Kickstarter Project

by Daniel Müller, Dominik Hohmann & Johannes Pfohl

Overview

01 TASK

02 EDA

03 MODEL

04 FUTURE WORK

TASK

Predict success/fail of a **Kickstarter** Project

Daniel Müller, Dominik Hohmann, Johannes Pfohl

EDA

- Data range (year) 2009–2019
- Number of lines 209222
- Number of columns 37
- slightly unbalanced 61% vs. 39%

EDA

Special focus on:

- backers count
- country
- currency
- pledged in US dollar
- category
- parent category
- location type
- state

Daniel Müller, Dominik Hohmann, Johannes Pfohl

EDA

Special focus on:

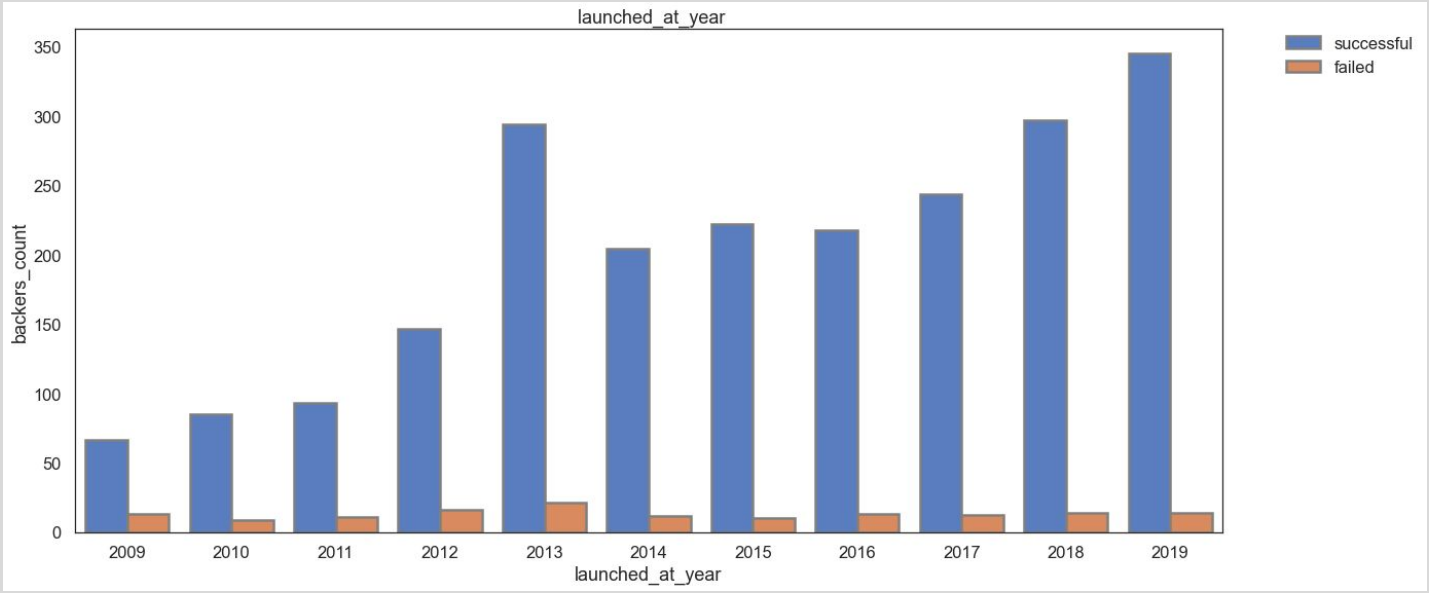
- backers count
- country
- currency
- pledged in US dollar
- category
- parent category
- location type
- state

Engineered columns:

- duration
- goal in US dollar
- launched at month
- launched at year
- created at month
- created at year
- blurb length

Daniel Müller, Dominik Hohmann, Johannes Pfohl

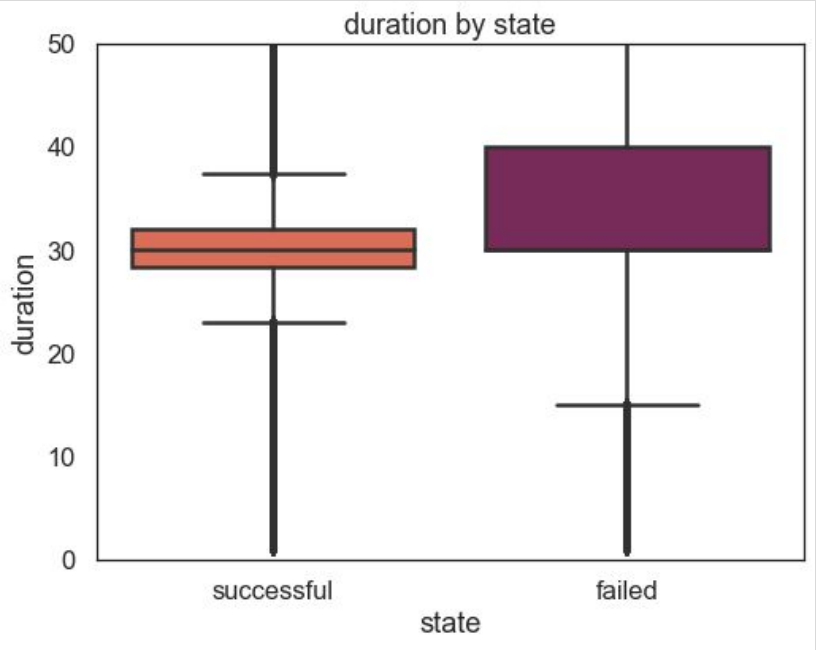
EDA



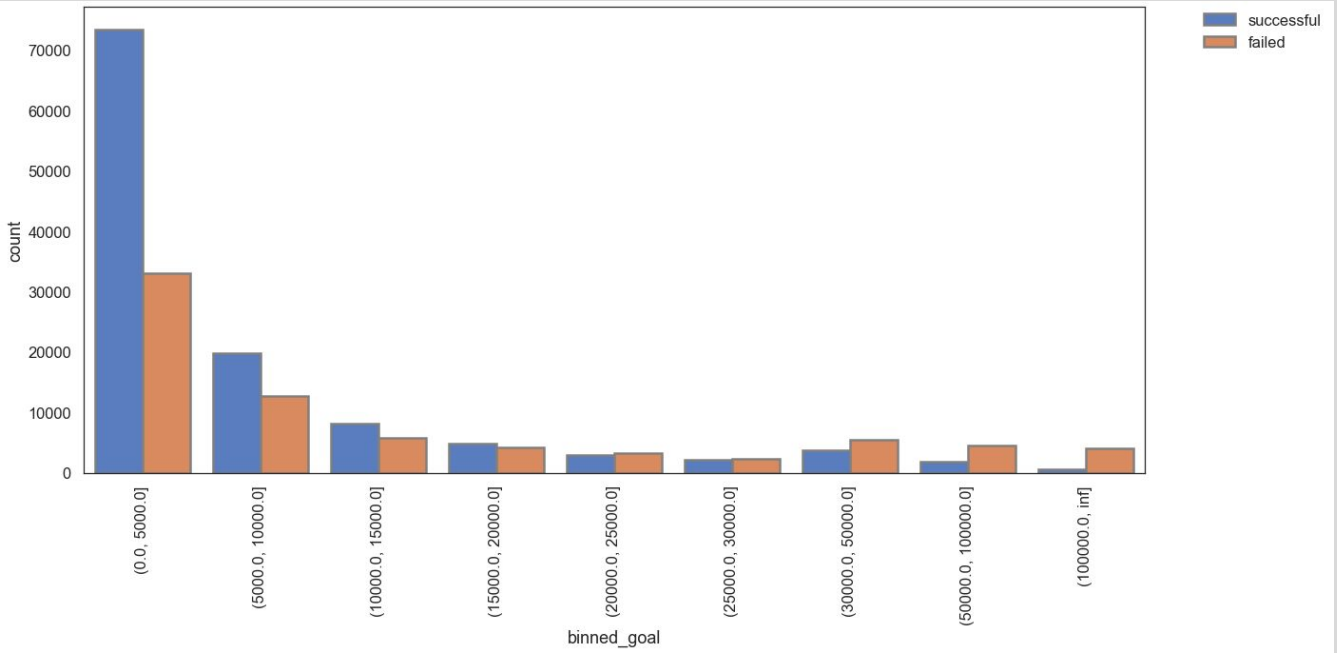
- **Continuously increasing trend in the number of backers**
- **Increasing popularity can also expected in the future**

EDA

- The longer a project, the less likely it is to be successful
- Majority of all projects set to 30 days
- Recommended by kickstarter duration of 30 days

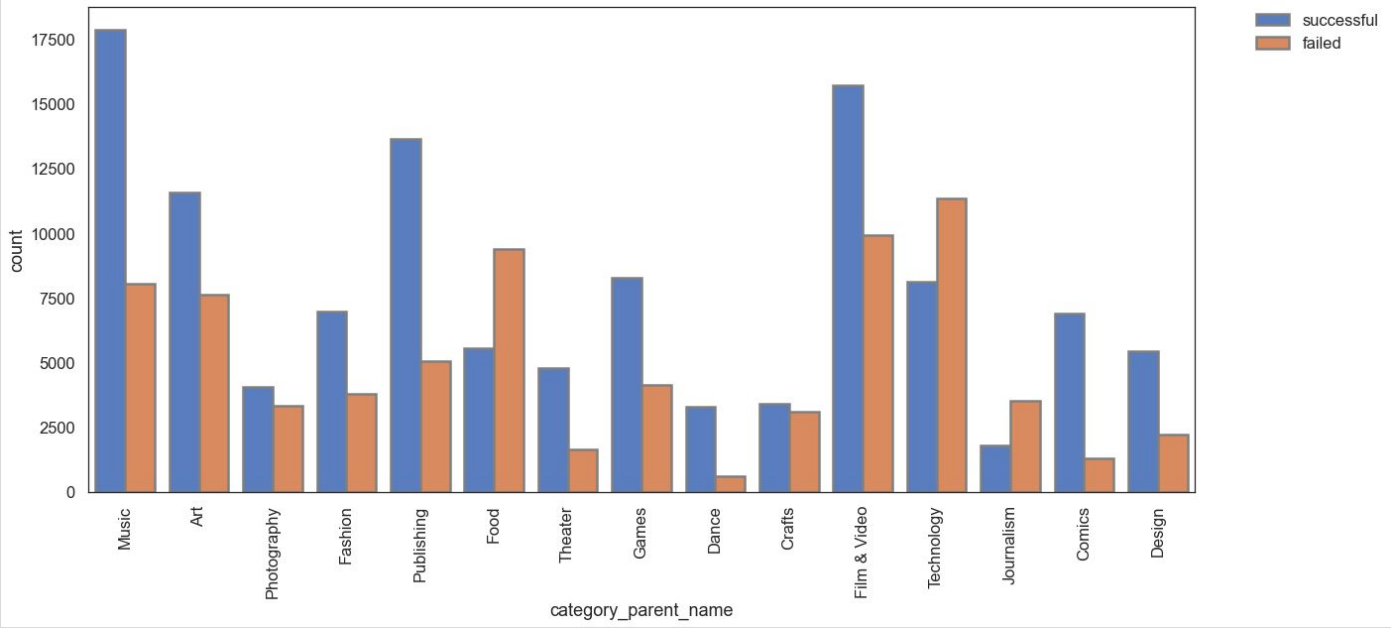


EDA



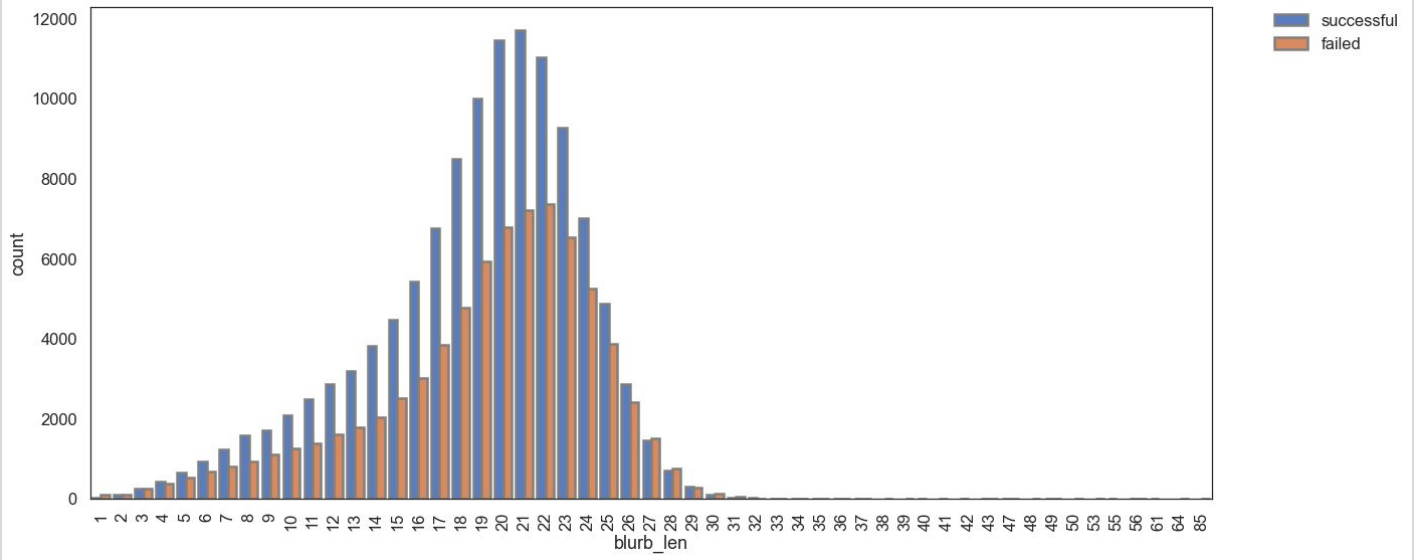
- **At best, choose a low goal**
- **The higher the goal, the lower the probability of success**

EDA



- **Prefer to avoid projects in the areas: Food, Journalism, Technology**
- **Projects in Music, Publishing, Games, Comics, Dance and Design most promising**

EDA



- Avoid a very small or large number of words
- At best, choose between 15 to 23 words in the blurb

MODEL

Features:

- country
- currency
- category name
- location type
- duration
- goal in US dollar
- launched at month
- created at month
- blurb length

Target

- state

MODEL

Features:

- country -> categorical
- currency -> categorical
- category name -> categorical
- location type -> categorical
- duration
- goal in US dollar
- launched at month
- created at month
- blurb length

Target:

- state

MODEL

Compared models:	Accuracy
• Random Forest	0.80
• Extra Trees	0.79
• Light GBM	0.79
• XGBoost	0.78
• AdaBoost	0.77
• Support Vector Machine	0.74
• Quadratic Discriminant Analysis	0.67
• K-nearest Neighbor	0.65
• Stochastic Gradient Descent	0.61
• DummyClassifier	0.61 (Baseline Model)

FUTURE WORK

- Further analysis on **subcategories**.
- Does a **creator with many projects** achieve more successful pledges?
- What are the factors that make the project being **picked by the staff**?
- Do specific **words in the blurb** correlate with successful projects?
- How does kickstarter.com decide if a project is **presented on the front-page**?

THANKS FOR YOUR ATTENTION!