# Kickstarter Project Analysis

Machine Learning Project
René Ebrecht, Jens Reich, Max Langer

# What is KICKSTARTER ?

- Kickstarter is a funding platform for creative projects
  - allows creators to present a project idea to a broad audience
  - everybody interested in a project can pledge money (backer)

- if the funding goal is reached,
  - the creators get the pledge money
  - and the project is realized

- examples of Kickstarter projects are:
  - Pebble Time, Coolest Cooler, Legend of Vox Machina



kickstarter.com

# How does Kickstarter make money?

- Kickstarter charges a 5% fee of the funds raised
  - when a project is successfully funded

- our job is to help creators realize projects
  - more successful creators means higher revenue for Kickstarter

- increase in creator satisfaction
  - higher success rate of their projects
  - less time wasted on unsuccessful projects

kickstarter.com

# The Data Set

- data from 2009 to 2019
  - only for the first few months of 2019

- distributed over 55 csv files

- a total of 180000 projects

- large amount different features
  - mostly categorical
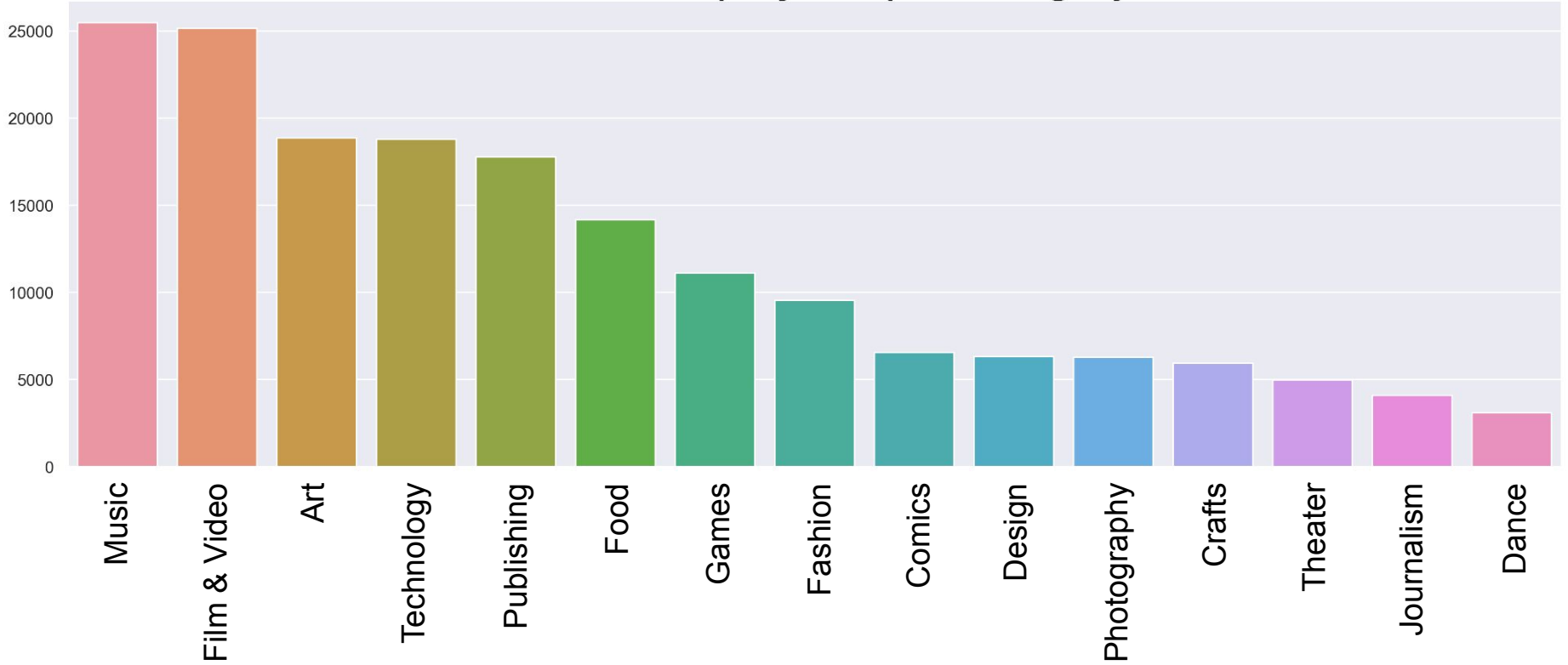  - like currencies, text, unix dates, …

kickstarter.com

# What makes a project successful?

- What are popular categories?

- Influence of the funding goal?

- Influence of the length of the funding period?
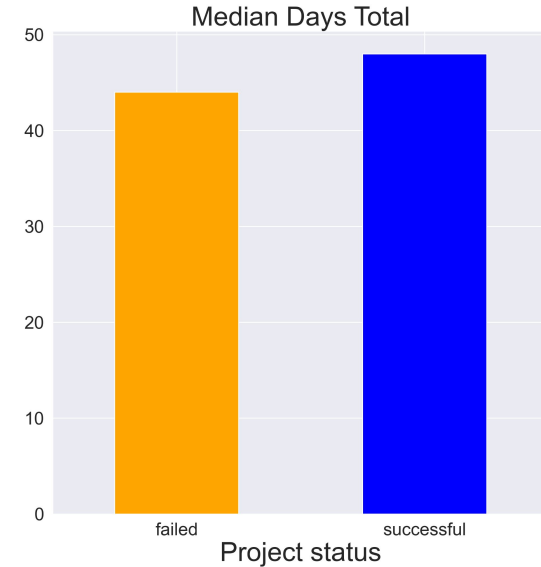
- Which other factors are important?
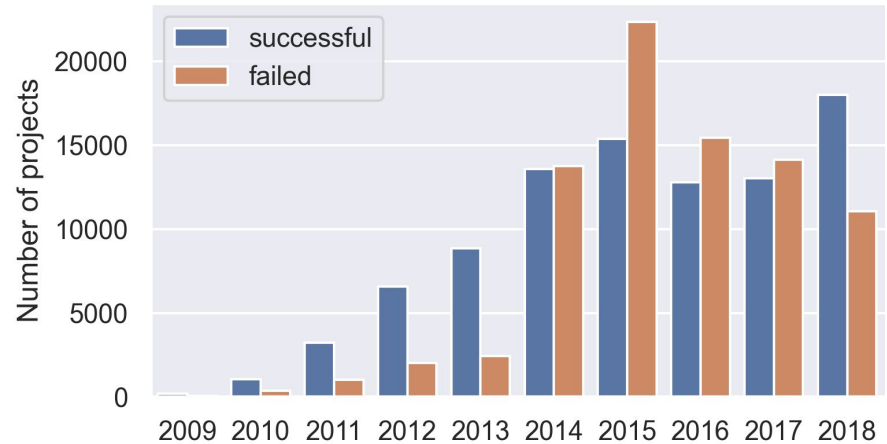
# Number of projects per category

# Other factors for project success?

- funding goals
  - high funding goals = more failure
  - lower goals = more success

- duration of the project
  - longer projects are more successful
  - difference is small (effect too?)



Median project goal (USD)



Median Days Total

7

# Baseline Model

- <u>DummyClassifier</u> is an extremely simple model

- projects are classified based on what the most common classification was in the dataset

- the DummyClassifier classifies all projects as successful, since most projects were successful
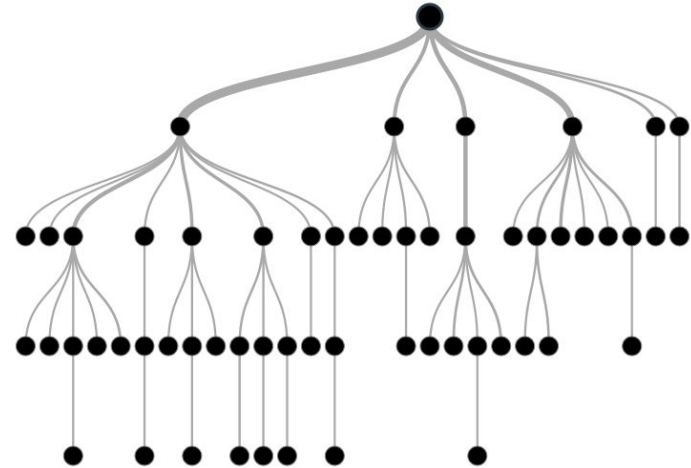
- Correct predictions: 43%



Baseline model prediction of kickstarter projects
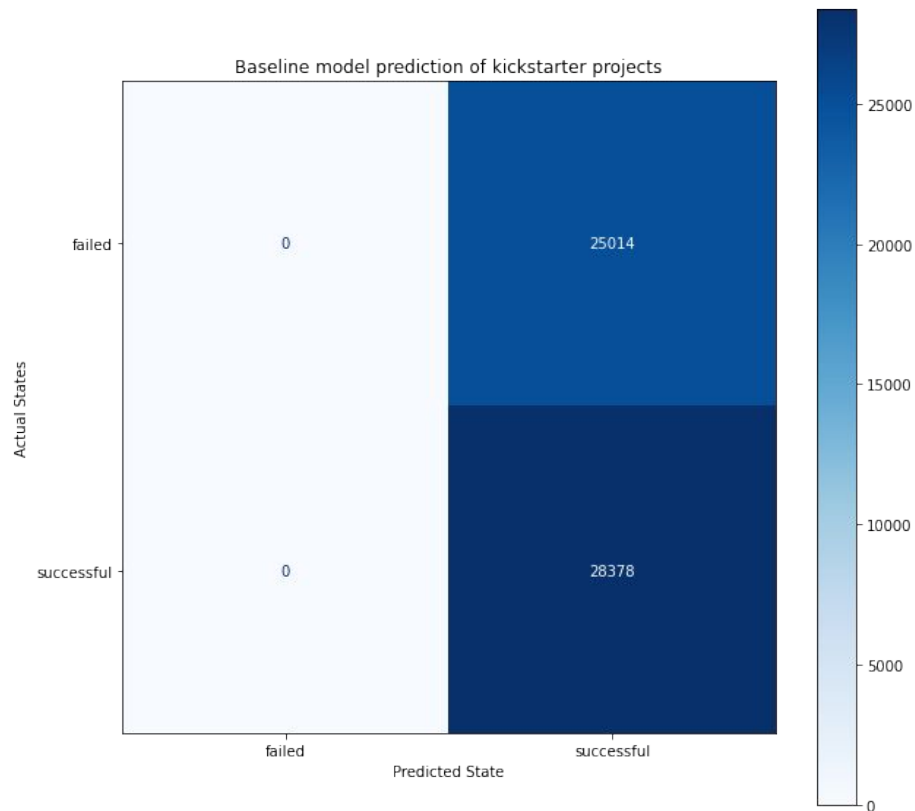
# Confusion matrix, simply explained

- Correct predictions

- Mispredictions



Baseline model prediction of kickstarter projects
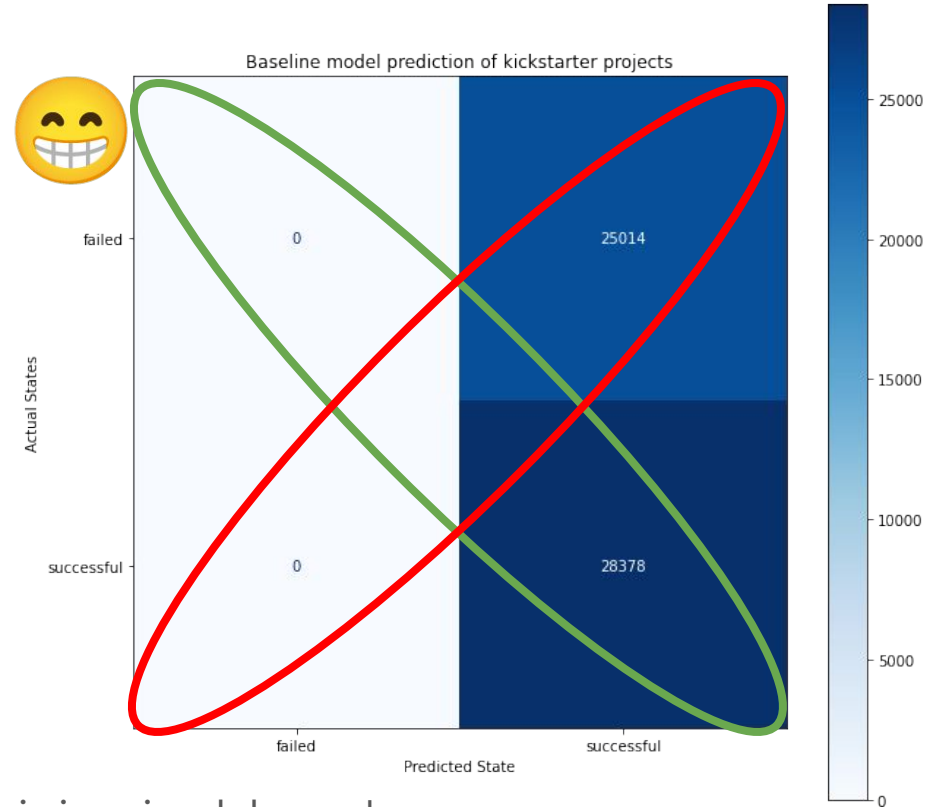
We want to minimize missed chances!
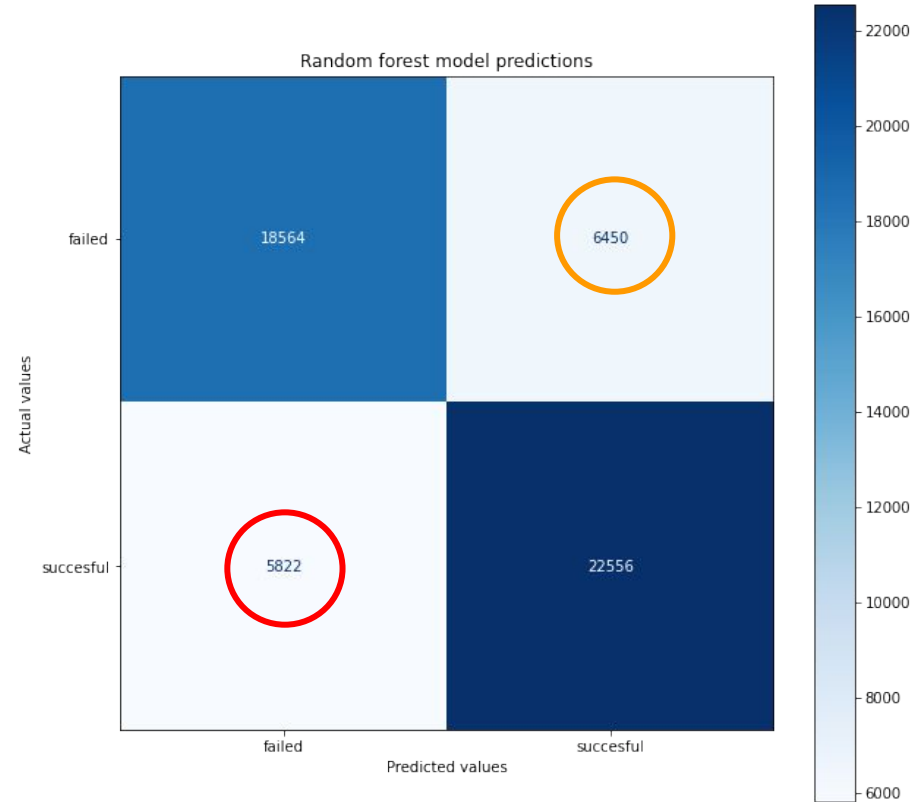So we score our model accordingly

# The model selection

- after cleaning and restructuring of the data
  - we had 406 features to model

- a decision tree model was chosen (random forest)
  - since we had a lot of categorical data
  - interpretability is important for goal-based consulting

- be able to tell our stakeholders
  - not only **if** projects have a high chance of being successful
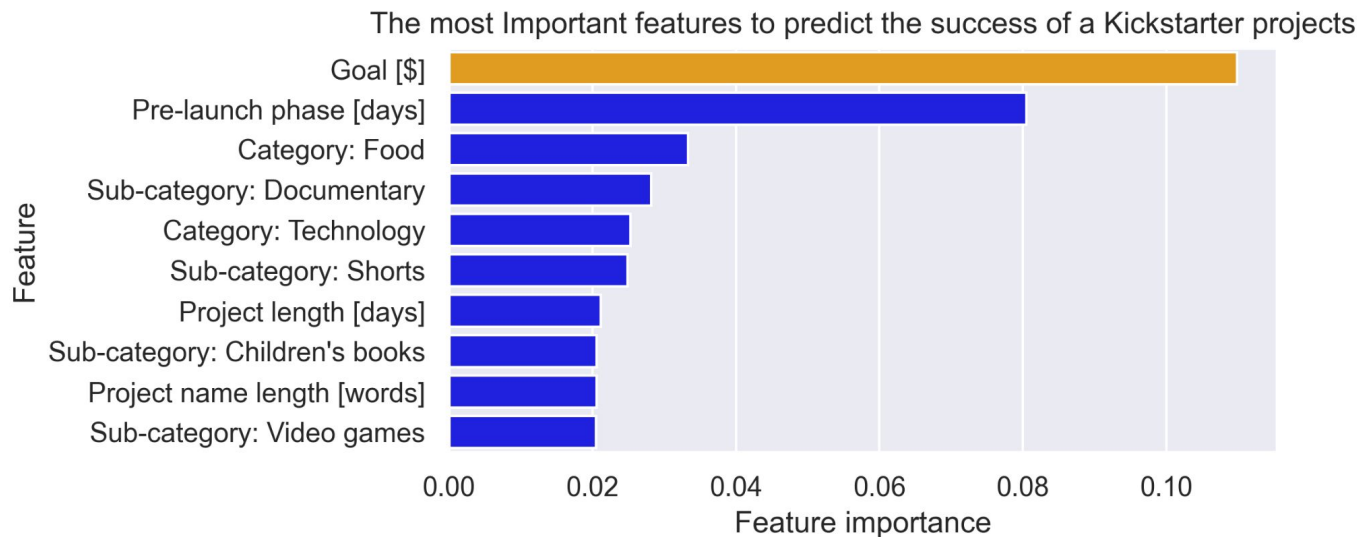  - but also **why → feature importance!**

wiki.pathmind.com

# Random Forest Model

- Correct predictions: 77%
  - 79% increase compared to our baseline model!
- Fewer missed opportunities compared to wrongly predicted successful projects
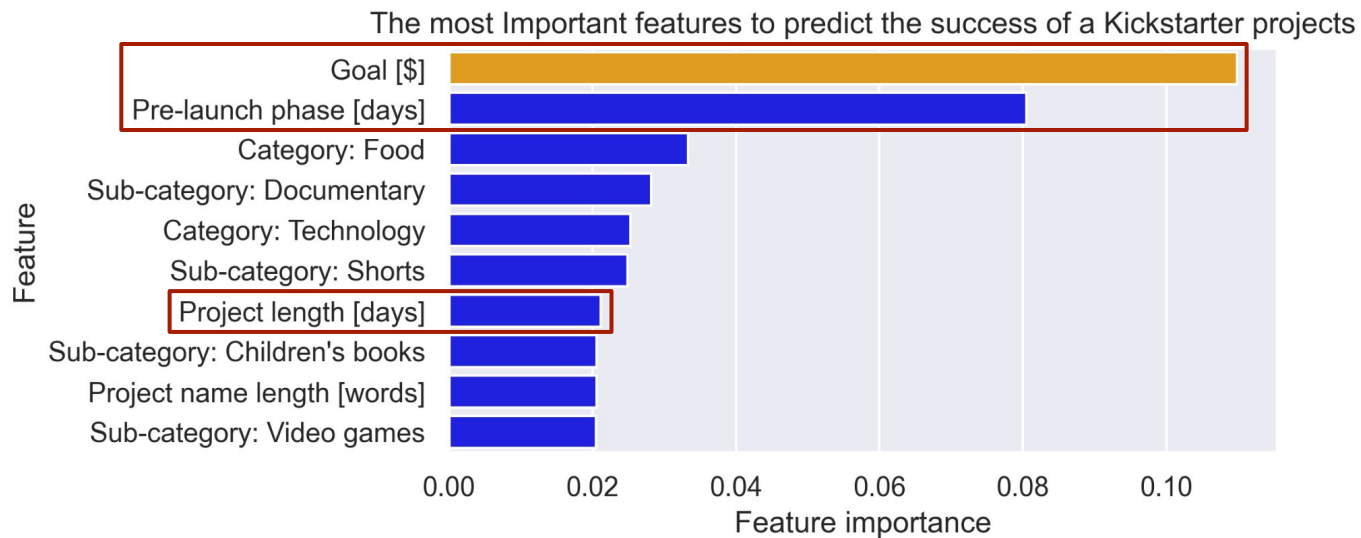  → desired outcome ✅



Random forest model predictions

# The most important factors

The most Important features to predict the success of a Kickstarter projects

# The most important factors

The most Important features to predict the success of a Kickstarter projects
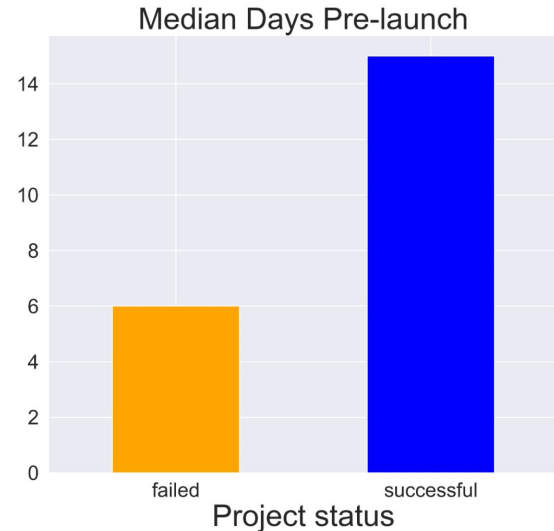
# A closer look at the pre-launch phase

- the pre-launch phase is
  - the phase between project creation
  - and the launch of the project

- this phase seems to be critical
  - a longer phase leads to more successful projects

- Kickstarter can support creators by
  - social media marketing
  - promotion videos
  - influencer marketing



Median Days Pre-launch

Project status

Number of projects per category
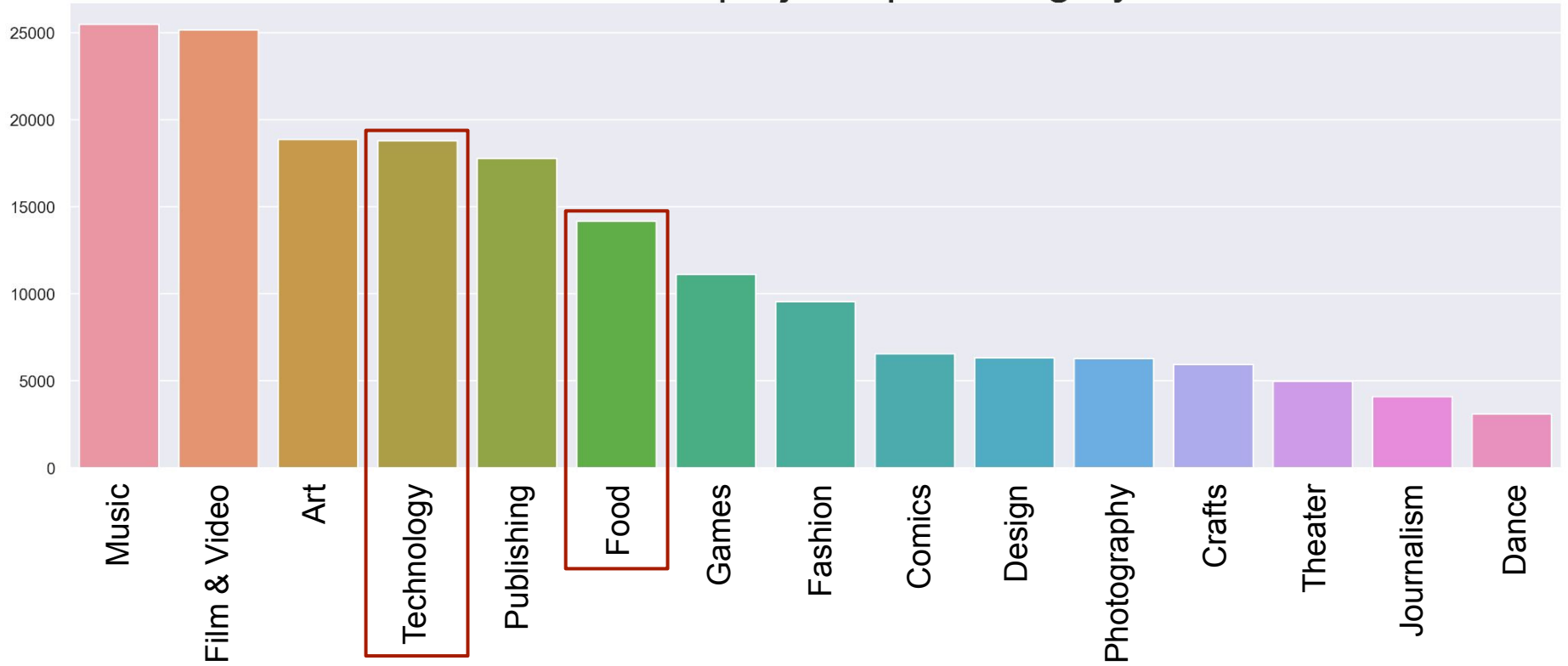
# The recommendations

- the funding goal is a crucial factor
  - lower funding goals lead more often to successful projects
  - is probably more realistic for the backers

- the pre-launch phase is very important
  - longer phases lead to more successful projects

- projects in certain categories are more successful
  - for example food and technology

- the project length has an influence
  - but it is only marginal
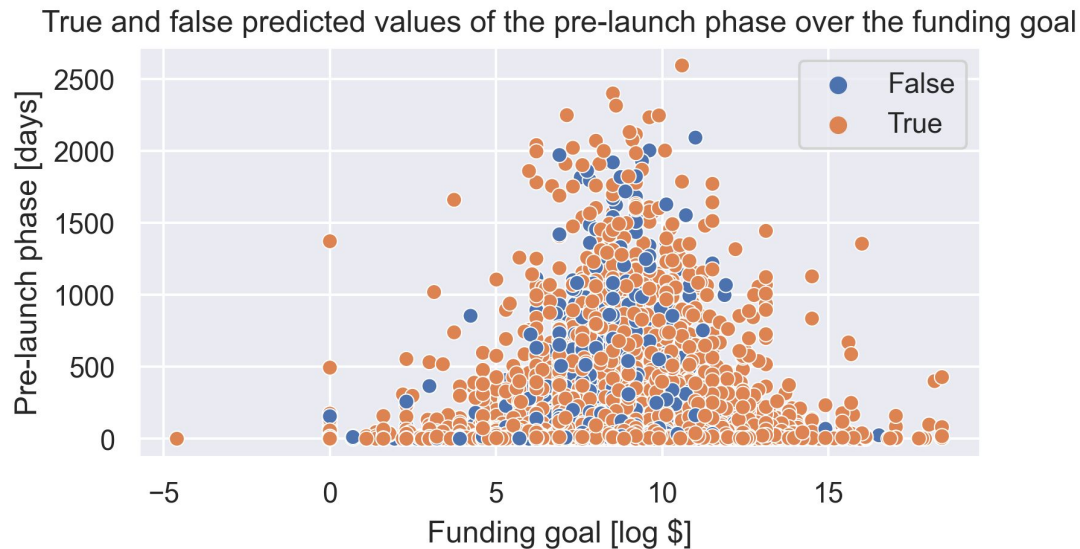
unsplash.com

# Outlook

- another model option would be the XGBoost (gradient boosting)
  - performs well for large number of (maybe irrelevant) features

- look at the different ranges of funding goals
  - there are a lot of projects with low funding goals (under $10 000)
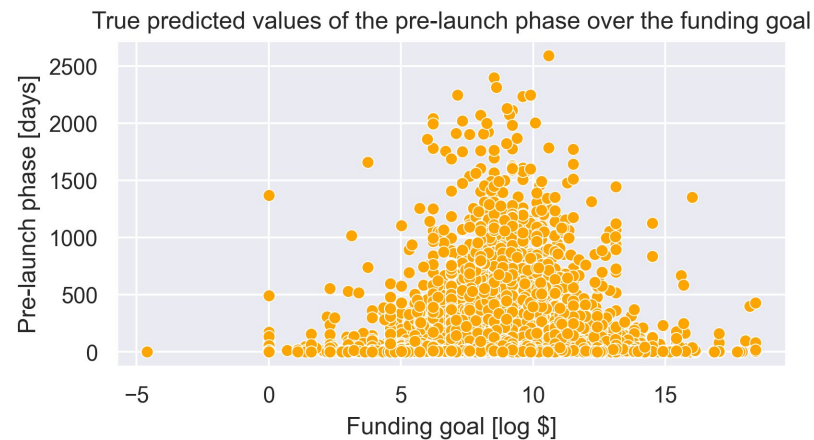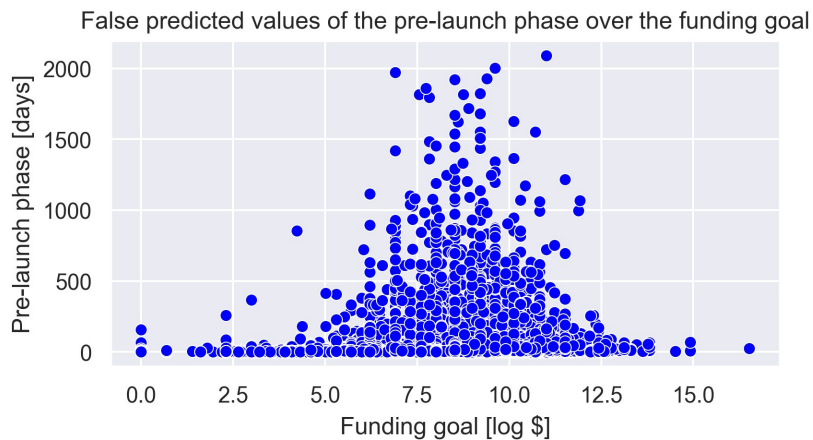  - but also some with much higher goals

unsplash.com

# Error analysis - True vs false predicted

True and false predicted values of the pre-launch phase over the funding goal

# Error analysis - True vs false predicted



False predicted values of the pre-launch phase over the funding goal



True predicted values of the pre-launch phase over the funding goal

# Error analysis - Numerical variables