal sqrt\_pledged sqrt\_backers sqrt\_usd\_pledge\_real log\_usd\_goal\_real duration d\_food d\_games d\_photography d\_art d\_fashion d\_film\_video d\_music d\_technology d\_publishing d\_crafts d\_comics d\_design d\_dance d\_journalism d\_theater / selection= backward stb rsquare;*

*run;*

*/\*All the model parameters are significant and fits the criteria\*/*

**Python Code File**

