Success factors of Kickstarter campaigns

...

Seokhyun Ryu, A13884401
Jieun Lee, A16551797
Gauri Samith, A15828385
Aditi Krishnakumar, A15914778

What is Kickstarter?

Introduction to the platform

- American public benefit corporation based in Brooklyn, New York,
- global crowdfunding platform focused on creativity.
- Mission: "help bring creative projects to life".

https://www.slashgear.com/is-sony-poisoning-the-kickstarter-well-25390489/

What is Kickstarter?

How it works

- Every project creator sets their project's funding goal and deadline.
- If people like the project, they can pledge money to make it happen.
- If the project falls short of its funding goal, no one is charged.

Research Question:

What is the relationship between funding goal, main category, and location of a Kickstarter campaign and its success?

Hypothesis:

We hypothesize that the main category, funding goal, and location of a Kickstarter campaign will determine the success of a campaign.

Predictions:

- lower funding goal → higher chance of success
- Western countries → higher chance of success
- Technology campaigns → higher chance of success.

Understanding the dataset

- Dataset Name: ks-projects-201801.csv
- Number of observations: 378660
- Link to the dataset: https://www.kaggle.com/kemical/kickstarter-projects

Preview of dataset (first few rows)

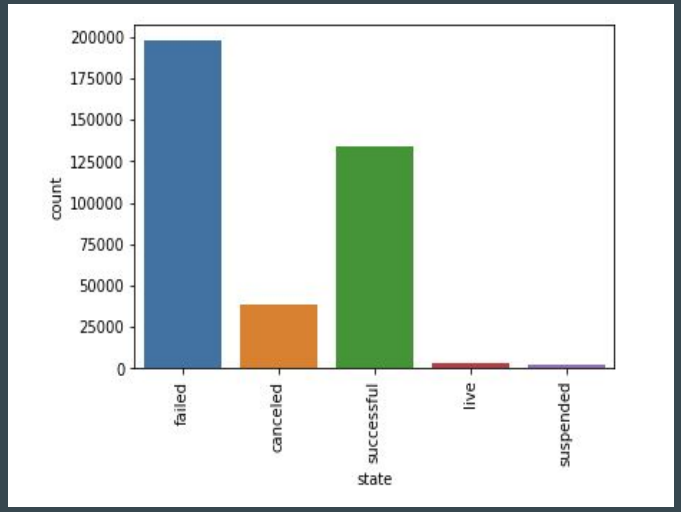
ID	name	category	main_cate...	currency	deadline	goal	launched	pledged	state	backers	country	usd pledged	usd_pledg...	usd_goal_r...
1000002330	The Songs of Adelaide & Abullah	Poetry	Publishing	GBP	2015-10-09	1000.00	2015-08-11 12:12:28	0.00	failed	0	GB	0.00	0.00	1533.95
1000003930	Greeting From Earth: ZGAC Arts Capsule For ET	Narrative Film	Film & Video	USD	2017-11-01	30000.00	2017-09-02 04:43:57	2421.00	failed	15	US	100.00	2421.00	30000.00
1000004038	Where is Hank?	Narrative Film	Film & Video	USD	2013-02-26	45000.00	2013-01-12 00:20:50	220.00	failed	3	US	220.00	220.00	45000.00
1000007540	ToshiCapital Rekordz Needs Help to Complete Album	Music	Music	USD	2012-04-16	5000.00	2012-03-17 03:24:11	1.00	failed	1	US	1.00	1.00	5000.00
1000011046	Community Film Project: The Art of Neighborhood Filmmaking	Film & Video	Film & Video	USD	2015-08-29	19500.00	2015-07-04 08:35:03	1283.00	canceled	14	US	1283.00	1283.00	19500.00
1000014025	Monarch Espresso Bar	Restaurants	Food	USD	2016-04-01	50000.00	2016-02-26 13:38:27	52375.00	successful	224	US	52375.00	52375.00	50000.00
1000023410	Support Solar Roasted Coffee & Green Energy! SolarCoffee.co	Food	Food	USD	2014-12-21	1000.00	2014-12-01 18:30:44	1205.00	successful	16	US	1205.00	1205.00	1000.00
1000030581	Chaser Strips. Our Strips make Shots their B*tch!	Drinks	Food	USD	2016-03-17	25000.00	2016-02-01 20:05:12	453.00	failed	40	US	453.00	453.00	25000.00

https://www.kaggle.com/kemical/kickstarter-projects

Columns/Variables we are interested in

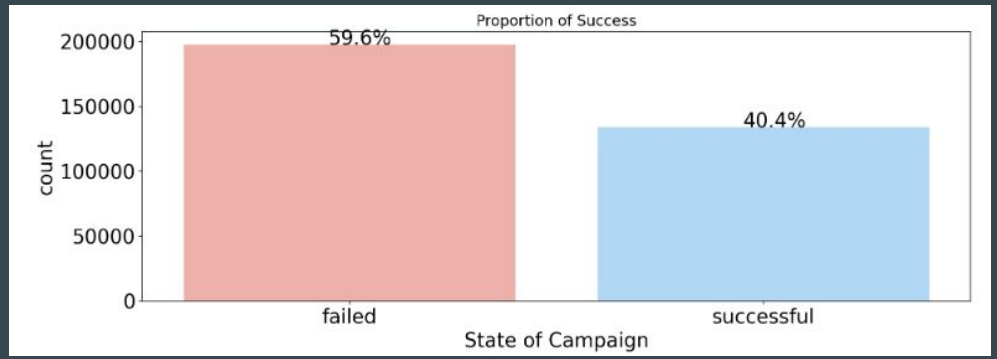
- Main_category
- Country
- Usd_goal_real
- State

State



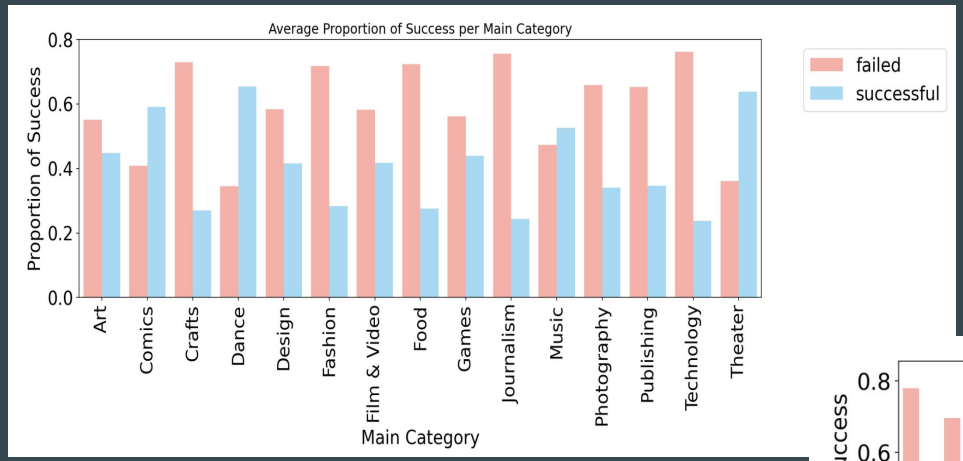
failed	197719
successful	133956
canceled	38779
undefined	3562
live	2799
suspended	1846

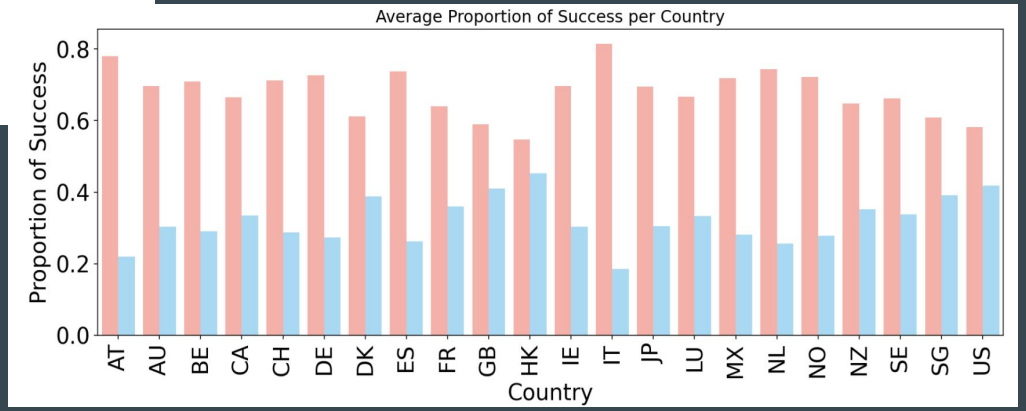
- 5 different states
- Focusing on failed and Successful



https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

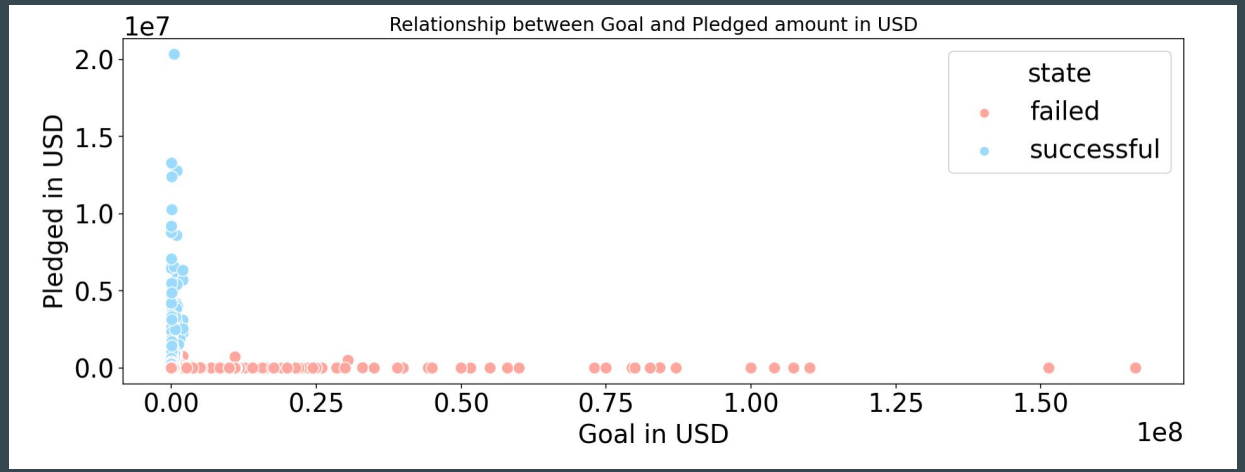
Exploratory Highlights





https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Usd_goal_real



- The successful campaigns have much lower goal amounts than the failed campaigns

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Analysis : relationship between variables

- T-test
- Linear Regression
- Logistic Regression
- ANOVA
- Chi-Squared test

T-test - Goal (US$)

```
success = ks_projects_2[ks_projects_2.get('state') == 'successful']['usd_goal_real']

fail = ks_projects_2[ks_projects_2.get('state') == 'failed']['usd_goal_real']
t_val, p_val = stats.ttest_ind(success, fail)
print("T-value: ", t_val)
print("P-value: ", p_val)
```

```
T-value:  -13.933903139342407
P-value:  4.0569274163690045e-44
```

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Chi-Squared Test : State and Location(country)

- **Null Hypothesis** : The two categorical variables, State & Location, are independent to each other.
- **Alternate Hypothesis :** The two categorical variables, State & Location, are dependent to each other.
- P-value: 0.000
    - Reason: Large Dataset

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Chi-Squared Test : State and Main Category

- **Null Hypothesis :** The two categorical variables, State & Location, are independent to each other.
- **Alternate Hypothesis :** The two categorical variables, State & Location, are dependent to each other.
- P-value: 0.000
    - Reason: Large Dataset

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Logistic regression for country and state

Exponential of Coefficients of countries

```
country_AU    0.436387
country_BE    0.409704
country_CA    0.501943
country_CH    0.402151
country_DE    0.374950
country_DK    0.636042
country_ES    0.356264
country_FR    0.563275
country_GB    0.694144
country_HK    0.827586
country_IE    0.434874
country_IT    0.227461
country_JP    0.437500
country_LU    0.500000
country_MX    0.390148
country_NL    0.343924
country_NO    0.385714
country_NZ    0.542373
country_SE    0.509000
country_SG    0.644928
country_US    0.718793
dtype: float64
```

- Hong Kong has the highest odds ratio → Campaign in Hong Kong have the highest proportion of success compared to other countries.
- R-squared value is low which means that location is not a good predictor overall of success.

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Logistic regression for main_category and state

Exponential of Coefficients of countries

```
main_category_Comics            1.447473
main_category_Crafts            0.370857
main_category_Dance             1.893117
main_category_Design            0.712097
main_category_Fashion           0.394401
main_category_Film & Video      0.717908
main_category_Food              0.381075
main_category_Games             0.782277
main_category_Journalism        0.322704
main_category_Music             1.111034
main_category_Photography       0.517701
main_category_Publishing        0.532168
main_category_Technology        0.312085
main_category_Theater           1.762136
dtype: float64
```

- Dance has the highest odds ratio which is consistent with our EDA results, however the R-squared value is low which means that category is not a good predictor overall of success.

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Linear modeling - pledged_prop ~ main_category

OLS Regression Results

Dep. Variable:	pledged_prop	R-squared:	0.000
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	2.592
Date:	Wed, 17 Mar 2021	Prob (F-statistic):	0.000947
Time:	20:54:59	Log-Likelihood:	-2.3411e+06
No. Observations:	331460	AIC:	4.682e+06
Df Residuals:	331445	BIC:	4.682e+06
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.7107	1.765	1.536	0.125	-0.749	6.170
main_category[T.Comics]	4.2360	3.347	1.266	0.206	-2.324	10.796
main_category[T.Crafts]	0.0070	3.651	0.002	0.998	-7.149	7.163
main_category[T.Dance]	-1.8647	5.047	-0.369	0.712	-11.756	8.027
main_category[T.Design]	-0.4258	2.503	-0.170	0.865	-5.331	4.480
main_category[T.Fashion]	-1.6643	2.675	-0.622	0.534	-6.907	3.578
main_category[T.Film & Video]	-1.0887	2.128	-0.512	0.609	-5.260	3.082
main_category[T.Food]	-1.7942	2.596	-0.691	0.489	-6.881	3.293
main_category[T.Games]	5.5919	2.432	2.299	0.022	0.825	10.359
main_category[T.Journalism]	-2.1453	4.730	-0.454	0.650	-11.416	7.125
main_category[T.Music]	6.1248	2.204	2.779	0.005	1.804	10.445
main_category[T.Photography]	-2.0729	3.370	-0.615	0.539	-8.678	4.533
main_category[T.Publishing]	-0.7926	2.317	-0.342	0.732	-5.335	3.750
main_category[T.Technology]	0.3400	2.463	0.138	0.890	-4.488	5.168
main_category[T.Theater]	-1.6515	3.304	-0.500	0.617	-8.126	4.823

Omnibus:	1713518.184	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	74990957830431.078
Skew:	243.251	Prob(JB):	0.00
Kurtosis:	73689.010	Cond. No.	15.7

- Significant p-value for Games and Music
- R-squared value is 0 => model does not explain the variability along the mean well.

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Linear modeling - pledged_prop ~ country

OLS Regression Results

Dep. Variable:	pledged_prop	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.000
Method:	Least Squares	F-statistic:	0.2492
Date:	Wed, 17 Mar 2021	Prob (F-statistic):	1.00
Time:	20:59:48	Log-Likelihood:	-2.3411e+06
No. Observations:	331460	AIC:	4.682e+06
Df Residuals:	331438	BIC:	4.683e+06
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.6355	12.834	0.050	0.961	-24.518	25.789
country[T.AU]	0.6152	13.296	0.046	0.963	-25.444	26.675
country[T.BE]	0.0159	17.817	0.001	0.999	-34.905	34.936
country[T.CA]	1.7312	13.083	0.132	0.895	-23.911	27.373
country[T.CH]	0.3329	16.948	0.020	0.984	-32.884	33.550
country[T.DE]	0.3734	13.710	0.027	0.978	-26.497	27.244
country[T.DK]	0.2313	15.842	0.015	0.988	-30.819	31.281
country[T.ES]	0.3208	14.400	0.022	0.982	-27.902	28.544
country[T.FR]	1.7073	14.014	0.122	0.903	-25.761	29.175
country[T.GB]	0.5441	12.939	0.042	0.966	-24.816	25.904
country[T.HK]	1.4584	18.226	0.080	0.936	-34.263	37.180
country[T.IE]	0.0094	16.783	0.001	1.000	-32.884	32.903
country[T.IT]	0.3753	14.086	0.027	0.979	-27.233	27.984
country[T.JP]	-0.1435	60.314	-0.002	0.998	-118.358	118.071
country[T.LU]	-0.0556	39.574	-0.001	0.999	-77.620	77.509
country[T.MX]	0.6199	14.877	0.042	0.967	-28.538	29.778
country[T.NL]	0.2135	14.065	0.015	0.988	-27.354	27.781
country[T.NO]	-0.0431	17.377	-0.002	0.998	-34.101	34.015
country[T.NZ]	1.0768	15.080	0.071	0.943	-28.480	30.633
country[T.SE]	0.3397	14.753	0.023	0.982	-28.575	29.255
country[T.SG]	0.7391	18.457	0.040	0.968	-35.436	36.914
country[T.US]	3.4444	12.846	0.268	0.789	-21.733	28.621

Omnibus:	1713534.337	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	74994622639331.719
Skew:	243.260	Prob(JB):	0.00
Kurtosis:	73690.811	Cond. No.	175.

- No significant p-value
- R-squared value is 0 => model does not explain the variability along the mean well.

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

ANOVA

- **Null Hypothesis** : There is no significant difference between mean proportion pledged per country
- **Alternate Hypothesis** : There is a significant difference between mean proportion pledged per country

	sum_sq	df	F	PR(>F)
country	4.180067e+05	21.0	0.249182	0.999808
Residual	2.647578e+10	331438.0	NaN	NaN

Null Hypothesis : There is no significant difference between mean proportion pledged per country

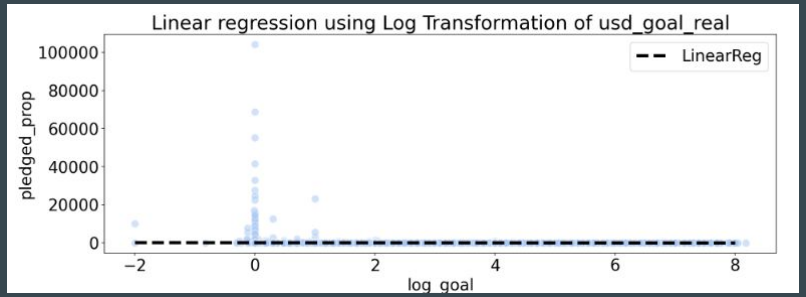
https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Log Transformation of usd_goal_real

OLS Regression Results

Dep. Variable:	pledged_prop	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.002
Method:	Least Squares	F-statistic:	717.8
Date:	Wed, 17 Mar 2021	Prob (F-statistic):	5.86e-158
Time:	20:46:47	Log-Likelihood:	-2.3408e+06
No. Observations:	331460	AIC:	4.682e+06
Df Residuals:	331458	BIC:	4.682e+06
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	70.8798	2.562	27.668	0.000	65.859	75.901
log_goal	-18.0772	0.675	-26.792	0.000	-19.400	-16.755



$$ProportionPledged = 70.8798 - 18.0772 * log_{10}(Goal)$$

- used logarithmic scaling on our 'usd_goal_real' variable before conducting linear regression
- For every 1 unit increase in Goal, there is a 0.180722 unit decrease in Proportion Pledged.

https://github.com/COGS108/group072_wi21/blob/main/recovered_FinalProject_group072.ipynb

Ethics

- **Permission**:
  - We are using “Kickstarter Projects” dataset from **Kaggle**, an online community of data scientists and machine learning practitioners, which allows the public to explore, publish, and access datasets.
- **Privacy**:
  - We are focusing on specific variables from the dataset, such as “main_category”, “state”, “country”, “usd_pledged_real”, and “usd_goal_real”. According to Kickstarter’s policy on “Trust & Safety” states that details about the project and information, including variables stated above, are publicly provided for the backers to explore different projects. Furthermore, these data does not access confidential information.

Recall: Hypothesis

We predicted that a lower funding goal will lead to a higher chance of success and that campaigns located in western countries in the category of Technology will have a higher chance of success.

Result

- Campaigns with lower funding locals had a much higher success rate.
- Main category and location do not act as significant predictors of success.
- Technology had the highest proportion of failed campaigns.
- Hong Kong had the highest proportion of success.