

WHAT'S KICKING IN

KICKSTARTER?



ISSS616 Applied Statistical
Analysis with R

Group 6

Archie Quiambao **DOLIT**

MIN Xiaoqi

Kevin Magic Rialubin **SUNGA**

WANG Zhifei

Lynette **WONG** Yik Lynn

WONG Sook Xian Sophia

AGENDA

01

WHAT DO WE KNOW ABOUT KICKSTARTER?

02

WHAT ARE THE PROBLEMS FACED BY CREATORS AND BACKERS?

03

HOW CAN OUR R SHINY APP HELP CREATORS AND BACKERS?

04

WHAT IS OUR ANALYTICAL PROCESS?

05

LET'S WALK YOU THROUGH OUR SHINY APP!

06

CONCLUSION

KICKSTARTER



Successfully
Funded Projects

199,326

\$5,700,456,256

total dollars pledged to Kickstarter projects

Less than \$1,000 Raised	\$1,000 to \$9,999 Raised	\$10,000 to \$19,999 Raised	\$20,000 to \$99,999 Raised	\$100 K to \$999,999 Raised	\$1 M Raised
25,936	107,071	28,650	29,465	7,670	534

- Online platform for creative ideas



- Funders/start-ups present ideas → supported by people worldwide (Crowd funding)

Since its launch on 28 April 2009,



Backers

19.4
million



Pledged

\$5.7
billion

FLOW OF KICK STARTER

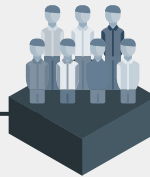
STAGE 1

Creators **set monetary goal & deadline**



STAGE 2

Backers **pledge money** if they like the project



STAGE 3

Campaign status **(All-or Nothing)**:

Success: when goal amount met by set deadline

Failed: does not meet funding goal, creators receives no money



DISPERSED INFO & LIMITED RESOURCES

- Accessed from different websites
- Limited data analysis
- Past project performance

TOO MANY FACTORS TO CONSIDER IN PROJECT PLANNING

- One-man show or a very small team of creators
→ difficult to make decisions

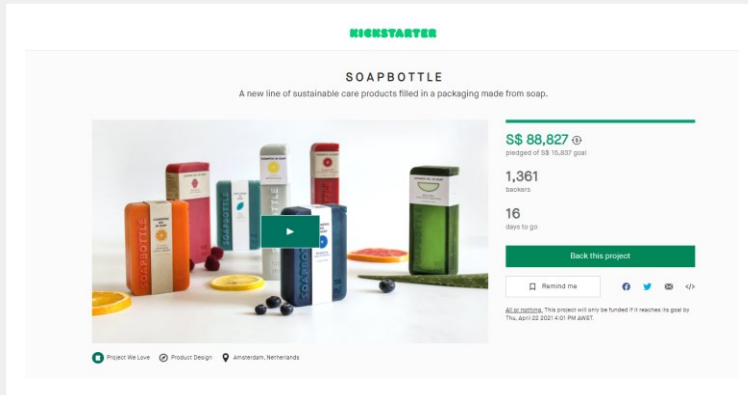
MAKING THE CHOICE

- Range of variables to quantify
- Backer: successful project, historical data
- Creator: What's the optimal amount, date, topic, success rate etc.

PROBLEMS FACED BY CREATORS AND BACKERS?

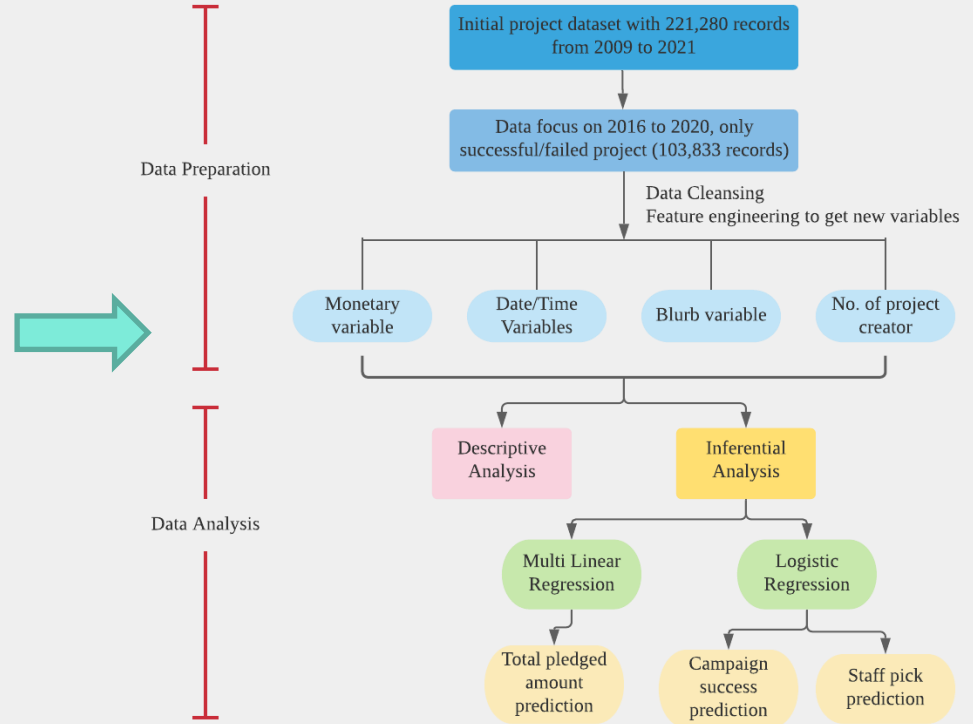


KICKSTARTER



Data **crawling** from Kickstarter website

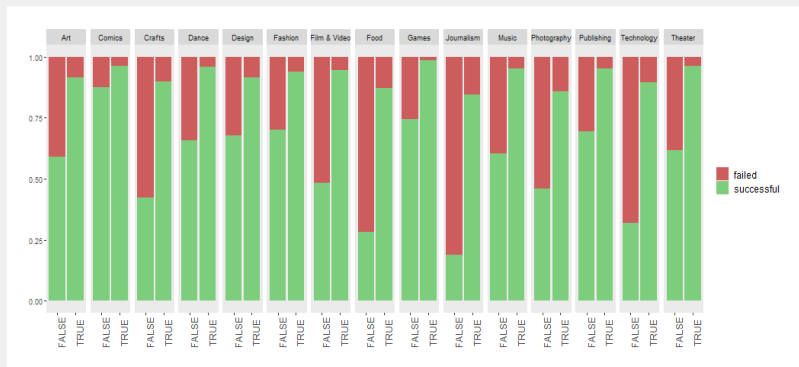
ANALYTICAL APPROACH



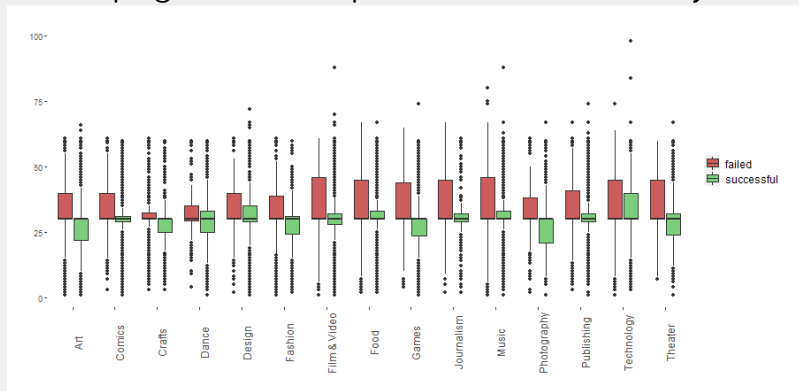
TRENDS AND INSIGHTS

Success Rate [%]

Staff Pick: higher success rate across all categories

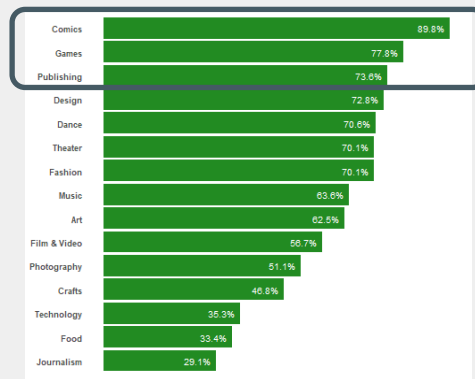


Campaign Duration: optimal duration <= 30 days



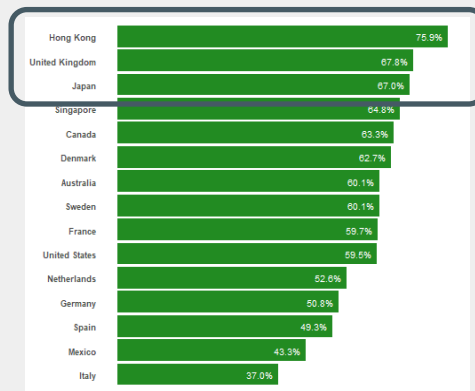
Category

Top 3

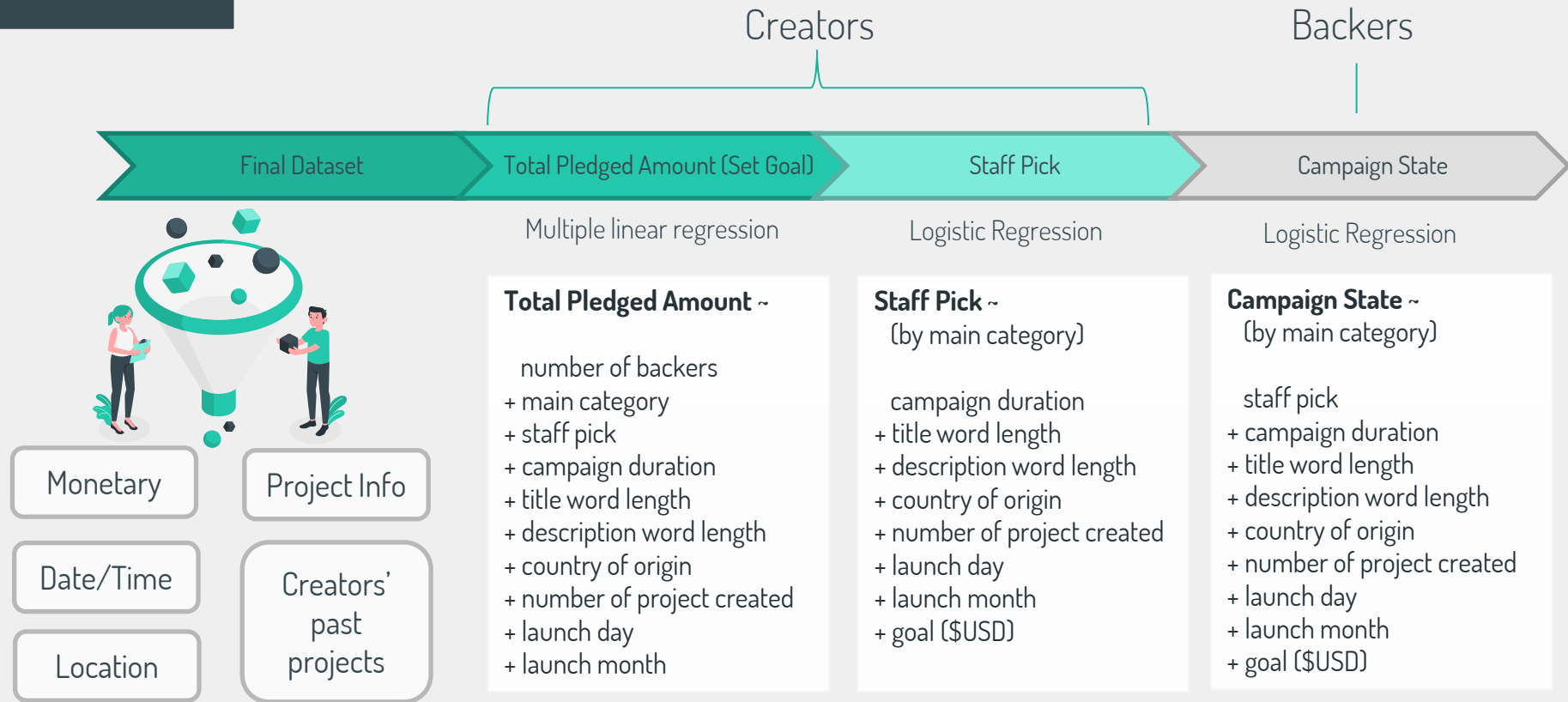


Location

Top 3



FINAL DATASET AND MODEL



HOW CAN OUR R SHINY APP HELP CREATORS & BACKERS?



DECISION MAKING

Creators can understand contributing factors of a staff-picked project and determine suitable goal . Backers can decide whether to pledge based on probability of success.



RESOURCE PLANNING

Insights gained from dashboard and predictions based on historical data can help in resource planning



TRACKING PROGRESS

Once project has been launched, predict pledged amount based on moving indicators.

APP DEMO VIA ROLE PLAY



APP DEVELOPERS



CREATOR

BACKER

CONCLUSION



PROBLEMS AND SOLUTIONS

PROBLEMS

DISPERSED INFO & LIMITED RESOURCES

- Accessed from different websites
- Limited data analysis
- Backer: successful project, historical data

TOO MANY FACTORS TO CONSIDER IN PROJECT PLANNING

- Different factors to consider

MAKING THE CHOICE

- Backer: successful project, historical data
- Creator: What's the optimal amount, date, topic, success rate etc.

SOLUTIONS

A COMPREHENSIVE DASHBOARD

- One stop shop
- Rich exploratory data analysis
- Useful to creators and backers

ISOLATE THE VARIABLES THAT MATTER

- Through exploratory data analysis and modelling, creators can narrow down the factors that have the greatest impact on their project's success.

DATA-DRIVEN DECISION MAKING

- Backers can use our logit models to determine a project's likelihood of success
- Creators can use our MLR and logit models to help set goals and other aspects of their campaign

WHAT DIFFERENTIATES US?



DATA

Our study and app is based on 5 years worth of data and insights (2016-2020).



ANALYTICAL APPROACH

Our approach is unique: we predict pledge amount, staff pick, and success rate using variables like project locations.



EXPLORATORY DATA VISUALIZATION

Our app is the most comprehensive out in the market, and it makes it easy for users to visualize and explore historical data.



ROBUST AND FLEXIBLE MODELS

Our models are flexible and allow creators and backers to change variables over the lifetime of a project to get more up to date predictions.

IMPROVE THE PREDICTION ACCURACY AND FUTURE WORK

FUTURE WORK

- Text analytics or text data mining to derive additional information from project name and blurb
- Variable on “Number of Projects by the creator” would need to be **calibrated** as it currently takes into account the total number of projects launched by a creator during the 5-year period.
- Further analysis can be performed to determine how creators of certain project category (e.g. games) with a succession of successful projects could have an impact on the outcome.
- Consider **partitioning** the data into training, validation and testing

ADDITIONAL INFORMATION

- Social analytics – impact of online ads (in particular targeted and hyper personalized marketing campaigns)
- Backers' demographics
- Results of search engine optimization



THANK YOU VERY MUCH!

ANY QUESTIONS?



SCAN THE QR CODE TO VIEW THE APP



[HTTPS://ADOLIT.SHINYAPPS.IO/KICKSTARTER-ASAR-G2-GROUP6/](https://adolit.shinyapps.io/kickstarter-asar-g2-group6/)



H_0 : No association exists between Campaign Duration by Project Success State

H_1 : Association exists between Campaign Duration by Project Success State

```
Success Fail
<= 30d 43031 25072
> 30d 19002 16728
>
> Result <- chisq.test(Matrix,correct=FALSE)
> Result
```

Pearson's Chi-squared test

```
data: Matrix
X-squared = 974.98, df = 1, p-value < 2.2e-16

> Result$observed - Result$expected
      Success      Fail
<= 30d 2344.191 -2344.191
> 30d -2344.191 2344.191
```

Chi-Square Test Result of Campaign Duration by Project Success State

STATISTICAL TESTS

CHI-SQUARE TEST OF ASSOCIATION

H_0 : No association exists between Main Category and Country of Origin for Staff Picked Projects

H_1 : Association exists between Main Category and Country of Origin for Staff Picked Projects

Pearson's Chi-squared test

```
data: Matrix
X-squared = 1962.4, df = 196, p-value < 2.2e-16
```

> Result\$observed - Result\$expected

	Australia	Canada	Denmark	France	Germany	Hong Kong	Italy	Japan	Mexico	Netherlands	Singapore	Spain	Sweden	United Kingdom	United States
Art - Staff Pick	-1.9197540	-14.27117602	-0.4570331	3.002613	-8.0439662	-2.2877786	6.76295158	1.3524981	14.2422752	-2.8674865	-1.42598002	3.4963874	-3.503613	13.4007686	-7.480707
Comics - Staff Pick	-3.1629516	-64.65411222	-5.7156034	-32.231360	-32.8724058	-7.3466564	0.04458109	-14.0299769	-35.5073021	-10.7901614	-2.28823982	-10.3566487	-10.356649	-29.0092237	128.968486
Crafts - Staff Pick	1.7403536	-2.63981553	1.5305150	-1.933743	3.0617986	-1.4136818	-1.71575711	4.8192160	27.8482706	-0.9345119	3.20015373	-2.4739431	-1.473943	4.5136049	-34.128517
Dance - Staff Pick	-3.9076095	-7.76049193	-1.3480400	-3.443351	-5.3647963	-1.2968486	-1.49131437	-2.9179093	-4.3953882	-1.7746349	-0.73374327	-0.2694850	-1.269485	-11.6362798	47.609377
Design - Staff Pick	3.5445042	3.88408916	11.3930822	20.435357	17.9698693	13.5300538	3.33405073	-4.8073789	-18.4633359	3.2516526	1.03674097	6.9275942	1.927594	0.3515757	-64.315450
Fashion - Staff Pick	-0.5511914	12.13850884	6.3950038	3.481091	9.7004612	-0.5060723	1.18570331	-0.6386626	-3.2235204	5.5706380	1.58209070	0.6143736	4.614374	-5.1349731	-35.227825
Film & Video - Staff Pick	-17.6239047	-38.99508071	-5.2196003	-7.266487	-21.2520369	-6.8315142	-11.88685626	9.8790930	145.3538816	-7.4536510	3.56256726	-7.2051499	-5.205150	-10.8372022	-11.893774
Food - Staff Pick	-8.1375865	-6.81091468	0.7778632	-15.086856	-12.6564181	-2.0238278	-5.65103766	-6.3036126	-17.57226749	4.1252882	1.15757110	-5.7916987	-2.791699	-24.5549577	101.270561
Games - Staff Pick	7.1374327	29.88039969	-2.8172175	33.510223	24.5338970	-4.5963105	11.24919293	18.4083013	-15.2287471	3.3418909	1.83366641	27.2064566	8.206457	-34.5205227	-108.145119
Journalism - Staff Pick	-1.1749424	3.07394312	-1.7852421	-1.208762	-0.4290546	-0.7174481	-3.29930822	3.8642583	0.8817832	1.6498078	-0.97171407	3.0055342	1.994466	4.1033051	4.752959
Music - Staff Pick	-11.5094543	-21.72667179	1.5797079	-7.962952	-4.0350500	-5.1385088	-0.71345119	-6.0616449	-11.2124520	-0.7684858	-0.03889316	-4.4923905	11.507610	-6.1420446	66.714681
Photography - Staff Pick	3.2727902	-13.30115296	0.3342813	13.235895	25.4137586	-0.5644889	4.07348194	2.2299001	-2.6016910	3.4906995	0.54903920	1.5121445	2.512145	49.4399693	-89.596772
Publishing - Staff Pick	26.3716372	0.01998463	-6.3608762	-5.988855	-10.8051499	-4.8155265	-4.54035357	-15.0849347	-47.4373559	-6.9054573	-1.81667948	-4.1771714	-5.177171	67.9738663	18.744043
Technology - Staff Pick	11.0285165	1.11053036	3.8002306	9.965488	22.7276710	26.0356649	5.54219831	21.5802460	-32.6112221	11.8382782	2.62544197	0.5624135	2.562414	-42.4983859	-44.269485
Theater - Staff Pick	-5.1078401	-9.25626441	-2.1070715	-8.508301	-7.9485780	-2.0270561	-2.89408148	-4.5608762	-0.1225211	-1.7738663	-1.14688701	-2.5473482	-3.547348	24.5504996	26.997540

Chi-Square Test Result of Main Category by Country of Origin for Staff Picked Projects

TECH ONLY

```
call:
glm(formula = staff_pick ~ location_country + usd_goal + campaign_duration +
    num_project_creator + launch_month + launch_day + name_word_len +
    blurb_word_len, family = binomial(link = "logit"), data = ks_data_tech)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.2449	-0.4585	-0.3529	-0.2584	3.0510

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.048e+00	3.101e-01	-13.054	< 2e-16 ***
location_countryCA	-2.774e-01	2.490e-01	-1.114	0.265204
location_countryDE	2.566e-01	2.449e-01	1.048	0.294823
location_countryDK	5.981e-01	3.953e-01	1.513	0.130325
location_countryES	-7.310e-01	3.654e-01	-2.001	0.045424 *
location_countryFR	1.927e-01	2.653e-01	0.726	0.467659
location_countryGB	3.553e-01	2.222e-01	1.599	0.109794
location_countryHK	-1.824e-01	2.703e-01	-0.675	0.499773
location_countryIT	-6.474e-01	3.188e-01	-2.030	0.042319 *
location_countryJP	1.458e+00	2.816e-01	5.179	2.23e-07 ***
location_countryMX	-1.031e+00	4.586e-01	-2.248	0.024565 *
location_countryNL	3.209e-01	3.086e-01	1.040	0.298269
location_countrySE	6.307e-01	3.608e-01	1.748	0.080427 .
location_countrySG	-3.702e-01	4.660e-01	-0.794	0.426994
location_countryUS	1.689e-01	2.001e-01	0.844	0.398739
usd_goal	-2.104e-07	2.217e-07	-0.949	0.342474
campaign_duration	-1.214e-03	2.828e-03	-0.429	0.667779
num_project_creator	4.352e-02	1.463e-02	2.974	0.002938 **
launch_monthaug	-1.639e-01	1.673e-01	-0.979	0.327454
launch_monthdec	-4.233e-01	1.998e-01	-2.118	0.034157 *
launch_monthfeb	-1.268e-01	1.725e-01	-0.735	0.462292
launch_monthjan	-5.023e-02	1.675e-01	-0.300	0.764299
launch_monthjul	-1.685e-01	1.694e-01	-0.995	0.319809
launch_monthjun	-1.679e-02	1.626e-01	-0.103	0.917762
launch_monthmar	-5.260e-02	1.627e-01	-0.323	0.746411
launch_monthmay	1.570e-01	1.575e-01	0.997	0.318774
launch_monthnov	9.859e-03	1.555e-01	0.063	0.949442
launch_monthoct	-5.369e-02	1.592e-01	-0.337	0.735960
launch_monthsep	-1.186e-01	1.643e-01	-0.722	0.470541
launch_daymon	5.615e-01	1.524e-01	3.684	0.000229 ***
launch_daysat	-2.982e-01	2.449e-01	-1.218	0.223231
launch_daysun	-2.858e-02	2.581e-01	-0.111	0.911824
launch_daythu	4.435e-01	1.778e-01	2.810	0.004955 **
launch_daytue	1.135e+00	1.797e-01	8.231	< 2e-16 ***
launch_daywed	7.779e-01	1.456e-01	5.344	9.10e-08 ***
name_word_len	1.533e-01	1.343e-02	11.407	< 2e-16 ***
blurb_word_len	-8.347e-03	6.847e-03	-1.219	0.222783

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 7097.9 on 12656 degrees of freedom
AIC: 6715.4 on 12620 degrees of freedom

Number of Fisher Scoring iterations: 7

Lower AIC Value (6715.4 vs 74024)
Better model fit

ALL ONLY

```
call:
glm(formula = staff_pick ~ location_country + category_parent_name_recode +
    usd_goal + campaign_duration + num_project_creator + launch_month +
    launch_day + name_word_len + blurb_word_len, family = binomial(link = "logit"),
    data = ks_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3694	-0.5489	-0.4410	-0.3463	2.7915

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.298e+00	9.967e-02	-33.088	< 2e-16 ***
location_countryCA	2.594e-01	8.360e-02	3.103	0.001914 **
location_countryDE	6.977e-01	9.073e-02	7.690	1.47e-14 ***
location_countryDK	5.587e-01	1.419e-01	3.938	8.23e-05 ***
location_countryES	-3.445e-03	1.152e-01	-0.030	0.976148
location_countryFR	7.262e-01	9.367e-02	7.752	9.02e-15 ***
location_countryGB	5.569e-01	7.490e-02	7.424	1.14e-13 ***
location_countryHK	-6.070e-02	1.397e-01	-0.435	0.663892
location_countryIT	-3.589e-02	1.120e-01	-0.320	0.748656
location_countryJP	1.457e+00	1.135e-01	12.615	< 2e-16 ***
location_countryMX	1.017e+00	8.733e-02	11.642	< 2e-16 ***
location_countryNL	4.389e-01	1.271e-01	3.453	0.000554 ***
location_countrySE	4.451e-01	1.181e-01	3.768	0.000164 ***
location_countrySG	-1.346e-01	1.744e-01	-0.772	0.440312
location_countryUS	3.964e-01	7.140e-02	5.552	2.83e-08 ***
category_parent_name_recodeeconomics	1.528e+00	4.923e-02	31.043	< 2e-16 ***
category_parent_name_recodecrafts	-7.077e-02	7.727e-02	-0.916	0.359731
category_parent_name_recodedance	1.247e+00	8.587e-02	14.518	< 2e-16 ***
category_parent_name_recodedesign	6.535e-01	5.939e-02	11.003	< 2e-16 ***
category_parent_name_recodefashion	-2.415e-01	6.325e-02	-3.818	0.000134 ***
category_parent_name_recodetech & video	6.467e-01	4.717e-02	13.711	< 2e-16 ***
category_parent_name_recodetech & food	3.857e-01	5.329e-02	7.238	4.55e-13 ***
category_parent_name_recodegames	5.089e-01	5.217e-02	9.753	< 2e-16 ***
category_parent_name_recodejournalism	8.996e-01	7.504e-02	11.988	< 2e-16 ***
category_parent_name_recodemusic	5.154e-01	4.958e-02	10.396	< 2e-16 ***
category_parent_name_recodephotography	7.977e-01	6.511e-02	12.253	< 2e-16 ***
category_parent_name_recodepublishing	1.279e+00	4.539e-02	28.171	< 2e-16 ***
category_parent_name_recodetechnology	7.958e-02	5.128e-02	1.552	0.120702
category_parent_name_recodetheater	5.457e-01	6.979e-02	7.819	5.31e-15 ***
usd_goal	-6.695e-08	4.786e-08	-1.399	0.161890
campaign_duration	-4.444e-03	8.580e-04	-9.958	< 2e-16 ***
num_project_creator	-6.177e-03	3.488e-03	-1.771	0.076582 .
launch_monthaug	5.413e-02	4.718e-02	1.147	0.251217
launch_monthdec	-2.175e-01	5.893e-02	-3.693	0.000222 ***
launch_monthfeb	5.132e-02	4.758e-02	1.079	0.280798
launch_monthjan	3.200e-02	4.782e-02	0.669	0.503433
launch_monthjul	-1.698e-02	4.813e-02	-0.353	0.724256
launch_monthjun	7.176e-02	4.740e-02	0.362	0.717380
launch_monthmar	2.506e-03	4.695e-02	0.053	0.957428
launch_monthmay	-3.663e-02	4.708e-02	-0.778	0.436616
launch_monthnov	1.366e-01	4.632e-02	2.949	0.003187 **
launch_monthoct	1.281e-01	4.482e-02	2.859	0.004352 **
launch_monthsep	1.032e-01	4.640e-02	2.223	0.026194 *
launch_daymon	1.461e-01	3.591e-02	9.637	< 2e-16 ***
launch_daysat	-2.085e-01	4.900e-02	-4.255	2.09e-05 ***
launch_daysun	-1.081e-01	5.039e-02	-2.146	0.031878 *
launch_daythu	2.533e-01	3.750e-02	6.807	9.99e-12 ***
launch_daytue	5.790e-01	3.357e-02	17.245	< 2e-16 ***
launch_daywed	3.438e-01	3.607e-02	9.533	< 2e-16 ***
name_word_len	6.915e-02	3.678e-03	18.802	< 2e-16 ***
blurb_word_len	-6.060e-03	1.575e-03	-3.450	0.000561 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 78362 on 103832 degrees of freedom
AIC: 74024 on 103782 degrees of freedom

Number of Fisher Scoring iterations: 6

WHY PER CATEGORY?