

KICKSTARTER

Success Predictor

By: Binaryxx Sune

Intro

Kickstarter is a crowdfunding platform where a creator can pitch an idea on the website and receive donations to fund for its creation. Kickstarter employs an “all or nothing” policy in which projects that fail to meet its goal by its deadline will not receive any of the pledged money by the donators. Donators will not be charged for unsuccessful programs.

The goal is to predict the probability that a project will meet its set goal based on details such as the category the project is listed as, the amount of money set as the goal, time allocated to reach the goal, and the country in which the project was started.

Intro

The effect of each attribute to the success rate of the project will be analyzed through a variety of models. From these models, the one with the performance will be selected as a basic to use when making predictions on a creator's chances of success.

As an aspiring creator, it is important to know what steps can be taken to ensure that their project will be successful in raising in money.

By understanding which categories are most successful in Kickstarter, they are able to decide whether Kickstarter is the platform where they will have the most success.

By understanding the typical time interval and goals of successful projects are, they are able to set a realistic goal that will allow them to raise as much money without the risk of not receiving any of the pledged money.

Data Set Descriptions

The data set, Kickstarter Project, is from Kaggle from the user Mickaël Mouillé. The csv file named ks-projects-201801 will be used.

The data set consists of 378,661 projects that launched from May 21st, 2009 to January 2nd, 2018. The data set has 15 attributes in total.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ID	name	category	main_category	currency	deadline	goal	launched	pledged	state	backers	country	usd pledged	usd_pledged_real	usd_goal_real			
2	1000002330	The Songs of Adelaide & Abullah	Poetry	Publishing	GBP	10/09/15	1000	08/11/15 12:12	0	failed	0	GB	0	0	1533.95			
3	1000000930	Greeting From Earth: ZGAC Arts Capsule For ET	Narrative Film	Film & Video	USD	11/01/17	30000	08/02/17 04:43	2421	failed	15	US	100	2421	30000			
4	1000004038	Where is Hank?	Narrative Film	Film & Video	USD	02/26/13	45000	01/12/13 00:20	220	failed	3	US	220	220	45000			
5	1000007540	ToshiCapital Records Needs Help to Complete Album	Music	Music	USD	04/16/12	5000	03/17/12 03:24	1	failed	1	US	1	1	5000			
6	1000011046	Community Film Project: The Art of Neighborhood Filmmaking	Film & Video	Film & Video	USD	08/29/15	19500	07/04/15 08:35	1283	canceled	14	US	1283	1283	19500			
7	1000014025	Monarch Espresso Bar	Restaurants	Food	USD	04/01/16	50000	02/26/16 13:38	52375	successful	224	US	52375	52375	50000			
8	1000023410	Support Solar Roasted Coffee & Green Energy! SolarCoffee.co	Food	Food	USD	12/21/14	1000	12/01/14 18:30	1205	successful	16	US	1205	1205	1000			
9	1000030581	Chaser Strips. Our Strips make Shots their B*tch!	Drinks	Food	USD	03/17/16	25000	02/01/16 20:05	453	failed	40	US	453	453	25000			
10	1000034518	SPIN - Premium Retractable In-Ear Headphones with Mic	Product Design	Design	USD	05/29/14	125000	04/24/14 18:14	8233	canceled	58	US	8233	8233	125000			
11	100004195	STUDIO IN THE SKY - A Documentary Feature Film (Canceled)	Documentary	Film & Video	USD	08/10/14	65000	07/11/14 21:55	6240.57	canceled	43	US	6240.57	6240.57	65000			
12	100004721	Of Jesus and Madmen	Nonfiction	Publishing	CAD	10/09/13	2500	09/09/13 18:19	0	failed	0	CA	0	0	2406.39			
13	100005484	Lisa Lim New CD!	Indie Rock	Music	USD	04/08/13	12500	03/09/13 06:42	12700	successful	100	US	12700	12700	12500			
14	1000055792	The Cottage Market	Crafts	Crafts	USD	10/02/14	5000	09/02/14 17:11	0	failed	0	US	0	0	5000			
15	1000056157	G-Spot Place for Gamers to connect with eachother & go pro!	Games	Games	USD	03/25/16	200000	02/09/16 23:01	0	failed	0	US	0	0	200000			
16	1000057089	Tombstone: Old West tabletop game and miniatures in 32mm.	Tabletop Games	Games	GBP	05/03/17	5000	04/05/17 19:44	94175	successful	761	GB	57763.78	121857.33	6469.73			
17	1000064368	Survival Rings	Design	Design	USD	02/28/15	2500	01/29/15 02:10	664	failed	11	US	664	664	2500			
18	1000064918	The Beard	Comic Books	Comics	USD	11/08/14	1500	10/09/14 22:27	395	failed	16	US	395	395	1500			
19	1000068480	Notes From London: Above & Below	Art Books	Publishing	USD	05/10/15	3000	04/10/15 21:20	789	failed	20	US	789	789	3000			
20	1000070642	Mike Corey's Darkness & Light Album	Music	Music	USD	08/17/12	250	08/02/12 14:11	250	successful	7	US	250	250	250			
21	1000071625	Boco Tea	Food	Food	USD	06/02/12	5000	05/03/12 17:24	1781	failed	40	US	1781	1781	5000			
22	1000072011	CMUK. Shoes: Take on Life Feet First.	Fashion	Fashion	USD	12/30/13	20000	11/25/13 07:06	34268	successful	624	US	34268	34268	20000			
23	1000081649	Mikely clothing brand fundraiser	Childrenswear	Fashion	AUD	09/07/17	2500	08/08/17 01:20	1	failed	1	AU	0	0.81	2026.1			
24	1000082254	Alice in Wonderland in G Minor	Theater	Theater	USD	06/15/14	3500	05/16/14 10:10	650	failed	12	US	650	650	3500			
25	1000087442	Mountain brew: A quest for alcohol sustainability	Drinks	Food	NOK	02/25/15	500	01/26/15 19:17	48	failed	3	NO	6.18	6.29	65.55			
26	1000091520	The Book Zoo - A Mini-Comic	Comics	Comics	USD	11/12/14	175	10/23/14 17:15	701.66	successful	66	US	701.66	701.66	175			
27	1000102741	Matt Cavanaugh & Jenny Powers make their 1st album!	Music	Music	USD	01/06/11	10000	12/07/10 23:16	15827	successful	147	US	15827	15827	10000			
28	1000103948	Superhero Teddy Bear	DIY	Crafts	GBP	01/05/16	12000	12/06/15 20:09	0	failed	0	GB	0	0	17489.65			
29	1000104688	Permaculture Skills	Webseries	Film & Video	CAD	12/14/14	17757	11/14/14 18:02	48905	successful	571	CA	43203.25	42174.03	15313.04			
30	1000104953	Rebel Army Origins: The Heroic Story Of Major Gripes	Comics	Comics	GBP	01/28/16	100	12/29/15 16:59	112	successful	27	GB	167.7	160.6	142.91			
31	100011318	My Moon - Animated Short Film	Animation	Film & Video	USD	05/03/17	50000	04/03/17 17:11	57577.31	successful	840	US	10120	57577.31	50000			
32	1000115172	Daily Brew Coffee	Food Trucks	Food	GBP	03/31/15	3500	03/01/15 18:06	21	failed	1	GB	32.42	31.09	5181.12			
33	1000117861	Ledr workbook: one tough journal!	Product Design	Design	USD	10/08/16	1000	09/07/16 13:14	47266	successful	549	US	11253	47266	1000			
34	1000120151	Feather Cast Furlid Fly Fishing Leaders	Product Design	Design	AUD	08/22/15	2000	07/23/15 03:09	2000	successful	18	AU	1473.62	1427.35	1427.35			
35	1000120287	8B130A	Public Art	Art	USD	03/24/13	25000	02/12/13 01:07	1395	failed	30	US	1395	1395	25000			
36	100012079	Chris Eger Band - New Nashville Record!	Music	Music	USD	08/13/14	12000	07/14/14 22:35	13260	successful	92	US	13260	13260	12000			
37	1000129669	Squatch Watchers	Webseries	Film & Video	USD	08/23/13	10000	07/24/13 21:40	1373.37	failed	54	US	1373.37	1373.37	10000			
38	1000131947	Arrows & Sound Debut Album	Indie Rock	Music	USD	05/19/12	4000	04/19/12 01:04	8641.34	successful	157	US	8641.34	8641.34	4000			
39	1000132945	Zen in America: a film on Zen Buddhism in North America	Documentary	Film & Video	USD	10/21/13	10000	09/21/13 23:00	15663	successful	213	US	15663	15663	10000			
40	1000134913	Galaxy: Winnie and Friends	Illustration	Art	USD	01/27/16	3000	12/23/15 21:47	17	failed	2	US	17	17	3000			
41	1000146991	Help Scott Reynolds Make a New Record!	Indie Rock	Music	USD	09/02/12	15000	08/03/12 18:43	17499.99	successful	277	US	17499.99	17499.99	15000			
42	1000149007	Unschooling To University Book Project	Nonfiction	Publishing	CAD	01/19/18	3000	11/20/17 18:15	592	live	24	CA	185.65	472.88	2396.36			
43	1000153208	Phil Younger - Slow Down (Official Music Video)	Music	Music	USD	12/20/16	2100	12/09/16 20:27	2100	successful	34	US	83	2100	2100			
44	1000160286	"THE RETURN" book project	Photography	Photography	USD	12/13/13	15000	11/13/13 00:34	23907	successful	232	US	23907	23907	15000			
45	1000170964	Pennv Bineo Plavine Card Game fun for the whole family!	Tabletop Games	Games	USD	03/27/17	1500	03/02/17 04:01	856	failed	25	US	324	856	1500			

Attribute Description

ID: Unique identifier of project

Name: Name of the project

Category: Category of the project

Main Category: Main category is the umbrella category that groups all the categories into corresponding groups

Currency: Currency that the monetary goal was originally set by the user

Deadline: Date when the option to crowdfund ends

Goal: Money value set as a goal (set in the creator's local currency)

Attribute Description

Pledged: The amount of money that the project had at the time that the data was collected

State: States that describe the status of the project such as whether the project succeeded, failed, cancelled, etc.

Backers: Number of backers for the project

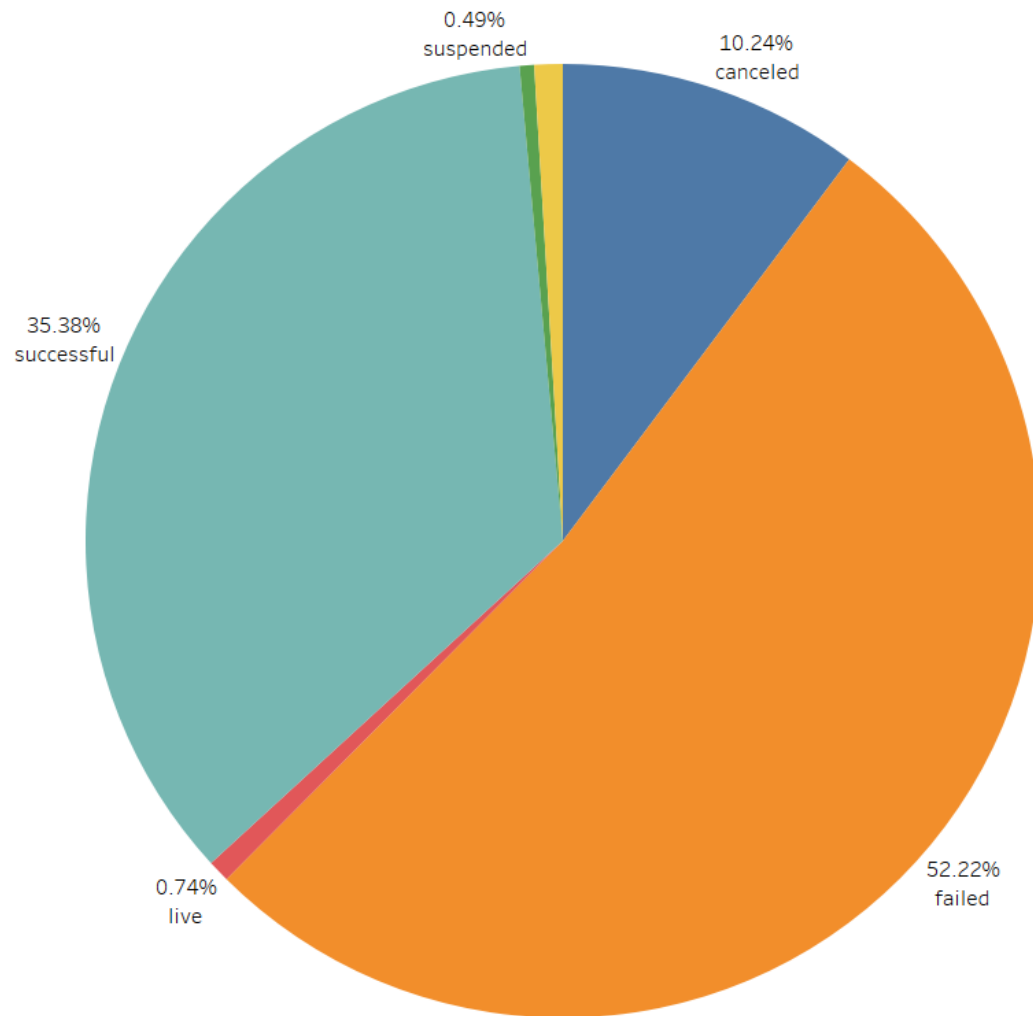
Country: Country where the project was submitted

USD Pledged: Conversion of currency to USD in the Pledged column (done by Kickstarter)

USD Pledged Real: Conversion of currency to USD in the Pledged column (Done by the Fixer.io API)

USD Goal Real: Conversion of currency to USD in the Goal column (Done by the Fixer.io API)

Attribute Visualization: State



Count of ks-projects-201801

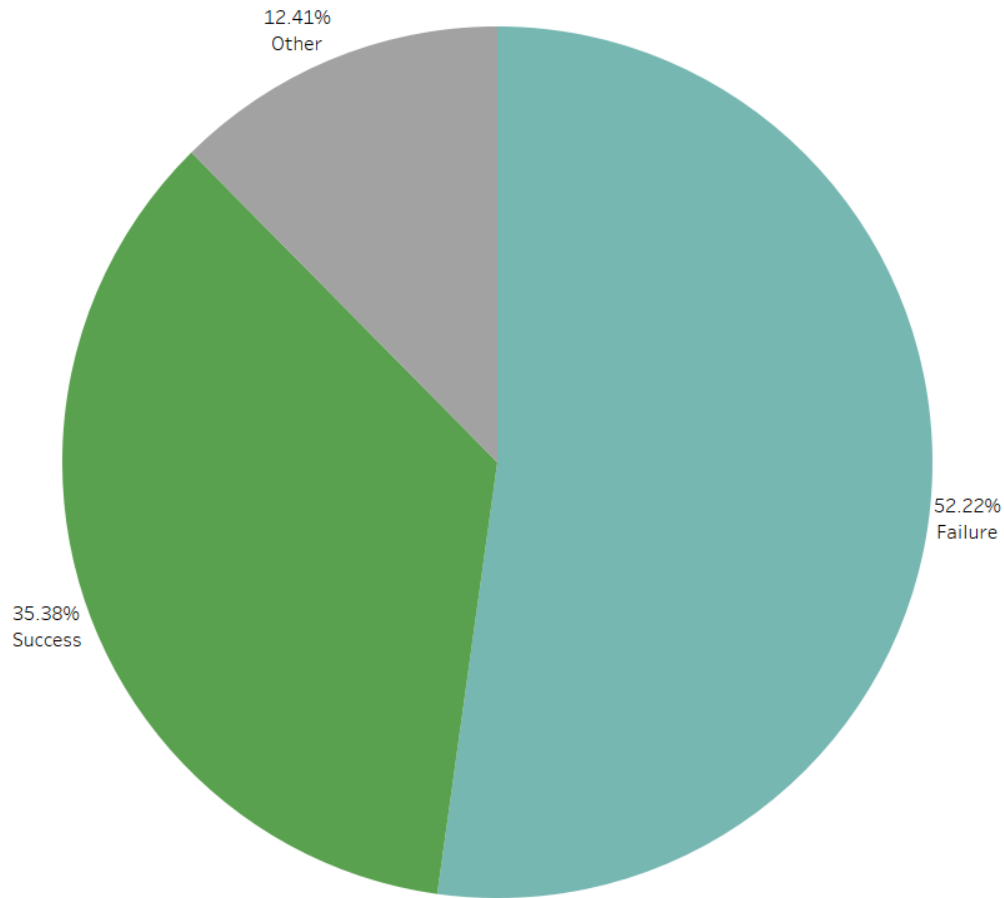
378,661

State

- canceled
- failed
- live
- successful
- suspended
- undefined

The data set originally had 6 states for the projects: Canceled, Failed, Live, Successful, Suspended, and Undefined

Attribute Visualization: State [v2]



Count of ks-projects-201801

378,661

State (group)

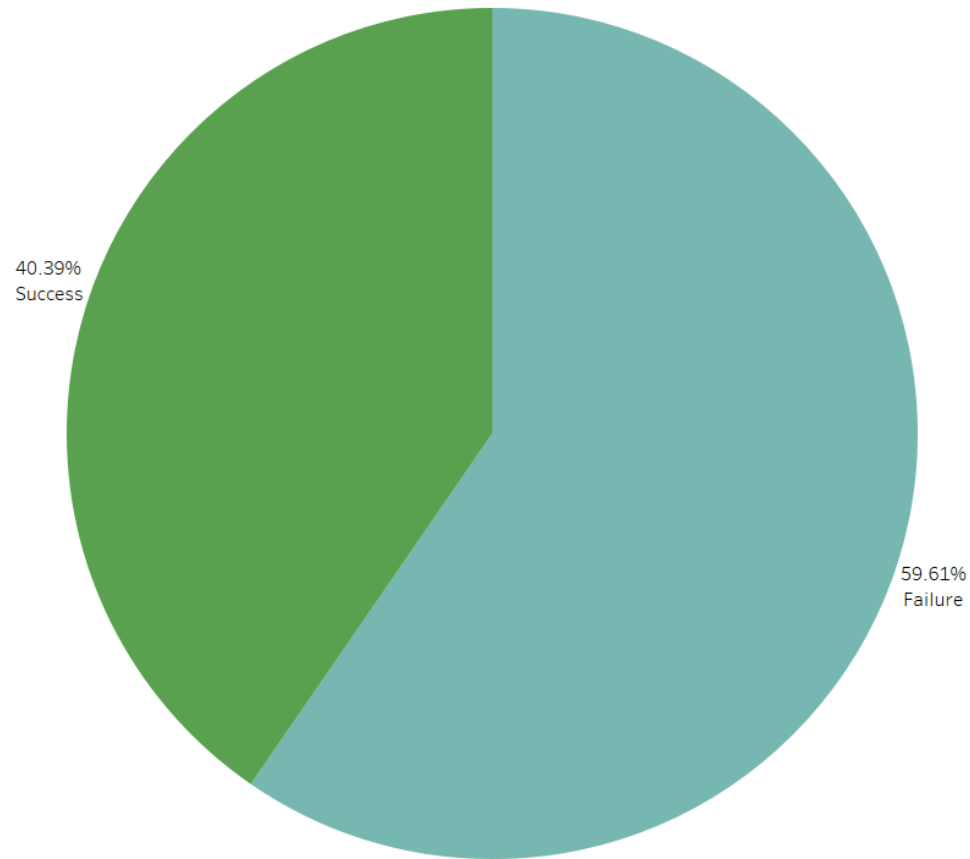
Failure

Success

Other

Edited the States attribute to group the values Live, Suspended, Canceled, and Undefined to be grouped into one. The main focus for this attribute is whether the project was a success or a failure by its deadline, assuming that the creator has no reason to cancel or be suspended.

Attributes Visualization: State [v3]



Count of ks-projects-201801

331,675

State (group)

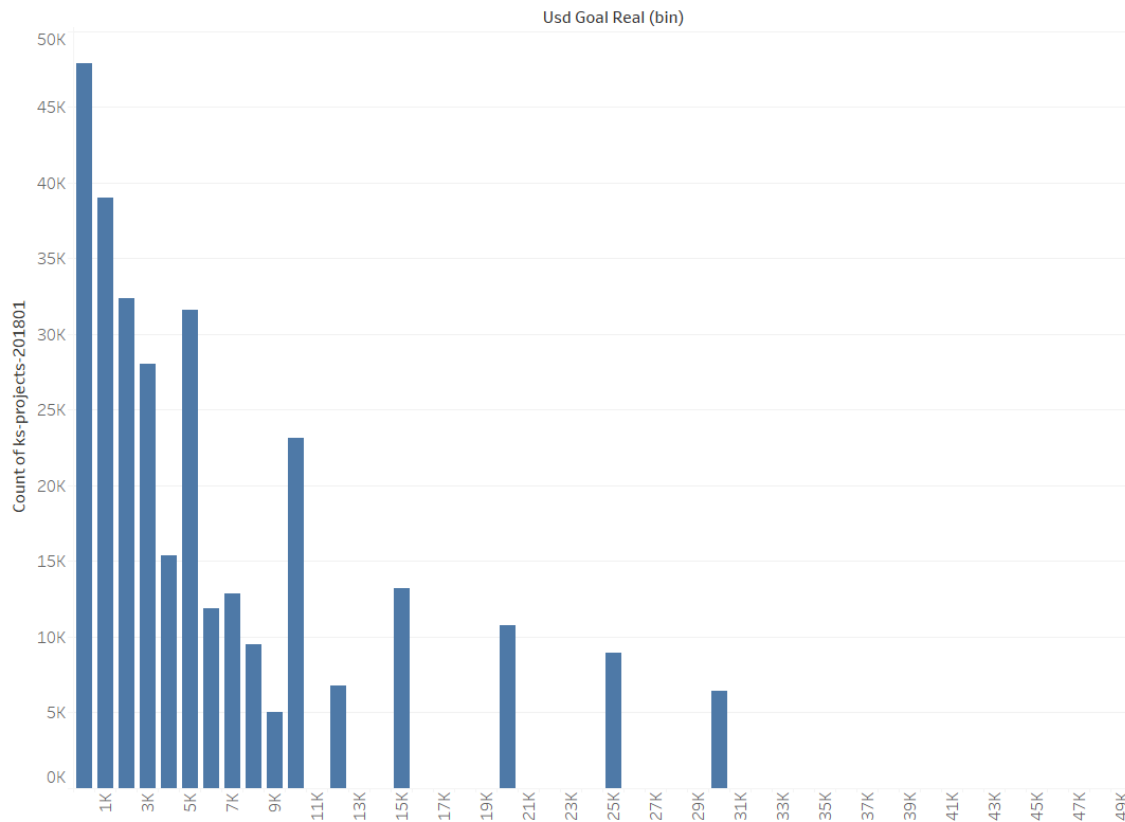
Failure

Success

After filtering out the other values, the percentage of successful projects assuming that it has reached its deadline without any complication is around 40.39%. Leaving the remaining 59.61% to be projects that failed to meet their goals.

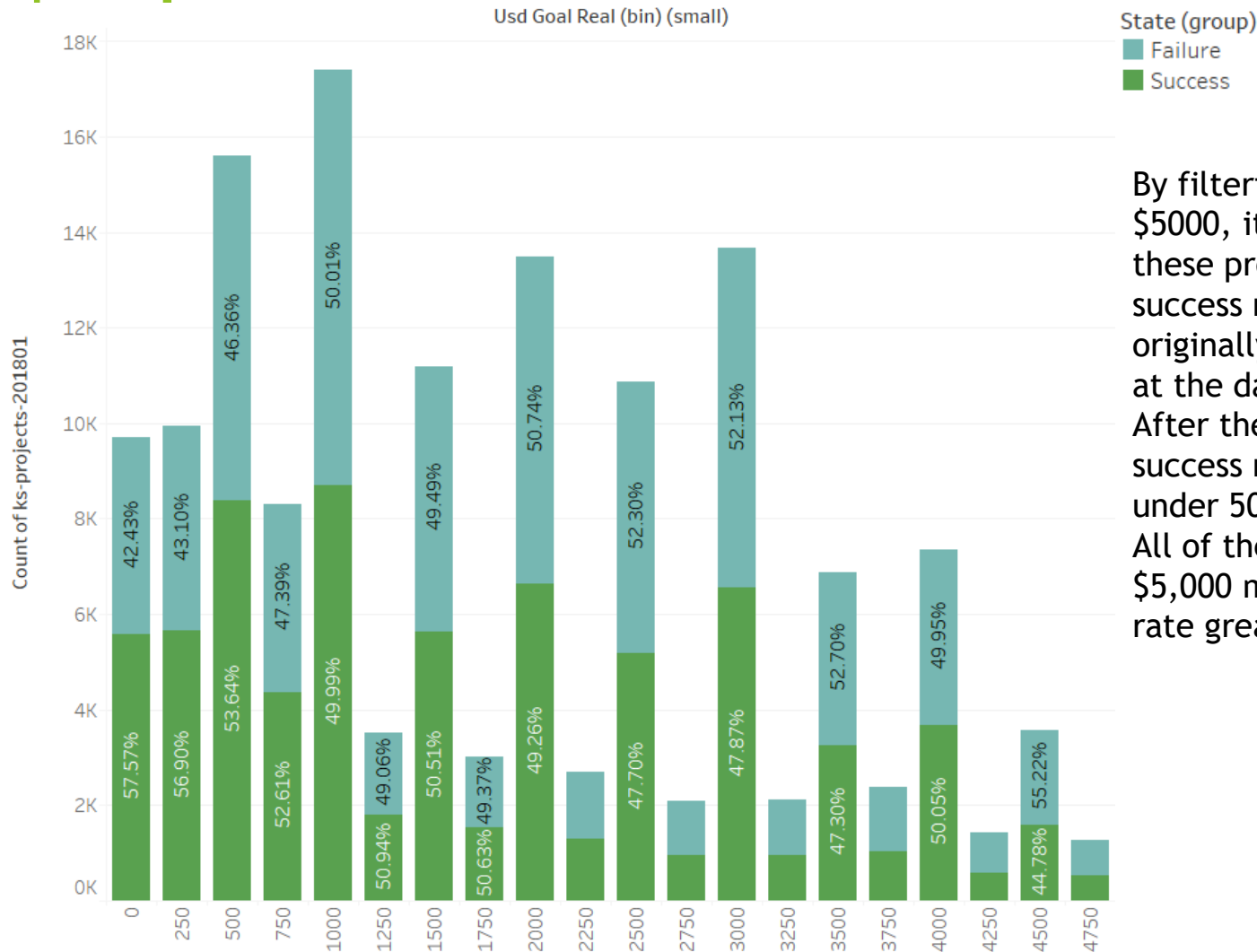
Attribute Visualization: Goal (USD)

[Overview]



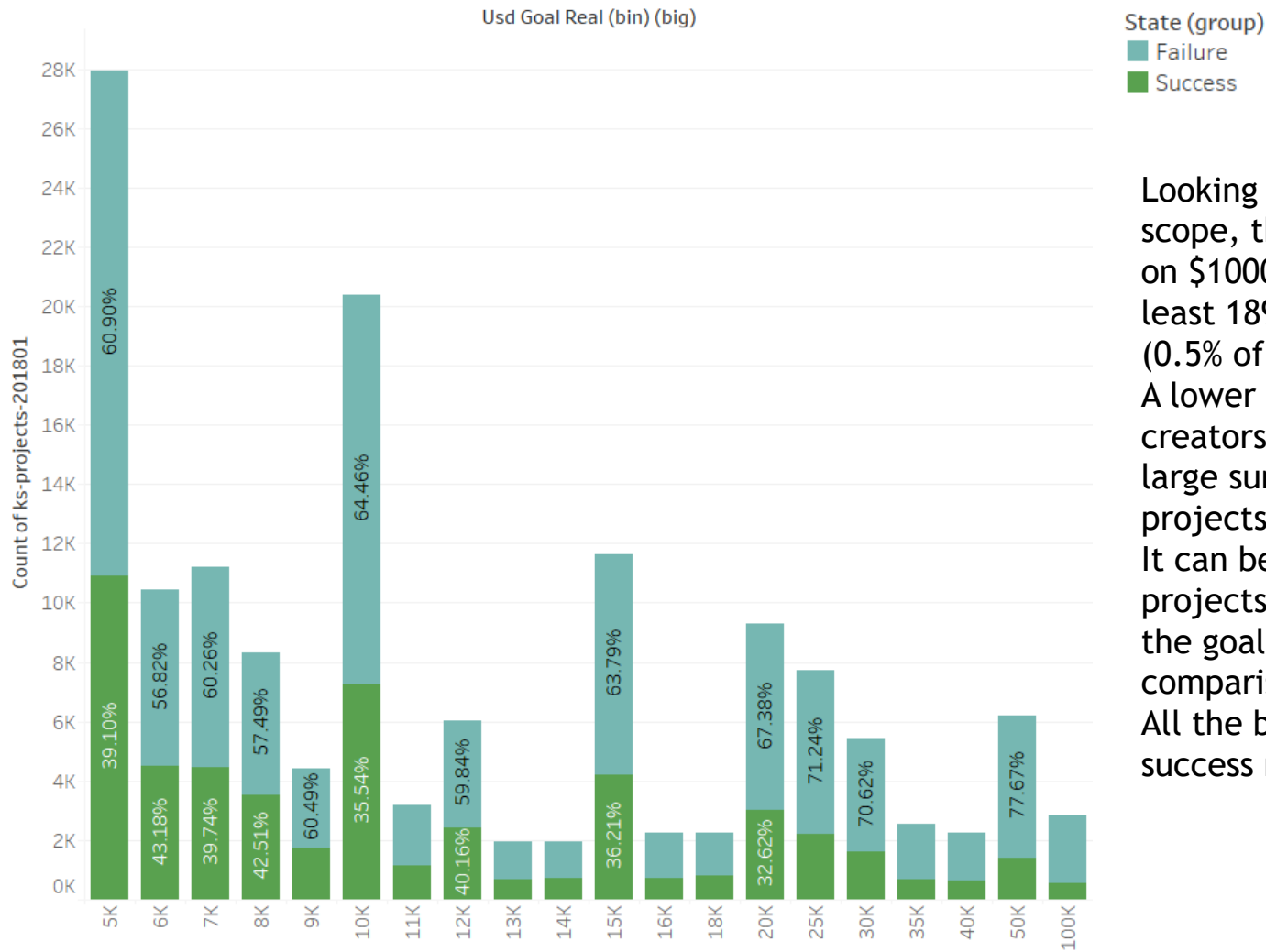
- Converted the Goal USD attribute from a string type to an integer type to allow calculations to be done.
- Since the suggested bins created a histogram that were not usable, a more readable histogram was made through trial and error.
- The current graph is from a bin size of \$1000 for goals and a filter requiring the bin to have at least 3,790 projects (~1% of the data).
- It can be observed that there are spikes in projects in intervals of \$5,000.

Attribute Visualization: Goal (USD) [< \$5k]



By filtering the goals to under \$5000, it can be seen that these projects have a higher success rate than what was originally seen when looking at the data set as a whole. After the \$2000 mark, the success rate begins to fall under 50%. All of the bins under the \$5,000 mark has a success rate greater than 45%.

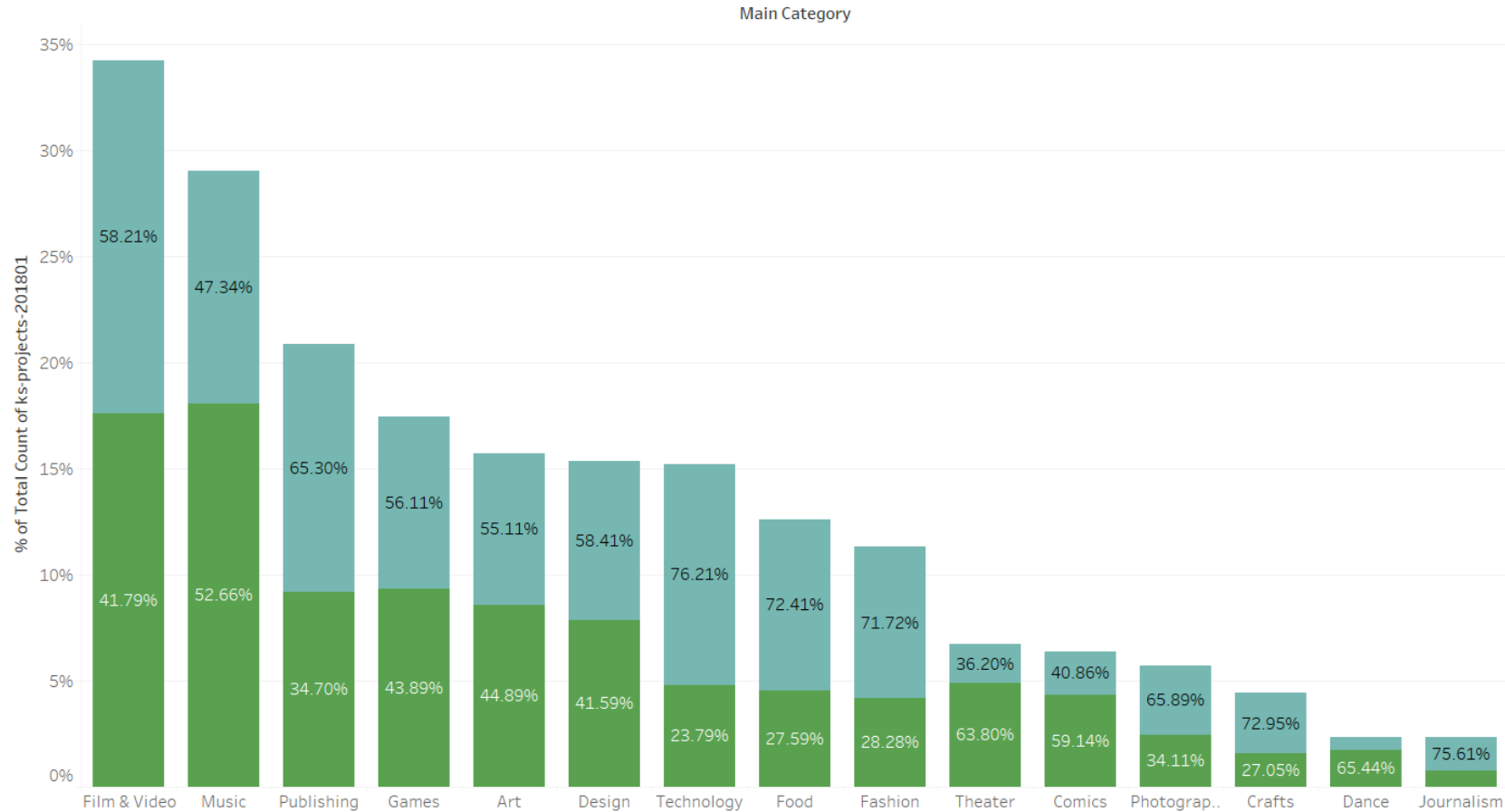
Attribute Visualization: Goal (USD) [> \$5k]



Looking into the data with a wider scope, the histogram here is based on \$1000 bins with a filter of at least 1895 projects in each one (0.5% of the original data set). A lower threshold is used since less creators would naturally ask for large sums of money such as these projects.

It can be observed that these projects have lower success rates as the goals go past \$5,000 in comparison to those under \$5,000. All the bins over the \$5,000 have a success rate lower than 45%.

Attribute Visualization: Main Category



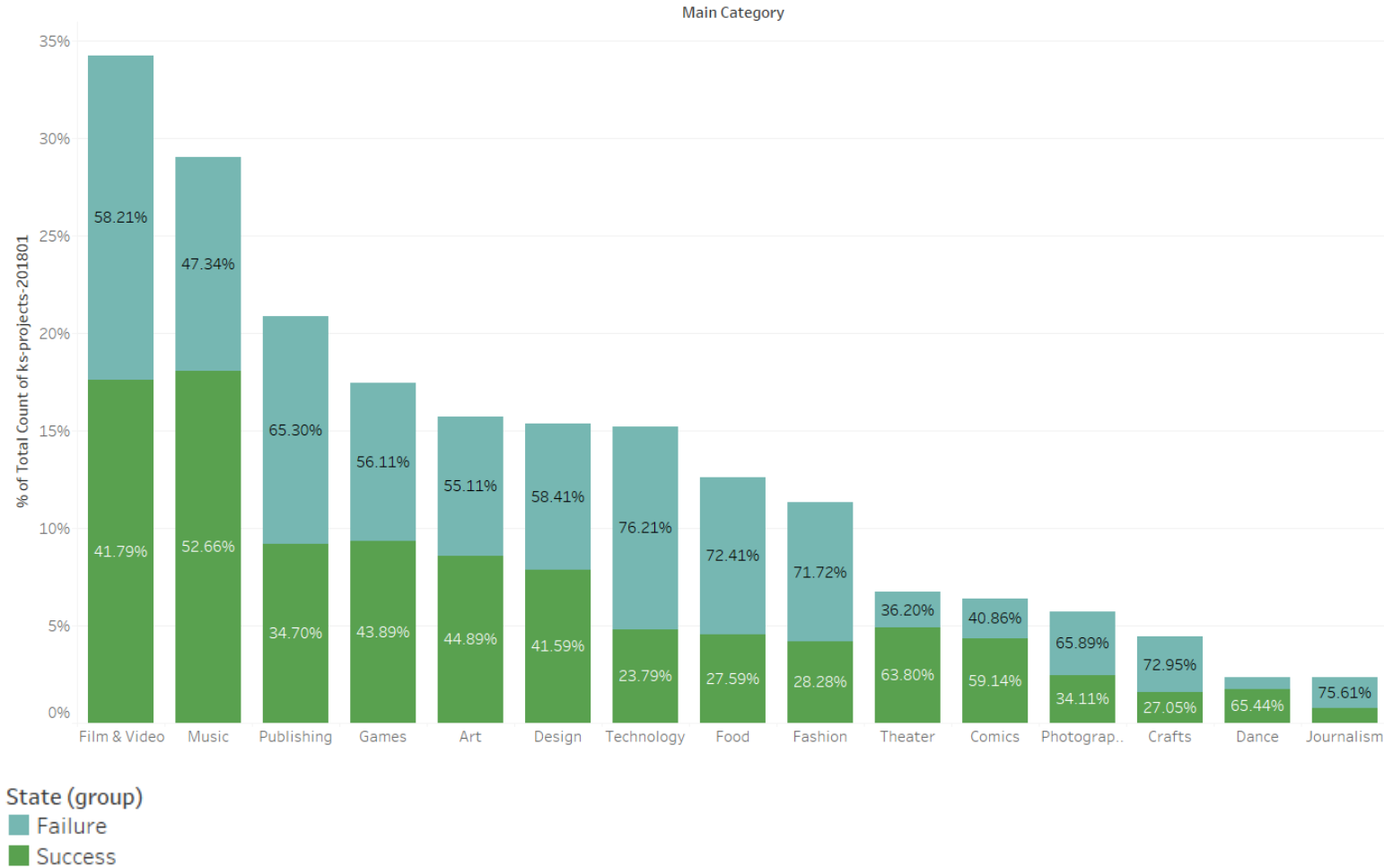
There are 15
Main
Categories:

- Film & Video
- Music
- Publishing
- Games
- Art
- Design
- Technology
- Food
- Theater
- Comics
- Photograph
- Crafts
- Dance
- Journalism

State (group)

- Failure
- Success

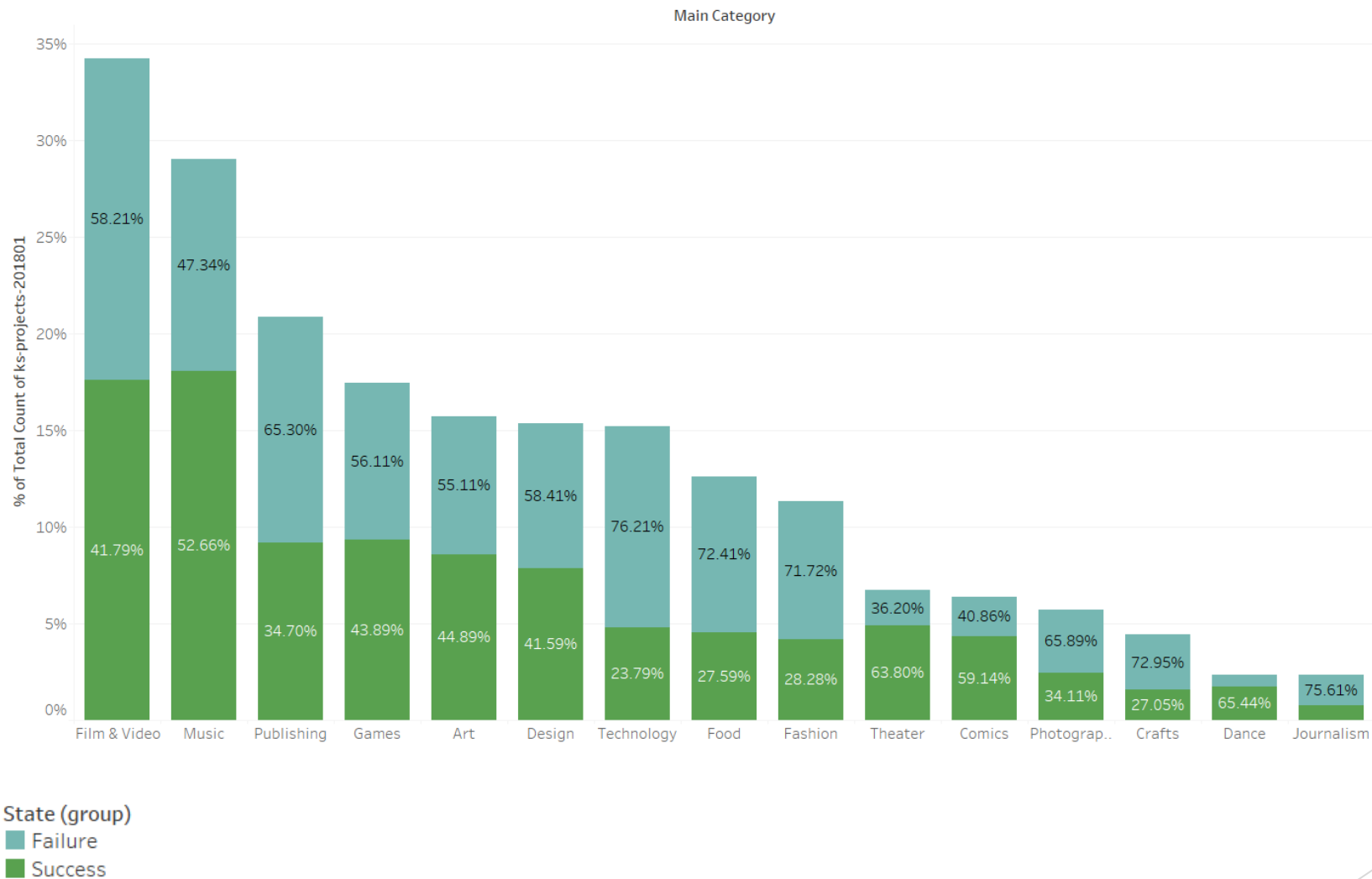
Attribute Visualization: Main Category



The top 5 main categories with the most projects are:

1. Film & Video
2. Music
3. Publishing
4. Games
5. Art

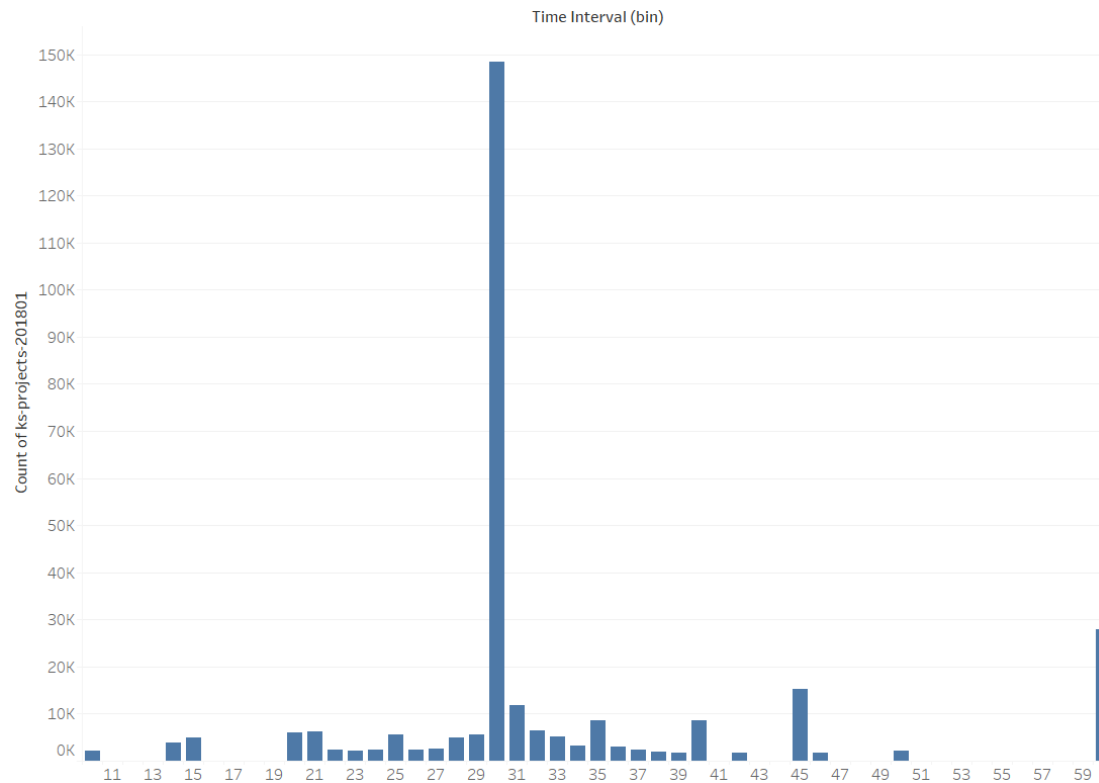
Attribute Visualization: Main Category



The top 5 main categories with the highest success rates are:

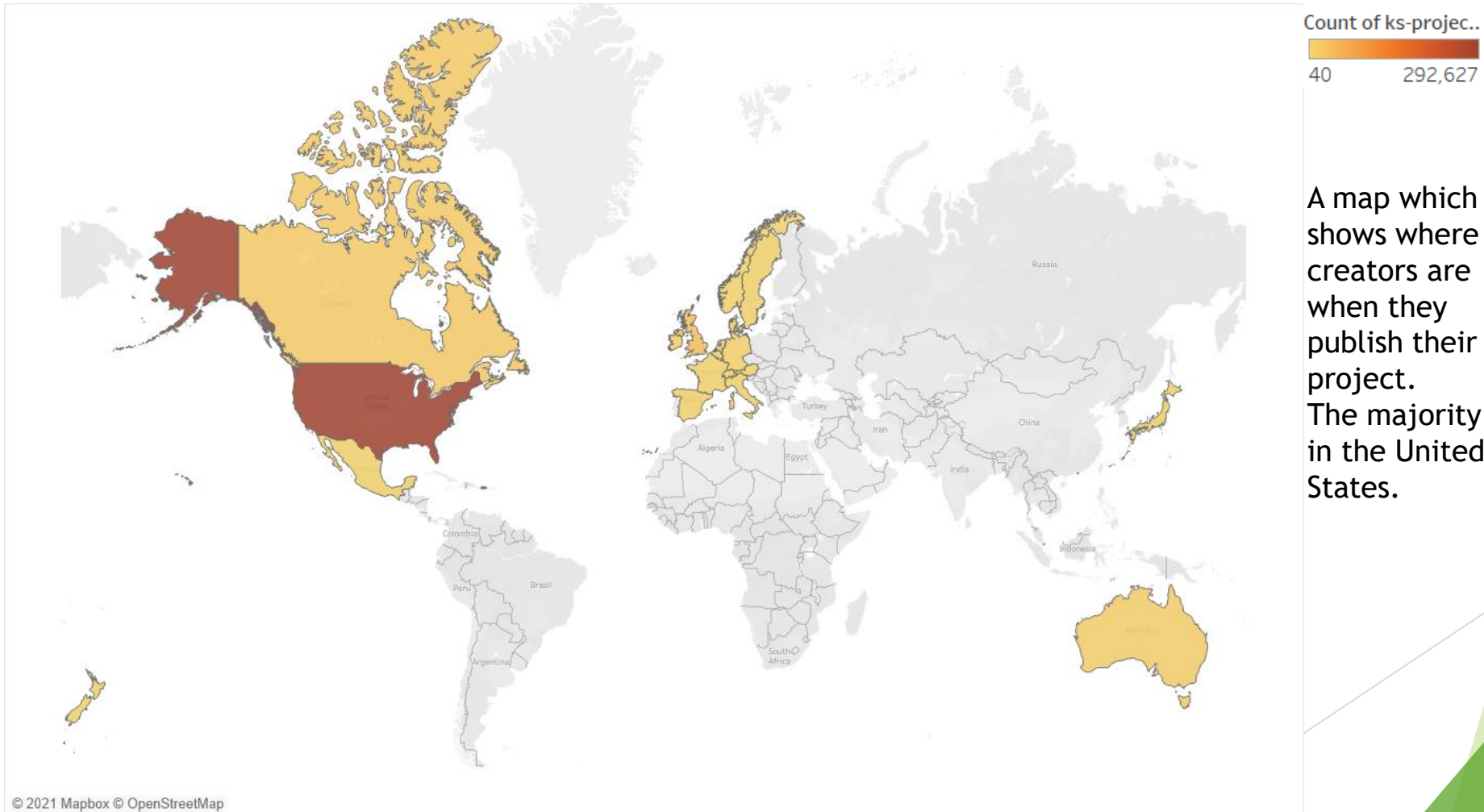
1. Dance
2. Theater
3. Comics
4. Music
5. Art

Attribute Visualization: Time Interval

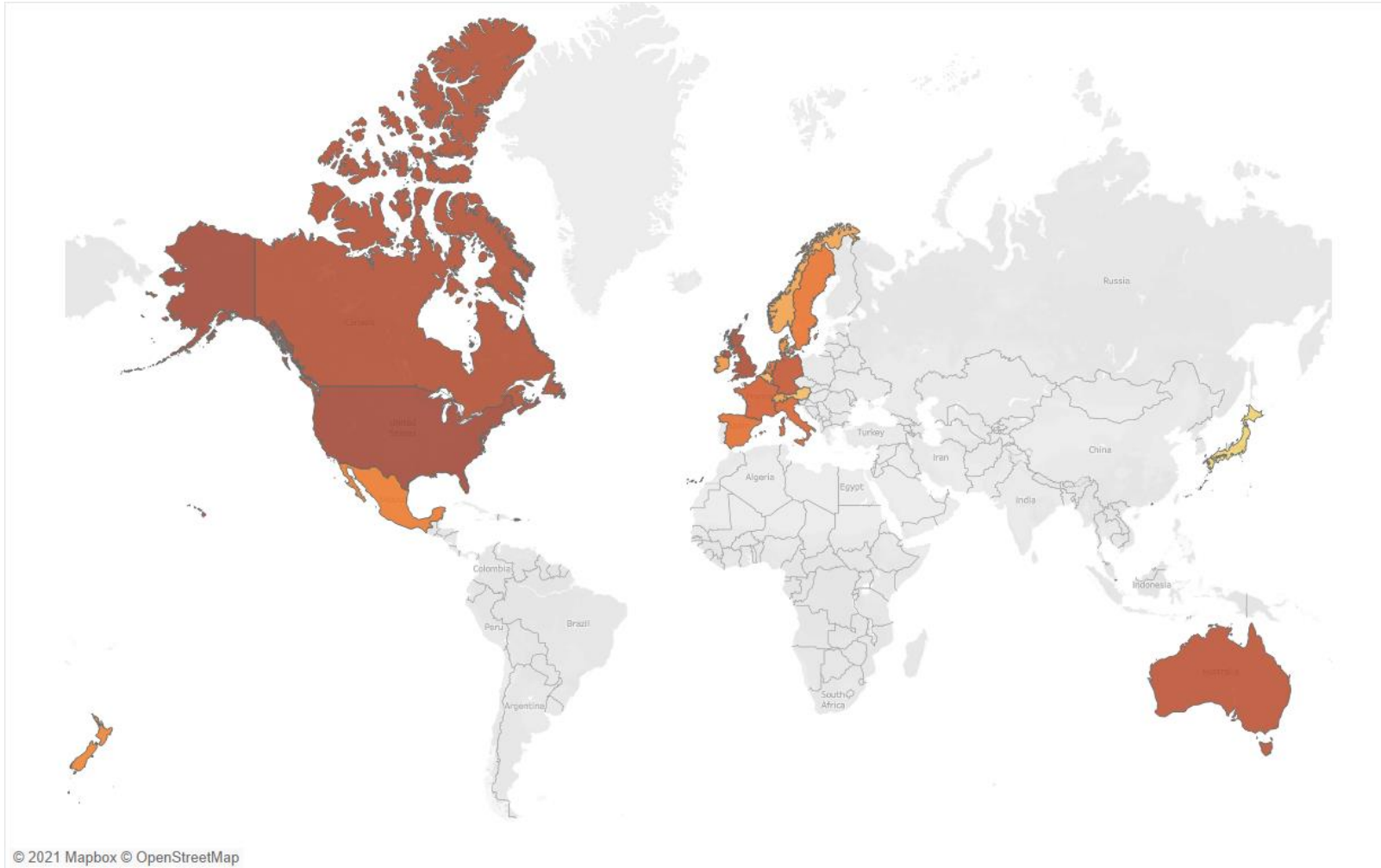


- ▶ Time Interval is the attribute created by calculating the difference of days between the Launch and Deadline attributes.
- ▶ The Histogram is showing the top 100 time ranges which is also filtered by bins with at least 1895 projects.
- ▶ The most common range is 30 days, followed by 60 days, and then 45 days.

Attribute Visualization: Country [# of projects]



Attribute Visualization: Country [Rank]



The countries of where the creator published their creation, ranked from most to least projects related to the location.

The top five countries are:

1. The US
2. The UK
3. Canada
4. Australia
5. Germany

Rank of Count of k..
1 23

Attribute Visualization: Country [Success %]

► It can be observed that the less popular your location is (meaning that there are less people submitting projects), the lower your chances of success is.

► Although whether this is because of the sample size in each location or whether it correlates to how the number of creators vs number of potential in your area affects your success rate (local projects may be shown first on the website) cannot be determined from just looking at the charts.

Success Rate per Country

State (gr..	Country				
	US	GB	CA	AU	DE
Success	41.82%	40.97%	33.42%	30.38%	27.27%
Failure	58.18%	59.03%	66.58%	69.62%	72.73%

Data Preprocessing

Got rid of attributes: Name, Category, Currency, Goal, Pledged, Backers, USD Pledged Real, USD Pledged.

Changed the formatting of Deadline DateTime to Date.

Removed all the apostrophes in order for WEKA to process it as a CSV file.

Got rid of instances in States: Canceled, Live, Undefined, and Suspended

Removed attributes Deadline and Launched after creating a new attribute called Time Interval which is the difference between the other two attributes. It should be noted that the Date values in Weka is in milliseconds.

There are no null values to address in the data set at this point.

Finding the Best Model

- ▶ For all the models used, 70% of the data set is used as a training set and the rest as the testing set
- ▶ Two metrics will be used to measure a model's performance: Accuracy and the Area under the ROC Curve
 - ▶ Overall accuracy will be the basis for comparing models since it determines how many projects were predicted correctly
 - ▶ Accuracy will be accompanied by the ROC Area in order to avoid any possible chance models that has gotten "lucky" in receiving a high accuracy rate
- ▶ I will be using the following models provided by Weka: ZeroR (Rule Based), J48 (Decision Tree), Decision Stump with AdaBoost, NaiveBayes, Logistic Regression, and Ibk (Nearest Neighbor).

ZeroR

- The ZeroR algorithm will be used as a baseline for all the other classifiers used for this data set.
- Since ZeroR predicts the majority class, in this case projects that has failed, correctly around 59.61% of the time, the best model to predict whether or not a project will fail should have a higher accuracy rate.
- A baseline ROC Area would be 0.50

=== Summary ===

Correctly Classified Instances	197719	59.6123 %
Incorrectly Classified Instances	133956	40.3877 %
Kappa statistic	0	
Mean absolute error	0.1605	
Root mean squared error	0.2833	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	331675	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.596	1.000	0.747	?	0.500	0.596	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.000	0.000	?	0.000	?	?	0.500	0.404	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.596	0.596	?	0.596	?	?	0.500	0.518	

```
=== Confusion Matrix ===
```

```

a      b      c      d      e      f      <-- classified as
197719 0      0      0      0      0      |      a = failed
0      0      0      0      0      0      |      b = canceled
133956 0      0      0      0      0      |      c = successful
0      0      0      0      0      0      |      d = live
0      0      0      0      0      0      |      e = undefined
0      0      0      0      0      0      |      f = suspended

```

Classification Algorithms: J48 Overview

[w/ MinBinSize: 3,790]

=== Summary ===

Correctly Classified Instances	215695	65.032 %
Incorrectly Classified Instances	115980	34.968 %
Kappa statistic	0.2428	
Mean absolute error	0.1468	
Root mean squared error	0.2709	
Relative absolute error	91.4428 %	
Root relative squared error	95.6415 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.793	0.560	0.676	0.793	0.730	0.249	0.668	0.717	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.440	0.207	0.590	0.440	0.504	0.249	0.668	0.562	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.650	0.417	0.641	0.650	0.639	0.249	0.668	0.655	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
156732	0	40987	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
74993	0	58963	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► This Decision Tree has one of the highest accuracy rates out of all the selected classifiers.

► But this can be mainly be attributed to its complexity.

► This decision tree has 33 leaves and a tree size of 52.

► This model has an accuracy rate of 65.03% and an ROC Area of .668.

► This tree uses the attributes Goal, Main Category, and Time Interval.

Classification Algorithms: J48 Overview

[w/ MinBinSize: 37,900]

=== Summary ===

Correctly Classified Instances	204696	61.7158 %
Incorrectly Classified Instances	126979	38.2842 %
Kappa statistic	0.1359	
Mean absolute error	0.1528	
Root mean squared error	0.2764	
Relative absolute error	95.2161 %	
Root relative squared error	97.5844 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.847	0.722	0.634	0.847	0.725	0.152	0.618	0.690	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.278	0.153	0.552	0.278	0.370	0.152	0.618	0.497	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.617	0.492	0.601	0.617	0.582	0.152	0.618	0.612	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
167402	0	30317	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
96662	0	37294	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► By adjusting the minimum number of instances a leaf node should have, the tree can be simplified at a cost of its accuracy.

► The tree has a size of 18 with an accurate rate of 61.72% and an ROC Area of 0.62.

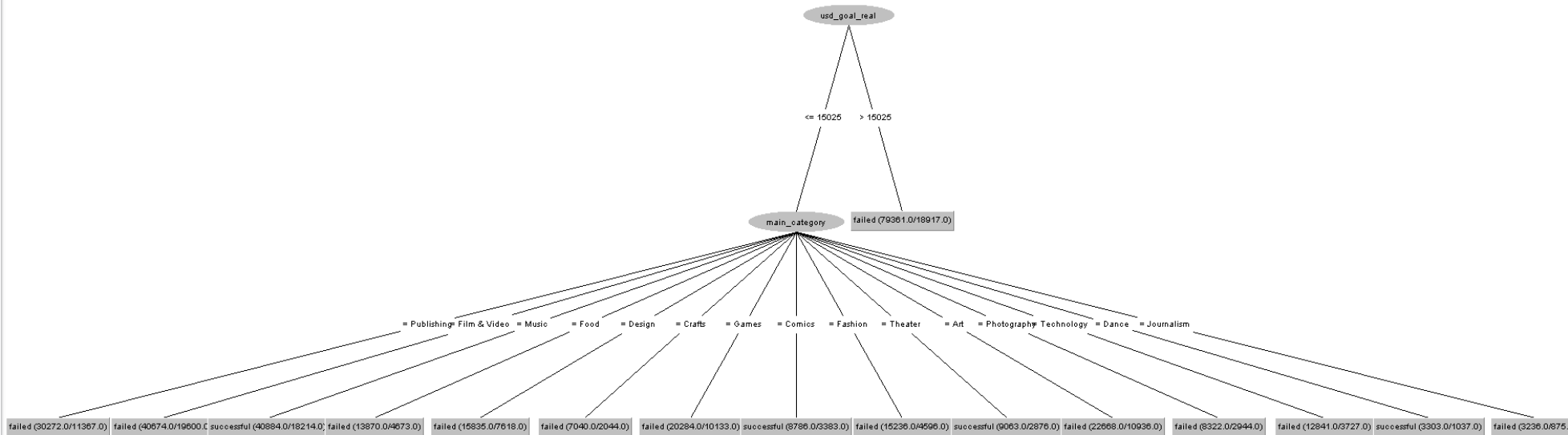
► This tree uses the attributes Goal and Main Category.

Classification Algorithms: J48 Tree

[Split w/ MinBinSize: 37,900]

Number of Leaves : 16

Size of the tree : 18



Classification Algorithms: J48 Overview

[w/ MinBinSize: 47,375*]

=== Summary ===

Correctly Classified Instances	204613	61.6908 %
Incorrectly Classified Instances	127062	38.3092 %
Kappa statistic	0.1211	
Mean absolute error	0.1532	
Root mean squared error	0.2767	
Relative absolute error	95.417 %	
Root relative squared error	97.6832 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.878	0.768	0.628	0.878	0.732	0.144	0.609	0.684	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.232	0.122	0.562	0.232	0.328	0.144	0.609	0.490	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.617	0.507	0.601	0.617	0.569	0.144	0.609	0.606	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
173555	0	24164	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
102898	0	31058	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► Increasing the minimum number of instances per leaf node further, the tree can be further simplified for only a small sacrifice in accuracy rate and ROC Area.

► This model has a tree size of 5, an accuracy rate of 61.69%, and an ROC Area of 0.609.

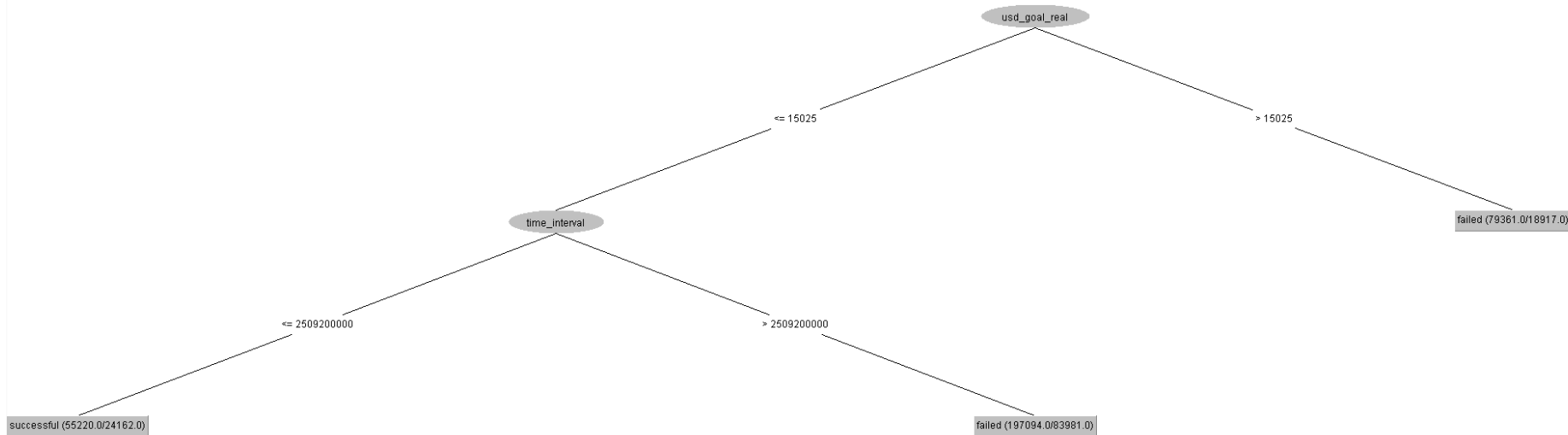
► This tree uses the attributes Goal and Time Interval.

Classification Algorithms: J48 Tree

[70:30 Split w/ MinBinSize: 47,375]

Number of Leaves : 3

Size of the tree : 5



Classification Algorithms: J48 Overview

[70:30 Split w/ MinBinSize: 56,850]

=== Summary ===

Correctly Classified Instances	202225	60.9708 %
Incorrectly Classified Instances	129450	39.0292 %
Kappa statistic	0.1203	
Mean absolute error	0.1576	
Root mean squared error	0.2807	
Relative absolute error	98.1825 %	
Root relative squared error	99.0947 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.838	0.727	0.630	0.838	0.719	0.134	0.556	0.627	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.273	0.162	0.533	0.273	0.361	0.134	0.556	0.460	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.610	0.499	0.591	0.610	0.574	0.134	0.556	0.559	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
165671	0	32048	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
97402	0	36554	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► To see how simplified the tree can get and to see its respective accuracy rate and ROC Area, the minimum bin size has been increased to 1.5% of the total size of the original data set.

► This tree only has 2 leaves and uses the attribute Goal.

► Although the difference in accuracy (60.97%) and ROC Area (0.556) is small compared to the last tree. These measures are very similar to the set baseline.

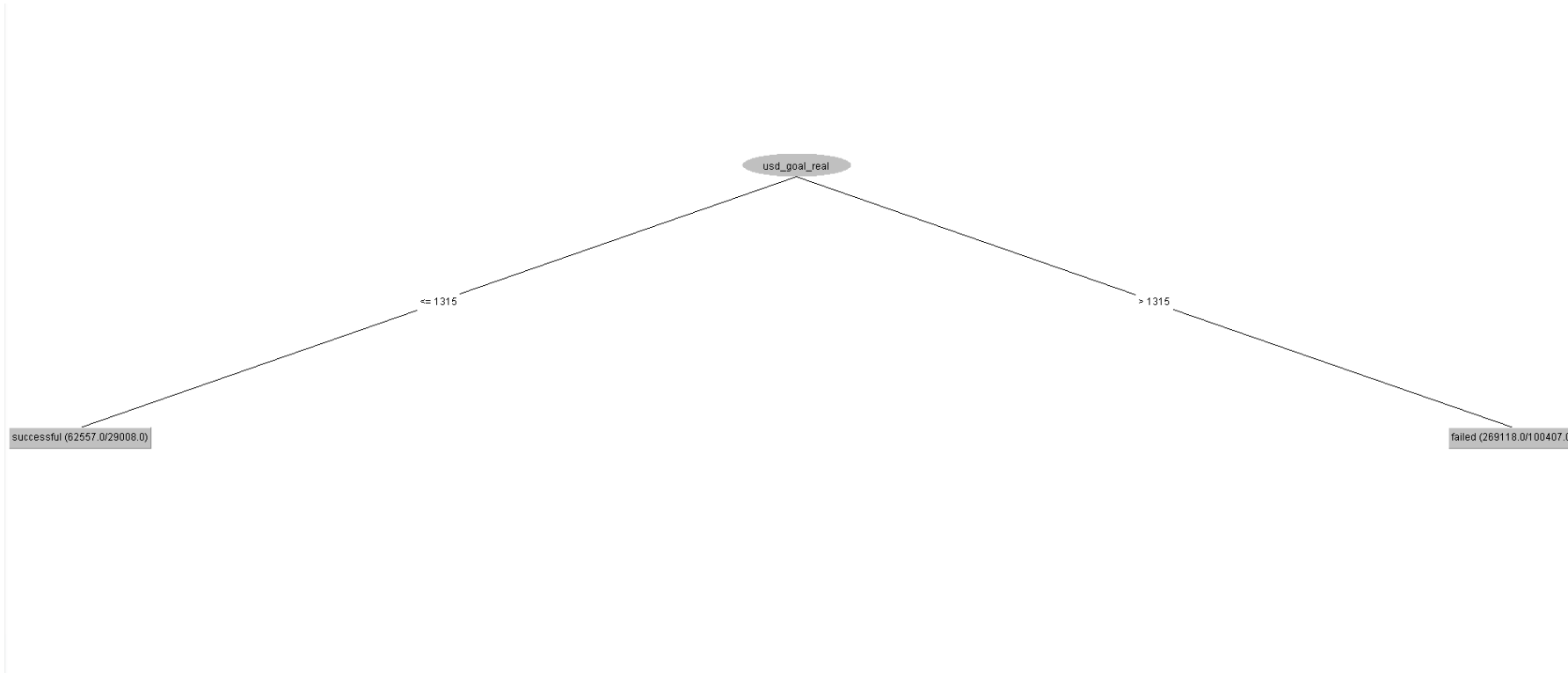
► This decision tree is essentially a decision stump.

Classification Algorithms: J48 Tree

[70:30 Split w/ MinBinSize: 56,850]

Number of Leaves : 2

Size of the tree : 3



Decision Stump w/ AdaBoost

=== Summary ===

Correctly Classified Instances	203724	61.4228 %
Incorrectly Classified Instances	127951	38.5772 %
Kappa statistic	0.1803	
Mean absolute error	0.2296	
Root mean squared error	0.3166	
Relative absolute error	143.0153 %	
Root relative squared error	111.7585 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.724	0.548	0.661	0.724	0.691	0.182	0.645	0.727	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.452	0.276	0.526	0.452	0.486	0.182	0.645	0.527	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.614	0.438	0.607	0.614	0.608	0.182	0.645	0.646	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
143131	0	54588	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
73363	0	60593	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► The accuracy for the Adaboosted Decision Stump was 61.42%

► The ROC Area was at 0.645

Decision Stump w/ AdaBoost

The Attribute Used for Comparison	Prediction \leq the Cutoff vs $>$ the Cutoff [$=$ vs \neq for nominal]	Weight
Goal: \$15,031	Failure / Failure	0.39
Goal: \$15,031	Success / Failure	0.33
Goal: \$15,031	Success / Failure	0.26
Time Interval: ~56 days	Failure / Failure	0.16
Time Interval: ~56 days	Success / Failure	0.13
Goal: \$101,101	Failure / Failure	0.10
Goal: \$101,101	Success / Failure	0.05
Time Interval: ~30 days	Success / Failure	0.16
Main Category: Music	Success / Failure	0.08
Goal: \$101,101	Failure / Failure	0.14

► The higher the weight given, the more likely it is to appear in the training set. The weight is set based on how much the algorithm misclassified, giving a chance to those incorrectly labeled to be processed once more for better accuracy.

► From here, it can be seen the attributes and their values that is most difficult to predict its result.

Naïve Bayes: Overview

=== Summary ===

Correctly Classified Instances	146029	44.0277 %
Incorrectly Classified Instances	185646	55.9723 %
Kappa statistic	0.0468	
Mean absolute error	0.1825	
Root mean squared error	0.4113	
Relative absolute error	113.7012 %	
Root relative squared error	145.1697 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.069	0.012	0.892	0.069	0.129	0.133	0.666	0.743	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.988	0.931	0.418	0.988	0.588	0.133	0.666	0.555	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.440	0.383	0.701	0.440	0.314	0.133	0.666	0.667	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
13733	0	183986	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
1660	0	132296	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► Very low accuracy rate can possibly be a result of using a simple model such as Naïve Bayes, which has a conditional independence assumption.

► However, the model does have a high ROC Area which might imply that the model may be affected by an imbalance of positive and negative classes since accuracy works on a set threshold while ROC covers a range of thresholds and the model's performance at each segment.

NaiveBayes

Naive Bayes Classifier

Attribute	Class failed (0.6)	canceled (0)	successful (0.4)
=====			
ID			
mean	1074638763.2147	0	1073727306.5258
std. dev.	619219908.2321	1079.1069	619149054.7248
weight sum	197719	0	133956
precision	6474.6415	6474.6415	6474.6415
main_category			
Publishing	23146.0	1.0	12301.0
Film & Video	32905.0	1.0	23624.0
Music	21753.0	1.0	24198.0
Food	15970.0	1.0	6086.0
Design	14815.0	1.0	10551.0
Crafts	5704.0	1.0	2116.0
Games	16004.0	1.0	12519.0
Comics	4037.0	1.0	5843.0
Fashion	14183.0	1.0	5594.0
Theater	3709.0	1.0	6535.0
Art	14132.0	1.0	11511.0
Photography	6385.0	1.0	3306.0
Technology	20617.0	1.0	6435.0
Dance	1236.0	1.0	2339.0
Journalism	3138.0	1.0	1013.0
[total]	197734.0	15.0	133971.0

country						
GB	17388.0	1.0	12068.0	1.0	1.0	1.0
US	152062.0	1.0	109300.0	1.0	1.0	1.0
CA	8237.0	1.0	4135.0	1.0	1.0	1.0
AU	4607.0	1.0	2011.0	1.0	1.0	1.0
NO	421.0	1.0	163.0	1.0	1.0	1.0
IT	1931.0	1.0	440.0	1.0	1.0	1.0
DE	2500.0	1.0	938.0	1.0	1.0	1.0
IE	477.0	1.0	208.0	1.0	1.0	1.0
MX	1016.0	1.0	397.0	1.0	1.0	1.0
ES	1382.0	1.0	493.0	1.0	1.0	1.0
N,0	106.0	1.0	106.0	1.0	1.0	1.0
SE	1001.0	1.0	510.0	1.0	1.0	1.0
FR	1613.0	1.0	909.0	1.0	1.0	1.0
NL	1795.0	1.0	618.0	1.0	1.0	1.0
NZ	827.0	1.0	449.0	1.0	1.0	1.0
CH	466.0	1.0	188.0	1.0	1.0	1.0
AT	379.0	1.0	108.0	1.0	1.0	1.0
DK	567.0	1.0	361.0	1.0	1.0	1.0
BE	372.0	1.0	153.0	1.0	1.0	1.0
HK	262.0	1.0	217.0	1.0	1.0	1.0
LU	39.0	1.0	20.0	1.0	1.0	1.0
SG	277.0	1.0	179.0	1.0	1.0	1.0
JP	17.0	1.0	8.0	1.0	1.0	1.0
[total]	197742.0	23.0	133979.0	23.0	23.0	23.0
usd_goal_real						
mean	63054.338	0	9399.3031	0	0	0
std. dev.	1435685.2394	620.9832	27990.0931	620.9832	620.9832	620.9832
weight sum	197719	0	133956	0	0	0
precision	3725.899	3725.899	3725.899	3725.899	3725.899	3725.899
time_interval						
mean	3039781352.5019	0	2779321530.6982	0	0	0
std. dev.	1138219774.6825	4910112.3596	1016906399.2454	4910112.3596	4910112.3596	4910112.3596
weight sum	197719	0	133956	0	0	0
precision	29460674.1573	29460674.1573	29460674.1573	29460674.1573	29460674.1573	29460674.1573

Logistic Regression: Overview

=== Summary ===

Correctly Classified Instances	214242	64.594 %
Incorrectly Classified Instances	117433	35.406 %
Kappa statistic	0.2228	
Mean absolute error	0.1462	
Root mean squared error	0.2699	
Relative absolute error	91.0688 %	
Root relative squared error	95.2591 %	
Total Number of Instances	331675	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.816	0.605	0.666	0.816	0.733	0.233	0.677	0.753	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.395	0.184	0.593	0.395	0.474	0.233	0.677	0.568	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.646	0.435	0.636	0.646	0.628	0.233	0.677	0.678	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
161385	0	36334	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
81099	0	52857	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

► Accuracy of 64.59% and ROC Area 0.677.

Logistic Regression: Coefficients

► The coefficients provided is the weights for the corresponding attribute for the logistic regression equation when determining if the project is a success or not.

► Since negative coefficients translate to an odds ratio lower than 1, it indicates that it has a lower chance to be that specific class.

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

Variable	Class failed	canceled	successful	live	undefined
ID	0	-0	0	-0	-0
main_category=Publishing	0.3364	-0.047	0.0279	-0.047	-0.047
main_category=Film & Video	-0.0078	-0.0027	0.1211	-0.0027	-0.0027
main_category=Music	-0.0866	-0.0056	0.3575	-0.0056	-0.0056
main_category=Food	0.4758	-0.1144	-0.0202	-0.1144	-0.1144
main_category=Design	-0.4295	0.1221	-0.2479	0.1221	0.1221
main_category=Crafts	1.6766	-0.4544	0.9352	-0.4544	-0.4544
main_category=Games	-0.4993	0.1451	-0.2877	0.1451	0.1451
main_category=Comics	-1.3227	0.3217	-0.6363	0.3217	0.3217
main_category=Fashion	0.5084	-0.0953	-0.0648	-0.0953	-0.0953
main_category=Theater	-1.3137	0.2704	-0.4152	0.2704	0.2704
main_category=Art	-0.1251	0.0584	-0.0364	0.0584	0.0584
main_category=Photography	0.0049	0.0558	-0.3305	0.0558	0.0558
main_category=Technology	0.3659	-0.1346	-0.1085	-0.1346	-0.1346
main_category=Dance	-0.9098	0.143	0.0167	0.143	0.143
main_category=Journalism	1.6906	-0.4997	0.9516	-0.4997	-0.4997
country=GB	-0.7039	0.2136	-0.6318	0.2136	0.2136
country=US	0.938	-0.2876	1.0781	-0.2876	-0.2876
country=CA	-0.7474	0.2361	-0.9084	0.2361	0.2361
country=AU	-0.8256	0.263	-1.0876	0.263	0.263
country=NO	-0.4912	0.1643	-0.9014	0.1643	0.1643
country=IT	-1.5453	0.4723	-2.3082	0.4723	0.4723
country=DE	-1.0059	0.3083	-1.3467	0.3083	0.3083
country=IE	-0.5329	0.1712	-0.8077	0.1712	0.1712
country=MX	-0.0509	0.054	-0.5877	0.054	0.054
country=ES	-0.6166	0.1968	-1.0451	0.1968	0.1968
country=N, 0	0.0715	0.0053	0.1598	0.0053	0.0053
country=SE	-0.534	0.1744	-0.6931	0.1744	0.1744
country=FR	-1.2783	0.3506	-1.164	0.3506	0.3506
country=NL	-0.7755	0.2397	-1.1856	0.2397	0.2397
country=NZ	-0.4115	0.1348	-0.5164	0.1348	0.1348
country=CH	-1.698	0.4764	-1.8083	0.4764	0.4764
country=AT	-0.2594	0.0681	-0.8199	0.0681	0.0681
country=DK	-0.7322	0.2028	-0.6414	0.2028	0.2028
country=BE	-1.147	0.3473	-1.3994	0.3473	0.3473
country=HK	-1.2001	0.2956	-0.7344	0.2956	0.2956
country=LU	-1.352	0.3792	-1.3121	0.3792	0.3792
country=SG	-0.7946	0.2232	-0.6452	0.2232	0.2232
country=JP	-1.0751	0.3619	-1.3668	0.3619	0.3619
usd_goal_real	0	0	-0	0	0
time_interval	0	-0	0	-0	-0
Intercept	24.2489	-3.7158	24.5434	-3.7158	-3.7158

Logistic Regression: Odds Ratio

Odds Ratios...

Variable	Class failed	canceled	successful	live	undefined
ID	1	1	1	1	1
main_category=Publishing	1.3999	0.9541	1.0283	0.9541	0.9541
main_category=Film & Video	0.9922	0.9973	1.1288	0.9973	0.9973
main_category=Music	0.9171	0.9944	1.4297	0.9944	0.9944
main_category=Food	1.6094	0.8919	0.98	0.8919	0.8919
main_category=Design	0.6508	1.1299	0.7804	1.1299	1.1299
main_category=Crafts	5.3475	0.6348	2.5478	0.6348	0.6348
main_category=Games	0.607	1.1561	0.75	1.1561	1.1561
main_category=Comics	0.2664	1.3795	0.5293	1.3795	1.3795
main_category=Fashion	1.6626	0.9091	0.9372	0.9091	0.9091
main_category=Theater	0.2688	1.3105	0.6602	1.3105	1.3105
main_category=Art	0.8824	1.0601	0.9643	1.0601	1.0601
main_category=Photography	1.0049	1.0574	0.7186	1.0574	1.0574
main_category=Technology	1.4418	0.8741	0.8972	0.8741	0.8741
main_category=Dance	0.4026	1.1537	1.0169	1.1537	1.1537
main_category=Journalism	5.4226	0.6067	2.5899	0.6067	0.6067
country=GB	0.4946	1.2382	0.5316	1.2382	1.2382
country=US	2.5547	0.7501	2.9392	0.7501	0.7501
country=CA	0.4736	1.2663	0.4032	1.2663	1.2663
country=AU	0.438	1.3008	0.337	1.3008	1.3008
country=NO	0.6119	1.1785	0.406	1.1785	1.1785
country=IT	0.2132	1.6037	0.0994	1.6037	1.6037
country=DE	0.3657	1.3611	0.2601	1.3611	1.3611
country=IE	0.5869	1.1867	0.4459	1.1867	1.1867
country=MX	0.9504	1.0555	0.5556	1.0555	1.0555
country=ES	0.5398	1.2175	0.3517	1.2175	1.2175
country=N, 0	1.0742	1.0053	1.1733	1.0053	1.0053
country=SE	0.5863	1.1906	0.5	1.1906	1.1906
country=FR	0.2785	1.4199	0.3122	1.4199	1.4199
country=NL	0.4605	1.2709	0.3056	1.2709	1.2709
country=NZ	0.6626	1.1443	0.5967	1.1443	1.1443
country=CH	0.183	1.6103	0.1639	1.6103	1.6103
country=AT	0.7715	1.0704	0.4405	1.0704	1.0704
country=DK	0.4808	1.2248	0.5266	1.2248	1.2248
country=BE	0.3176	1.4152	0.2467	1.4152	1.4152
country=HK	0.3012	1.3439	0.4798	1.3439	1.3439
country=LU	0.2587	1.4611	0.2693	1.4611	1.4611
country=SG	0.4518	1.25	0.5245	1.25	1.25
country=JP	0.3412	1.4361	0.2549	1.4361	1.4361
usd_goal_real	1	1	1	1	1
time_interval	1	1	1	1	1

If the Odd Ratio is higher than 1, that means there is a higher probability for that instance to be classified in that specific class. If it's lower than 1, then it has a lower probability. 1 means equal probability for either scenarios.

Film & Video, Music, and Dance for Main Category, and US for Country has a higher Odds Ratio for Success and is greater than 1.

These attributes are favorable when considering starting a project at Kickstarter.

On the other hand, Publishing, Food, Crafts, Fashion, Photography, Technology, and Journalism for Main Category are unfavorable.

Games, Comics, Theater, and Art for Main Categories and GB, FR, DK, HK, and SG for Country has a lower Odds Ratio for Failure and is lower than 1.

=== Summary ===

```
Correctly Classified Instances  187395      56.4996 %
Incorrectly Classified Instances 144280      43.5004 %
Kappa statistic                0.098
Mean absolute error            0.145
Root mean squared error        0.3808
Relative absolute error         90.34 %
Root relative squared error     134.4156 %
Total Number of Instances      331675
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.632	0.534	0.636	0.632	0.634	0.098	0.549	0.624	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.466	0.368	0.462	0.466	0.464	0.098	0.549	0.435	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.565	0.467	0.566	0.565	0.565	0.098	0.549	0.548	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
124955	0	72764	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
71516	0	62440	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

=== Summary ===

```
Correctly Classified Instances  63497      63.8148 %
Incorrectly Classified Instances 36005      36.1852 %
Kappa statistic                0.1929
Mean absolute error            0.1474
Root mean squared error        0.273
Relative absolute error         91.8028 %
Root relative squared error     96.3624 %
Total Number of Instances      99502
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.842	0.662	0.652	0.842	0.735	0.209	0.656	0.722	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.338	0.158	0.592	0.338	0.430	0.209	0.656	0.550	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.638	0.459	0.628	0.638	0.612	0.209	0.656	0.652	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
49915	0	9377	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
26628	0	13582	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

=== Summary ===

```
Correctly Classified Instances  62980      69.2952 %
Incorrectly Classified Instances 36522      36.7048 %
Kappa statistic                0.1755
Mean absolute error            0.1507
Root mean squared error        0.274
Relative absolute error         93.8599 %
Root relative squared error     96.7082 %
Total Number of Instances      99502
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.851	0.689	0.646	0.851	0.734	0.194	0.648	0.718	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.311	0.149	0.587	0.311	0.406	0.194	0.648	0.542	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.633	0.471	0.622	0.633	0.602	0.194	0.648	0.647	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
50479	0	8813	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
27709	0	12501	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

lbk (1 / 100 / 1000 nearest neighbor)

► The lbk algorithm was first used with the parameters of searching the for 1, 100, and 1000 nearest neighbor.

► These results will be used as a reference for figuring out the optimal k number of nearest neighbors to use

► The better performance was closer around the model that searched for the 100th nearest neighbor.

```
=== Summary ===
```

```
Correctly Classified Instances    63335          63.652 %
Incorrectly Classified Instances  36167          36.348 %
Kappa statistic                  0.1855
Mean absolute error              0.1488
Root mean squared error          0.2731
Relative absolute error          92.6779 %
Root relative squared error      96.4092 %
Total Number of Instances       99502
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.850	0.678	0.649	0.850	0.736	0.204	0.654	0.723	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.322	0.150	0.592	0.322	0.417	0.204	0.654	0.549	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.637	0.465	0.626	0.637	0.607	0.204	0.654	0.652	

```
=== Confusion Matrix ===
```

	a	b	c	d	e	f	<-- classified as
50380	0	8912	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
27255	0	12955	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

```
=== Summary ===
```

```
Correctly Classified Instances    63093          63.4088 %
Incorrectly Classified Instances  36409          36.5912 %
Kappa statistic                  0.1795
Mean absolute error              0.1498
Root mean squared error          0.2736
Relative absolute error          93.3394 %
Root relative squared error      96.5621 %
Total Number of Instances       99502
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.849	0.683	0.647	0.849	0.734	0.197	0.651	0.720	failed
	?	0.000	?	?	?	?	?	?	canceled
	0.317	0.151	0.588	0.317	0.412	0.197	0.651	0.545	successful
	?	0.000	?	?	?	?	?	?	live
	?	0.000	?	?	?	?	?	?	undefined
	?	0.000	?	?	?	?	?	?	suspended
Weighted Avg.	0.634	0.468	0.623	0.634	0.604	0.197	0.651	0.649	

```
=== Confusion Matrix ===
```

	a	b	c	d	e	f	<-- classified as
50338	0	8954	0	0	0	0	a = failed
0	0	0	0	0	0	0	b = canceled
27455	0	12755	0	0	0	0	c = successful
0	0	0	0	0	0	0	d = live
0	0	0	0	0	0	0	e = undefined
0	0	0	0	0	0	0	f = suspended

lbc (616/308* nearest neighbor)

- Using the square root of the number of total objects in the data set, I have found a slightly better performance in comparison to the previous 3 settings.
- To double check on how much more the model can be improved, I decided to use half of that as well to get a gauge of how much more it will improve its performance.
- The lbc model that uses the first 308th nearest neighbors performed the best although with not much improvement from the one that uses the 616th nearest neighbor.
- The best model had an accuracy of 63.41% and an ROC Area of 0.654

Conclusion

- ▶ The kind of creators that are debating on whether they should fundraise at Kickstarter should consider whether their project is within the realm of creative media designed for entertainment, has a fundraising goal more towards a short term coverage that is generally low (under \$1,400), and has around 30 days to wait for the fundraising to end but no more longer. If they meet these criteria, they are starting off on a good chance, given that their work is on par with its peers.
- ▶ Based on the decision trees created and the odds ratio from the logistic regression, these attributes has an impact on your chances of meeting your fundraising goals:
 - ▶ Time Interval: most models points toward aiming towards having around 30 days or less when fundraising for the project
 - ▶ Goal: The threshold for goal setting is debatable but setting a goal somewhere around \$1,400 seems to suggest a higher chance of succeeding. In general, anything higher than this can be considered as setting a high goal which leads to a problem of meeting it in the first place.
 - ▶ Main Category: Film & Video, Music, Games, Comics, Dance, Theatre, and Art experienced more chances of success. The rest tend to experience a less successful fundraising attempt.
- ▶ The most useful model out of the 6 models was Logistic Regression with an accuracy rate of 64.59% and ROC Curve Area of 0.677. This placement was followed by the IBk Lazy Learner with 308 nearest neighbors and Decision Tree with a minimum bin size of 47,375.