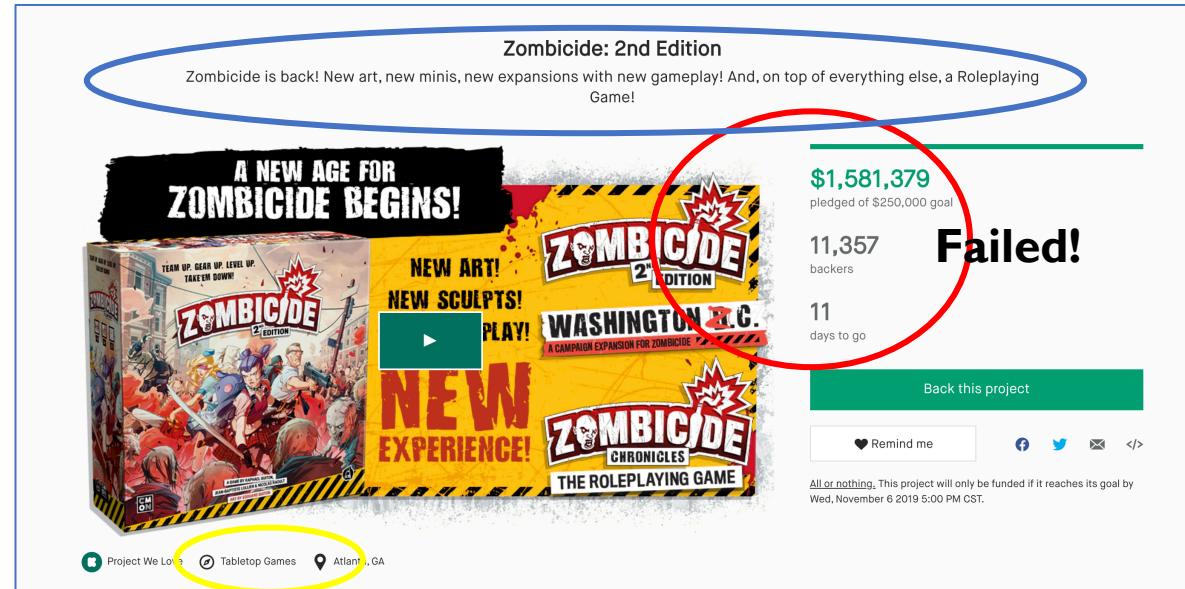
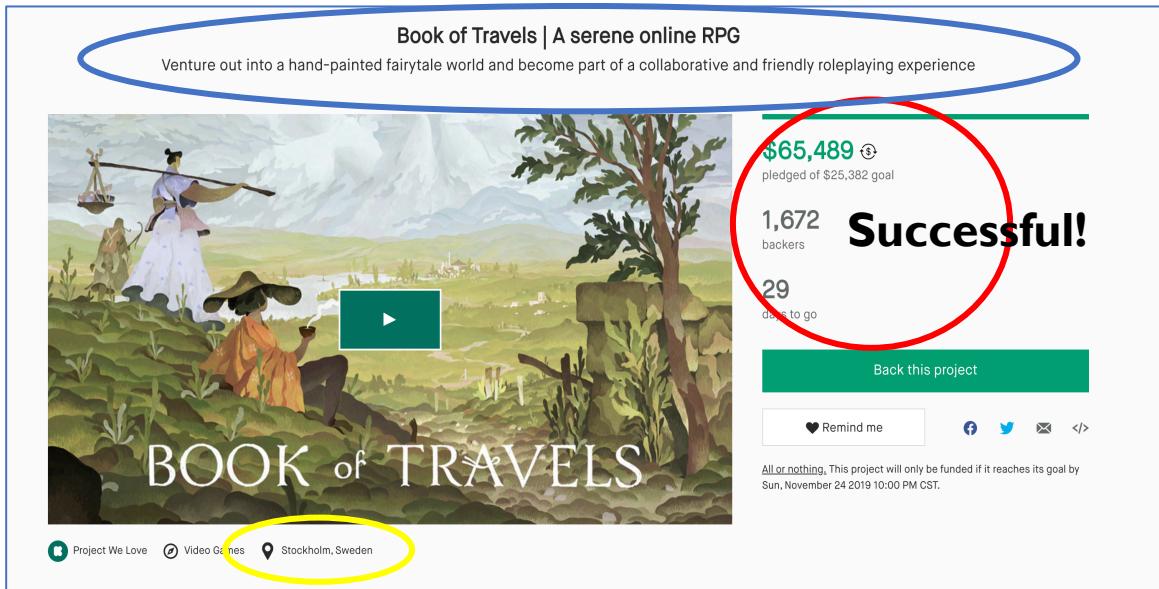


How to make your project successful? - An exploratory data analyses of Kickstarter dataset

Jie Zhang

10/27/2019

Introduction and Research Questions



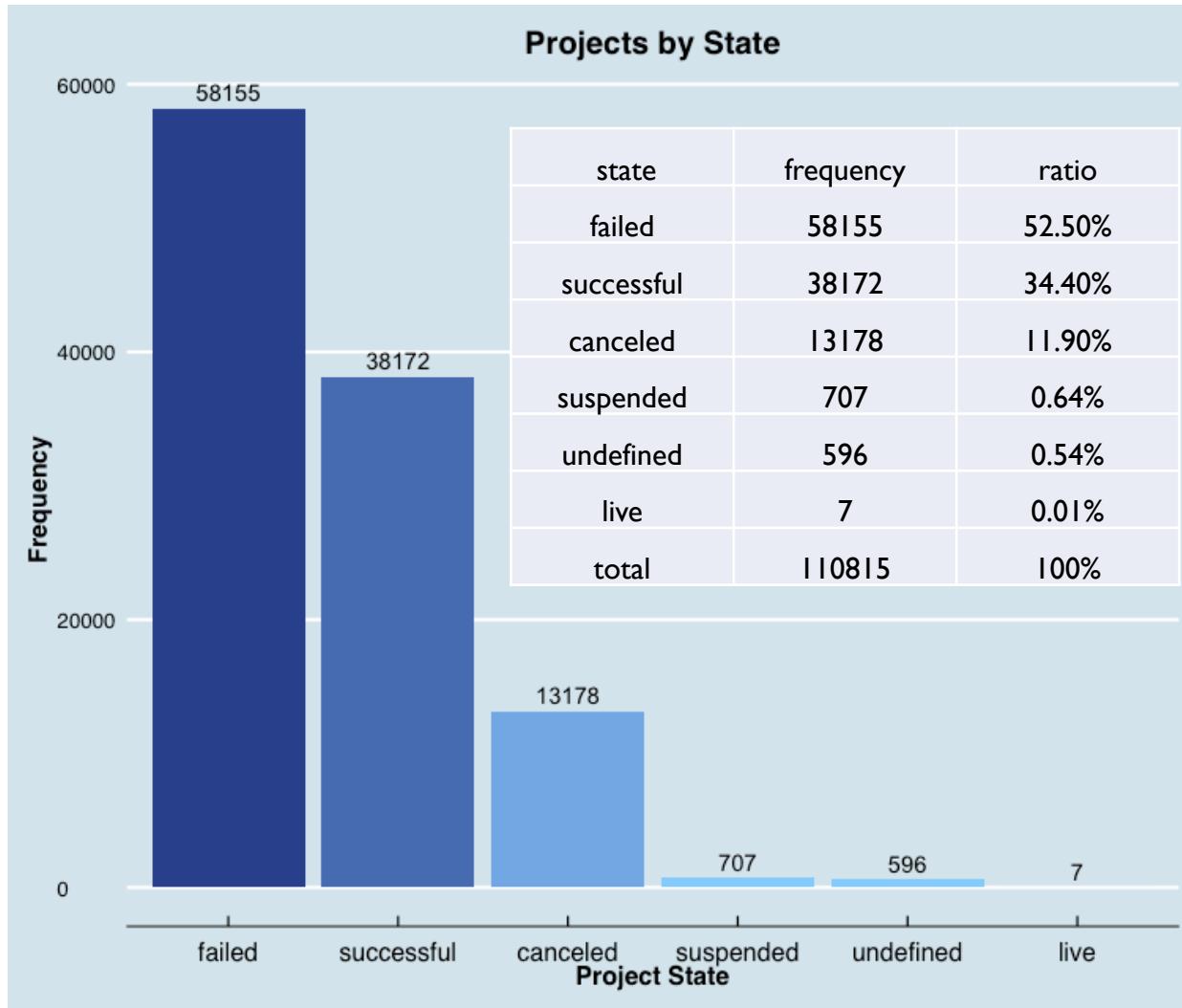
- I. Does the goal impact project success?**
- 2. Does project campaign length impact project success?**
- 3. Does other factors such as project name, country etc. impact project success? How?**

Outline

- I. Data Set: Kickstarter.com, 110815 observations, 16 variables; 2016 and 2017 Kickstarter campaigns.
2. Exploratory Data Analysis (EDA): Data Visualization, Descriptive Statistics.
3. Model: Regression, Machine learning models, etc.

```
> str(ks_data)
'data.frame': 110815 obs. of 15 variables:
 $ ID           : int 1015685046 1019043170 1036288991 1045749249 1048577059 ...
 $ name         : Factor w/ 110108 levels "", "__Shift : image frame", ...
 $ category     : Factor w/ 159 levels "3D Printing", ...
 $ main_category: Factor w/ 15 levels "Art", "Comics", ...
 $ currency     : Factor w/ 14 levels "AUD", "CAD", "CHF", ...
 $ deadline     : Factor w/ 731 levels "1/1/2016", "1/1/2017", ...
 $ goal          : int 6000 2000 5000 20000 40000 3650 2500 16000 25000 1500 ...
 $ launched      : Factor w/ 102442 levels "1/1/2016 0:26", ...
 $ pledged       : num 100 2102 5630 101 2246 ...
 $ state         : Factor w/ 6 levels "canceled", "failed", ...
 $ backers       : int 2 48 116 2 19 24 3 53 1093 0 ...
 $ country       : Factor w/ 23 levels "AT", "AU", "BE", ...
 $ usd.pledged   : num 74.8 2314 5630 101 3416.6 ...
 $ usd_pledged_real: num 70.9 2283 5630 101 3273.5 ...
 $ usd_goal_real : num 4254 2172 5000 20000 58299 ...
```

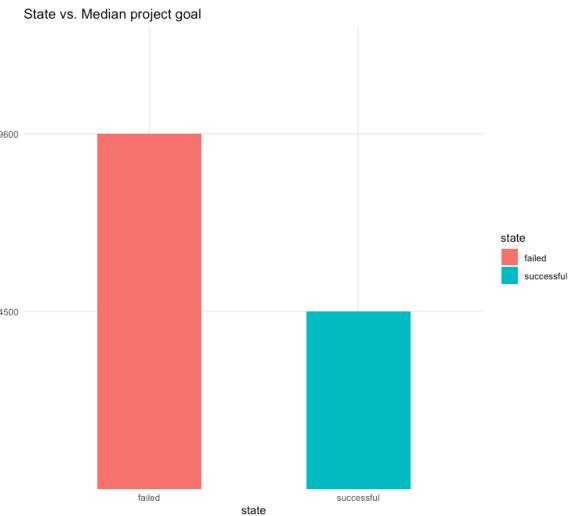
Success vs. Failure



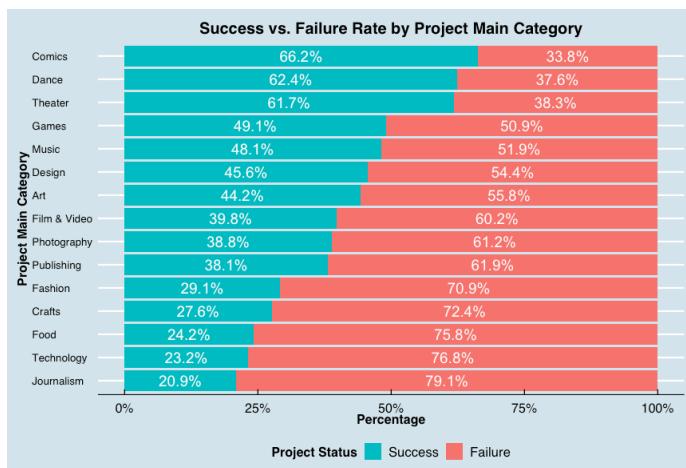
- **More than half of the projects failed!**
- Approximately 34% of the projects succeed.

Which projects tend to succeed?

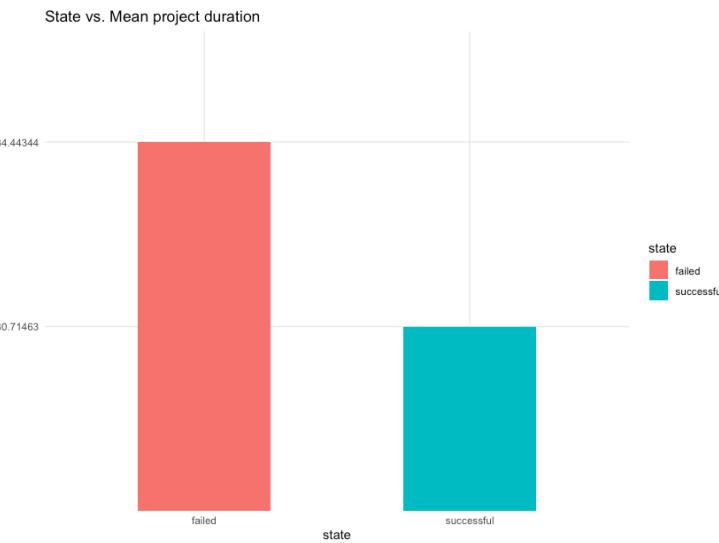
Projects have smaller goals



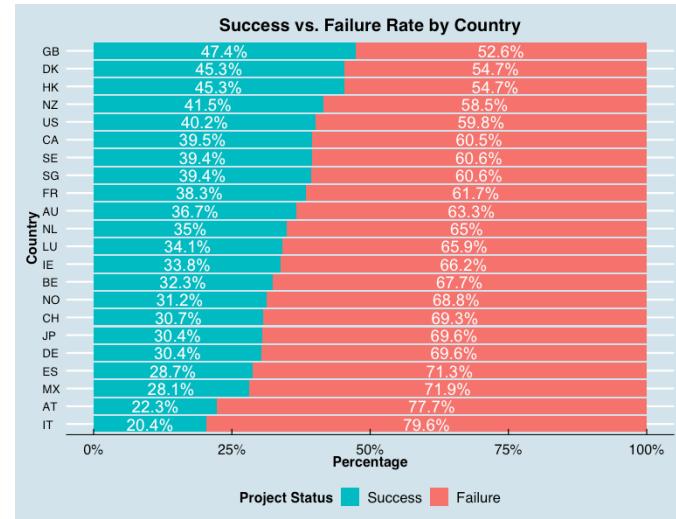
Projects from
Comics,
Dance and
Theater



Projects have shorter campaign

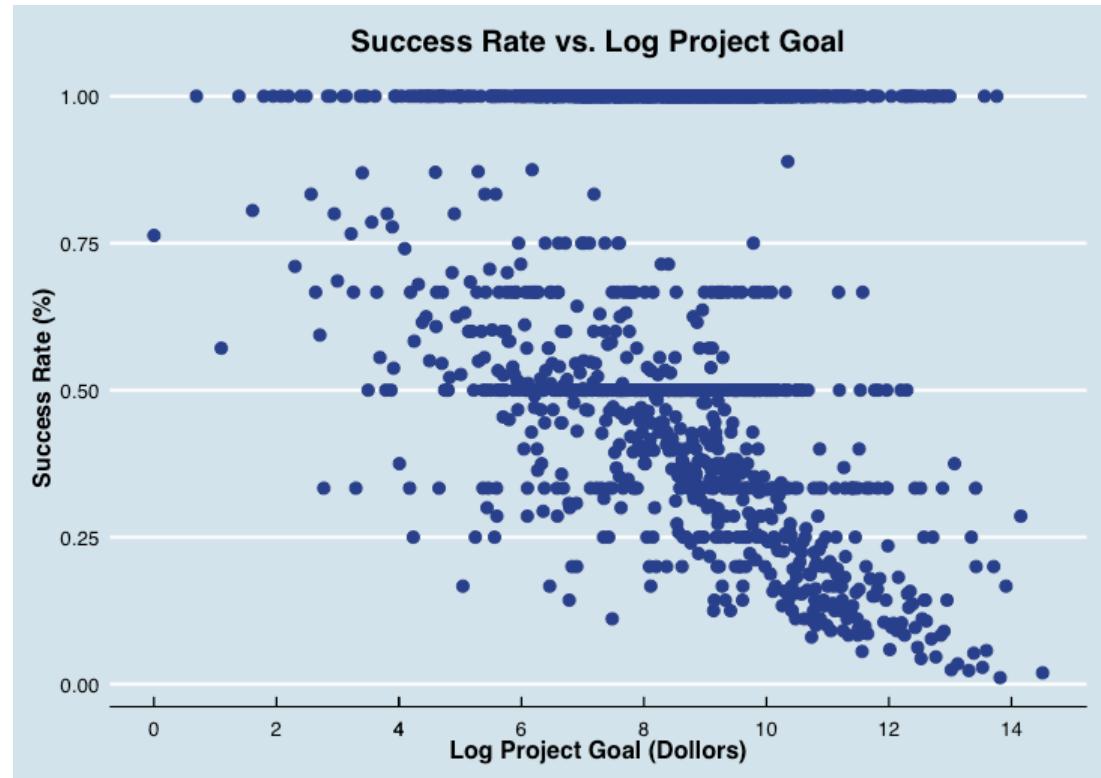


Projects from
Great Britain,
Denmark, Hong
Kong, New
Zealand, United
States



Project Goal vs. Project Success

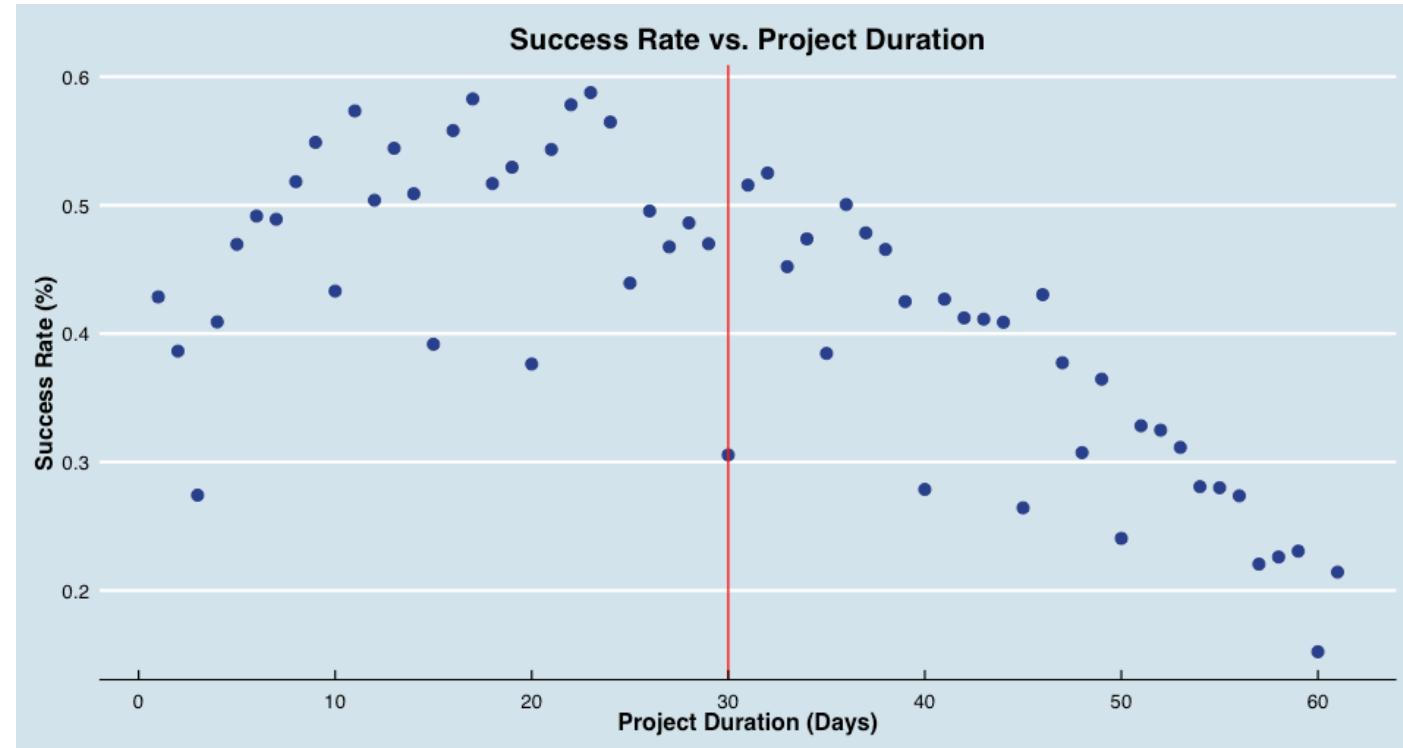
- Overall, there seems to be a negative relationship between project goal and project success rate.
 - ❖ A number of projects having success rate of 100% regardless of their goals.
 - ❖ It might be interesting to explore why they are always successful given different project goals (future study).



Note: we do log transformation of project goal because it's highly skewed.

Project Duration vs. Project Success

- Projects exceeding 30 days have a lower success rate than projects lasting fewer than 30 days.
- For projects exceeding 30 days, there seems to be a negative relationship between success rate and project duration.
 - Kickstarter's recommendation is to set project length to be less than 30 days.



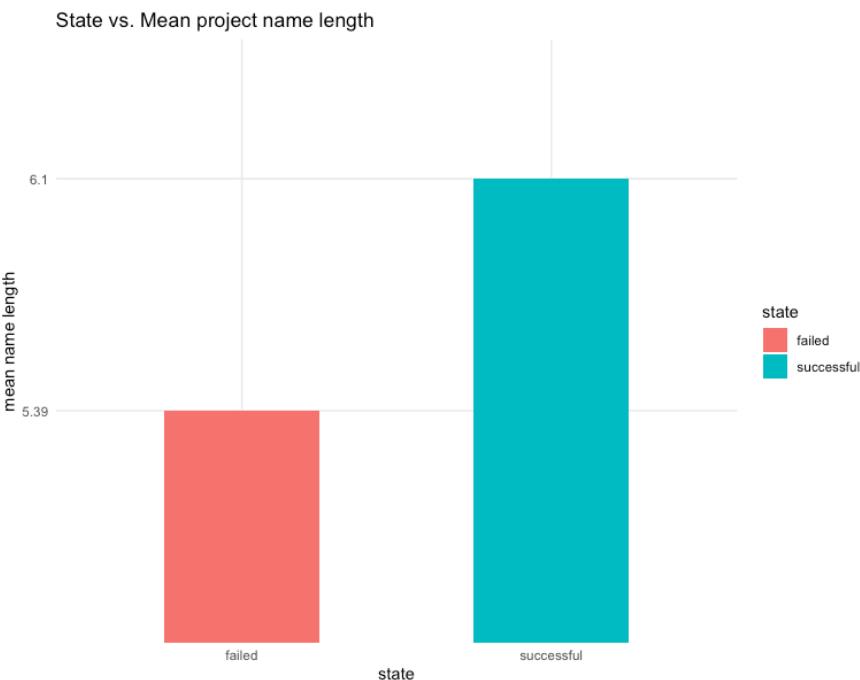
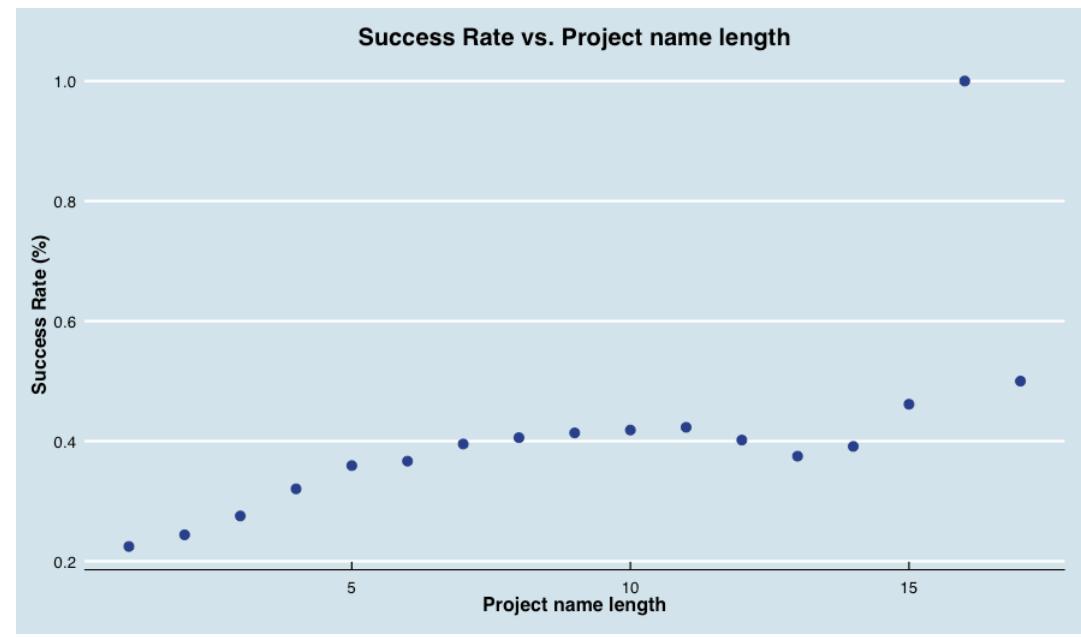
Results

Logistic Regression

- DV: project success rate
- IV: project goals, project duration, category and country
- Results:
 - Project goal impact: negative, very significant
 - Project duration impact: negative, very significant
 - Category impact (example): Project in comics (the best) have 45% higher chance than the ones in Journalism
 - Country impact (GB vs. US): Projects in GB have 7% higher chance than US to be successful
- Full Model Prediction accuracy: 67%

Other factors

- Projects have longer names are more likely to success.
- On average, the title of the successful projects have approximately one more word than failed projects.



Which factors are important for project success?

Factors had positive effect

- Most important
 - Smaller project goals
 - Shorter project campaigns
 - Comics, dance, theater
 - Projects from Hongkong, Great Britain, Denmark
- Less important
 - Project name length

Factors had negative effect

- Larger goals
- Longer campaigns
- Food, technology and journalism projects
- Projects from Italy

Future Research - Model Selection

Several other possible ways for prediction:

- Updated regression models: adding non-linear terms or interaction terms in IV
- Principal component analysis (PCA) with logistic regression
- Decision tree method
- Random forests with PCA or without PCA
- Non-parametric method such as k-Nearest Neighbors (kNN) algorithm

Questions?

Thank You!