



KICKSTARTER Projects: Is This Project Going to Reach its Crowdfunding Goal?

Lior Solomon, Eva Marciano, Omer Sarig and Yotam Aharony

TAU

What is KICKSTARTER?

Kickstarter is an American public-benefit corporation, that maintains a [global crowdfunding platform](#) focused on creativity and merchandising.



How does it **Work**?



Project

Set up a Finite work you'd like bring to life



Funding Goal

Set up the amount of money need to complete the project



Backers Pledge

Join the creators in bringing the project to life



all-or-nothing

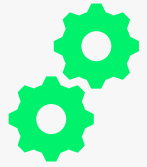
No charged for a pledged towards a project unless it reaches its funding GOAL



reward

A creator's change to share a piece of their project with their backers





Project **Problem** Formulation

Determine whether a new Kickstarter project is going to raise his crowdfunding goal by **Machine Learning Model** based on dataset of over 300k past projects

Our DataSet .



Over 300k projects (entries)



id internal kickstarter ID



name name of project | state current condition the project is



Category category | main_category category of the campaign



launched date launched | deadline deadline for crowdfunding



goal fundraising goal | pledged amount pledged by "crowd" | backers number of backers



country country pledged from | currency currency



Our DataSet .

Data preparation

```
In [2]: df = pd.read_csv("ks-projects-201801.csv")
df.head(5)
```

Out[2]:

name	category	main_category	currency	deadline	goal	launched	pledged	state	backers	country	usd_pledged	usd_pledged_real	usd_goal_real
ie Songs of Adelaide & Abullah	Poetry	Publishing	GBP	2015-10-09	1000.0	2015-08-11 12:12:28	0.0	failed	0	GB	0.0	0.0	1533.95
etring From arth: ZGAC rts Capsule For ET	Narrative Film	Film & Video	USD	2017-11-01	30000.0	2017-09-02 04:43:57	2421.0	failed	15	US	100.0	2421.0	30000.00
Where is Hank?	Narrative Film	Film & Video	USD	2013-02-26	45000.0	2013-01-12 00:20:50	220.0	failed	3	US	220.0	220.0	45000.00
oshiCapital Rekordz eds Help to Complete Album	Music	Music	USD	2012-04-16	5000.0	2012-03-17 03:24:11	1.0	failed	1	US	1.0	1.0	5000.00
Community ilm Project: The Art of ghborhoo...	Film & Video	Film & Video	USD	2015-08-29	19500.0	2015-07-04 08:35:03	1283.0	canceled	14	US	1283.0	1283.0	19500.00

```
In [89]: df.columns #Let's see which columns there are
```

```
Out[89]: Index(['name', 'category', 'main_category', 'deadline', 'launched', 'state',
               'country', 'usd_pledged_real', 'usd_goal_real', 'duration',
               'country_GDP'],
              dtype='object')
```

KICK
STARTER
STAFF
BOOK

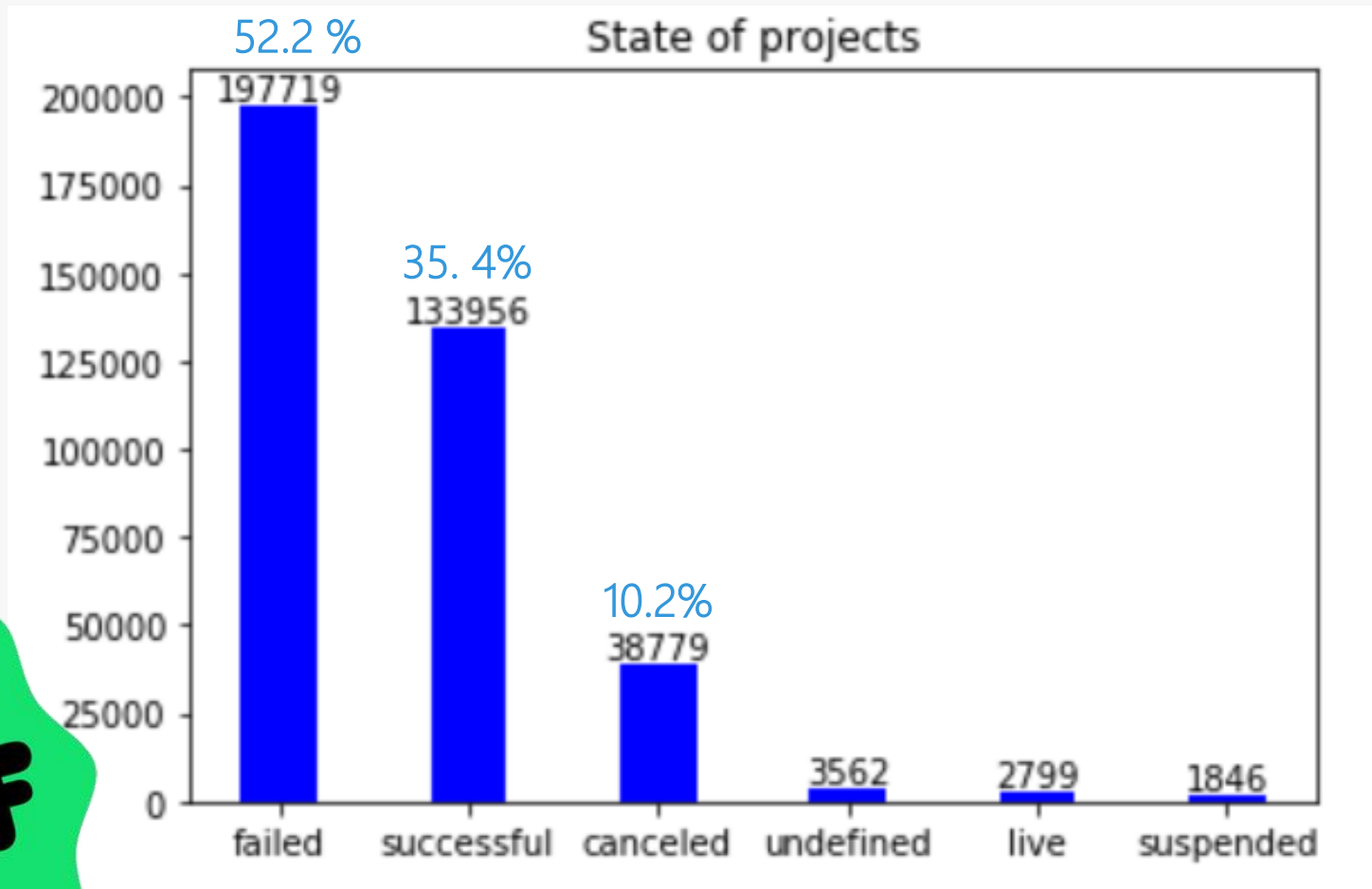
KICK
STARTER

Initial Analysis



Initial Analysis

#status

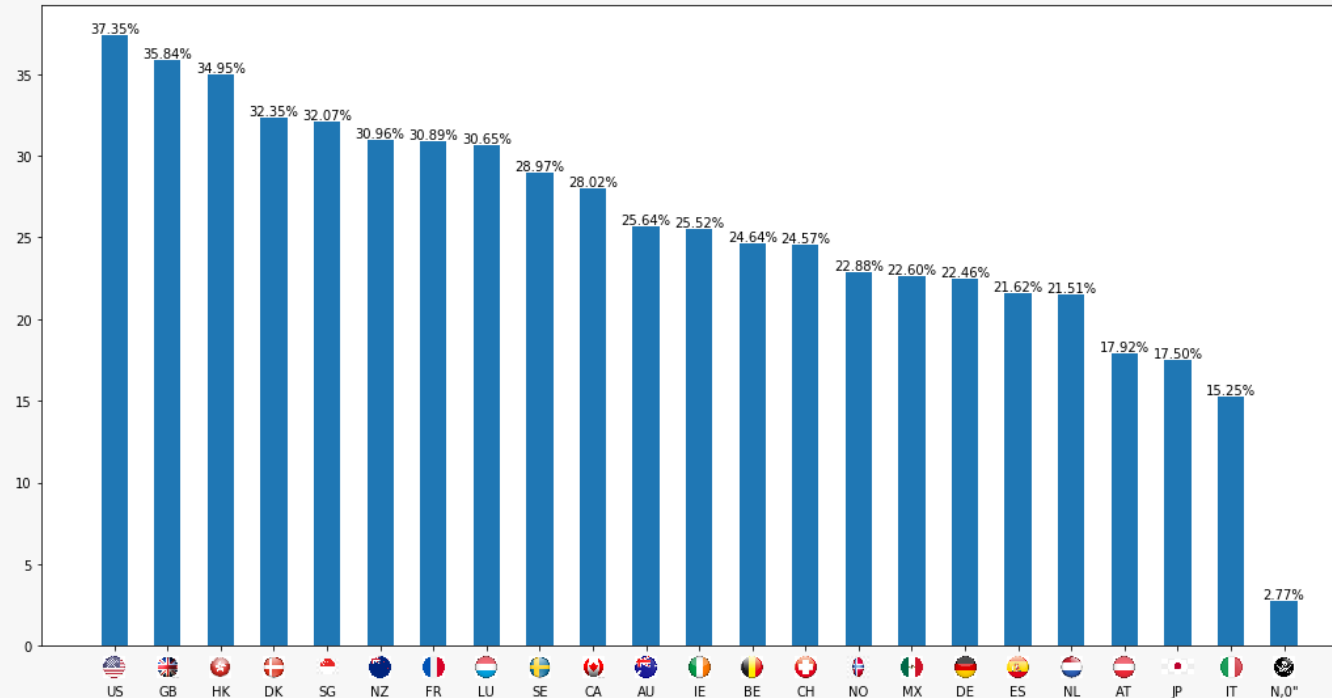


Undefined, live and suspended are 1% from all entries
Failure is obviously the most common status

Initial Analysis

#country

Percentile of successful projects according to country



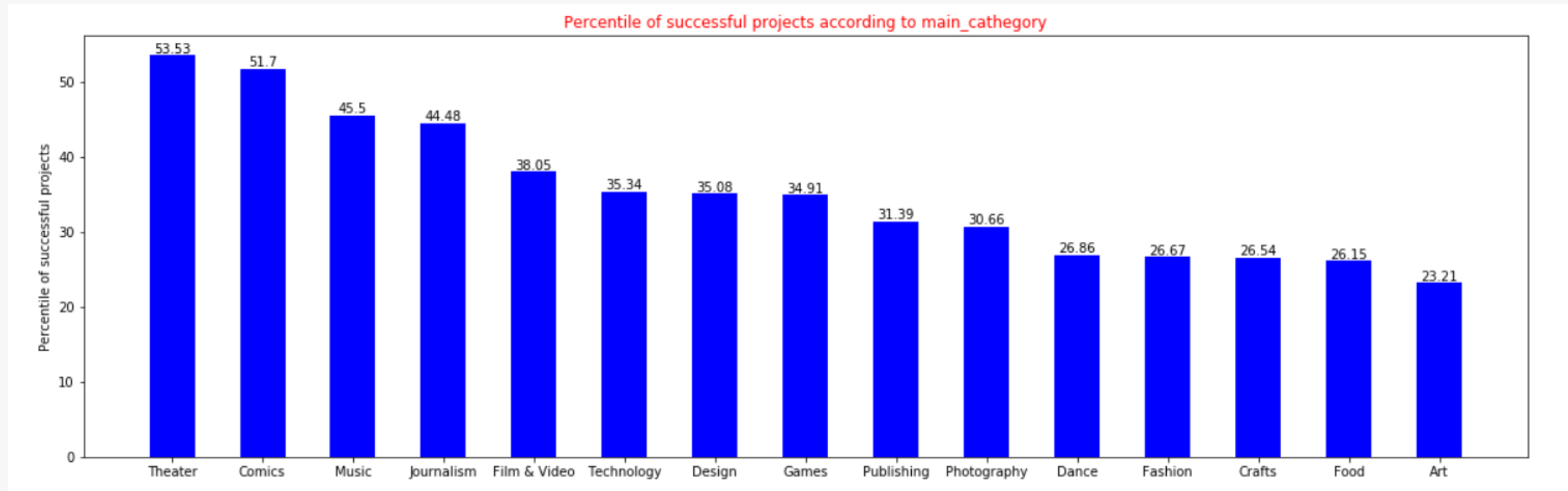
N,0'' country – needs to be cleaned

Understand why US, GB, HK have higher success rate then IT, AT, JP (for example)



Initial Analysis

#category



There are categories which the probability that the project will succeed is higher
TODO: cross country and category

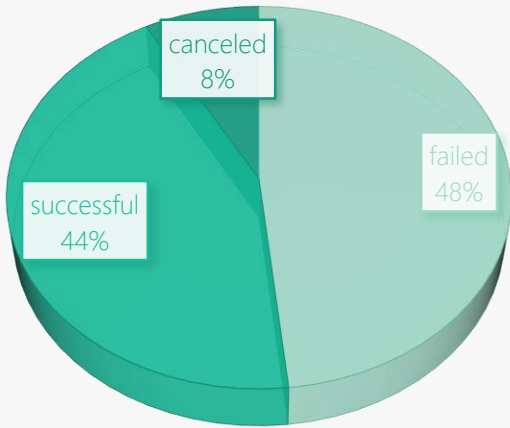




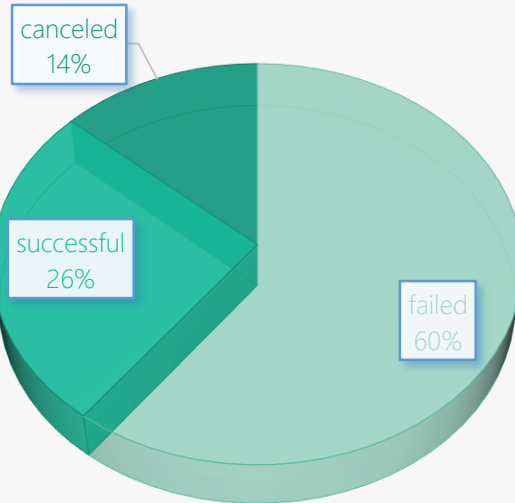
#goal

Initial Analysis

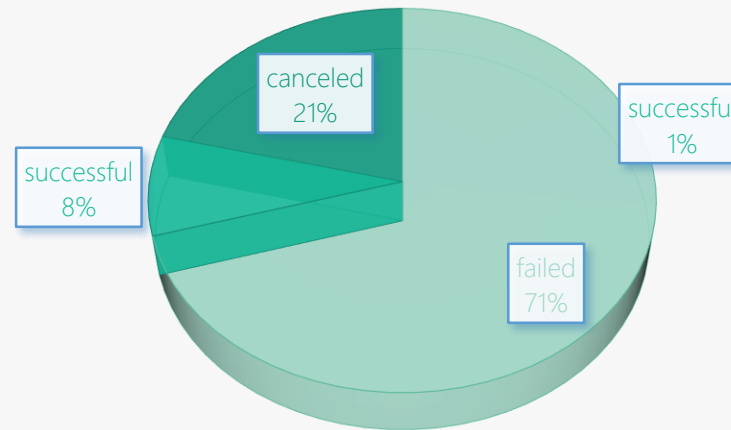
GOAL < 10K\$



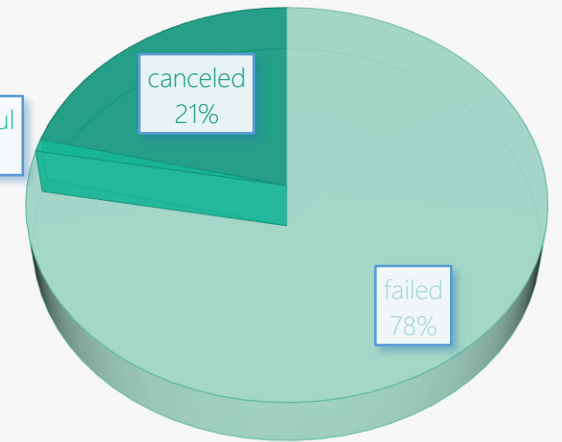
10K\$ < GOAL < 100K\$



100K\$ < GOAL < 1M\$



GOAL > 1M\$

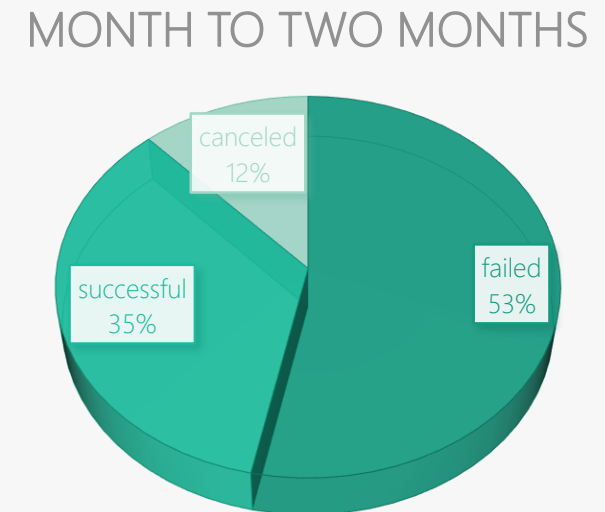
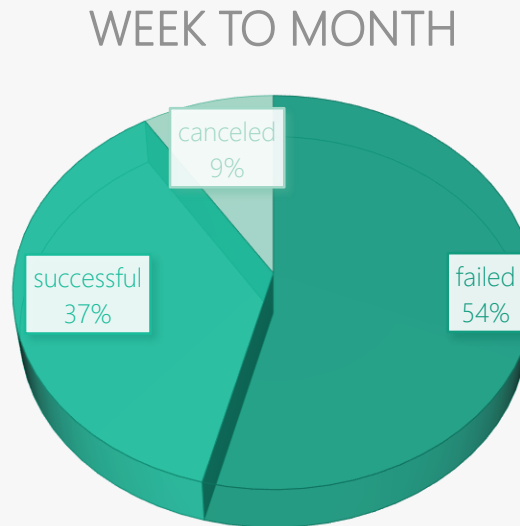
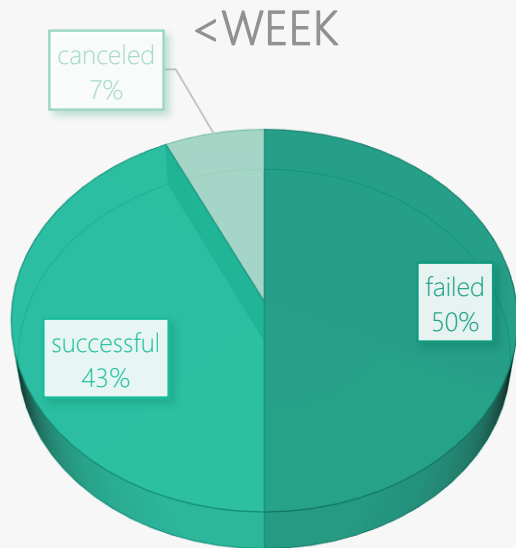


Greater Goal -> Lower success rate

KICKSTARTER

Initial Analysis

#duration

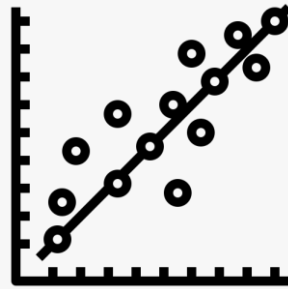


* Two months is the maximum amount of time

Seems like there's no correlation between success and duration. TODO: try finding a correlation between time on the year (month, holiday, special event) and success rate



baseline MODEL



LinearSVC

based on goal, duration, country, category and main_category

Statistics	
Train Score	65.63%
Test Score	65.56%
Precision	59.36%
Recall	14.3%
fScore	23.04%

Low Recall!

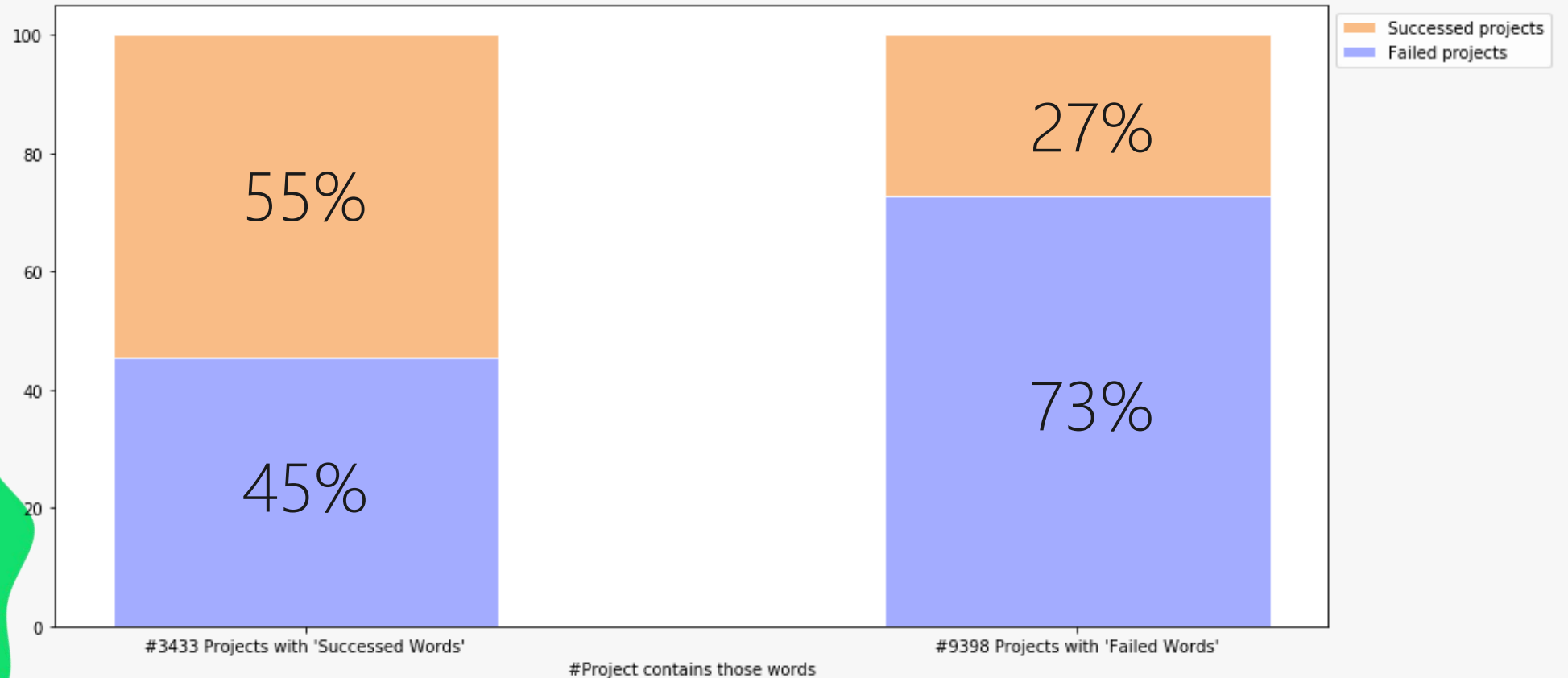
Feature Extraction



Any CLUES?

Top50 but not Top100: ['edition', 'dance', 'release', 'presents', 'recording', 'full']

Top70 but not Top100: ['app', 'clothing', 'mobile', 'fashion', 'dream', 'design', 'apparel', 'journey', 'photography', 'system', 'way', 'online', 'social']



KICKSTARTER
funded with

new INFO



Key words



GDP



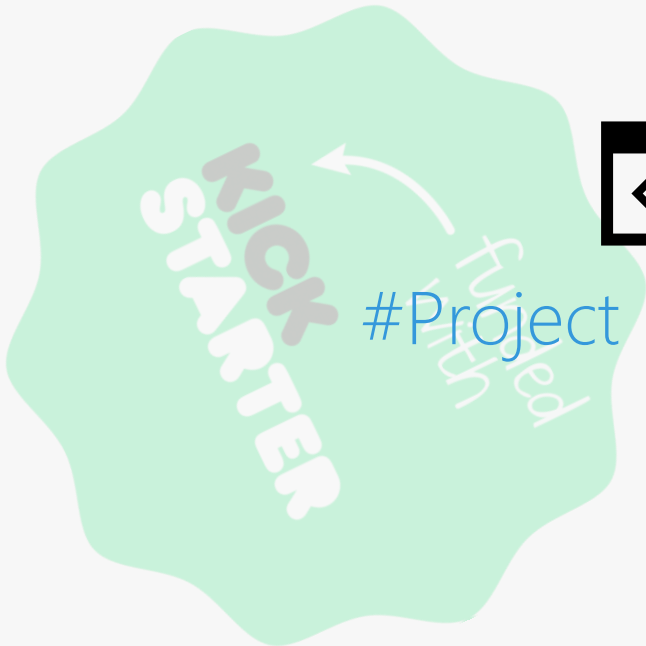
Happiness



#Project past 7days



Time since last project
in the category



KICKSTARTER

enhanced MODEL

#Better Performance



RandomForest

based on goal, duration, country, category and main_category

Statistics	
Train Score	74.42%
Test Score	67.45%
Precision	53.87%
Recall	65.86%
fScore	59.27%

TODO:

- Add our new features
- Try different methods

KICKSTARTER

More to come...

References

- Main Kickstarter projects DATASET
 - <https://www.kaggle.com/kemical/kickstarter-projects>
- Country / Gross Domestic Product DATASET
 - <https://www.kaggle.com/stieranka/predicting-gdp-world-countries/data>
- World Happiness Report DATASET
 - <https://www.kaggle.com/unsdsn/world-happiness#2019.csv>
- Previous Works:
 - <https://www.kaggle.com/ratatman/data-cleaning-challenge-scale-and-normalize-data>
 - <https://www.kaggle.com/shivamb/an-insightful-story-of-crowdfunding-projects>

