

GROUP 17 – PROJECT KICKSTARTER

This report presents in-depth coverage of the course project for BUAN 6337.

The report is about Kickstarter data and analysis of success factors for a crowdsourcing campaign. The report covers in detail the controlling factors and scenarios that a company seeking crowdfunding should consider in order to maximize the chances of success.

BUAN 6337: DATE-
5/6/2019

FACULTY MENTOR
PROF. SHERVIN TEHRANI

Dongwook Chun
Pavithra Gunasekaran
Ga Hyun Lee
Anurag Sethi
Seung Ah Shin
Huimin Tang

Contents

1. Executive summary

- Research question

2. Data preprocessing

1) Reason for Selection

2) Dataset Description

3) Data Feature Engineering

3. Exploratory Data Analysis (EDA)

1) Descriptive analysis

2) Hypothesis Test

4. Models & Interpretations (Logit model)

5. Business Solution

6. Conclusion

1. Executive summary

Online crowdfunding industry is a new but rapid growing industry. The industry size USD718M and annual growth rate 7%. Main players in the industry include Kickstarter, Patreon, Indigogo, RocketHub etc.

Kickstarter is one of the biggest platforms based in Brooklyn, New York with its mission "help bring creative projects to life".¹ Till now, Kickstarter has successfully helped to raise fund for over 257,000 projects around the world, in categories of film, music, art, theater, games, comics, design, photography, and more. Kickstarter has mainly two kinds of users: project creators and backers. The major business model is project creators post their project on Kickstarter with a defined deadline and a funding goal. If the goal is not met by the deadline, no funds are collected. If the goal is met, project creators can use the money on their projects. Backers are those who choose to fund their selected Kickstarter projects, offered tangible rewards or experiences in exchange for their pledges.

From historical data, the average success rate of the projects posted is only around 41%. The rate is contributed to many factors, such as goal amount, duration and launch time etc. To successfully raise the goal amount, there's something that a project creator can do, such as deciding a realistic goal amount, and choosing when to launch the project and the duration. There's also a lot of things that a project creator cannot control, such as the category and location etc.

It is extremely valuable to research from the perspective of project creator and explore what affects the success of a project, and what project creator do to increase the chance of success. With this consideration, the main research question was constituted as

“How to successfully achieve crowdfunding goal?”

To do the research, we used the dataset from Kickstarter, with more than 330,000 posted projects regardless of success or failure during April 2009 to December 2017. The main techniques used in our research includes but not limited to descriptive analysis, hypothesis testing, statistical analysis, feature engineering and logistic regression implemented in SAS.

The report will explain in detail regarding our exploration, findings and suggestions for Kickstarter project creators, to help improve the success rate on their projects.

¹ <https://en.wikipedia.org/wiki/Kickstarter>

2.Data preprocessing

1) Reason for Selection

For the project a rich assortment of the datasets was presented to the group for selection, there were many datasets on ecommerce, financial credit, sports etc. There were multiple factors that were the key drivers in the selection of Kickstarter dataset among the other presented datasets. They have been listed below.

- Raising capital is very essential for the success of a start-up, the traditional avenues to raise money are venture capital, angel investors. Crowdfunding is a relatively new phenomenon which has grown into significance only recently in last one and half decades. The key characteristic difference between crowdfunding and traditional financing avenues is that in a venture capital setup, startups usually part some part of their equity, this is generally not the case in the crowdfunding setup. **Crowdfunding is more novel, more democratic as it provides a more equal platform for all participants irrespective of their financial strength.**
- There are various studies that cite the **explosive growth of crowdfunding**. According to Statistica.com crowdfunding market will register a CAGR of more than 17% by 2022. At present, In the year 2019 in the United States the total amount raised through crowdfunding is expected to be over USD 718M. Kickstarter is the foremost trend setter in this industry, however in recent years other competitors like Patreon, GoFundMe, Indiegogo have grown their financial footprint.
- The **motto and business model of Kickstarter** is interesting and appeals to us as a group, “Help bring creative projects to life” does make it clear that the company does cater to the art community and promotes originality in value creation. Such sectors of economy like art, music may not have same access to venture capital as for instance tech industry. This is one of the reasons Kickstarter platforms was appealing to us as a group. The business model of Kickstarter is interesting, it sets up an all or nothing scenario for the company seeking crowdfunding. One needs to be successful in raising the full amount or they end up with nothing.
- **Business motivations:** We, as a group imagine ourselves as a consulting company, our client is an individual or a start-up seeking to begin their crowdfunding campaign, there are certain factors like macro-economic scenario that are not in our control. But there are certain factors that our client can control and change, like the goal amount sought, duration of the campaign, timing of the campaign launch etc. Making intelligent decisions for these controllable features can be a difference between a successful and unsuccessful campaign. **The report has this aspect as the central theme.**

2) Dataset Description

The given dataset required some pre-processing to be made '**analysis ready**'. We were provided with two different files with a lot of common projects, our first steps were separate out **failed and successful projects**, as they were the the basis of the analysis.

A small fraction of observations had incorrect and illogical entries (< 0.01%) like symbols and names in numeric fields, they were subsequently removed.

In the data we observed that there were observations which had very low values for the Project goal amount. These observations although very small in proportions had the potential to make the data biased. These projects also had a disproportionately high success rate, compared to their category average. To overcome this situation, bin wise analysis was done for each category. We analyzed the number of projects falling in the goal amount categories, 0-10 USD, 10-20 USD, 20-50 USD, 50-100 USD. Naturally, the projects the success rate is very high for the lowest bucket as the goal amount sometimes even touches zero. For each of the main category we set the lower threshold for the projects to be excluded from our analysis. For example, for projects in main category Art, when the goal amount is less than 10 USD the success rate is more than twice the overall category average. Hence 10 USD is the lower threshold chosen for this category. Similarly, different lower thresholds are chosen for different categories. After this pre-processing step another 2609 observations are removed from the dataset.

Our final data set used for analysis and modelling, has data for 330k+ Kick Starter Projects launched during the time April 2009 – December 2017.

Original Set of Feature Variables

1. ID : Internal Kickstarter id
2. Name: Name of project, A project is a finite work with a clear goal that you'd like to bring to life. Think albums, books, or films.
3. Category: Project Category
4. Main_category : Category of campaign, a higher level categorization
5. Currency : Currency used to support
6. Deadline : deadline for crowdfunding
7. Goal: fundraising goal, The funding goal is the amount of money that a creator needs to complete their project.
8. Launched: Date launched
9. Pledged : Amount pledged by "crowd"
10. State: Current condition the project is in
11. Backers: Number of backers
12. Country: Country pledged from
13. USD pledged: Amount of money pledged in USD
14. USD Goal - Goal amount in USD

3) Data Feature Engineering

For the requirement of the analysis our team engineered new variables from the existing set of variables. **Feature engineering is one of the most important and consequential part of the project.** The newly created features and their description are given below. Some of them have been used in hypothesis tests, some of them in the model.

Variable-Name	Description
Duration	The length of the campaign in days.
Launched_month	Month of the year when the campaign was launched, labelled 1-12
Launched_year	Year when the campaign was launched, labelled 2009-2017
Launched_week	Week of the year when the campaign was launched, labelled 1-52
Launched_day_type	Whether day of the week was weekday or weekend
Quarter	Quarter of the year, labelled 1-4 project was launched
Competitors	Calculated for each project ass, number of projects launched in the same week as the current project that belong to the same main category of the given project.
Avg_Amount_Pledged_Backer	Average amount pledged per backer
Level_USD_Goal*	Project Goal in USD divided into four quantiles using the projects in that main category as reference
Sub.region	Continent where the project was launched
Sentiment	The sentiments whether positive, negative or neutral extracted from the name of the project using the bag of words approach
Name_count	Number of words in the name of the project
All_goal_level	Project Goal in USD divided into four quantiles using the all projects in the dataset.

* Table for Goal Level by Each Category is attached in Appendix

3.Exploratory Data Analysis

1)Descriptive analysis

a) The change of success rates by each category

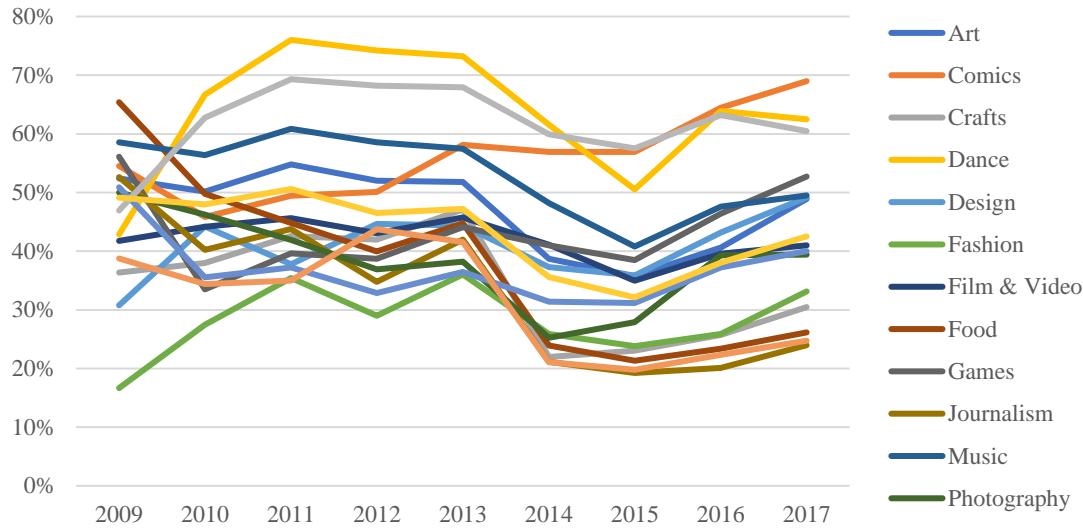


Figure 1 – success rate of main category by year

Category	2009	2010	2011	2012	2013	2014	2015	2016	2017	Grand Total
Art	142	760	2,166	3,235	3,244	4,546	4,733	3,205	3,419	25,450
Comics	22	237	539	1,092	1,291	1,432	1,709	1,678	1,760	9,760
Crafts	11	50	108	299	449	1,718	2,278	1,555	1,299	7,767
Dance	7	162	391	496	530	666	578	410	325	3,565
Design	26	156	799	1,701	2,825	4,592	5,681	4,922	4,584	25,286
Fashion	24	153	520	1,489	2,137	3,699	4,543	3,522	3,414	19,501
Film & Video	296	3,330	7,244	8,721	8,786	8,719	8,556	6,000	4,622	56,274
Food	26	241	576	1,702	2,111	5,407	5,434	3,357	2,723	21,577
Games	38	205	714	2,421	3,351	4,709	5,702	5,439	5,275	27,854
Journalism	38	167	153	252	191	775	1,255	714	501	4,046
Music	267	2,292	6,166	8,429	6,527	6,461	6,358	4,873	4,229	45,602
Photography	61	406	784	1,080	1,145	2,079	1,877	1,158	873	9,463
Publishing	111	625	2,028	4,975	5,337	6,467	6,291	4,929	4,484	35,247
Technology	48	189	399	739	1,614	5,334	8,002	5,798	4,649	26,772
Theater	49	572	1,387	1,703	1,413	1,678	1,560	1,049	801	10,212
Grand Total	1,166	9,545	23,974	38,334	40,951	58,282	64,557	48,609	42,958	328,376

Figure 1 shows the change in success rates over time. Comics show a steadily increasing trend. However, the success rates of most categories have fallen sharply from 2013 to 2015, as projects have increased sharply at that time.

b) Total success rate and the number of projects by main category

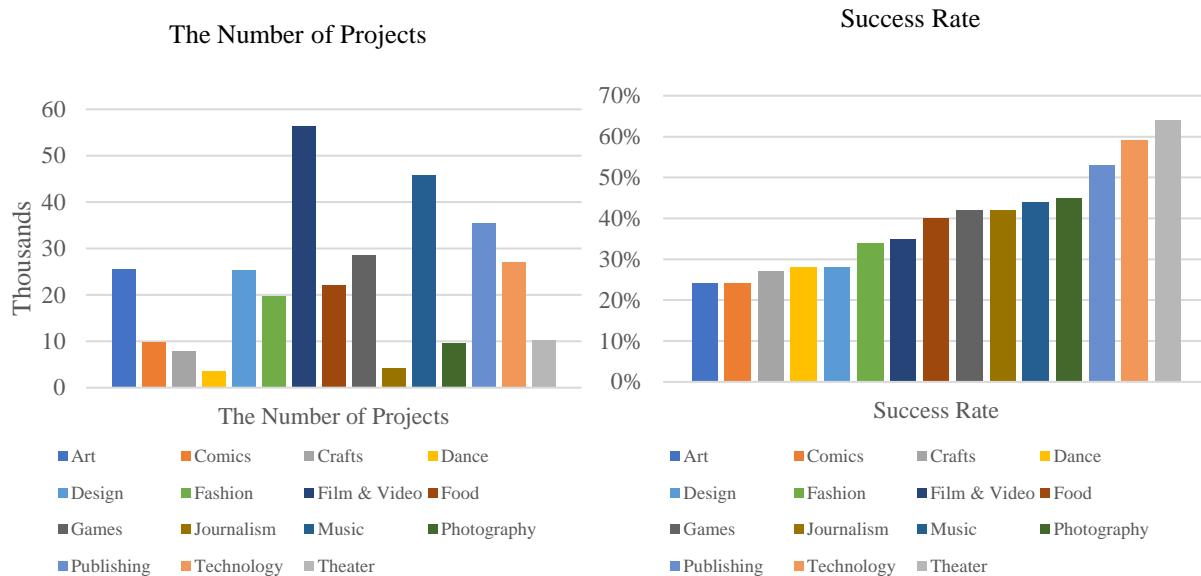


Figure 2 - Success rate by main category

Figure 3- The number of projects by main category

Figure 2 shows the total number of each project. We can see which categories of projects went on frequently. We can also see the overall success rate of each category in the figure 3.

c) Total success rate and the number of projects by Duration Level

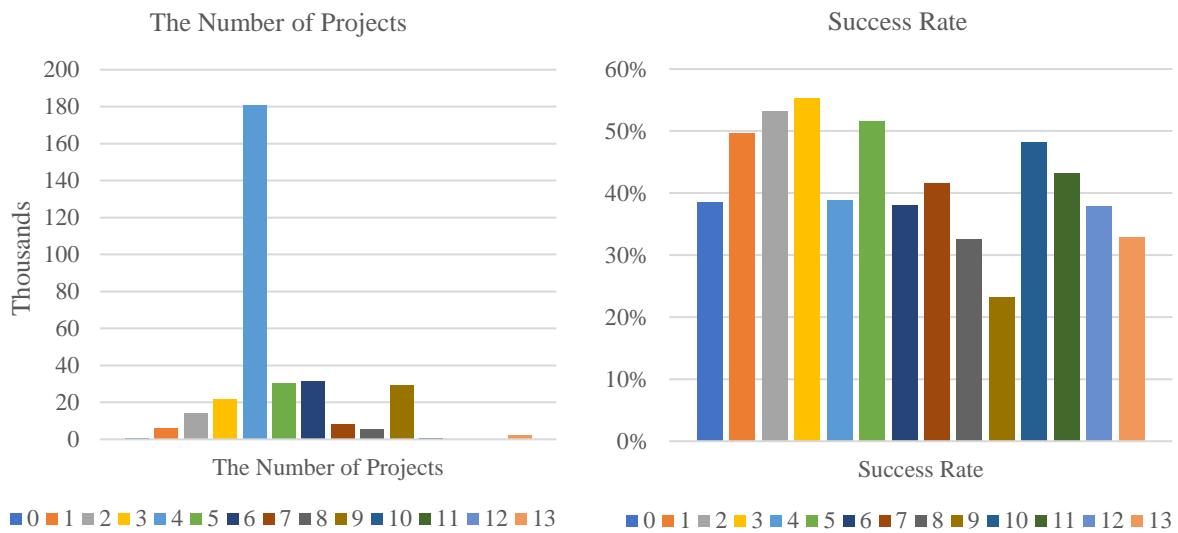


Figure 4 - Success rate by duration level

Figure 5 - The number of projects by duration level

It shows the distribution and success rate according to the duration of the project. Most projects have a period of about four weeks, with three weeks showing the best success rate.

d) The number of projects and total successful rate and per sub region

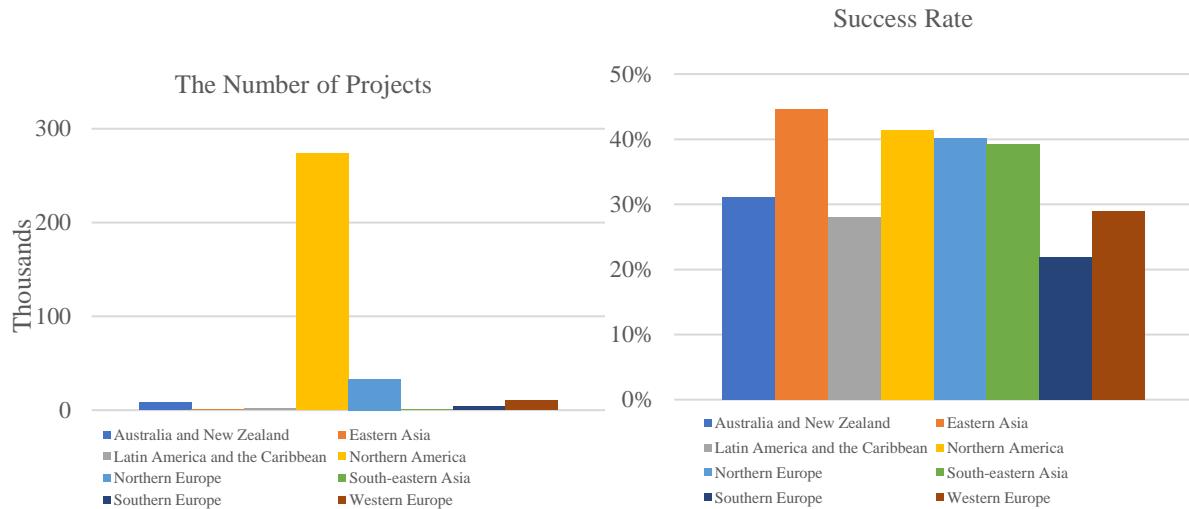


Figure 6 - The number of projects by sub region

Figure 7 – Success rates of each sub region

Most projects are from North America. However, the success rate of eastern Asia is higher than others, on the other hands, Southern Europe has lowest success rate.

e) The Success rate by launched quarter

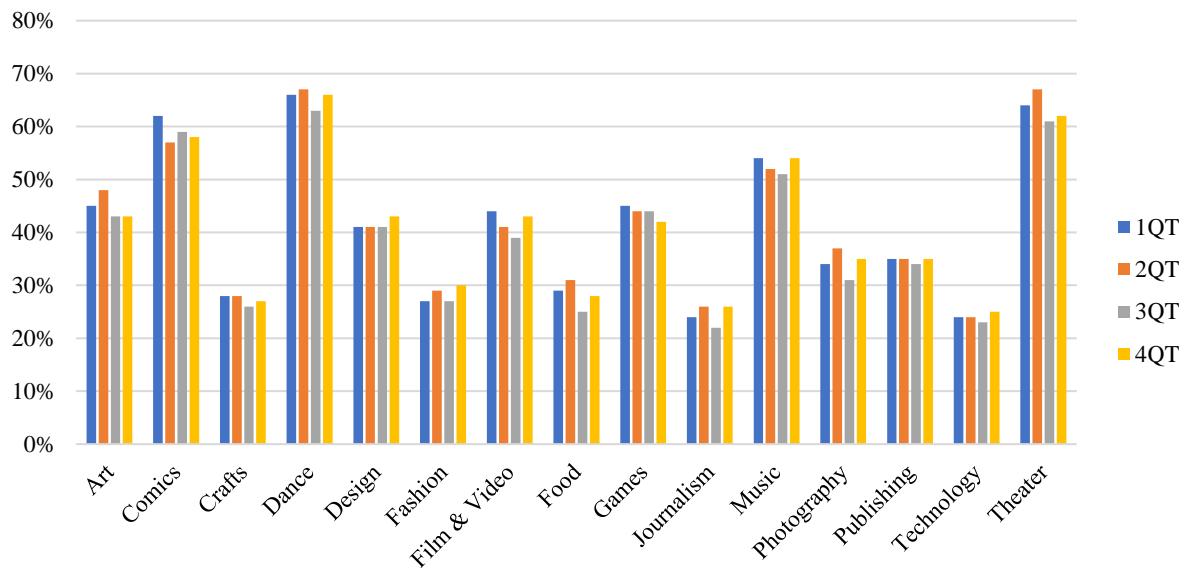


Figure 8 - Success rates by launched quarter and category

Figure 8 shows the success rates for each category and launched quarter. It seems that there is no big difference.

f) The Success rate by launched day type

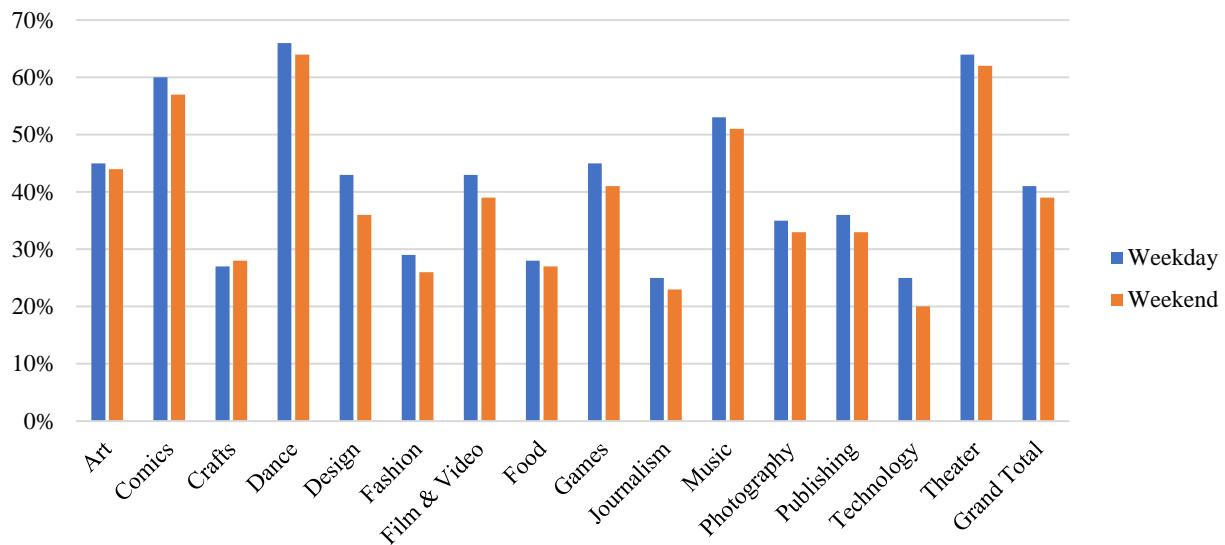


Figure 9 - Success rate by launched date type and category

Figure 9 shows the success rates for launched date type and category. Weekday usually shows better success rates.

g) Number of Backers by Main Category

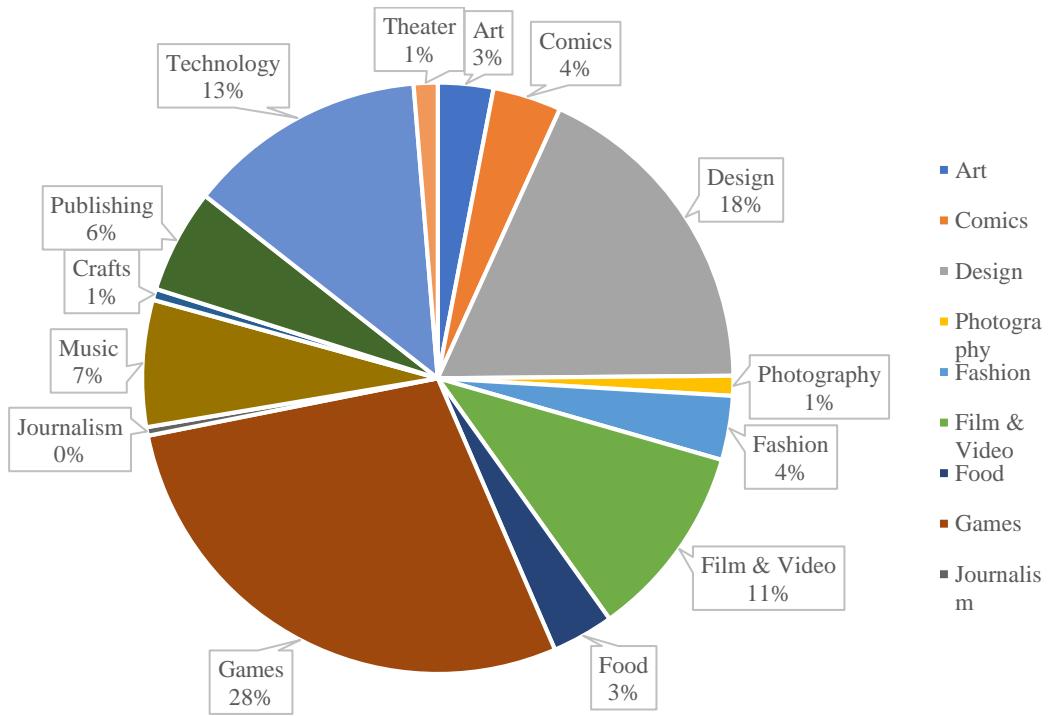


Figure 10 - The number of backers by category

We can see there are a lot of people who participate in funding of game and design project in figure 10.

h) Average Goal amounts per main category

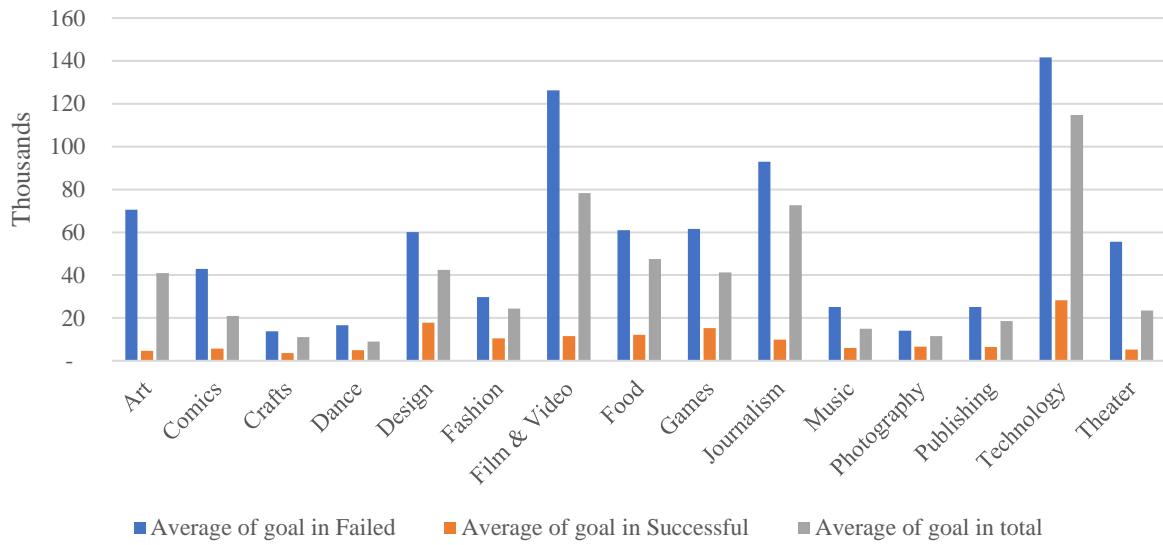


Figure 11 - Average goal by category and state

Failed projects have a higher average goal. On the other hands, the average goal for successful projects is lower than total average goal.

i) Average Pledged amounts per main category

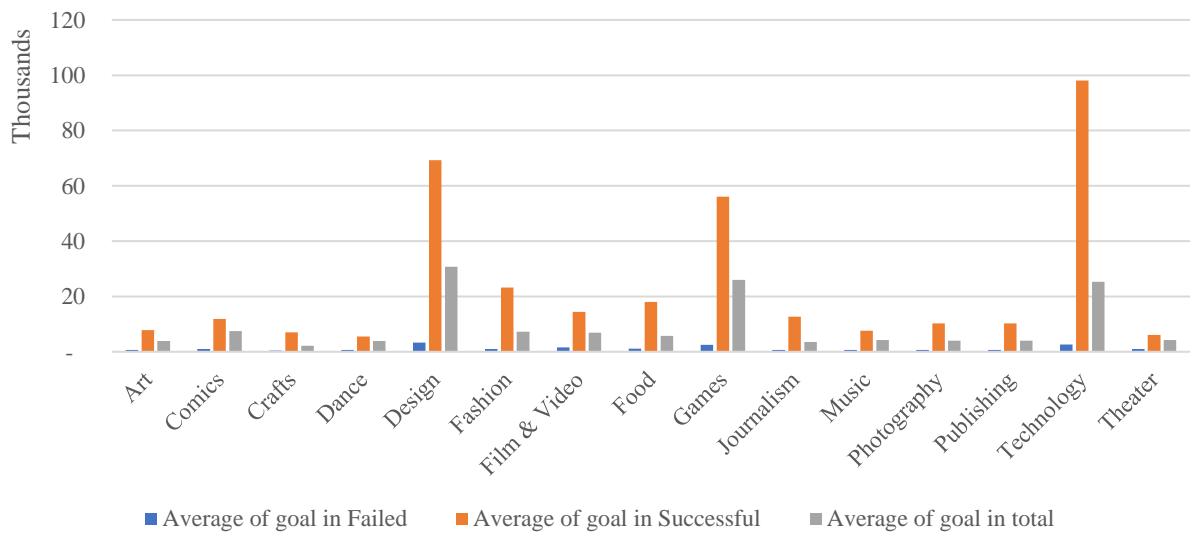


Figure 12 - Average pledged by category

Technology, Design and Games have higher pledged amount compare to other categories.

j) Change of Competitors per main category over time

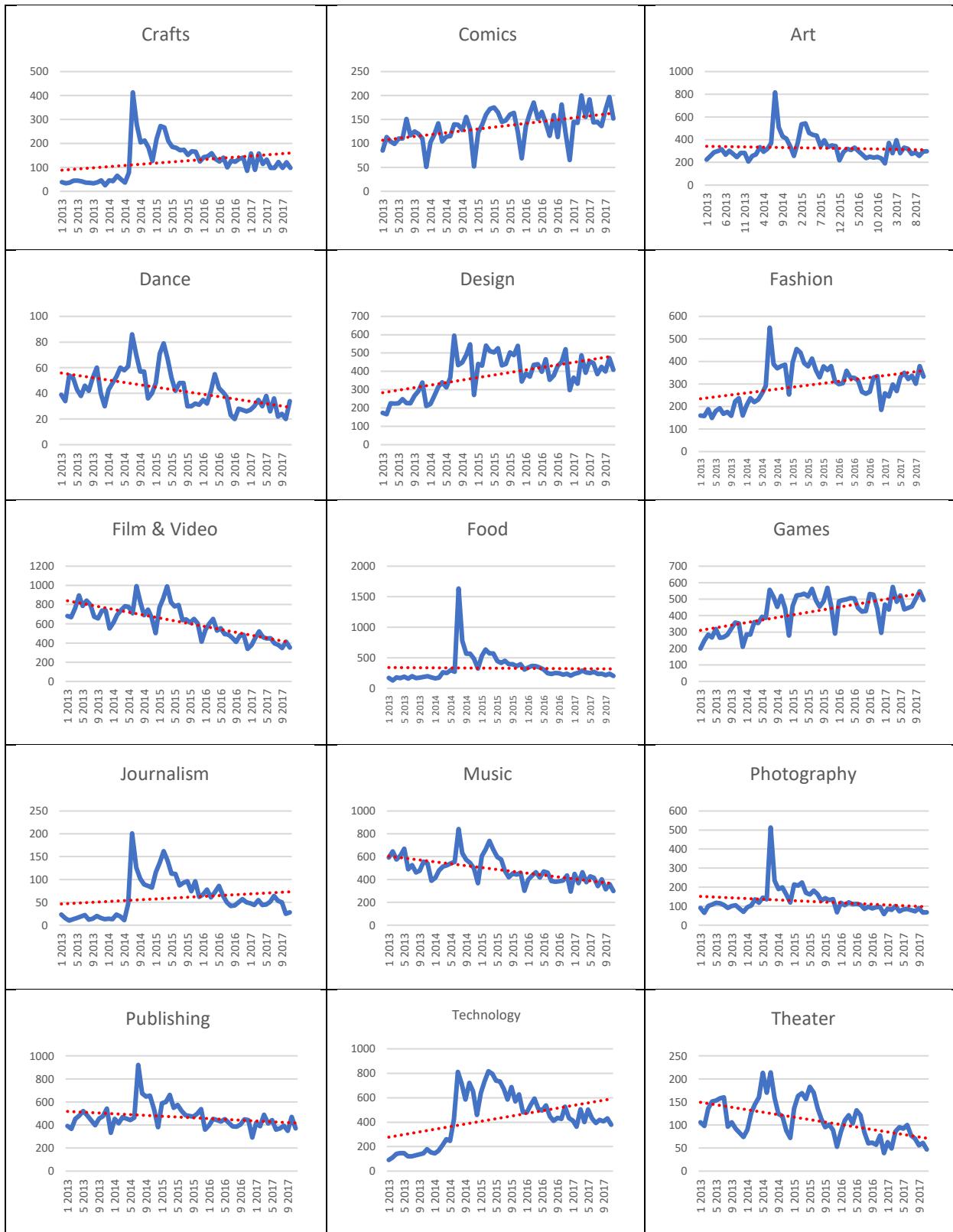


Figure 13 – Change of competitors over time

Figure 13 shows the change of competitors number over time. Some categories have increasing trend over time, but the others have decreasing trend.

k) Correlation of each quantitative variable

Pearson Correlation Coefficients, N = 328595 Prob > r under H0: Rho=0									
	backers	duration	Competitors	Avg_Amount_Pledged_Backer	usd_pledged_real	usd_goal_real	goal_achievement	name_count	
backers backers	1.00000 0.8179	-0.00040 0.0012	-0.00564 <.0001	0.01724 <.0001	0.75337 <.0001	0.00519 0.0030	0.31584 <.0001	0.01044 <.0001	
duration duration	-0.00040 0.8179	1.00000 <.0001	-0.03129 <.0001	0.02042 <.0001	0.00820 <.0001	0.02160 <.0001	-0.01882 <.0001	0.00376 0.0311	
Competitors Competitors	-0.00564 0.0012	-0.03129 <.0001	1.00000 <.0001	0.01255 <.0001	0.00078 0.6534	0.01534 <.0001	-0.01521 <.0001	-0.03315 <.0001	
Avg_Amount_Pledged_Backer Avg_Amount_Pledged_Backer	0.01724 <.0001	0.02042 <.0001	0.01255 <.0001	1.00000 <.0001	0.09260 <.0001	0.01341 <.0001	0.03123 <.0001	0.04902 <.0001	
usd_pledged_real usd_pledged_real	0.75337 <.0001	0.00820 <.0001	0.00078 0.6534	0.09260 <.0001	1.00000 <.0003	0.00632 0.0003	0.26600 <.0001	0.01360 <.0001	
usd_goal_real usd_goal_real	0.00519 0.0030	0.02160 <.0001	0.01534 <.0001	0.01341 <.0001	0.00632 0.0003	1.00000 <.0001	-0.00417 0.0169	-0.00321 0.0662	
goal_achievement	0.31584 <.0001	-0.01882 0.0001	-0.01521 <.0001	0.03123 <.0001	0.26600 <.0001	-0.00417 0.0169	1.00000 <.0001	0.01274 <.0001	
name_count name_count	0.01044 <.0001	0.00376 0.0311	-0.03315 <.0001	0.04902 <.0001	0.01360 <.0001	-0.00321 0.0662	0.01274 <.0001	1.00000 <.0001	

The number of backers is obviously highly correlated pledged amount.

2) Hypothesis Test

a) T-test

To determine whether there is enough evidence to support our claims of models, we use t test to prove it. In this report, the significant level ($=\alpha$) is equal to 0.05. Only when p-value is less than α , we say that the estimate is statistically significant.

i) Goal Amount

Difference of goal based on state(success or fail)

The TTEST Procedure

Variable: usd_goal_real (usd_goal_real)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	63496.4	1440054	3248.5	10.7300	1.6636E8
successful	132080	9651.5	28128.1	77.3966	10.6900	2015609
Diff (1-2)		53844.9	1113787	3962.9		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		63496.4	57129.4	69863.3	1440054
successful		9651.5	9499.8	9803.2	28128.1
Diff (1-2)	Pooled	53844.9	46077.7	61612.1	1113787
Diff (1-2)	Satterthwaite	53844.9	47476.1	60213.7	1111101

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	13.59	<.0001
Satterthwaite	Unequal	196737	16.57	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	196514	132079	2621.06	<.0001

The goal is to compare the means of goal amount by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of goal amount by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of goal amount by state are equal

Alternative hypothesis (H_1): At least one of the variances of goal amount by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of goal amount by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The goal amount has equal means by state

Alternative hypothesis (H_1): At least one of goal amount means by state is different

ii) Pledged Amount

Difference of pledged based on state(success or fail)

The TTEST Procedure

Variable: usd_pledged_real (usd_pledged_real)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	1327.1	6839.2	15.4279	0	757353
successful	132080	22916.6	151910	418.0	0	20338986
Diff (1-2)		-21589.5	96455.8	343.2		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		1327.1	1296.9	1357.4	6839.2 6817.9 6860.7
successful		22916.6	22097.4	23735.9	151910 151333 152492
Diff (1-2)	Pooled	-21589.5	-22262.2	-20916.8	96455.8 96223.2 96689.6
Diff (1-2)	Satterthwaite	-21589.5	-22409.3	-20769.7	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	-62.91	<.0001
Satterthwaite	Unequal	132439	-51.62	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	132079	196514	493.36	<.0001

The goal is to compare the means of pledged amount by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of Pledged amount by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of pledged amount by state are equal

Alternative hypothesis (H_1): At least one of the variances of pledged amount by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of Pledged amount by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The pledged amount has equal means by state

Alternative hypothesis (H_1): At least one of pledged amount means by state is different

iii) Duration

Difference of duration based on state(success or fail)

The TTEST Procedure

Variable: duration (duration)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	35.1925	13.1734	0.0297	1.0000	92.0000
successful	132080	32.2213	11.7287	0.0323	1.0000	92.0000
Diff (1-2)		2.9711	12.6126	0.0449		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		35.1925	35.1342	35.2507	13.1734 13.1324 13.2147
successful		32.2213	32.1581	32.2846	11.7287 11.6841 11.7736
Diff (1-2)	Pooled	2.9711	2.8832	3.0591	12.6126 12.5822 12.6432
Diff (1-2)	Satterthwaite	2.9711	2.8852	3.0571	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	66.21	<.0001
Satterthwaite	Unequal	304081	67.73	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	196514	132079	1.26	<.0001

The goal is to compare the means of duration by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of duration by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of duration by state are equal

Alternative hypothesis (H_1): At least one of the variances of duration by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of duration by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The duration has equal means by state

Alternative hypothesis (H_1): At least one of duration means by state is different

iv) Competitors

Difference of Competitors based on state(success or fail)

The TTEST Procedure

Variable: Competitors (Competitors)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	102.0	62.5955	0.1412	0	728.0
successful	132080	96.0354	55.0858	0.1516	0	728.0
Diff (1-2)		5.9405	59.6906	0.2124		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		102.0	101.7	102.3	62.4004 62.7918
successful		96.0354	95.7383	96.3325	55.0858 54.8765 55.2967
Diff (1-2)	Pooled	5.9405	5.5242	6.3568	59.6906 59.5467 59.8353
Diff (1-2)	Satterthwaite	5.9405	5.5345	6.3465	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	27.97	<.0001
Satterthwaite	Unequal	305938	28.68	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	196514	132079	1.29	<.0001

The goal is to compare the means of competitors by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of competitors by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of competitors by state are equal

Alternative hypothesis (H_1): At least one of the variances of competitors by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of competitors by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The competitors has equal means by state

Alternative hypothesis (H_1): At least one of competitors means by state is different

v) Name Counts (The number of words in project name)

Difference of Name Counts based on state (success or fail)

The TTEST Procedure

Variable: name_count (name_count)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	5.3193	4.4645	0.0101	1.0000	1163.0
successful	132080	5.9511	6.5606	0.0181	1.0000	1264.0
Diff (1-2)		-0.6317	5.4056	0.0192		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		5.3193	5.2996	5.3391	4.4645
successful		5.9511	5.9157	5.9864	6.5606
Diff (1-2)	Pooled	-0.6317	-0.6694	-0.5940	5.4056
Diff (1-2)	Satterthwaite	-0.6317	-0.6722	-0.5912	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	-32.85	<.0001
Satterthwaite	Unequal	213208	-30.56	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	132079	196514	2.16	<.0001

The goal is to compare the means of name counts by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of name counts by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of name counts by state are equal

Alternative hypothesis (H_1): At least one of the variances of name counts by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of name counts by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The name counts has equal means by state

Alternative hypothesis (H_1): At least one of name counts means by state is different

vi) Backers

Difference of backers based on state(success or fail)

The TTEST Procedure

Variable: backers (backers)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	16.4987	71.4174	0.1611	0	6550.0
successful	132080	266.6	1514.2	4.1664	1.0000	219382
Diff (1-2)		-250.1	961.6	3.4214		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		16.4987	16.1829	16.8145	71.4174
successful		266.6	258.4	274.8	1514.2
Diff (1-2)	Pooled	-250.1	-256.8	-243.4	961.6
Diff (1-2)	Satterthwaite	-250.1	-258.3	-241.9	959.3
					963.9

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	-73.10	<.0001
Satterthwaite	Unequal	132474	-59.98	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	132079	196514	449.53	<.0001

The goal is to compare the means of backers by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of backers by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of backers by state are equal

Alternative hypothesis (H_1): At least one of the variances of backers by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of backers by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The backers has equal means by state

Alternative hypothesis (H_1): At least one of backers means by state is different

vii) Sentiment

Difference of sentiment based on state(success or fail)

The TTEST Procedure

Variable: sentiment (sentiment)

state	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed	196515	0.0711	0.6181	0.00139	-10.0000	6.0000
successful	132080	0.0394	0.6303	0.00173	-12.0000	6.0000
Diff (1-2)		0.0317	0.6230	0.00222		

state	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
failed		0.0711	0.0684	0.0739	0.6181 0.6162 0.6201
successful		0.0394	0.0360	0.0428	0.6303 0.6279 0.6327
Diff (1-2)	Pooled	0.0317	0.0274	0.0361	0.6230 0.6215 0.6246
Diff (1-2)	Satterthwaite	0.0317	0.0274	0.0361	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	328593	14.31	<.0001
Satterthwaite	Unequal	279539	14.25	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	132079	196514	1.04	<.0001

The goal is to compare the means of sentiment by state. First, we need to know if the variances of two distribution are equal or not equal. First test is Equality of Variance test. In this test the null hypothesis is variance of sentiment by state are equal. In corresponding result, we can see the result that the p-value is less than 0.0001 which is less than 0.05. we conclude that we can reject the null hypothesis, the variances are not equal.

Null hypothesis (H_0): The variances of sentiment by state are equal

Alternative hypothesis (H_1): At least one of the sentiment of backers by state is not equal

And now we move to check the equality of mean. Since the variance are not equal, we check the result corresponding the result of Satterthwaite method. In this case the null hypothesis is mean of sentiment by state are equal. The p-value is significant, and its values is less than 0.001. Therefore, we can conclude that we can reject the null hypothesis.

Null hypothesis (H_0): The sentiment has equal means by state

Alternative hypothesis (H_1): At least one of sentiment means by state is different

b) ANOVA (For unbalanced data)

We use ANOVA to compare means of more than two levels of variables.

i) Goal amount across category

Difference of goal based on main category																	
The GLM Procedure																	
Dependent Variable: usd_goal_real usd_goal_real																	
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F												
Model	14	2.5047587E14	1.7891134E13	14.42	<.0001												
Error	328580	4.0760505E17	1.2405047E12														
Corrected Total	328594	4.0785553E17															
<table border="1"> <thead> <tr> <th>R-Square</th><th>Coeff Var</th><th>Root MSE</th><th>usd_goal_real Mean</th><th></th><th></th></tr> </thead> <tbody> <tr> <td>0.000614</td><td>2661.157</td><td>1113779</td><td>41853.21</td><td></td><td></td></tr> </tbody> </table>						R-Square	Coeff Var	Root MSE	usd_goal_real Mean			0.000614	2661.157	1113779	41853.21		
R-Square	Coeff Var	Root MSE	usd_goal_real Mean														
0.000614	2661.157	1113779	41853.21														
Source	DF	Type I SS	Mean Square	F Value	Pr > F												
main_category	14	2.5047587E14	1.7891134E13	14.42	<.0001												
Source	DF	Type III SS	Mean Square	F Value	Pr > F												
main_category	14	2.5047587E14	1.7891134E13	14.42	<.0001												

Level of main_category	N	usd_goal_real	
		Mean	Std Dev
Art	25473	39670.3712	1371416.12
Comics	9761	20426.1212	1022449.26
Crafts	7770	10003.6596	122492.86
Dance	3565	8345.3832	37311.91
Design	25288	38326.3441	698362.57
Fashion	19526	22151.1741	735580.56
Film & Video	56274	76383.5845	1706997.56
Food	21596	49093.9855	1366102.71
Games	27941	39737.1987	882108.30
Journalism	4054	71724.4021	1725773.00
Music	45605	14489.2003	341509.44
Photography	9486	11328.9170	106364.45
Publishing	35244	17628.2131	589580.44
Technology	26799	98795.1853	1598647.72
Theater	10213	23185.5255	467161.92

Null hypothesis (H_0): Goal amount has significantly equal means across category

Alternative hypothesis (H_1): At least one of them is different

The p-value for this is below than 0.0001 and it is significant (95% of significant level). Therefore, we can reject the null hypothesis. We can conclude that means of goal amount across categories are different from each other.

ii) Pledged amount across category

Difference of pledged based on main category																	
The GLM Procedure																	
Dependent Variable: usd_pledged_real usd_pledged_real																	
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F												
Model	14	2.7011829E13	1.9294164E12	206.71	<.0001												
Error	328580	3.0669432E15	9333931432.9														
Corrected Total	328594	3.093955E15															
<table border="1"> <thead> <tr> <th>R-Square</th><th>Coeff Var</th><th>Root MSE</th><th>usd_pledged_real Mean</th><th></th><th></th></tr> </thead> <tbody> <tr> <td>0.008731</td><td>965.6307</td><td>96612.27</td><td>10005.10</td><td></td><td></td></tr> </tbody> </table>						R-Square	Coeff Var	Root MSE	usd_pledged_real Mean			0.008731	965.6307	96612.27	10005.10		
R-Square	Coeff Var	Root MSE	usd_pledged_real Mean														
0.008731	965.6307	96612.27	10005.10														
Source	DF	Type I SS	Mean Square	F Value	Pr > F												
main_category	14	2.7011829E13	1.9294164E12	206.71	<.0001												
Source	DF	Type III SS	Mean Square	F Value	Pr > F												
main_category	14	2.7011829E13	1.9294164E12	206.71	<.0001												

Level of main_category	N	usd_pledged_real	
		Mean	Std Dev
Art	25473	3493.6391	22920.045
Comics	9761	7219.8569	25608.680
Crafts	7770	1791.4920	8529.842
Dance	3565	3615.0462	5869.541
Design	25288	27857.8568	233960.913
Fashion	19526	6404.8566	32137.301
Film & Video	56274	6735.4402	43870.831
Food	21596	5673.5409	33204.182
Games	27941	25537.1208	188827.652
Journalism	4054	3005.2755	11225.341
Music	45605	4164.6595	13550.643
Photography	9486	3966.8465	20428.196
Publishing	35244	3695.8487	15623.602
Technology	26799	24035.4727	135389.220
Theater	10213	4171.6988	10639.997

Null hypothesis (H_0): Pledged amount has significantly equal means across category
Alternative hypothesis (H_1): At least one of them is different

The p-value for this is below than 0.0001 and it is significant (95% of significant level). Therefore, we can reject the null hypothesis. We can conclude that means of pledged amount across categories are different from each other.

iii) Goal achievement across launched quarter and day type

Difference of goal_achievement based on launced quater and day type

The GLM Procedure					
Dependent Variable: goal_achievement					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	3186.78	455.25	10.02	<.0001
Error	328587	14935341.22	45.45		
Corrected Total	328594	14938528.00			
<hr/>					
R-Square	Coeff Var	Root MSE	goal_achievement Mean		
0.000213	683.0383	6.741901	0.987046		
<hr/>					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Quarter	3	225.892372	75.297457	1.66	0.1740
launched_day_type	1	2803.208182	2803.208182	61.67	<.0001
Quarter*launched_day	3	157.680090	52.560030	1.16	0.3248
<hr/>					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Quarter	3	301.571538	100.523846	2.21	0.0845
launched_day_type	1	2756.623129	2756.623129	60.65	<.0001
Quarter*launched_day	3	157.680090	52.560030	1.16	0.3248

Level of Quarter	Level of launched_day_type	N	goal_achievement	
			Mean	Std Dev
1	Weekday	55113	1.06439691	5.89051970
1	Weekend	25015	0.92139606	7.33174783
2	Weekday	60755	1.01062393	6.81079297
2	Weekend	25302	0.79524906	2.83665343
3	Weekday	61663	1.06883786	9.02433696
3	Weekend	25048	0.79839476	3.08019800
4	Weekday	53812	1.04406442	7.17961027
4	Weekend	21887	0.86884890	5.33775698

Null hypothesis (H_0): Goal achievement has significantly equal means across launched day type and quarter

Alternative hypothesis (H_1): At least one of them is different

Goal achievement means ratio of pledged and goal. If it is over 1, it means it achieve its goal. Otherwise, it did not achieve its goal.

The p-value for this is below than 0.0001 and it is significant (95% of significant level). Therefore, we can reject the null hypothesis. We can conclude that means of goal achievement across launched quarter and day type are different from each other.

It seems that weekend have less mean of goal achievement than weekday.

iv) Goal achievement across sub region

Difference of goal_achievement based on sub region						
The GLM Procedure						
Dependent Variable: goal_achievement						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	7	1371.31	195.90	4.31	<.0001	
Error	328587	14937156.69	45.46			
Corrected Total	328594	14938528.00				
R-Square	Coeff Var	Root MSE	goal_achievement Mean			
0.000092	683.0799	6.742311	0.987046			
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
sub_region	7	1371.312287	195.901755	4.31	<.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
sub_region	7	1371.312287	195.901755	4.31	<.0001	

Level of sub_region	N	goal_achievement	
		Mean	Std Dev
Australia and New Zealand	7780	0.96555792	5.6491886
Eastern Asia	498	2.01642581	8.7094125
Latin America and the Caribbean	1402	0.52205890	2.1115426
Northern America	271349	0.99321862	6.8074634
Northern Europe	32900	0.99125036	5.5344035
South-eastern Asia	447	1.37462527	4.0949832
Southern Europe	4209	0.66022761	2.5540764
Western Europe	10010	0.95662243	10.1327734

Null hypothesis (H_0): Goal achievement has significantly equal means across sub region

Alternative hypothesis (H_1): At least one of them is different.

The p-value for this is below than 0.0001 and it is significant (95% of significant level). Therefore, we can reject the null hypothesis. We can conclude that means of goal achievement across sub region are different from each other.

It seems that eastern Asia and South-eastern Asia have higher means of goal achievement.

4. Model

Now, we want to know about important factors to achieve the goal amount. Thus, we made 16 logit models: one for the whole observations and 15 for each category. The reason why we made the models for each category is that we cannot control the category—a client will already have one in their mind. Also, based on our previous hypothesis testing results, each category has significantly different range of goal amount and pledged amount of money from one another. First, we will examine the tendency of successful projects through the total model using the whole observations and interpret about the important factors of each category. Furthermore, we will check each model's performance about how accurate our model is. In this case, sensitivity of our model is important because we want to know how well our model catches success of projects.

1) Total model

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-6.4297	0.0575	12506.2914	<.0001
All_goal_level	L1	1	5.2314	0.0296	31321.5830	<.0001
All_goal_level	L2	1	3.8890	0.0259	22581.0094	<.0001
All_goal_level	L3	1	2.0977	0.0215	9561.2311	<.0001
All_goal_level	L4	1	-1.3869	0.0283	2400.3203	<.0001
main_category	Art	1	0.4710	0.0476	98.0286	<.0001
main_category	Comics	1	-0.00431	0.0964	0.0020	0.9643
main_category	Crafts	1	-0.2525	0.0671	14.1437	0.0002
main_category	Dance	1	1.0623	0.1203	77.9690	<.0001
main_category	Design	1	-0.7776	0.0747	108.2827	<.0001
main_category	Fashion	1	-0.2926	0.0679	18.5570	<.0001
main_category	Film & Video	1	0.4793	0.0521	84.6989	<.0001
main_category	Food	1	-0.00544	0.0424	0.0165	0.8978
main_category	Games	1	-1.2056	0.0862	195.5560	<.0001
main_category	Journalism	1	-0.1028	0.0949	1.1725	0.2789
main_category	Music	1	0.4013	0.0501	64.2291	<.0001
main_category	Photography	1	0.1388	0.0596	5.4299	0.0198
main_category	Publishing	1	0.0124	0.0562	0.0485	0.8257
main_category	Technology	1	-0.7957	0.0773	105.8618	<.0001
backers		1	0.0509	0.000253	40372.2576	<.0001
duration		1	-0.0112	0.000520	467.1667	<.0001
Competitors		1	-0.00484	0.000888	29.7591	<.0001

1) Total model – Cont'd

Competito*main_categ	Art	1	0.000245	0.00103	0.0565	0.8120	
Competito*main_categ	Comics	1	0.00110	0.00311	0.1264	0.7222	
Competito*main_categ	Crafts	1	-0.00293	0.00173	2.8692	0.0903	
Competito*main_categ	Dance	1	-0.0144	0.0105	1.8694	0.1715	
Competito*main_categ	Design	1	0.00673	0.00117	33.1984	<.0001	
Competito*main_categ	Fashion	1	0.00216	0.00126	2.9458	0.0861	
Competito*main_categ	Film & Video	1	0.00384	0.000937	16.8307	<.0001	
Competito*main_categ	Food	1	0.00285	0.000925	9.4885	0.0021	
Competito*main_categ	Games	1	0.00433	0.00121	12.7440	0.0004	
Competito*main_categ	Journalism	1	-0.0123	0.00447	7.5900	0.0059	
Competito*main_categ	Music	1	0.00476	0.000953	24.9228	<.0001	
Competito*main_categ	Photography	1	-0.00768	0.00185	17.2367	<.0001	
Competito*main_categ	Publishing	1	0.00237	0.00101	5.5384	0.0186	
Competito*main_categ	Technology	1	0.00346	0.00107	10.5272	0.0012	
Avg_Amount_Pledged_B		1	0.0147	0.000111	17345.7775	<.0001	
sub_region	Australia and New Zealand	1	-0.0503	0.0535	0.8846	0.3469	
sub_region	Eastern Asia	1	0.2447	0.1544	2.5132	0.1129	
sub_region	Latin America and the Caribbean	1	-0.1774	0.0880	4.0618	0.0439	
sub_region	Northern America	1	0.3184	0.0369	74.3240	<.0001	
sub_region	Northern Europe	1	0.1754	0.0403	18.9822	<.0001	
sub_region	South-eastern Asia	1	-0.2933	0.1669	3.0865	0.0789	
sub_region	Southern Europe	1	-0.2516	0.0720	12.2055	0.0005	
sentiment		1	-0.0115	0.0108	1.1361	0.2865	
name_count		1	0.00853	0.00170	25.0672	<.0001	
launched_day_type	Weekday	1	0.00236	0.00707	0.1111	0.7389	
Quarter	1	1	0.0463	0.0122	14.2955	0.0002	
Quarter	2	1	0.00703	0.0121	0.3385	0.5607	
Quarter	3	1	-0.0451	0.0122	13.6641	0.0002	
launched_day*Quarter	Weekday	1	1	0.0133	0.0122	1.1860	0.2761
launched_day*Quarter	Weekday	2	1	0.00640	0.0120	0.2843	0.5939
launched_day*Quarter	Weekday	3	1	-0.00126	0.0121	0.0109	0.9170

1) Total model – Cont'd

Inference

From the above analysis, we can see that the lower the goal level, the higher the log ratio of odds (= the ratio) to achieve the goal amount is on average. Also, in Main category, Film & Video seems to have higher ratio than all other categories. On the other hand, Game seems to have lower ratio than others.

Duration seems to have a negative impact on the log ratio of odds to achieve the goal.

It can be assumed naturally, but still, the more Backers, the ratio to achieve the goal will increase, the more Competitors, the ratio to achieve the goal will decrease.

We estimated the interaction between Competitors and Main Category of a Kickstarter. Interestingly, Design is the only Category that has a positive impact on the ratio to achieve the goal with additional Competitors.

In general, launching a Kickstarter Project in Weekday and the 1st Quarter will increase the log ratio of odds to achieve the goal. In location perspective, Northern America seems to be a better place to launch a Kickstarter project to achieve the goal amount rather than South Eastern Asia.

We will further take a look at each category's model.

Prediction

Confusion Matrix for train

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	F_state(From: state)	I_state(Into: state)		
		failed	successful	Total
failed	149237	8158	157395	
	56.77	3.10	59.87	
	94.82	5.18		
	94.54	7.77		
	successful	8617	96864	105481
		3.28	36.85	40.13
		8.17	91.83	
		5.46	92.23	
	Total	157854	105022	262876
		60.05	39.95	100.00

Confusion Matrix for testset

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	F_state(From: state)	I_state(Into: state)		
		failed	successful	Total
failed	37147	1972	39119	
	56.52	3.00	59.53	
	94.96	5.04		
	94.28	7.49		
	successful	2254	24345	26599
		3.43	37.04	40.47
		8.47	91.53	
		5.72	92.51	
	Total	39401	26317	65718
		59.95	40.05	100.00

- Train Sensitivity: 92.23%, Test sensitivity: 92.51%
- Total Train Accuracy: 93.62%, Total Test Accuracy: 93.62%

2) Film & Videos

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-5.2179	6.2274	0.7021	0.4021	
Level_USD_Goal	L1	1	3.4380	0.0850	1634.8905	<.0001	
Level_USD_Goal	L2	1	1.8884	0.0861	480.5579	<.0001	
Level_USD_Goal	L3	1	-0.2049	0.1003	4.1764	0.0410	
backers		1	0.0587	0.000686	7324.2351	<.0001	
duration		1	-0.0177	0.00184	93.1423	<.0001	
Competitors		1	-0.00149	0.000351	18.0525	<.0001	
Avg_Amount_Pledged_B		1	0.00971	0.000200	2366.6595	<.0001	
sub_region	Australia and New Zealand	1	1.1832	6.2273	0.0361	0.8493	
sub_region	Eastern Asia	1	-7.8977	43.5725	0.0329	0.8562	
sub_region	Latin America and the Caribbean	1	1.4143	6.2285	0.0516	0.8204	
sub_region	Northern America	1	1.4804	6.2265	0.0565	0.8121	
sub_region	Northern Europe	1	1.5035	6.2266	0.0583	0.8092	
sub_region	South-eastern Asia	1	0.4616	6.3108	0.0053	0.9417	
sub_region	Southern Europe	1	0.6921	6.2292	0.0123	0.9115	
sentiment		1	0.0175	0.0259	0.4531	0.5009	
name_count		1	-0.00601	0.00459	1.7140	0.1905	
launched_day_type	Weekday	1	0.0153	0.0161	0.8982	0.3433	
Quarter	1	1	0.1016	0.0276	13.6011	0.0002	
Quarter	2	1	-0.0580	0.0276	4.4337	0.0352	
Quarter	3	1	-0.0794	0.0284	7.8357	0.0051	
launched_day*Quarter	Weekday	1	1	-0.0207	0.0272	0.5794	0.4466
launched_day*Quarter	Weekday	2	1	0.0388	0.0272	2.0271	0.1545
launched_day*Quarter	Weekday	3	1	0.00578	0.0282	0.0420	0.8377
duration*Level_USD_G	L1	1	0.000724	0.00214	0.1146	0.7349	
duration*Level_USD_G	L2	1	0.00525	0.00225	5.4367	0.0197	
duration*Level_USD_G	L3	1	0.00914	0.00266	11.8240	0.0006	

2) Film & Videos – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 3.44 than L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.02.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.001.
- Sub Region: Although All sub region is not significant at 5% significance level, if a campaign for Kickstarter launches in Northern Europe, the log ratio of odds to achieve goal amount will decrease by 1.50 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Name Count: Although it is not significant at 5% significance level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.001.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount increases by 0.02 compared to during the weekends.
- Quarter: If a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount increases by 0.1 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount will increase by 0.01 compared to L4. This is the largest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Film&Video for train

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	F_state(From: state)	I_state(Into: state)		
		failed	successful	Total
	failed	24944 55.41 94.83 95.13	1361 3.02 5.17 7.24	26305 58.43
	successful	1276 2.83 6.82 4.87	17439 38.74 93.18 92.76	18715 41.57
	Total	26220 58.24	18800 41.76	45020 100.00

Confusion Matrix of Film&Video for test

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	F_state(From: state)	I_state(Into: state)		
		failed	successful	Total
	failed	6164 54.77 94.83 94.90	336 2.99 5.17 7.06	6500 57.76
	successful	331 2.94 6.96 5.10	4423 39.30 93.04 92.94	4754 42.24
	Total	6495 57.71	4759 42.29	11254 100.00

- Train Sensitivity: 92.76%, Test sensitivity: 92.94%
- Total Train Accuracy: 94.14%, Total Test Accuracy: 94.07%

3) Games

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-3.6117	0.1773	414.7738	<.0001	
Level_USD_Goal	L1	1	2.7942	0.1303	459.7822	<.0001	
Level_USD_Goal	L2	1	1.4031	0.1384	102.8109	<.0001	
Level_USD_Goal	L3	1	-0.00092	0.1622	0.0000	0.9955	
backers		1	0.0142	0.000250	3225.5291	<.0001	
duration		1	-0.0130	0.00320	16.5648	<.0001	
Competitors		1	-0.00020	0.000634	0.0987	0.7534	
Avg_Amount_Pledged_B		1	0.0125	0.000421	876.9975	<.0001	
sub_region	Australia and New Zealand	1	-0.0798	0.1388	0.3306	0.5653	
sub_region	Eastern Asia	1	0.4659	0.3644	1.6348	0.2010	
sub_region	Latin America and the Caribbean	1	-0.2072	0.2894	0.5128	0.4739	
sub_region	Northern America	1	0.2302	0.0981	5.5095	0.0189	
sub_region	Northern Europe	1	0.1848	0.1068	2.9954	0.0835	
sub_region	South-eastern Asia	1	-0.4577	0.4647	0.9699	0.3247	
sub_region	Southern Europe	1	-0.0692	0.1457	0.2254	0.6349	
sentiment		1	-0.0776	0.0316	6.0262	0.0141	
name_count		1	0.0921	0.00770	143.2226	<.0001	
launched_day_type	Weekday	1	-0.0344	0.0220	2.4397	0.1183	
Quarter	1	1	0.0249	0.0387	0.4148	0.5195	
Quarter	2	1	0.0155	0.0377	0.1685	0.6814	
Quarter	3	1	0.0409	0.0375	1.1931	0.2747	
launched_day*Quarter	Weekday	1	1	-0.0376	0.0387	0.9428	0.3316
launched_day*Quarter	Weekday	2	1	0.0147	0.0377	0.1522	0.6965
launched_day*Quarter	Weekday	3	1	-0.00316	0.0375	0.0071	0.9329
duration*Level_USD_G	L1	1	-0.00830	0.00373	4.9449	0.0262	
duration*Level_USD_G	L2	1	-0.00074	0.00403	0.0338	0.8540	
duration*Level_USD_G	L3	1	-0.00065	0.00479	0.0185	0.8917	

3) Games – Cont'd

Inference

- Level-USD-Goal: Although not all levels are significant at 5% significance level, if the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 2.79 than L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.01.
- Competitors: Although it is not significant at 5% significance level, if the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.0002.
- Sub Region: Although not all levels are significant at 5% significance level, if a campaign for Kickstarter launches in Eastern Asia, the log ratio of odds to achieve goal amount will increase by 0.47 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: If sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.07.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.09.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will decrease by 0.03 compared to during the weekends.
- Quarter: Although all levels are not significant at 5% significance level, if a project launches during the 3rd quarter, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: Although all levels are not significant at 5% significance level, if the duration increases by 1 day when Level_USD_Goal is L1, the log ratio of odds to achieve the goal amount will decrease by 0.008 compared to L4. (L1 is the largest decreasing.)

Prediction

Confusion Matrix of games for train

Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	F_state(From: state)	I_state(Into: state)		
		failed	successful	Total
failed	12028 53.81 94.45 90.16	707 3.16 5.55 7.84		12735 56.97
successful	1312 5.87 13.64 9.84	8306 37.16 86.36 92.16		9618 43.03
Total	13340 59.68	9013 40.32		22353 100.00

Confusion Matrix of games for testset

Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	F_state(From: state)	I_state(Into: state)		
		failed	successful	Total
failed	2924 52.33 93.75 89.75		195 3.49 6.25 8.37	3119 55.82
successful	334 5.98 13.53 10.25		2135 38.21 86.47 91.63	2469 44.18
Total	3258 58.30		2330 41.70	5588 100.00

- Train Sensitivity: 92.16%, Test sensitivity: 92.63%
- Total Train Accuracy: 91.00%, Total Test Accuracy: 90.53%

4) Design

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.8699	0.1693	522.3105	<.0001
Level_USD_Goal	L1	1	2.8848	0.1362	448.7070	<.0001
Level_USD_Goal	L2	1	0.9717	0.1640	35.0978	<.0001
Level_USD_Goal	L3	1	-0.7293	0.1799	16.4361	<.0001
backers		1	0.0290	0.000533	2964.6141	<.0001
duration		1	-0.0226	0.00311	52.8313	<.0001
Competitors		1	0.000728	0.000698	1.0887	0.2968
Avg_Amount_Pledged_B		1	0.0102	0.000280	1328.7234	<.0001
sub_region	Australia and New Zealand	1	-0.0362	0.1296	0.0779	0.7802
sub_region	Eastern Asia	1	0.4422	0.2220	3.9682	0.0464
sub_region	Latin America and the Caribbean	1	0.0354	0.2748	0.0166	0.8975
sub_region	Northern America	1	0.0433	0.0756	0.3274	0.5672
sub_region	Northern Europe	1	0.1482	0.0908	2.6658	0.1025
sub_region	South-eastern Asia	1	-0.1848	0.2855	0.4189	0.5175
sub_region	Southern Europe	1	-0.5077	0.1832	7.6780	0.0056
sentiment		1	-0.0225	0.0364	0.3828	0.5361
name_count		1	0.0298	0.00912	10.7154	0.0011
launched_day_type	Weekday	1	0.0418	0.0269	2.4165	0.1201
Quarter	1	1	-0.00721	0.0474	0.0232	0.8790
Quarter	2	1	-0.0389	0.0459	0.7184	0.3967
Quarter	3	1	-0.0135	0.0463	0.0850	0.7706
launched_day*Quarter	Weekday	1	0.00806	0.0471	0.0292	0.8643
launched_day*Quarter	Weekday	2	0.0705	0.0459	2.3605	0.1244
launched_day*Quarter	Weekday	3	-0.0101	0.0463	0.0481	0.8264
duration*Level_USD_G	L1	1	0.00273	0.00379	0.5163	0.4724
duration*Level_USD_G	L2	1	0.0102	0.00469	4.6903	0.0303
duration*Level_USD_G	L3	1	0.0132	0.00505	6.8659	0.0088

4) Design – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 2.88 compared to L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.02.
- Competitors: Although it is not significant at 5% significance level, if the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.00007.
- Sub Region: Although not all levels are significant at 5% significance level, if a campaign for Kickstarter launches in Eastern Asia, the log ratio of odds to achieve goal amount will increase by 0.44 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: If sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.02.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.03.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to during the weekends.
- Quarter: Although all levels are not significant at 5% significance level, if a project launches during the 2nd quarter, the log ratio of odds to achieve the goal amount will decrease by 0.04 compared to 4th quarter. This is the highest decreasing amount across the quarters.
- Duration*Level_USD_Goal: Although not all levels are significant at 5% significance level, if the duration increases by 1 day when Level_USD_Goal is L3 the log ratio of odds to achieve the goal amount will increase by 0.01 compared to L4. L3 is the largest decreasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Design for train

		Table of F_state by I_state		
		I_state(Into: state)		
		failed	successful	Total
F_state(From: state)	failed	11142 55.07 94.08 89.57	701 3.46 5.92 9.00	11843 58.54
	successful	1298 6.42 15.47 10.43	7090 35.05 84.53 91.00	8388 41.46
	Total	12440 61.49	7791 38.51	20231 100.00

Confusion Matrix of Design for testset

		Table of F_state by I_state		
		I_state(Into: state)		
		failed	successful	Total
F_state(From: state)	failed	2756 54.50 93.58 89.71	189 3.74 6.42 9.52	2945 58.24
	successful	316 6.25 14.96 10.29	1796 35.52 85.04 90.48	2112 41.76
	Total	3072 60.75	1985 39.25	5057 100.00

- Train Sensitivity: 91.00%, Test sensitivity: 90.48%
- Total Train Accuracy: 90.19%, Total Test Accuracy: 90.01%

5) Food

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-6.9627	27.9312	0.0621	0.8031
Level_USD_Goal	L1	1	3.6705	0.1979	343.9768	<.0001
Level_USD_Goal	L2	1	1.5306	0.2200	48.3975	<.0001
Level_USD_Goal	L3	1	-1.0272	0.2481	17.1424	<.0001
backers		1	0.0661	0.00137	2330.6941	<.0001
duration		1	-0.0327	0.00478	46.9324	<.0001
Competitors		1	-0.00074	0.000307	5.7964	0.0161
Avg_Amount_Pledged_B		1	0.0131	0.000438	892.5772	<.0001
sub_region	Australia and New Zealand	1	0.9866	27.9312	0.0012	0.9718
sub_region	Eastern Asia	1	-7.0669	195.5	0.0013	0.9712
sub_region	Latin America and the Caribbean	1	1.4924	27.9332	0.0029	0.9574
sub_region	Northern America	1	1.7663	27.9304	0.0040	0.9496
sub_region	Northern Europe	1	1.2483	27.9306	0.0020	0.9644
sub_region	South-eastern Asia	1	0.5572	27.9553	0.0004	0.9841
sub_region	Southern Europe	1	0.0150	27.9348	0.0000	0.9996
sentiment		1	0.0652	0.0506	1.6583	0.1978
name_count		1	-0.00287	0.0132	0.0475	0.8274
launched_day_type	Weekday	1	-0.0784	0.0348	5.0825	0.0242
Quarter	1	1	0.0467	0.0601	0.6028	0.4375
Quarter	2	1	0.0340	0.0605	0.3169	0.5735
Quarter	3	1	0.0251	0.0609	0.1691	0.6809
launched_day*Quarter	Weekday	1	1 -0.00043	0.0598	0.0001	0.9943
launched_day*Quarter	Weekday	2	1 -0.0503	0.0598	0.7076	0.4002
launched_day*Quarter	Weekday	3	1 0.0306	0.0576	0.2824	0.5951
duration*Level_USD_G	L1	1	0.0142	0.00544	6.8192	0.0090
duration*Level_USD_G	L2	1	0.0162	0.00638	6.4585	0.0110
duration*Level_USD_G	L3	1	0.0169	0.00714	5.6244	0.0177

5) Food – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 3.67 compared to L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.03.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.0007.
- Sub Region: Although all levels are not significant at 5% significance level, if a campaign for Kickstarter launches in Northern America, the log ratio of odds to achieve goal amount will increase by 1.77 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significant level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will increase by 0.07.
- Name Count: Although it is not significant at 5% significant level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.003.
- Launched Day Type: If a project launches during the weekday, the log ratio of odds to achieve the goal amount will increase by 0.08 compared to during the weekends.
- Quarter: Although all levels are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.05 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L3 the log ratio of odds to achieve the goal amount will increase by 0.02 compared to L4. L1 is the largest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Food for train

		Table of F_state by I_state		
		I_state(Into: state)		
		failed	successful	Total
F_state(From: state)	failed	12203 70.63 97.39 95.34	327 1.89 2.61 7.30	12530 72.52
	successful	596 3.45 12.56 4.66	4151 24.03 87.44 92.70	4747 27.48
	Total	12799 74.08	4478 25.92	17277 100.00

Confusion Matrix of Food for testset

The FREQ Procedure

		Table of F_state by I_state		
		I_state(Into: state)		
		failed	successful	Total
F_state(From: state)	failed	3103 71.85 97.70 95.21	73 1.69 2.30 6.89	3176 73.54
	successful	156 3.61 13.65 4.79	987 22.85 86.35 93.11	1143 26.46
	Total	3259 75.46	1060 24.54	4319 100.00

- Train Sensitivity: 92.70%, Test sensitivity: 93.11%
- Total Train Accuracy: 94.66%, Total Test Accuracy: 94.70%

6) Fashion

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-6.1309	0.2834	468.0765	<.0001
Level_USD_Goal	L1	1	4.1681	0.2207	356.8165	<.0001
Level_USD_Goal	L2	1	1.4548	0.2335	38.8116	<.0001
Level_USD_Goal	L3	1	-1.0139	0.3121	10.5534	0.0012
backers		1	0.0971	0.00212	2102.6118	<.0001
duration		1	-0.0234	0.00524	19.9464	<.0001
Competitors		1	-0.00267	0.00123	4.7296	0.0296
Avg_Amount_Pledged_B		1	0.0171	0.000536	1019.6646	<.0001
sub_region	Australia and New Zealand	1	-0.3028	0.2173	1.9428	0.1634
sub_region	Eastern Asia	1	0.4957	0.5686	0.7599	0.3833
sub_region	Latin America and the Caribbean	1	-0.2806	0.5795	0.2344	0.6282
sub_region	Northern America	1	0.0292	0.1422	0.0422	0.8373
sub_region	Northern Europe	1	0.0666	0.1608	0.1714	0.6789
sub_region	South-eastern Asia	1	-0.3080	0.4282	0.5173	0.4720
sub_region	Southern Europe	1	-0.0717	0.2875	0.0622	0.8031
sentiment		1	-0.0196	0.0527	0.1385	0.7098
name_count		1	0.0172	0.0141	1.4948	0.2215
launched_day_type	Weekday	1	0.0305	0.0376	0.6570	0.4176
Quarter	1	1	0.0269	0.0667	0.1629	0.6865
Quarter	2	1	-0.0564	0.0668	0.7120	0.3988
Quarter	3	1	-0.0780	0.0631	1.5298	0.2161
launched_day*Quarter	Weekday	1	0.0638	0.0665	0.9216	0.3370
launched_day*Quarter	Weekday	2	0.00521	0.0668	0.0061	0.9378
launched_day*Quarter	Weekday	3	-0.00132	0.0629	0.0004	0.9832
duration*Level_USD_G	L1	1	0.00624	0.00608	1.0541	0.3046
duration*Level_USD_G	L2	1	0.0216	0.00685	9.9324	0.0016
duration*Level_USD_G	L3	1	0.0114	0.00919	1.5435	0.2141

6) Fashion – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 4.17 than L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount decreases by 0.02.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.003
- Sub Region: Although all levels are not significant at 5% significance level, if a campaign for Kickstarter launches in Eastern Asia, the log ratio of odds to achieve goal amount will increase by 0.5 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significant level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.02.
- Name Count: Although it is not significant at 5% significant level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.02.
- Launched Day Type: Although it is not significant at 5% significant level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount increases by 0.03 compared to during the weekends.
- Quarter: Although all levels are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount increases by 0.03 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: Although not all levels are significant at 5% significance level, if the duration increases by 1 day when Level_USD_Goal is L2, the log ratio of odds to achieve the goal amount increase by 0.02 compared to L4.

Prediction

Confusion Matrix of Fashion for train

		Table of F_state by I_state		
		I_state(Into: state)		
F_state(From: state)		failed	successful	Total
failed	10940	290	11230	
	70.03	1.86	71.89	
	97.42	2.58		
	94.97	7.07		
successful	580	3811	4391	
	3.71	24.40	28.11	
	13.21	86.79		
	5.03	92.93		
Total		11520	4101	15621
		73.75	26.25	100.00

Confusion Matrix of Fashion for testset

		Table of F_state by I_state		
		I_state(Into: state)		
F_state(From: state)		failed	successful	Total
failed	2756		78	2834
	70.58		2.00	72.57
	97.25		2.75	
	95.73		7.60	
successful	123		948	1071
	3.15		24.28	27.43
	11.48		88.52	
	4.27		92.40	
Total		2879	1026	3905
		73.73	26.27	100.00

- Train Sensitivity: 92.93%, Test sensitivity: 92.40%
- Total Train Accuracy: 94.43%, Total Test Accuracy: 94.85%

7) Crafts

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-6.4694	0.5090	161.5611	<.0001
Level_USD_Goal	L1	1	4.6992	0.4051	134.5923	<.0001
Level_USD_Goal	L2	1	1.9500	0.4181	21.7542	<.0001
Level_USD_Goal	L3	1	-1.1518	0.5518	4.3571	0.0369
backers		1	0.1604	0.00584	755.6950	<.0001
duration		1	-0.0449	0.0104	18.7549	<.0001
Competitors		1	-0.00329	0.00222	2.1836	0.1395
Avg_Amount_Pledged_B		1	0.0309	0.00152	415.7197	<.0001
sub_region	Australia and New Zealand	1	0.00808	0.3980	0.0004	0.9838
sub_region	Eastern Asia	1	0.4292	1.1832	0.1316	0.7168
sub_region	Latin America and the Caribbean	1	-0.5095	0.8571	0.3534	0.5522
sub_region	Northern America	1	-0.1949	0.2606	0.5596	0.4544
sub_region	Northern Europe	1	-0.3295	0.2891	1.2991	0.2544
sub_region	South-eastern Asia	1	-0.1425	0.6056	0.0554	0.8139
sub_region	Southern Europe	1	0.8191	0.8225	0.9917	0.3193
sentiment		1	-0.0333	0.0965	0.1188	0.7303
name_count		1	0.00669	0.0239	0.0785	0.7794
launched_day_type	Weekday	1	0.0375	0.0619	0.3665	0.5449
Quarter	1	1	0.0825	0.1031	0.6403	0.4236
Quarter	2	1	-0.0613	0.1145	0.2866	0.5924
Quarter	3	1	0.0281	0.1042	0.0730	0.7871
launched_day*Quarter	Weekday	1	0.0123	0.1032	0.0143	0.9049
launched_day*Quarter	Weekday	2	0.0402	0.1131	0.1260	0.7226
launched_day*Quarter	Weekday	3	-0.0760	0.1014	0.5612	0.4538
duration*Level_USD_G	L1	1	0.0301	0.0114	7.0182	0.0081
duration*Level_USD_G	L2	1	0.0409	0.0128	10.2855	0.0013
duration*Level_USD_G	L3	1	0.0298	0.0162	3.3897	0.0656

7) Crafts – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 4.70 compared to L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.05.
- Competitors: Although it is not significant at 5% significant level, if the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.003
- Sub Region: Although all levels are not significant at 5% significance level, if a campaign for Kickstarter launches in Southern Europe, the log ratio of odds to achieve goal amount will increase by 0.82 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significant level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.03.
- Name Count: Although it is not significant at 5% significant level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.007.
- Launched Day Type: Although it is not significant at 5% significant level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to during the weekends.
- Quarter: Although all levels are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.08 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L2, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to L4.

Prediction

Confusion Matrix of Crafts for train

		Table of F_state by I_state		
		I_state(Into: state)		Total
F_state(From: state)		failed	successful	
failed	4421	118	4539	
	71.12	1.90	73.02	
	97.40	2.60		
	95.49	7.44		
successful	209	1468	1677	
	3.36	23.62	26.98	
	12.46	87.54		
	4.51	92.56		
Total		4630	1586	6216
		74.49	25.51	100.00

Confusion Matrix of Crafts for testset

		Table of F_state by I_state		
		I_state(Into: state)		Total
F_state(From: state)		failed	successful	
failed	1107		36	1143
	71.24		2.32	73.55
	96.85		3.15	
	96.43		8.87	
successful	41		370	411
	2.64		23.81	
	9.98		90.02	
	3.57		91.13	
Total		1148	406	1554
		73.87	26.13	100.00

- Train Sensitivity: 92.56%, Test sensitivity: 91.13%
- Total Train Accuracy: 94.74%, Total Test Accuracy: 95.04%

8) Dance

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-8.4797	70.3555	0.0145	0.9041	
Level_USD_Goal	L1	1	5.4100	0.4687	133.2170	<.0001	
Level_USD_Goal	L2	1	2.0871	0.4571	20.8447	<.0001	
Level_USD_Goal	L3	1	-2.4942	0.5928	17.7041	<.0001	
backers		1	0.1733	0.00864	402.0418	<.0001	
duration		1	-0.00031	0.00861	0.0013	0.9712	
Competitors		1	-0.0106	0.0187	0.3225	0.5701	
Avg_Amount_Pledged_B		1	0.0398	0.00261	233.2497	<.0001	
sub_region	Australia and New Zealand	1	1.1359	70.3551	0.0003	0.9871	
sub_region	Latin America and the Caribbean	1	1.8753	70.3631	0.0007	0.9787	
sub_region	Northern America	1	0.8371	70.3533	0.0001	0.9905	
sub_region	Northern Europe	1	1.0008	70.3536	0.0002	0.9887	
sub_region	South-eastern Asia	1	-6.5205	422.1	0.0002	0.9877	
sub_region	Southern Europe	1	0.7087	70.3632	0.0001	0.9920	
sentiment		1	0.1309	0.1750	0.5597	0.4544	
name_count		1	-0.0630	0.0345	3.3454	0.0674	
launched_day_type	Weekday	1	-0.0615	0.0946	0.4234	0.5152	
Quarter	1	1	0.0599	0.1678	0.1273	0.7212	
Quarter	2	1	0.0234	0.1598	0.0214	0.8837	
Quarter	3	1	-0.1013	0.1576	0.4127	0.5206	
launched_day*Quarter	Weekday	1	1	0.1013	0.1676	0.3653	0.5456
launched_day*Quarter	Weekday	2	1	0.0872	0.1573	0.3075	0.5792
launched_day*Quarter	Weekday	3	1	-0.1759	0.1575	1.2480	0.2639
duration*Level_USD_G	L1	1	-0.0142	0.0110	1.6679	0.1965	
duration*Level_USD_G	L2	1	-0.00386	0.0131	0.0870	0.7680	
duration*Level_USD_G	L3	1	0.0513	0.0177	8.4044	0.0037	

8) Dance – Cont'd

Inference - Dance

- Level-USD-Goal: All levels are significant at 5% significance level If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 5.41 compared to L4.
- Duration: Although it is not significant at 5% significance level, if the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.00031
- Competitors: Although it is not significant at 5% significance level, if the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.01
- Sub Region: Although all levels are not significant at 5% significance level, if a campaign for Kickstarter launches in Eastern Asia, the log ratio of odds to achieve goal amount will increase by 1.88 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significant level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will increase by 0.13.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.06.
- Launched Day Type: Although it is not significant at 5% significant level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount decreases by 0.06 compared to during the weekends.
- Quarter: Although it is not significant at 5% significant level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount increases by 0.06 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: Although all levels are not significant at 5% significance level, if the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount increase by 0.05 compared to L4.

predictive

Confusion Matrix of Dance for train

		Table of F_state by I_state		
		I_state(Into: state)		
		failed	successful	Total
F_state(From: state)	failed	934 32.75 93.49 97.80	65 2.28 6.51 3.43	999 35.03
	successful	21 0.74 1.13 2.20	1832 64.24 98.87 96.57	1853 64.97
	Total	955 33.49	1897 66.51	2852 100.00

Confusion Matrix of Dance for testset

		Table of F_state by I_state		
		I_state(Into: state)		
		failed	successful	Total
F_state(From: state)	failed	215 30.20 91.49 97.29	20 2.81 8.51 4.07	235 33.01
	successful	6 0.84 1.26 2.71	471 66.15 98.74 95.93	477 66.99
	Total	221 31.04	491 68.96	712 100.00

- Train Sensitivity: 96.57%, Test sensitivity: 95.93%
- Total Train Accuracy: 96.99%, Total Test Accuracy: 96.21%

9) Journalism

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-6.3389	77.7392	0.0066	0.9350
Level_USD_Goal	L1	1	3.5093	0.3477	101.8679	<.0001
Level_USD_Goal	L2	1	1.9623	0.3563	30.3278	<.0001
Level_USD_Goal	L3	1	-0.2739	0.4171	0.4312	0.5114
backers		1	0.0565	0.00286	389.6784	<.0001
duration		1	-0.0191	0.00727	6.9180	0.0085
Competitors		1	-0.0173	0.00520	11.1081	0.0009
Avg_Amount_Pledged_B		1	0.0190	0.00133	203.6024	<.0001
sub_region	Australia and New Zealand	1	1.3853	77.7390	0.0003	0.9858
sub_region	Eastern Asia	1	-6.9242	466.4	0.0002	0.9882
sub_region	Latin America and the Caribbean	1	-0.4078	77.7454	0.0000	0.9958
sub_region	Northern America	1	1.5909	77.7383	0.0004	0.9837
sub_region	Northern Europe	1	1.4274	77.7385	0.0003	0.9854
sub_region	Southern Europe	1	1.4879	77.7393	0.0004	0.9847
sentiment		1	-0.1897	0.1222	2.4106	0.1205
name_count		1	0.0390	0.0254	2.3531	0.1250
launched_day_type	Weekday	1	-0.0250	0.0725	0.1195	0.7296
Quarter	1	1	-0.0613	0.1302	0.2212	0.6382
Quarter	2	1	0.1124	0.1240	0.8205	0.3650
Quarter	3	1	-0.0183	0.1235	0.0218	0.8825
launched_day*Quarter	Weekday	1	0.1150	0.1306	0.7751	0.3786
launched_day*Quarter	Weekday	2	0.0323	0.1238	0.0679	0.7945
launched_day*Quarter	Weekday	3	-0.0441	0.1214	0.1322	0.7161
duration*Level_USD_G	L1	1	0.00474	0.00867	0.2988	0.5846
duration*Level_USD_G	L2	1	0.0127	0.00914	1.9307	0.1647
duration*Level_USD_G	L3	1	0.0153	0.0109	1.9546	0.1621

9) Journalism – Cont'd

Inference

- Level-USD-Goal: Although not all levels are significant at 5% significance level, If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 3.51 compared to L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.02.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.02
- Sub Region: Although All sub region is not significant at 5% significance level, if a campaign for Kickstarter launches in Northern America, the log ratio of odds to achieve goal amount will increase by 1.59 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.19.
- Name Count: Although it is not significant, if sentiment increase by 1, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.04.
- Launched Day Type: If a project launches during the weekday, the log ratio of odds to achieve the goal amount will decrease by 0.03 compared to during the weekends.
- Quarter: Although all the quarters are not significant at 5% significance level, if a project launches during the 2nd quarter, the log ratio of odds to achieve the goal amount will increase by 0.11 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount will increase by 0.02 compared to L4.

Prediction

Confusion Matrix of Journalism for train

The FREQ Procedure			
	Table of F_state by I_state		
	I_state(Into: state)		Total
F_state(From: state)	failed	successful	
failed	2417 74.51 98.01 94.82	49 1.51 1.99 7.05	2466 76.02
successful	132 4.07 16.97 5.18	646 19.91 83.03 92.95	778 23.98
Total	2549 78.58	695 21.42	3244 100.00

Confusion Matrix of Journalism for testset

The FREQ Procedure			
	Table of F_state by I_state		
	I_state(Into: state)		Total
F_state(From: state)	failed	successful	
failed	602 74.32 97.73 95.10	14 1.73 2.27 7.91	616 76.05
successful	31 3.83 15.98 4.90	163 20.12 84.02 92.09	194 23.95
Total	633 78.15	177 21.85	810 100.00

- Train Sensitivity: 92.95%, Test sensitivity: 92.09%
- Total Train Accuracy: 94.42%, Total Test Accuracy: 94.44%

10) Music

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-6.1116	0.5970	104.7992	<.0001	
Level_USD_Goal	L1	1	4.3614	0.1086	1613.7114	<.0001	
Level_USD_Goal	L2	1	2.0546	0.1139	325.4804	<.0001	
Level_USD_Goal	L3	1	-0.4961	0.1377	12.9724	0.0003	
backers		1	0.1249	0.00159	6143.2866	<.0001	
duration		1	-0.00361	0.00203	3.1635	0.0753	
Competitors		1	0.000449	0.000547	0.6740	0.4117	
Avg_Amount_Pledged_B		1	0.0195	0.000435	2010.9003	<.0001	
sub_region	Australia and New Zealand	1	-0.0416	0.6077	0.0047	0.9454	
sub_region	Eastern Asia	1	2.4932	3.5221	0.5011	0.4790	
sub_region	Latin America and the Caribbean	1	-0.6178	0.6548	0.8902	0.3454	
sub_region	Northern America	1	-0.0378	0.5838	0.0042	0.9484	
sub_region	Northern Europe	1	-0.0995	0.5878	0.0287	0.8655	
sub_region	South-eastern Asia	1	-0.7225	2.0927	0.1192	0.7299	
sub_region	Southern Europe	1	-0.7809	0.6532	1.4293	0.2319	
sentiment		1	-0.00541	0.0385	0.0197	0.8884	
name_count		1	0.00662	0.00463	2.0458	0.1526	
launched_day_type	Weekday	1	0.0449	0.0226	3.9336	0.0473	
Quarter	1	1	0.0405	0.0390	1.0762	0.2995	
Quarter	2	1	0.0180	0.0380	0.2229	0.6368	
Quarter	3	1	-0.0584	0.0390	2.2449	0.1341	
launched_day*Quarter	Weekday	1	1	0.0110	0.0386	0.0820	0.7747
launched_day*Quarter	Weekday	2	1	-0.0769	0.0378	4.1276	0.0422
launched_day*Quarter	Weekday	3	1	0.00930	0.0389	0.0570	0.8112
duration*Level_USD_G	L1	1	-0.00789	0.00247	10.2181	0.0014	
duration*Level_USD_G	L2	1	-0.00145	0.00289	0.2503	0.6169	
duration*Level_USD_G	L3	1	0.00216	0.00355	0.3693	0.5434	

10) Music – Cont'd

Inference

- Level-USD-Goal: If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 4.36 compared to L4.
- Duration: Although it is not significant at 5% significance level, if the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.004.
- Competitors: Although it is not significant at 5% significance level, If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.00004.
- Sub Region: Although All sub region is not significant at 5% significance level, if a campaign for Kickstarter launches in Eastern Asia, the log ratio of odds to achieve goal amount will increase by 2.49 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significance level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.005.
- Name Count: Although it is not significant at 5% significance level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.007.
- Launched Day Type: If a project launches during the weekday, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to during the weekends.
- Quarter: Although all the quarters are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: Although all duration*Level_USD_Goal is not significant at 5% significance level, If the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount will increase by 0.002 compared to L4.

Prediction

Confusion Matrix of Music for train

		The FREQ Procedure		
		Table of F_state by I_state		
		I_state(Into: state)		
Frequency		F_state(From: state)	failed	successful
Percent				Total
Row Pct		failed	16424 45.02 94.71 96.45	917 2.51 5.29 4.71
Col Pct		successful	604 1.66 3.16 3.55	18539 50.81 96.84 95.29
		Total	17028 46.67	19456 53.33
			36484	100.00

Confusion Matrix of Music for test

		The FREQ Procedure		
		Table of F_state by I_state		
		I_state(Into: state)		
Frequency		F_state(From: state)	failed	successful
Percent				Total
Row Pct		failed	4084 44.78 94.89 96.55	220 2.41 5.11 4.50
Col Pct		successful	146 1.60 3.03 3.45	4671 51.21 96.97 95.50
		Total	4230 46.38	4891 53.62
			9121 100.00	

- Train Sensitivity: 95.29%, Test sensitivity: 95.50%
- Total Train Accuracy: 95.83%, Total Test Accuracy: 95.99%

11) Photography

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-7.8153	70.6753	0.0122	0.9119	
Level_USD_Goal	L1	1	3.8436	0.2172	313.2158	<.0001	
Level_USD_Goal	L2	1	1.5206	0.2212	47.2450	<.0001	
Level_USD_Goal	L3	1	-0.5441	0.2796	3.7860	0.0517	
backers		1	0.0891	0.00258	1195.1777	<.0001	
duration		1	-0.00973	0.00444	4.8073	0.0283	
Competitors		1	-0.00943	0.00193	23.7693	<.0001	
Avg_Amount_Pledged_B		1	0.0164	0.000809	410.1693	<.0001	
sub_region	Australia and New Zealand	1	2.8478	70.6753	0.0016	0.9679	
sub_region	Eastern Asia	1	-7.7965	420.6	0.0003	0.9852	
sub_region	Latin America and the Caribbean	1	2.4387	70.6804	0.0012	0.9725	
sub_region	Northern America	1	3.2456	70.6749	0.0021	0.9634	
sub_region	Northern Europe	1	3.2026	70.6750	0.0021	0.9639	
sub_region	South-eastern Asia	1	-9.4190	269.9	0.0012	0.9722	
sub_region	Southern Europe	1	2.5645	70.6758	0.0013	0.9711	
sentiment		1	-0.0434	0.0780	0.3097	0.5779	
name_count		1	-0.0471	0.0171	7.5700	0.0059	
launched_day_type	Weekday	1	-0.0475	0.0452	1.1073	0.2927	
Quarter	1	1	0.00498	0.0812	0.0038	0.9511	
Quarter	2	1	0.0399	0.0768	0.2696	0.6036	
Quarter	3	1	-0.00510	0.0760	0.0045	0.9464	
launched_day*Quarter	Weekday	1	1	0.0511	0.0810	0.3968	0.5287
launched_day*Quarter	Weekday	2	1	-0.00364	0.0767	0.0023	0.9621
launched_day*Quarter	Weekday	3	1	-0.0506	0.0743	0.4637	0.4959
duration*Level_USD_G	L1	1	-0.00661	0.00533	1.5363	0.2152	
duration*Level_USD_G	L2	1	0.0116	0.00582	3.9494	0.0469	
duration*Level_USD_G	L3	1	0.0102	0.00755	1.8195	0.1774	

11) Photography – Cont'd

Inference

- Level-USD-Goal: Although not all Level_USD_Goal is significant at 5% significance level, if the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 3.84 compared to L4.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.01.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.01.
- Sub Region: Although All sub region is not significant at 5% significance level, if a campaign for Kickstarter launches in Northern America, the log ratio of odds to achieve goal amount will increase by 3.25 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significance level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.04.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.05.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will decrease by 0.05 compared to during the weekends.
- Quarter: Although all the quarters are not significant at 5% significance level, if a project launches during the 2nd quarter, the log ratio of odds to achieve the goal amount will increase by 0.04 compared to the 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: Although not all duration*Level_USD_Goal is significant at 5% significance level, If the duration increases by 1 day when Level_USD_Goal is L2, the log ratio of odds to achieve the goal amount will increase by 0.01 compared to L4.

Prediction

Confusion Matrix of Photography for train

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	I_state(Into: state)			Total
F_state(From: state)	failed	successful		
failed	4895 64.50 96.47 94.99	179 2.36 3.53 7.35		5074 66.86
successful	258 3.40 10.26 5.01	2257 29.74 89.74 92.65		2515 33.14
Total	5153 67.90	2436 32.10		7589 100.00

Confusion Matrix of Photography for testset

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of F_state by I_state			
	I_state(Into: state)			Total
F_state(From: state)	failed	successful		
failed	1167 61.52 95.66 94.11			1220 64.31 8.07
successful	73 3.85 10.78 5.89			604 31.84 89.22 91.93
Total	1240 65.37			657 34.63 1897 100.00

- Train Sensitivity: 92.65%, Test sensitivity: 91.93%
- Total Train Accuracy: 94.24%, Total Test Accuracy: 93.36%

12) Technology

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-4.3439	0.2070	440.1812	<.0001	
Level_USD_Goal	L1	1	2.4312	0.1746	193.7969	<.0001	
Level_USD_Goal	L2	1	0.4760	0.1948	5.9670	0.0146	
Level_USD_Goal	L3	1	-1.1113	0.2438	20.7863	<.0001	
backers		1	0.0233	0.000486	2300.1138	<.0001	
duration		1	-0.0275	0.00429	41.0000	<.0001	
Competitors		1	-0.00210	0.000491	18.3560	<.0001	
Avg_Amount_Pledged_B		1	0.00567	0.000180	991.6672	<.0001	
sub_region	Australia and New Zealand	1	-0.0195	0.1493	0.0170	0.8963	
sub_region	Eastern Asia	1	1.0465	0.2891	13.1072	0.0003	
sub_region	Latin America and the Caribbean	1	-0.4846	0.2996	2.6170	0.1057	
sub_region	Northern America	1	0.1864	0.0942	3.9182	0.0478	
sub_region	Northern Europe	1	-0.1759	0.1173	2.2484	0.1338	
sub_region	South-eastern Asia	1	-0.4192	0.4365	0.9223	0.3369	
sub_region	Southern Europe	1	-0.2307	0.1814	1.6185	0.2033	
sentiment		1	0.0450	0.0434	1.0760	0.2996	
name_count		1	0.1044	0.0101	107.0320	<.0001	
launched_day_type	Weekday	1	0.0471	0.0306	2.3681	0.1238	
Quarter	1	1	0.0460	0.0531	0.7494	0.3867	
Quarter	2	1	-0.0433	0.0532	0.6629	0.4155	
Quarter	3	1	-0.0765	0.0524	2.1282	0.1446	
launched_day*Quarter	Weekday	1	0.0471	0.0530	0.7881	0.3747	
launched_day*Quarter	Weekday	2	1	-0.00910	0.0531	0.0293	0.8640
launched_day*Quarter	Weekday	3	1	0.00840	0.0523	0.0258	0.8725
duration*Level_USD_G	L1	1	0.0171	0.00481	12.6774	0.0004	
duration*Level_USD_G	L2	1	0.0286	0.00542	27.8653	<.0001	
duration*Level_USD_G	L3	1	0.0242	0.00669	13.0704	0.0003	

12) Technology – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level. If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 2.43 than L4. This is the highest increasing amount across the levels.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.03.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.0021.
- Sub Region: Although not all sub region is significant at 5% significance level, if a campaign for Kickstarter launches in Eastern Asia, the log ratio of odds to achieve goal amount will increase by 1.05 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significance level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will increase by 0.05.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.10.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will increase by 0.05 compared to during the weekends.
- Quarter: Although all the quarters are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.05 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L2, the log ratio of odds to achieve the goal amount will increase by 0.03 compared to L4. L2 is the highest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Technology for train

The FREQ Procedure			
Frequency Percent Row Pct Col Pct	Table of F_state by I_state		
	I_state(Into: state)		Total
F_state(From: state)	failed	successful	
failed	16069 74.95 97.89 92.48	346 1.61 2.11 8.51	16415 76.56
successful	1307 6.10 26.01 7.52	3718 17.34 73.99 91.49	5025 23.44
Total	17376 81.04	4064 18.96	21440 100.00

Confusion Matrix of Technology for testset

The FREQ Procedure			
Frequency Percent Row Pct Col Pct	Table of F_state by I_state		
	I_state(Into: state)		Total
F_state(From: state)	failed	successful	
failed	4010 74.83 98.00 92.18	82 1.53 2.00 8.13	4092 76.36
successful	340 6.34 26.84 7.82	927 17.30 73.16 91.87	1267 23.64
Total	4350 81.17	1009 18.83	5359 100.00

- Train Sensitivity: 91.49%, Test Sensitivity: 91.87%
- Total Train Accuracy: 92.29%, Total Test Accuracy: 94.85%

13) Publishing

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-5.8678	0.2771	448.4614	<.0001	
Level_USD_Goal	L1	1	4.2131	0.1349	975.2903	<.0001	
Level_USD_Goal	L2	1	1.9515	0.1429	186.4407	<.0001	
Level_USD_Goal	L3	1	-0.4671	0.1773	6.9401	0.0084	
backers		1	0.0836	0.00125	4452.3752	<.0001	
duration		1	-0.00255	0.00302	0.7124	0.3987	
Competitors		1	-0.00234	0.000613	14.6229	0.0001	
Avg_Amount_Pledged_B		1	0.0149	0.000419	1274.1792	<.0001	
sub_region	Australia and New Zealand	1	0.1602	0.2565	0.3901	0.5323	
sub_region	Eastern Asia	1	0.4771	1.2019	0.1576	0.6914	
sub_region	Latin America and the Caribbean	1	-0.4202	0.4378	0.9211	0.3372	
sub_region	Northern America	1	0.5213	0.2258	5.3289	0.0210	
sub_region	Northern Europe	1	0.0919	0.2344	0.1539	0.6949	
sub_region	South-eastern Asia	1	-0.8957	0.9130	0.9625	0.3266	
sub_region	Southern Europe	1	0.0820	0.3311	0.0613	0.8045	
sentiment		1	0.00622	0.0331	0.0353	0.8510	
name_count		1	0.0165	0.00847	3.7885	0.0516	
launched_day_type	Weekday	1	-0.0290	0.0241	1.4417	0.2299	
Quarter	1	1	0.0265	0.0417	0.4060	0.5240	
Quarter	2	1	-0.0199	0.0412	0.2339	0.6287	
Quarter	3	1	-0.00018	0.0417	0.0000	0.9965	
launched_day*Quarter	Weekday	1	0.00365	0.0416	0.0077	0.9302	
launched_day*Quarter	Weekday	2	1	-0.0510	0.0412	1.5353	0.2153
launched_day*Quarter	Weekday	3	1	0.0168	0.0414	0.1636	0.6859
duration*Level_USD_G	L1	1	-0.0121	0.00342	12.4900	0.0004	
duration*Level_USD_G	L2	1	-0.00297	0.00388	0.5862	0.4439	
duration*Level_USD_G	L3	1	0.00654	0.00490	1.7800	0.1821	

13) Publish – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level. If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 4.21 than L4. This is the highest. This is the highest increasing amount across the levels.
- Duration: Although it is not significant at 5% significance level, if the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.0025.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.0023.
- Sub Region: Although not all sub region is significant at 5% significance level, if a campaign for Kickstarter launches in Northern America, the log ratio of odds to achieve goal amount will increase by 0.52 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significance level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will increase by 0.006.
- Name Count: Although it is not significant at 5% significance level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.017
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will decrease by 0.029 compared to during the weekends.
- Quarter: Although all the quarters are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.03 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: Although not all levels are significant at 5% significance level, if the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount will increase by 0.0065 compared to L4. L3 is the highest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of publishing for train

Frequency Percent Row Pct Col Pct	Table of F_state by l_state			
	F_state(From: state)	l_state(Into: state)		
		failed	successful	Total
failed	17884 63.43 96.74 94.70	603 2.14 3.26 6.48	18487 65.57	
successful	1000 3.55 10.30 5.30	8709 30.89 89.70 93.52	9709 34.43	
Total	18884 66.97	9312 33.03	28196 100.00	

Confusion Matrix of publishing for test

Frequency Percent Row Pct Col Pct	Table of F_state by l_state			
	F_state(From: state)	l_state(Into: state)		
		failed	successful	Total
failed	4403 62.47 96.54 94.28	158 2.24 3.46 6.64	4561 64.71	
successful	267 3.79 10.74 5.72	2220 31.50 89.26 93.36	2487 35.29	
Total	4670 66.26	2378 33.74	7048 100.00	

- Train Sensitivity: 93.52%, Test Sensitivity: 93.36%
- Total Train Accuracy: 94.31%, Total Test Accuracy: 93.97%

14) Theater

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi Sq
Intercept		1	-4.4241	0.8452	27.3981	<.0001
Level_USD_Goal	L1	1	3.8058	0.1924	351.4119	<.0001
Level_USD_Goal	L2	1	1.7621	0.1942	82.2966	<.0001
Level_USD_Goal	L3	1	-0.6580	0.2279	8.3320	0.0039
backers		1	0.1178	0.00317	1384.0144	<.0001
duration		1	-0.0172	0.00389	21.8619	<.0001
Competitors		1	-0.00055	0.00414	0.0177	0.8942
Avg_Amount_Pledged_B		1	0.0192	0.000896	460.5473	<.0001
sub_region	Australia and New Zealand	1	-0.7778	0.9155	0.7213	0.3957
sub_region	Latin America and the Caribbean	1	-0.9018	0.9100	0.9820	0.3217
sub_region	Northern America	1	-0.1047	0.8187	0.0164	0.8982
sub_region	Northern Europe	1	0.2218	0.8222	0.0728	0.7873
sub_region	South-eastern Asia	1	2.2260	4.7961	0.2154	0.6426
sub_region	Southern Europe	1	-0.2889	1.1185	0.0667	0.7962
sentiment		1	-0.0436	0.0676	0.4157	0.5191
name_count		1	0.00521	0.00319	2.6665	0.1025
launched_day_type	Weekday	1	-0.0111	0.0465	0.0573	0.8108
Quarter	1	1	0.0736	0.0819	0.8078	0.3688
Quarter	2	1	0.0697	0.0786	0.7867	0.3751
Quarter	3	1	-0.0989	0.0781	1.6033	0.2054
launched_day*Quarter	Weekday	1	0.0148	0.0816	0.0327	0.8565
launched_day*Quarter	Weekday	2	0.0451	0.0732	0.3803	0.5374
launched_day*Quarter	Weekday	3	-0.0365	0.0780	0.2194	0.6395
duration*Level_USD_G	L1	1	0.00269	0.00486	0.3070	0.5795
duration*Level_USD_G	L2	1	0.00547	0.00522	1.0985	0.2946
duration*Level_USD_G	L3	1	0.0152	0.00632	5.7517	0.0165

14) Theater – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level. If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 3.61 than L4. This is the highest. This is the highest increasing amount across the levels.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.02.
- Competitors: Although it is not significant at 5% significance level, if the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.0006.
- Sub Region: Although all sub region is not significant at 5% significance level, if a campaign for Kickstarter launches in South-eastern Asia, the log ratio of odds to achieve goal amount will increase by 2.23 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significance level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.04.
- Name Count: Although it is not significant at 5% significance level, if the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.005.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will decrease by 0.01 compared to during the weekends.
- Quarter: Although all the quarters are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.07 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount will increase by 0.016 compared to L4. L3 is the highest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Theater for train

		Table of F_state by I_state		
		I_state(Into: state)		
F_state(From: state)		failed	successful	Total
failed	2682	304	2986	
	32.82	3.72	36.54	
	89.82	10.18		
	96.72	5.63		
successful	91	5094	5185	
	1.11	62.34	63.46	
	1.76	98.24		
	3.28	94.37		
Total		2773	5398	8171
		33.94	66.06	100.00

Confusion Matrix of Theater for testset

		Table of F_state by I_state		
		I_state(Into: state)		
F_state(From: state)		failed	successful	Total
failed	646	67	713	
	31.64	3.28	34.92	
	90.60	9.40		
	96.13	4.89		
successful	26	1303	1329	
	1.27	63.81	65.08	
	1.96	98.04		
	3.87	95.11		
Total		672	1370	2042
		32.91	67.09	100.00

- Train Sensitivity: 94.37%, Test Sensitivity: 95.11%
- Total Train Accuracy: 95.17%, Total Test Accuracy: 95.44%

15) Art

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-5.2323	0.1988	692.8713	<.0001	
Level_USD_Goal	L1	1	3.7484	0.1247	904.0928	<.0001	
Level_USD_Goal	L2	1	1.9501	0.1256	241.2330	<.0001	
Level_USD_Goal	L3	1	-0.2424	0.1504	2.5968	0.1071	
backers		1	0.1125	0.00191	3455.4873	<.0001	
duration		1	-0.0130	0.00248	27.2457	<.0001	
Competitors		1	-0.00325	0.000727	20.0126	<.0001	
Avg_Amount_Pledged_B		1	0.0212	0.000549	1492.1273	<.0001	
sub_region	Australia and New Zealand	1	0.2047	0.2068	0.9804	0.3221	
sub_region	Eastern Asia	1	0.2948	0.5138	0.3293	0.5661	
sub_region	Latin America and the Caribbean	1	-0.5611	0.3304	2.8836	0.0895	
sub_region	Northern America	1	0.3714	0.1401	7.0272	0.0080	
sub_region	Northern Europe	1	0.2436	0.1514	2.5873	0.1077	
sub_region	South-eastern Asia	1	-0.1996	0.7040	0.0804	0.7767	
sub_region	Southern Europe	1	-0.5172	0.2665	3.7647	0.0523	
sentiment		1	-0.0865	0.0440	3.8695	0.0492	
name_count		1	0.0168	0.00418	16.1746	<.0001	
launched_day_type	Weekday	1	0.0147	0.0267	0.3045	0.5811	
Quarter	1	1	-0.0123	0.0461	0.0718	0.7888	
Quarter	2	1	0.0888	0.0450	3.8906	0.0486	
Quarter	3	1	-0.0158	0.0458	0.1189	0.7303	
launched_day*Quarter	Weekday	1	0.0150	0.0460	0.1061	0.7446	
launched_day*Quarter	Weekday	2	1	-0.0176	0.0449	0.1540	0.6947
launched_day*Quarter	Weekday	3	1	0.0409	0.0452	0.8192	0.3654
duration*Level_USD_G	L1	1	0.00141	0.00321	0.1929	0.6605	
duration*Level_USD_G	L2	1	0.00441	0.00344	1.6386	0.2005	
duration*Level_USD_G	L3	1	0.000483	0.00418	0.0134	0.9080	

15) Art – Cont'd

Inference

- Level-USD-Goal: Although, not all levels are significant at 5% significance level. If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 3.75 than L4. This is the highest increasing amount across the levels.
- Duration: If the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will decrease by 0.01.
- Competitors: If the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.003.
- Sub Region: Although not all sub region is significant at 5% significance level, if a campaign for Kickstarter launches in Northern America, the log ratio of odds to achieve goal amount will increase by 0.37 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: If sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.09.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.02.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will increase by 0.01 compared to during the weekends.
- Quarter: Although not all quarters are significant at 5% significance level, if a project launches during the 2nd quarter, the log ratio of odds to achieve the goal amount will increase by 0.09 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L2, the log ratio of odds to achieve the goal amount will increase by 0.004 compared to L4. L2 is the highest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Art for train				Confusion Matrix of Art for testset			
The FREQ Procedure				The FREQ Procedure			
Frequency	Table of F_state by I_state			Frequency	Table of F_state by I_state		
	I_state(Into: state)	failed	successful		I_state(Into: state)	failed	successful
Percent	F_state(From: state)	failed	successful	Total	F_state(From: state)	failed	successful
	failed	10610 52.06 94.22 95.10	651 3.19 5.78 7.06	11261 55.26	failed	2665 52.32 94.44 94.37	157 3.08 5.56 6.92
		547 2.68 6.00 4.90	8571 42.06 94.00 92.94	9118 44.74		159 3.12 7.00 5.63	2113 41.48 93.00 93.08
		Total	11157 54.75	9222 45.25		2824 55.44	2270 44.56
				20379 100.00			5094 100.00

- Train Sensitivity: 92.94%, Test Sensitivity: 93.08%
- Total Train Accuracy: 94.12%, Total Test Accuracy: 93.80%

16) Comics

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept		1	-8.2071	2.0987	15.2932	<.0001	
Level_USD_Goal	L1	1	5.8437	0.2773	444.2483	<.0001	
Level_USD_Goal	L2	1	2.8377	0.2725	108.4663	<.0001	
Level_USD_Goal	L3	1	-0.8843	0.3283	7.2570	0.0071	
backers		1	0.0753	0.00222	1151.7670	<.0001	
duration		1	0.00262	0.00526	0.2479	0.6186	
Competitors		1	-0.00065	0.00411	0.0249	0.8745	
Avg_Amount_Pledged_B		1	0.0405	0.00183	489.9950	<.0001	
sub_region	Australia and New Zealand	1	0.8728	2.0867	0.1749	0.6758	
sub_region	Eastern Asia	1	0.6996	2.3449	0.0890	0.7654	
sub_region	Latin America and the Caribbean	1	0.9298	2.1604	0.1852	0.6669	
sub_region	Northern America	1	0.9380	2.0691	0.2055	0.6503	
sub_region	Northern Europe	1	0.8548	2.0734	0.1700	0.6802	
sub_region	South-eastern Asia	1	-5.0234	14.4099	0.1215	0.7274	
sub_region	Southern Europe	1	0.1405	2.1105	0.0044	0.9469	
sentiment		1	-0.0645	0.0675	0.9149	0.3388	
name_count		1	0.0512	0.0182	7.8937	0.0050	
launched_day_type	Weekday	1	-0.0291	0.0490	0.3523	0.5528	
Quarter	1	1	0.0154	0.0841	0.0334	0.8550	
Quarter	2	1	-0.0408	0.0845	0.2336	0.6289	
Quarter	3	1	-0.0166	0.0830	0.0400	0.8415	
launched_day*Quarter	Weekday	1	0.1195	0.0839	2.0289	0.1543	
launched_day*Quarter	Weekday	2	1	-0.0780	0.0845	0.8526	0.3558
launched_day*Quarter	Weekday	3	1	0.0914	0.0832	1.2055	0.2722
duration*Level_USD_G	L1	1	-0.0210	0.00639	10.7869	0.0010	
duration*Level_USD_G	L2	1	-0.0116	0.00714	2.6602	0.1029	
duration*Level_USD_G	L3	1	0.00943	0.00900	1.0981	0.2947	

16) Comics – Cont'd

Inference

- Level-USD-Goal: All levels are significant at 5% significance level. If the USD Goal Level is L1, the log ratio of odds to achieve the goal amount will increase by 5.84 than L4. This is the highest. This is the highest increasing amount across the levels.
- Duration: Although it is not significant at 5% significance level, if the Duration of a campaign increases by 1 day, the log ratio of odds to achieve the goal amount will increase by 0.003.
- Competitors: Although it is not significant at 5% significance level, if the number of Competitors of a KickStarter project campaign increases by 1, the log ratio of odds to achieve the goal amount will decrease by 0.001.
- Sub Region: Although all sub regions are not significant at 5% significance level, if a campaign for Kickstarter launches in Northern America, the log ratio of odds to achieve goal amount will increase by 0.94 compared to Western Europe (The reference group). This is the highest increasing amount across the sub-region.
- Sentiment: Although it is not significant at 5% significance level, if sentiment increase by 1, the log ratio of odds to achieve the goal amount will decrease by 0.06.
- Name Count: If the number of the words in a project's name increases by 1, the log ratio of odds to achieve the goal amount will increase by 0.05.
- Launched Day Type: Although it is not significant at 5% significance level, if a project launches during the weekday, the log ratio of odds to achieve the goal amount will decrease by 0.03 compared to during the weekends.
- Quarter: Although all quarters are not significant at 5% significance level, if a project launches during the 1st quarter, the log ratio of odds to achieve the goal amount will increase by 0.02 compared to 4th quarter. This is the highest increasing amount across the quarters.
- Duration*Level_USD_Goal: If the duration increases by 1 day when Level_USD_Goal is L3, the log ratio of odds to achieve the goal amount will increase by 0.01 compared to L4. L3 is the highest increasing amount across these Duration* Level_USD_Goal interaction terms.

Prediction

Confusion Matrix of Comics for train

Confusion Matrix of Comics for testset

The FREQ Procedure

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of F_state by I_state				Frequency Percent Row Pct Col Pct	Table of F_state by I_state				
	F_state(From: state)	I_state(Into: state)				F_state(From: state)	I_state(Into: state)			
		failed	successful	Total			failed	successful	Total	
	failed	2990 38.29 92.09 95.65	257 3.29 7.91 5.49	3247 41.58		failed	712 36.48 91.63 93.93	65 3.33 8.37 5.44	777 39.81	
	successful	136 1.74 2.98 4.35	4426 56.68 97.02 94.51	4562 58.42		successful	46 2.36 3.91 6.07	1129 57.84 96.09 94.56	1175 60.19	
	Total	3126 40.03	4683 59.97	7809 100.00		Total	758 38.83	1194 61.17	1952 100.00	

- Train Sensitivity: 94.51%, Test Sensitivity: 94.56%
- Total Train Accuracy: 94.97%, Total Test Accuracy: 94.31%

5. Business strategy for each category

Category	Goal Amount	Sub Region	Day Type	Quarter	Remarks
Film & Video	L1 (Maximum USD 980.65)	Northern Europe	Weekday	1 st	<ul style="list-style-type: none"> ➤ Considering the trendline Figure 13, it's descending. Also, the coefficient of competitors is negative. Therefore, it is good time to launch the project.
Games	L1 (Maximum USD 2,401.985)	Eastern Asia	Weekend	3 rd	<ul style="list-style-type: none"> ➤ Eastern Asia is the world's fast-growing region in game industry as well as its popularity. Therefore, we recommend Game projects should be launched in areas of high interest in games. ➤ The more words in project name, the higher chance to achieve the goal amount.
Design	L1 (Maximum USD 3,699.87)	Eastern Asia	Weekday	4 th	<ul style="list-style-type: none"> ➤ The more words in project name, the higher chance to achieve the goal amount.
Food	L1 (Maximum USD 3,779.58)	Eastern Asia	Weekend	1 st	<ul style="list-style-type: none"> ➤ Considering the trendline Figure 13, it's slightly descending. Also, the coefficient of competitors is negative. Therefore, it is good time to launch the project.
Fashion	L1 (Maximum USD 2,000)	Eastern Asia	Weekday	1 st	<ul style="list-style-type: none"> ➤ From competitors' point of view, we see from Figure 13 that the trend is increasing, so the company should hold off launch. Another insight is that first quarter is ideal for raising money, as this is the time fashion industry prepares for launch of its' spring collections.

Category	Goal Amount	Sub Region	Day Type	Quarter	Remarks
Crafts	L1 (Maximum USD 691.8975)	Southern Europe	Weekday	1 st	➤ People in South Europe seem to have relatively low work load, therefore they can have more vacations and free time than other regions for spending crafts creation.
Dance	L1 (Maximum USD 1,600)	Eastern Europe	Weekend	1 st	➤ Considering the trendline in Figure 13, it's slightly descending. Also, the coefficient of competitors is negative. Therefore, it is good time to launch the project.
Arts	L1 (Maximum 980.65 USD)	North America	Weekday	2 nd	➤ The trendline for competitors is flat and constant, we cannot make a recommendation with respect to competitors.
Comics	L1 (Maximum 1290 USD)	North America	Weekend	1 st	➤ The trendline for competitors is increasing, since in this category competition reduces success chances, it may not be ideal time to launch, also interestingly for this category longer duration is preferred to increase chances of success.
Publishing	L1 (Maximum USD 2000)	Northern America	Weekend	1 st	➤ English is the most commonly used language in the world. And Northern America have the largest English reader population. Book readers more likely to catch up the newest publishing information during weekends.
Technology	L1 (Maximum USD 5007)	Eastern Asia	Weekday	1 st	➤ Eastern Asia is fast-growing in manufacturing and high-tech industry. Business budget are usually allocated in the 1 st quarter of the year. People who are interested in technology projects are more likely to view during weekday work hours.

Category	Goal Amount	Sub Region	Day Type	Quarter	Remarks
Theatre	L1 (Maximum USD 1500)	South-Eastern Asia	Weekend	1 st	➤ People who are interested in new shows are more likely to research at weekends. Theatres are more likely to introduce new shows in the beginning of the year to attract more people.
Journalism	L1 (Maximum USD 980.65)	Northern America	Weekend	2 nd	➤ Northern America has a vital role in journalism, including freedom of speech and the press. Thus, there might be a lot of people actively participate in the journalism project.
Photography	L1 (Maximum USD 1500)	North America	Weekend	3 rd	➤ The number of words in projects' name have significant effect on success of projects. But it has negative coefficients that means if it has longer projects name, it tends to fail. It seems that for photography, its visual effect is more important than description.
Music	L1 (Maximum USD 1,581.9075)	Eastern Asia	Weekday	1 st	➤ Music industry of Eastern Asia is rapidly growing area of the world's largest music markets. Therefore, it seems that there is an incentive to invest their money in growing music area.

[Appendix]

<Summary Statistics of quantitative variable>

Summary Statistics

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
backers	backers	328595	117.02	969.37	0.00	2.00	15.00	64.00	219382.00
duration	duration	328595	34.00	12.70	1.00	30.00	30.00	36.00	92.00
Competitors	Competitors	328595	99.59	59.76	0.00	57.00	98.00	134.00	728.00
Avg_Amount_Pledged_Backer	Avg_Amount_Pledged_Backer	328595	66.53	132.80	0.00	17.17	42.67	77.95	10000.00
usd_pledged_real	usd_pledged_real	328595	10005.10	97034.73	0.00	50.00	803.69	4671.00	20338986.27
usd_goal_real	usd_goal_real	328595	41853.21	1114097.91	10.69	2000.00	5000.00	15000.00	166361390.71
goal_achievement		328595	0.99	6.74	0.00	0.01	0.20	1.09	1340.90
sentiment	sentiment	328595	0.06	0.62	-12.00	0.00	0.00	0.00	6.00
name_count	name_count	328595	5.57	5.41	1.00	3.00	5.00	8.00	1264.00

Summary Statistics by state

		N	Mean	StdDev	Min	P25	Median	P75	Max
state									
failed	backers	196515	16.50	71.42	0.00	1.00	3.00	12.00	6550.00
	duration	196515	35.19	13.17	1.00	30.00	30.00	40.00	92.00
	Competitors	196515	101.98	62.60	0.00	60.00	99.00	135.00	728.00
	Avg_Amount_Pledged_Backer	196515	49.46	134.02	0.00	1.57	25.00	55.25	10000.00
	usd_pledged_real	196515	1327.11	6839.21	0.00	3.00	100.00	670.00	757352.94
	usd_goal_real	196515	63496.37	1440054.04	10.73	2563.12	7500.00	20000.00	166361390.71
	goal_achievement	196515	0.09	0.15	0.00	0.00	0.02	0.11	1.11
successful	name_count	196515	5.32	4.46	1.00	3.00	5.00	7.00	1163.00
	backers	132080	266.59	1514.19	1.00	34.00	72.00	170.00	219382.00
	duration	132080	32.22	11.73	1.00	30.00	30.00	34.00	92.00
	Competitors	132080	96.04	55.09	0.00	51.00	96.00	132.00	728.00
	Avg_Amount_Pledged_Backer	132080	91.93	126.78	0.00	41.74	63.83	102.93	10000.00
	usd_pledged_real	132080	22916.61	151910.12	0.00	2032.00	5165.07	13448.87	20338986.27
	usd_goal_real	132080	9651.47	28128.12	10.69	1500.00	4000.00	10000.00	2015608.88
	goal_achievement	132080	2.32	10.49	0.00	1.05	1.16	1.61	1340.90
	name_count	132080	5.95	6.56	1.00	4.00	6.00	8.00	1264.00

<Summary statistic of categorical variable>

Summary Statistics of categorical variables

The FREQ Procedure

main_category		launched_year							
main_category	Frequency	Percent	Cumulative Frequency	Cumulative Percent	launched_year	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Art	25473	7.75	25473	7.75	2009	1166	0.35	1166	0.35
Comics	9761	2.97	35234	10.72	2010	9545	2.90	10711	3.26
Crafts	7770	2.36	43004	13.09	2011	23974	7.30	34685	10.56
Dance	3565	1.08	46569	14.17	2012	38335	11.67	73020	22.22
Design	25288	7.70	71857	21.87	2013	40967	12.47	113987	34.69
Fashion	19526	5.94	91383	27.81	2014	58357	17.76	172344	52.45
Film & Video	56274	17.13	147657	44.94	2015	64602	19.66	236946	72.11
Food	21596	6.57	169253	51.51	2016	48643	14.80	285589	86.91
Games	27941	8.50	197194	60.01	2017	43006	13.09	328595	100.00
Journalism	4054	1.23	201248	61.24					
Music	45605	13.88	246953	75.12					
Photography	9486	2.89	256339	78.01					
Publishing	35244	10.73	291583	88.74					
Technology	26799	8.16	318382	96.89					
Theater	10213	3.11	328595	100.00					

state				
state	Frequency	Percent	Cumulative Frequency	Cumulative Percent
failed	196515	59.80	196515	59.80
successful	132080	40.20	328595	100.00

launched_day_type				
launched_day_type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Weekday	231343	70.40	231343	70.40
Weekend	97252	29.60	328595	100.00

sub.region				
sub_region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Australia and New Zealand	7780	2.37	7780	2.37
Eastern Asia	498	0.15	8278	2.52
Latin America and the Caribbean	1402	0.43	9680	2.95
Northern America	271349	82.58	281029	85.52
Northern Europe	32900	10.01	313929	95.54
South-eastern Asia	447	0.14	314376	95.67
Southern Europe	4209	1.28	318585	96.95
Western Europe	10010	3.05	328595	100.00

<Correlation by state>

The CORR Procedure

state=failed

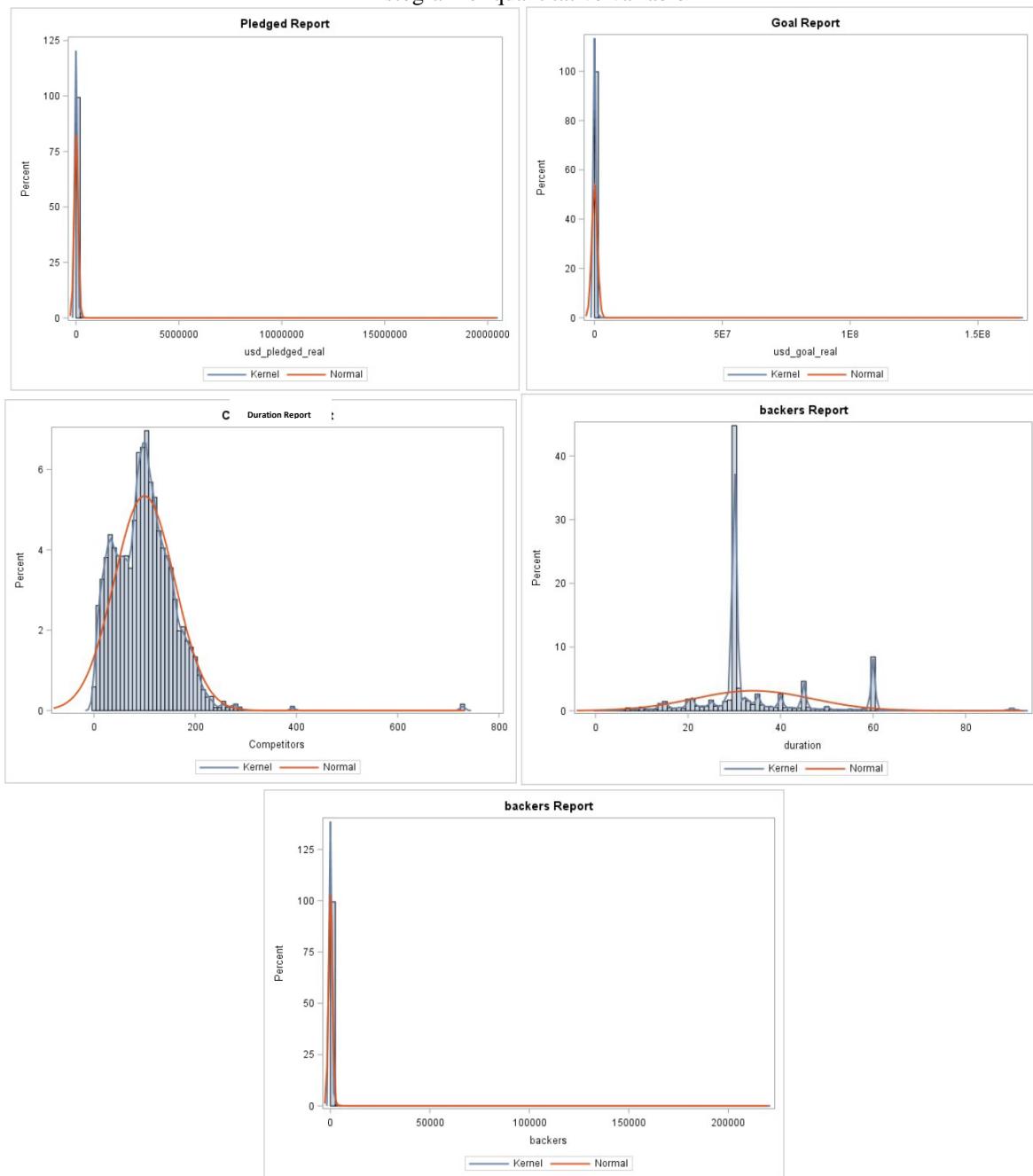
Pearson Correlation Coefficients, N = 196515 Prob > r under H0: Rho=0									
	backers	duration	Competitors	Avg_Amount_Pledged_Backer	usd_pledged_real	usd_goal_real	goal_achievement	name_count	
backers	1.00000	0.00079 0.7249	-0.01717 <.0001		0.05340 <.0001	0.73054 <.0001	0.00804 0.0004	0.31635 <.0001	0.02542 <.0001
duration		0.00079 0.7249	1.00000	-0.03060 <.0001		0.02503 <.0001	0.01358 <.0001	0.02286 <.0001	-0.05594 <.0001
Competitors			-0.01717 <.0001	-0.03060 <.0001	1.00000	0.00211 0.3504	-0.00174 0.4398	0.01705 <.0001	-0.05949 <.0001
Avg_Amount_Pledged_Backer				0.05340 <.0001	0.02503 0.3504	1.00000	0.23615 <.0001	0.01973 <.0001	0.20936 <.0001
usd_pledged_real					0.73054 <.0001	0.01358 0.4398	0.23615 <.0001	1.00000	0.02286 <.0001
usd_goal_real						0.01973 <.0001	0.02286 <.0001	1.00000	-0.02026 <.0001
goal_achievement							0.31193 <.0001	-0.02026 <.0001	0.1668
name_count								1.00000	0.06273 <.0001

The CORR Procedure

state=successful

Pearson Correlation Coefficients, N = 132080 Prob > r under H0: Rho=0									
	backers	duration	Competitors	Avg_Amount_Pledged_Backer	usd_pledged_real	usd_goal_real	goal_achievement	name_count	
backers	1.00000	0.02427 <.0001	0.00227 0.4096		-0.00829 0.0026	0.75010 <.0001	0.48733 <.0001	0.30193 <.0001	0.00300 0.2752
duration		0.02427 <.0001	1.00000	-0.04964 <.0001		0.06415 <.0001	0.03462 <.0001	0.05882 <.0001	0.00098 0.7215
Competitors			0.00227 0.4096	-0.04964 <.0001	1.00000	0.05331 <.0001	0.01065 0.0001	0.04031 <.0001	-0.01121 <.0001
Avg_Amount_Pledged_Backer				0.00829 0.0026	0.06415 <.0001	0.05331 <.0001	1.00000	0.10895 <.0001	0.17846 <.0001
usd_pledged_real					0.75010 <.0001	0.03462 0.0001		0.48237 <.0001	0.25327 <.0001
usd_goal_real						0.10895 <.0001	1.00000		0.00175 0.5240
goal_achievement								0.00175 0.5240	0.01123 0.1839
name_count									0.00366 0.1839

<Histogram of quantitative variable>



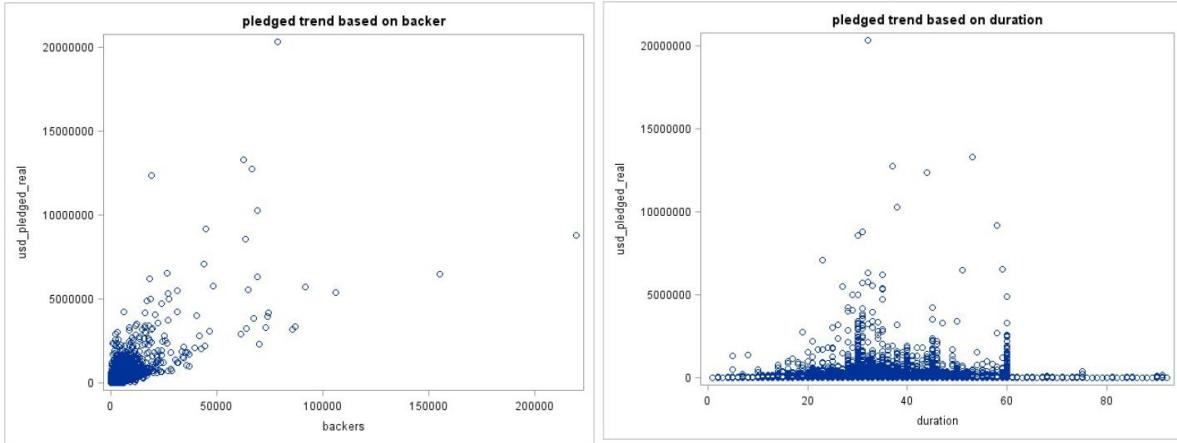


Table for Goal Level by Each Category (USD)

Goal Level	Art	Comics	Crafts	Dance	Design
L1	< 980.65	< 1,290.96	< 691.90	< 1,600.00	< 3,699.87
L2	< 2,971.24	< 3,500.00	< 2,179.74	< 3,200.00	< 10,000.00
L3	< 7,050.00	< 7,977.50	< 6,150.04	< 6,500.00	< 25,000.00
L4	7,050.00 ≤	7,977.50 ≤	6,150.04 ≤	6,500.00 ≤	25,000.00 ≤
Goal Level	Fashion	Film & Video	Food	Games	Journalism
L1	< 2,000.00	< 2,500.00	< 3,779.58	< 2,401.99	< 1,500.00
L2	< 5,500.00	< 6,400.00	< 10,000.00	< 7,500.00	< 5,000.00
L3	< 15,000.00	< 20,000.00	< 25,000.00	< 20,000.00	< 13,702.11
L4	15,000.00 ≤	20,000.00≤	25,000.00 ≤	20,000.00 ≤	13,702.11 ≤
Goal Level	Music	Photography	Publishing	Technology	Theater
L1	< 1,581.91	< 1,500.00	< 2,000.00	< 5,007.01	< 1,500.00
L2	< 4,000.00	< 3,926.50	< 4,995.04	< 18,000.00	< 3,103.45
L3	< 8,768.73	< 9,992.80	< 10,000.00	< 50,000.00	< 7,250.53
L4	8,768.73 ≤	9,992.80 ≤	10,000.00 ≤	50,000.00 ≤	7,250.53 ≤