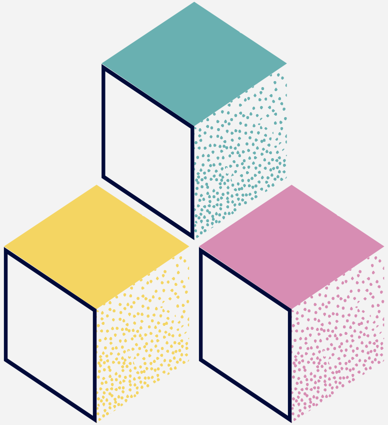
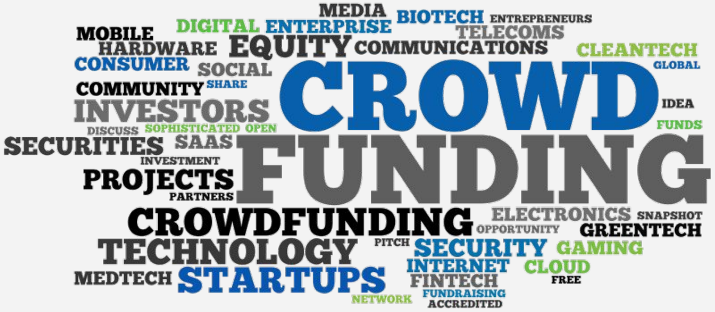


Predicting Kickstarter Project Success

Erick Cantu Perez, Su Wong

30.06.2022

Our stakeholder



Mr. Jacoby Stevens

- Retired entrepreneur
- Likes supporting crowdfunding projects
- Wants to give back to the community
- Wants to support the projects most likely to succeed
- Would like us to predict for him the likelihood of success or failure

About Kickstarter

- Founded in 2009, is a crowdfunding website
- Has an all-or-nothing funding model
- A project is only funded if it meets its goal amount; otherwise no money is given by backers to a project

About the dataset

- Dataset consists of 209222 projects between April 22 2009 and March 14 2019.

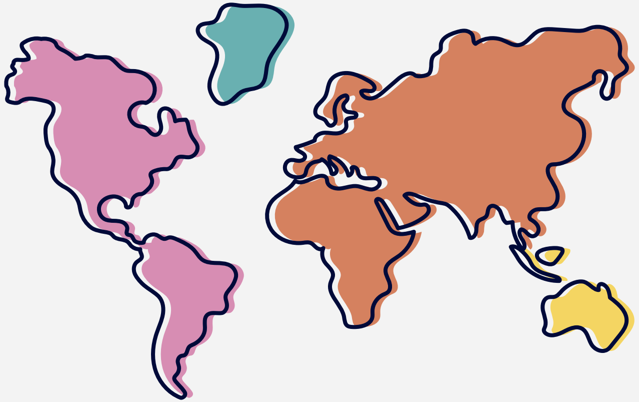
- 5 different outcomes for projects on Kickstarter: successful, failed, live, canceled, suspended. Only successful and failed projects were used to train our model.

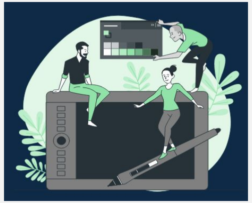
- Taking only successful/failed projects and eliminating duplicate projects left us with 168979 projects.

Exploratory Data Analysis findings

The predictors for our model are:

- Category (e.g. Art, Food, Technology)
- Country
- Fundraising goal in USD
- Campaign Duration
- Was the project selected by the staff as “staff pick”

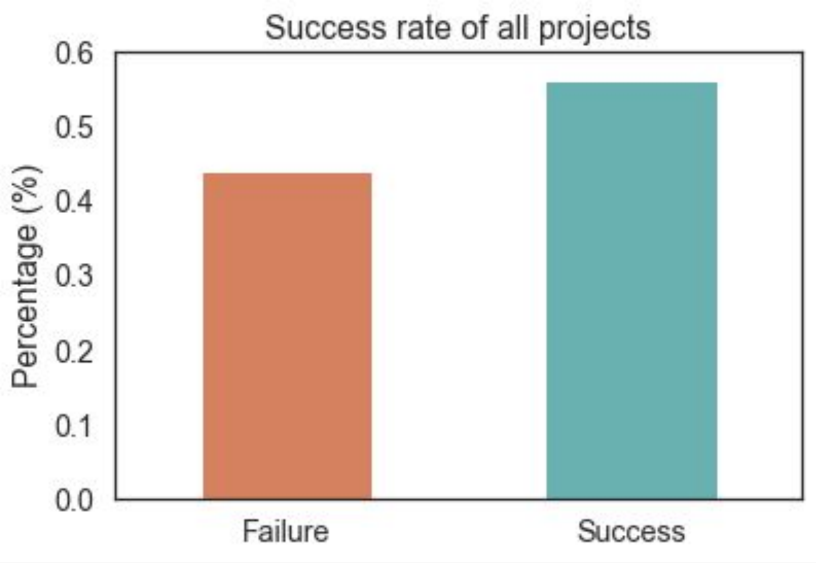




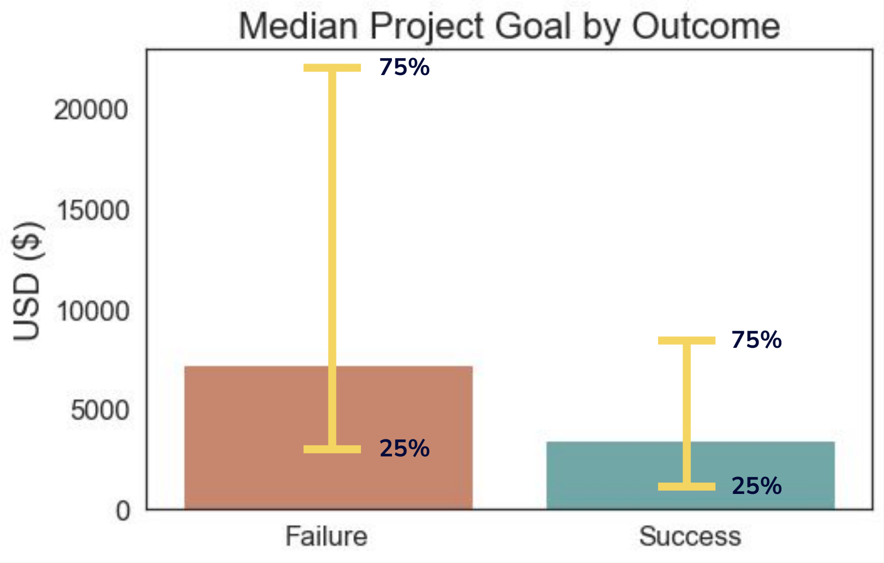
KICK STARTER STAFF PICK

Exploratory Data Analysis findings

Overall success rate of all projects is 56%



Goal ranged between 0.1 and 15.23M USD($)



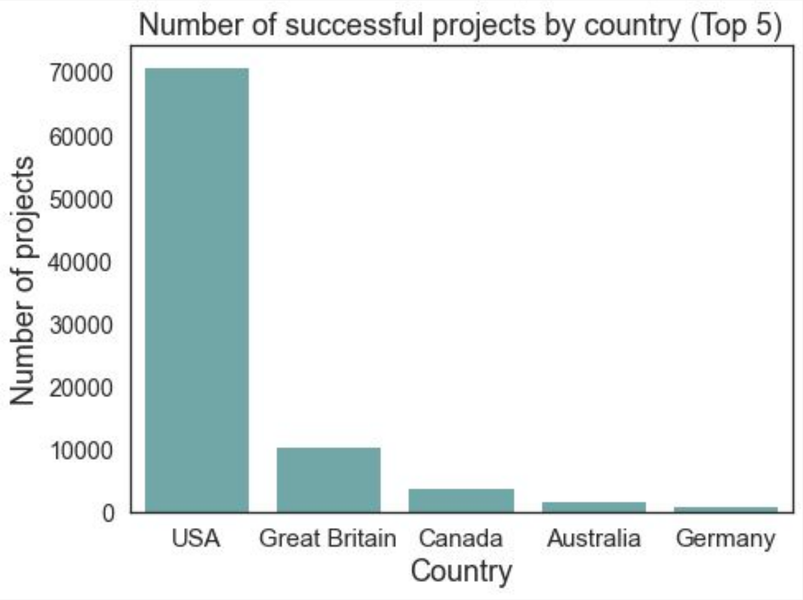
Baseline Model

Baseline Model:

We think that projects based in USA are likely to be successful

Evaluation Metric:

We train the model based on precision. Precision is the ratio of *correctly* predicted positives out of *all* of the results that were predicted positive.



Score for baseline model:

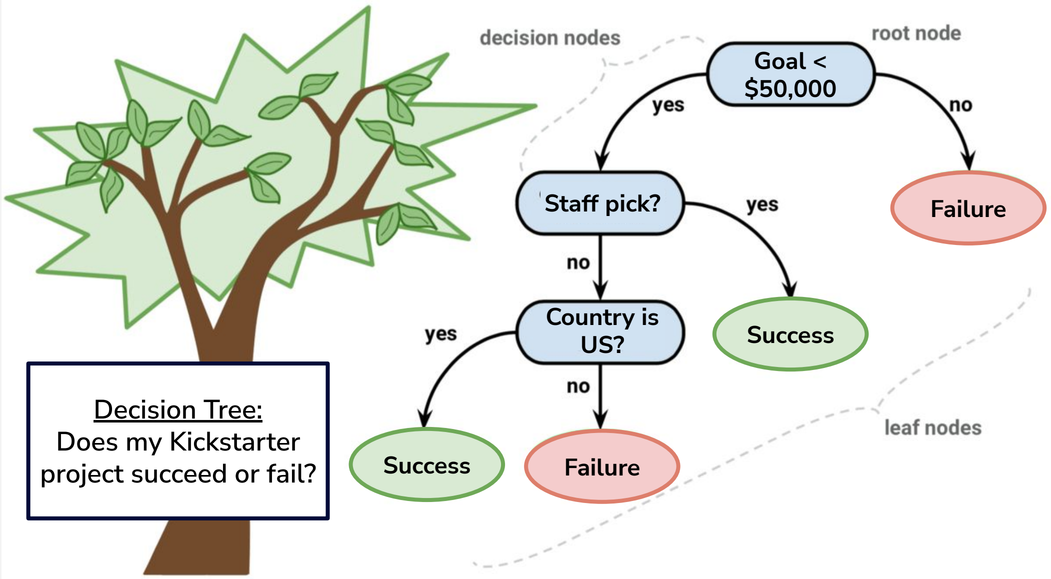
Precision = 57.8%, Accuracy* = 55.2%

*Accuracy is the percentage of all correctly predicted outcomes

Machine Learning models

Because our data has more categorical variables, we decided to focus more on tree-based supervised machine learning models:

1. Logistic regression
2. Decision Trees
3. Random forests
4. XGBoost
5. Adaboost



Decision Tree:
Does my Kickstarter project succeed or fail?

Results

XGBoost gives the best results “out of the box”

	Logistic Regression	Decision Tree	Random Forest	XGBoost	Adaboost
Precision*	69.5%	70.0%	70.5%	72.2%	69.4%
Accuracy**	69.6%	66.6%	68.4%	72.1%	67.7%

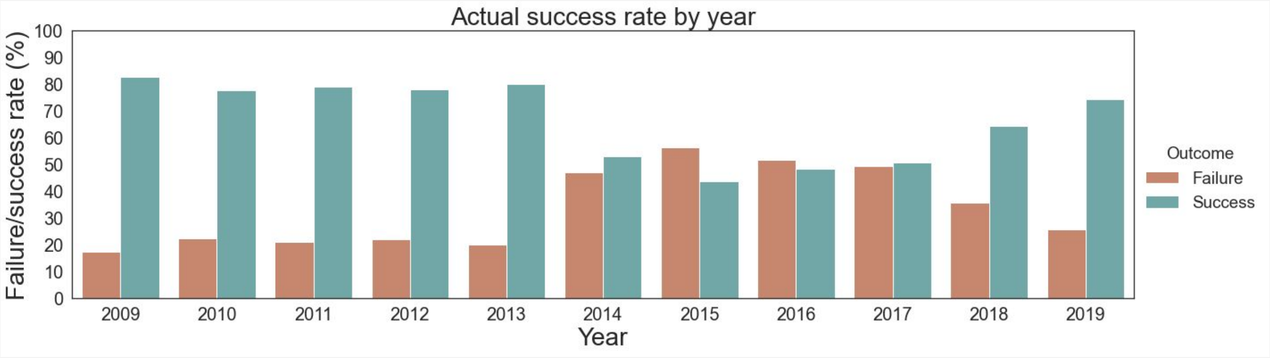
After optimizing the parameters:
Final precision: **72.1%** (Accuracy: **71.9%**)

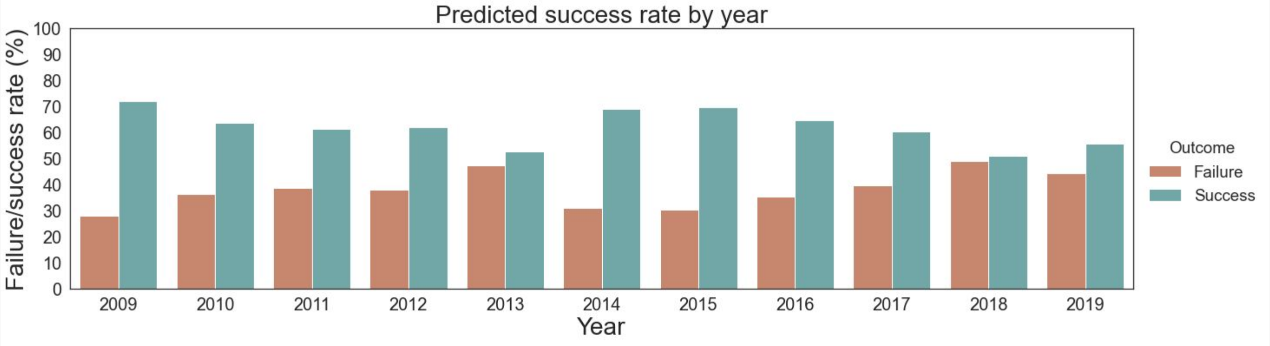
* Precision is the ratio of *correctly* predicted positives out of *all* of the results that were predicted positive.

** Accuracy is the percentage of all correctly predicted outcomes

Error Analysis

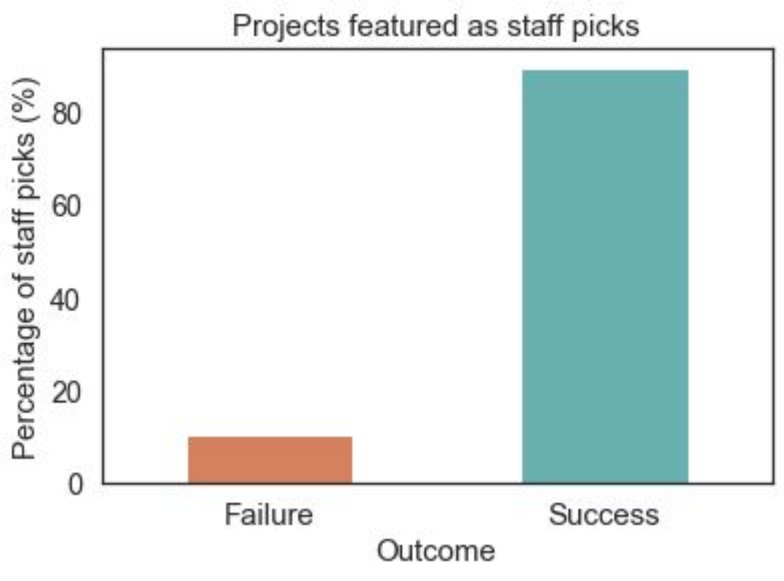
Our model is only 72% accurate. How can we improve this number?

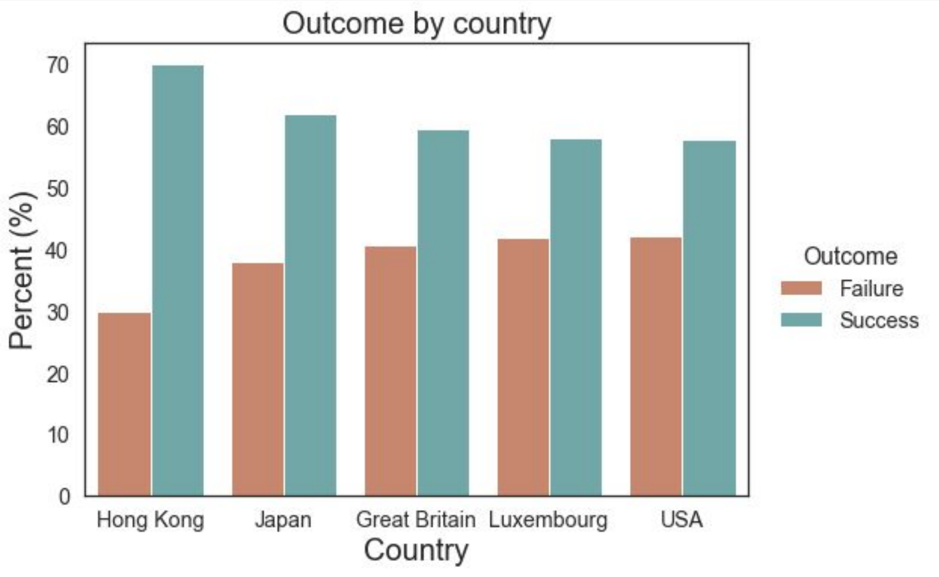




- Actual success rate from 2009 to 2013 is around 80%, which is not captured by the model.
- Reduction in success rate from 2013 to 2014 not captured
- Success rate from 2014 to 2017 is around 50%
- Success rate increases again in 2018 and 2019

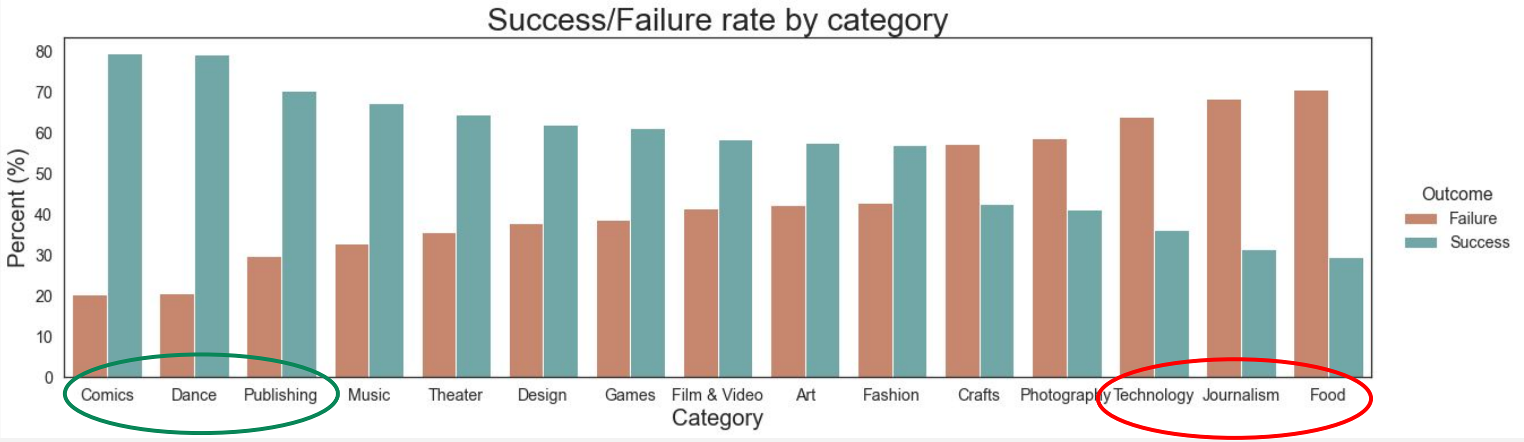
Recommendations

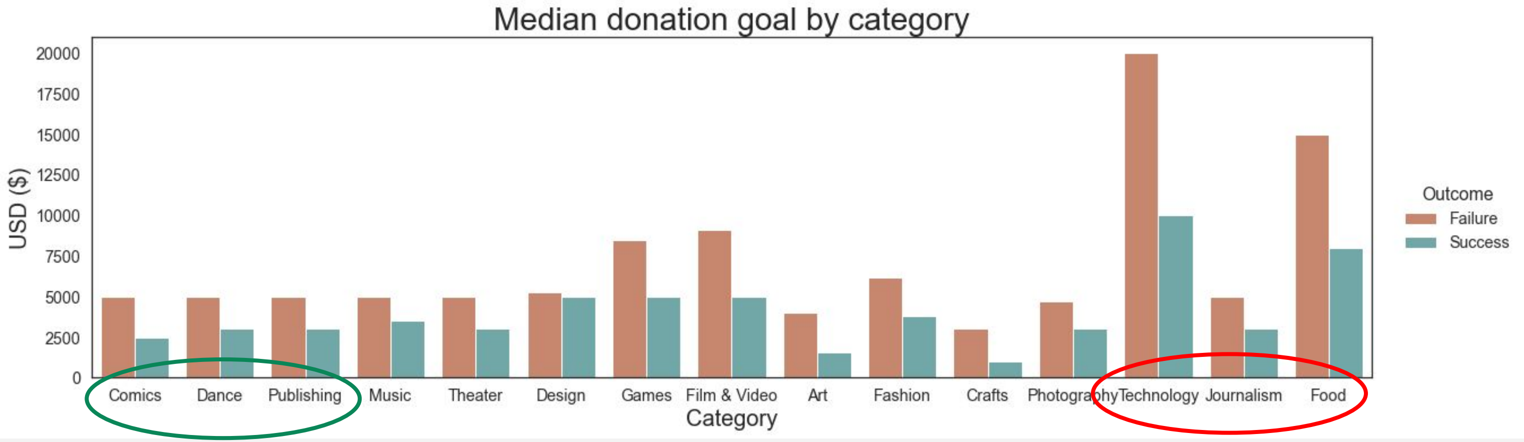




- Projects picked as “staff picks” have a high success rate
- Hong Kong, Japan, Great Britain are top 3 countries with highest success rate

Recommendations





Takeaways and Future Work

- Improvement in accuracy from 55.2% to 72.0% over the baseline model

- Best projects to invest in:
    - Comics, dance or publishing categories
    - Projects featured as “staff pick” when launched
    - Projects from Hong Kong, Japan or Great Britain

- Worst projects to invest in are technology and food and they also have the highest donation goals

- Future work: Explore underlying trends in the data by year or month, subcategory information

Thanks!

Questions?

Erick Cantu Perez
github.com/eaunaicr97

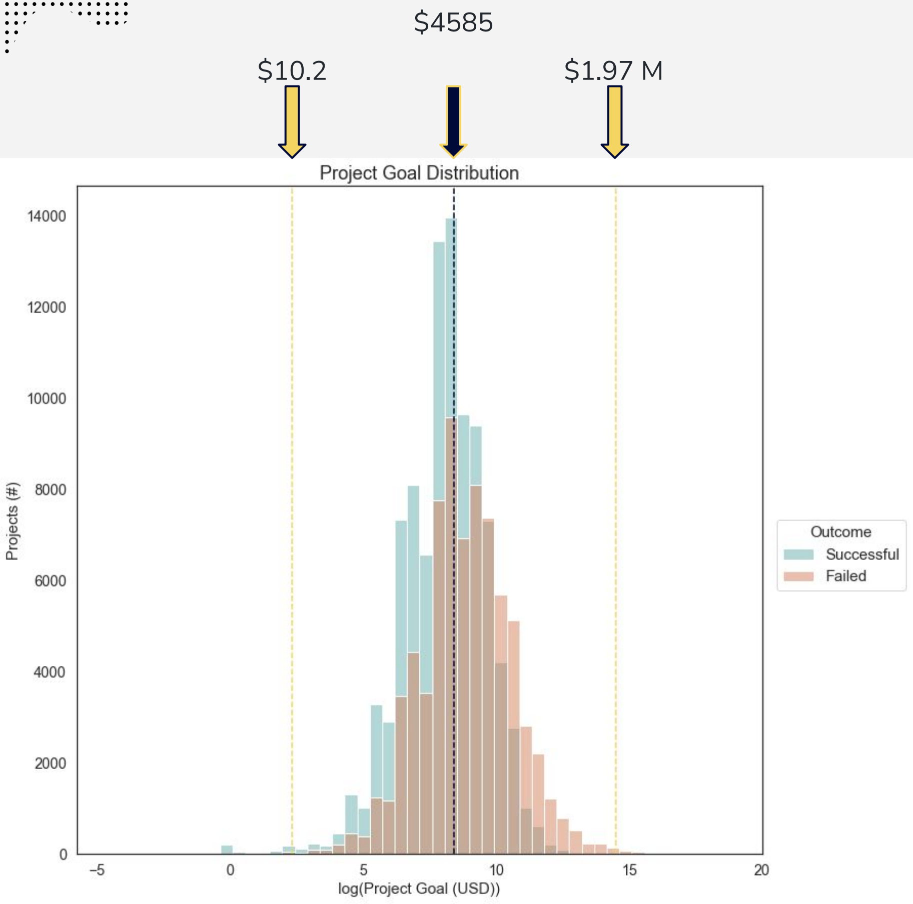
Su Leen Wong
github.com/suleenwong

Project repository:

https://github.com/eaunaicr97/ds-ml-project-kickstarter

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Exploratory Data Analysis findings



- Failed projects tend to have higher goals
- Project goal log transformed
- 0.523% observations are outside 3.5 std deviations
- 0.002 % successful projects >$1.97
- 0.049 % failed projects < 10.2

Error Analysis

Precision on train data: 0.738
Accuracy on train data :   0.738

Precision on test data: 0.7215
Accuracy on test data :   0.7185

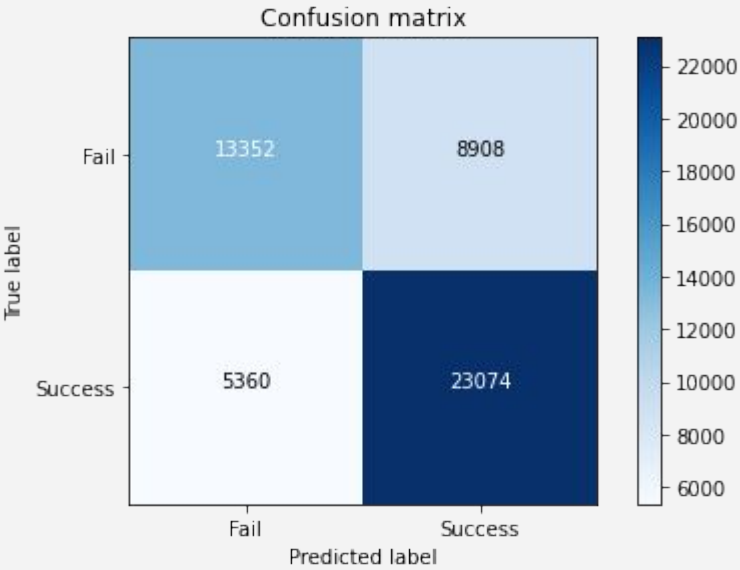
Accuracy on train and test data is quite similar, therefore there is no significant overfitting

Classification report:

```
              precision    recall  f1-score   support

           0       0.71      0.60      0.65     22260
           1       0.72      0.81      0.76     28434

    accuracy                           0.72     50694
   macro avg       0.72      0.71      0.71     50694
weighted avg       0.72      0.72      0.71     50694
```


Confusion matrix
True label
Fail
Success
Predicted label
Fail
Success