KICKSTARTER

Predicting Crowdfunding Project Outcomes

Institute of Data

Ruby Jang

Capstone Project

AGENDA

INTRODUCTION	Business context - Goals
DATA PREPARATION	Collect - clean - manipulate
ANALYSIS	Visualisation - Insights
MODELLING	Train - Evaluate
CONCLUSION	Recommendations - Next Steps

KICKSTARTER

Crowdfunding platform helping creators raise funds to bring creative projects to life

Creators bring creative project ideas

Kickstarter provides platform to raise funds

Backers pledge to projects of interest

Successful projects receive funds to bring ideas to life

Kickstarter gets 5% commission

Key Statistics

22M

Total Backers

233k

Projects Funded

40%

Success Rate

$7B

Total Pledged (USD)

How much has been pledged since 2009?

$7B

Total Pledged (USD)

$6.5B

Successful Dollars

5%

$325M

Profit made

$567M

Unsuccessful Dollars

5%

$28M

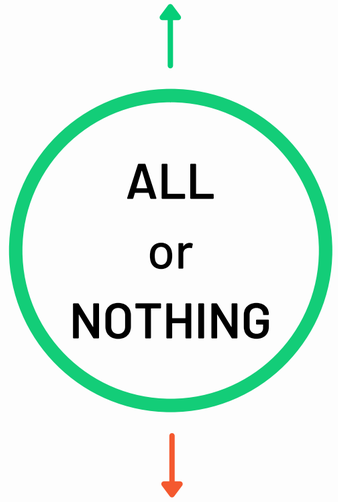
Profit missed

$2.2M

Profit missed per year

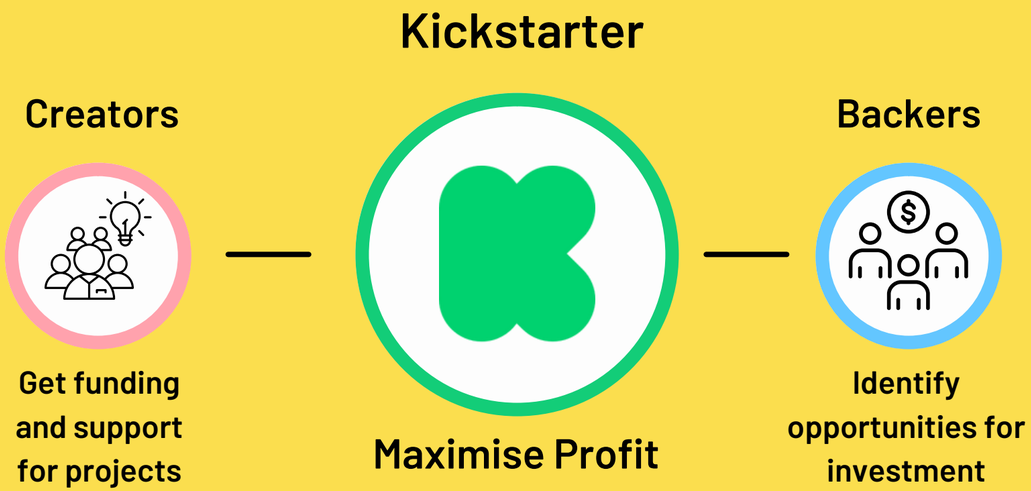
BUSINESS PROBLEM

For every successful project,
Kickstarter earns 5% commission.



For every unsuccessful project,
Kickstarter earns $0.

Who can benefit?



BUSINESS GOAL

Increase project success rate to maximise **profit.**

- Understand key factors behind success
- Identify projects to promote
- Guide creators and backers

PROJECT GOAL

- Predict outcome of a given project, from information available prior to launch
- Identify key factors behind success

DATA

- Source - Web Robots web scraper
- Data from every month in 2022 collected
- 542,385 projects, 40 columns

Project
ID
Name
Blurb
Category
Country
Creator
Number of backers
Staff pick
Spotlight
Status (target)

Currency
Currency
Currency symbol
Exchange rates
Goal (Local)
Pledged (Local)
Pledged (USD)
Current currency
Currency trailing code
USD type

Date Time
Created date
Launched date
Deadline
Status changed date

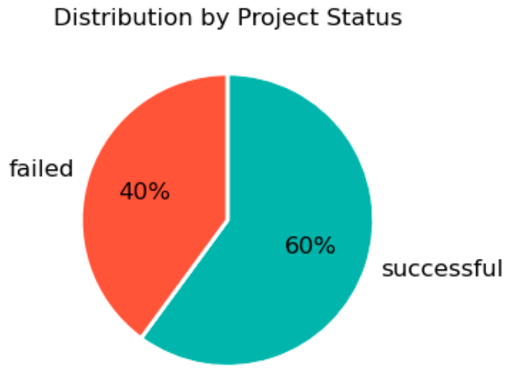
Other
Photo
Profile
Source URL

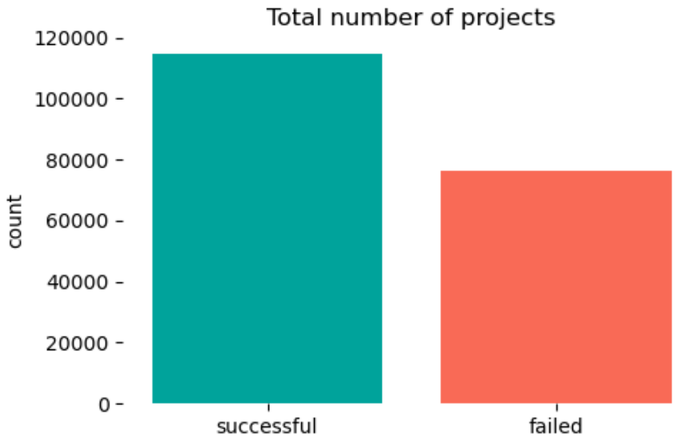
Dropped
is_starred*
is_backing*
permissions*
friends*
is_starrable**
disable_communication**

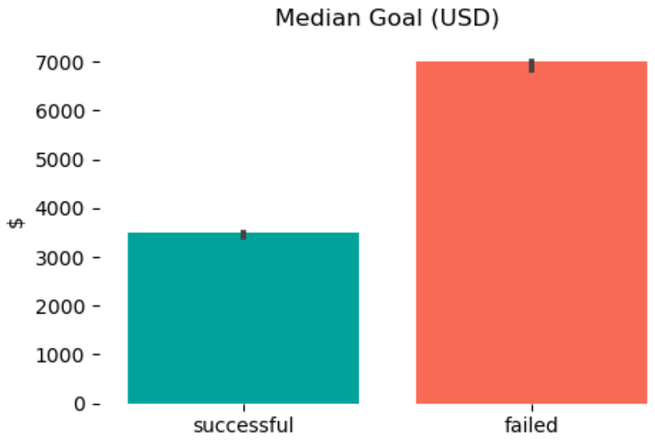
** Less than 0.05% complete*
*** Only contains single value*

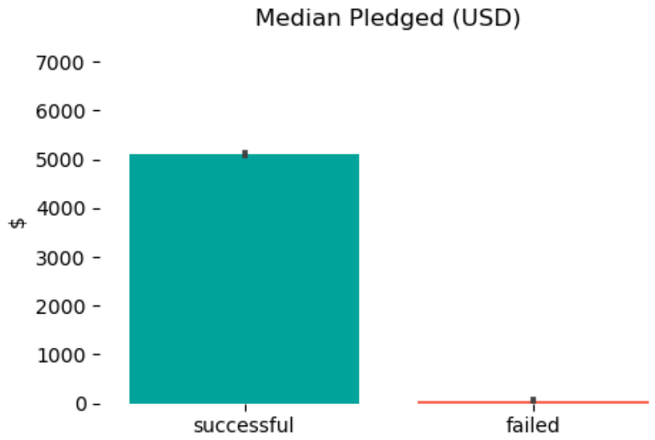
DATA CLEANING Remove duplicates, incomplete columns, date time & currency conversion

VISUALISATION BY STATUS





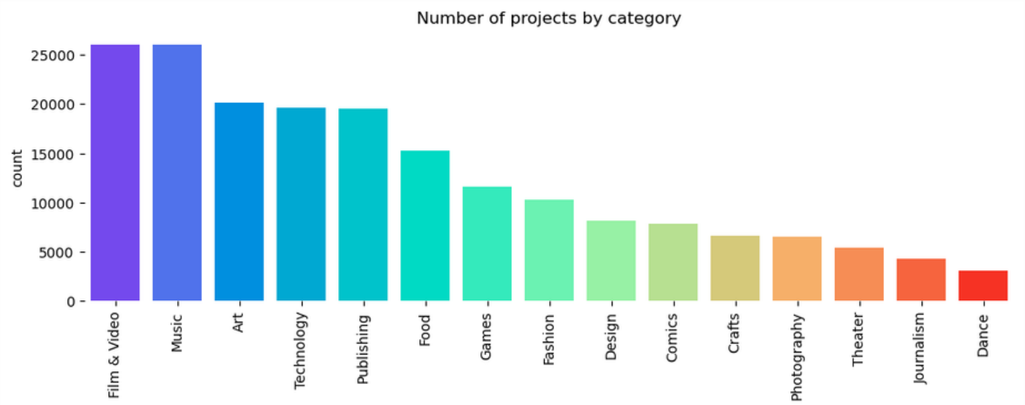


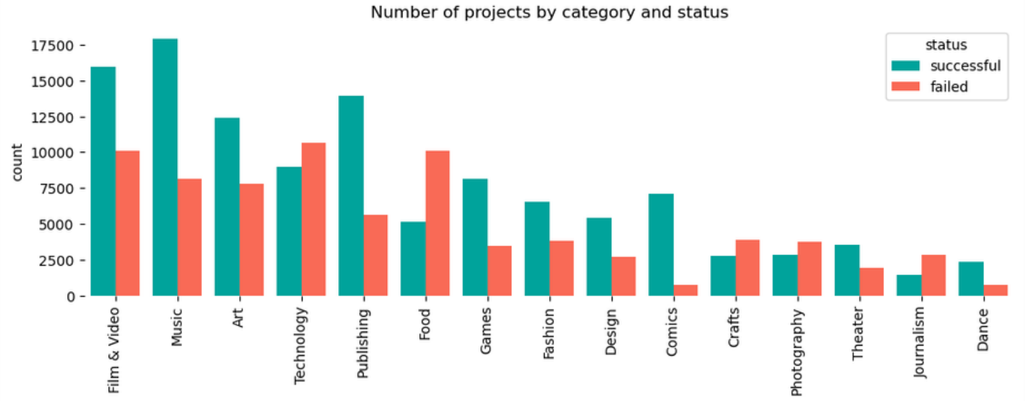


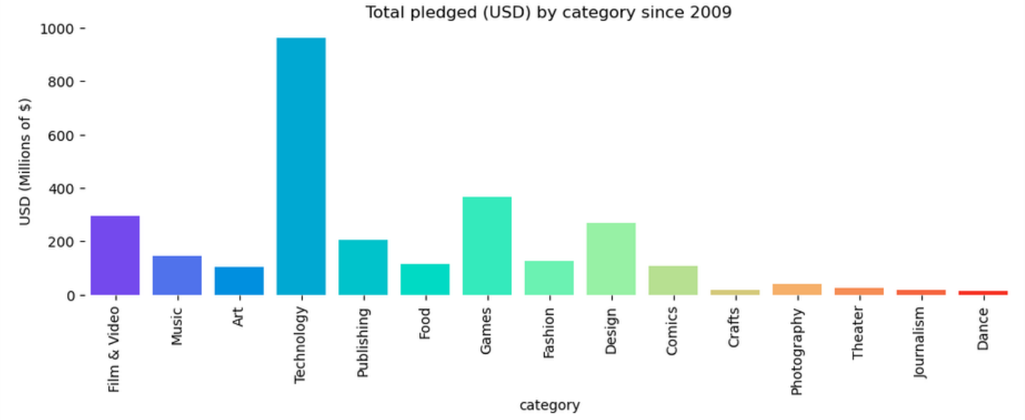
- **Successful** - 114,464 (60%)
- **Failed** - 76,134 (40%)

- Typical successful goal - $3500
  Exceeded goal by ~$1500
- Typical failed goal - $7000.
  Many got almost no funding.

BY CATEGORY





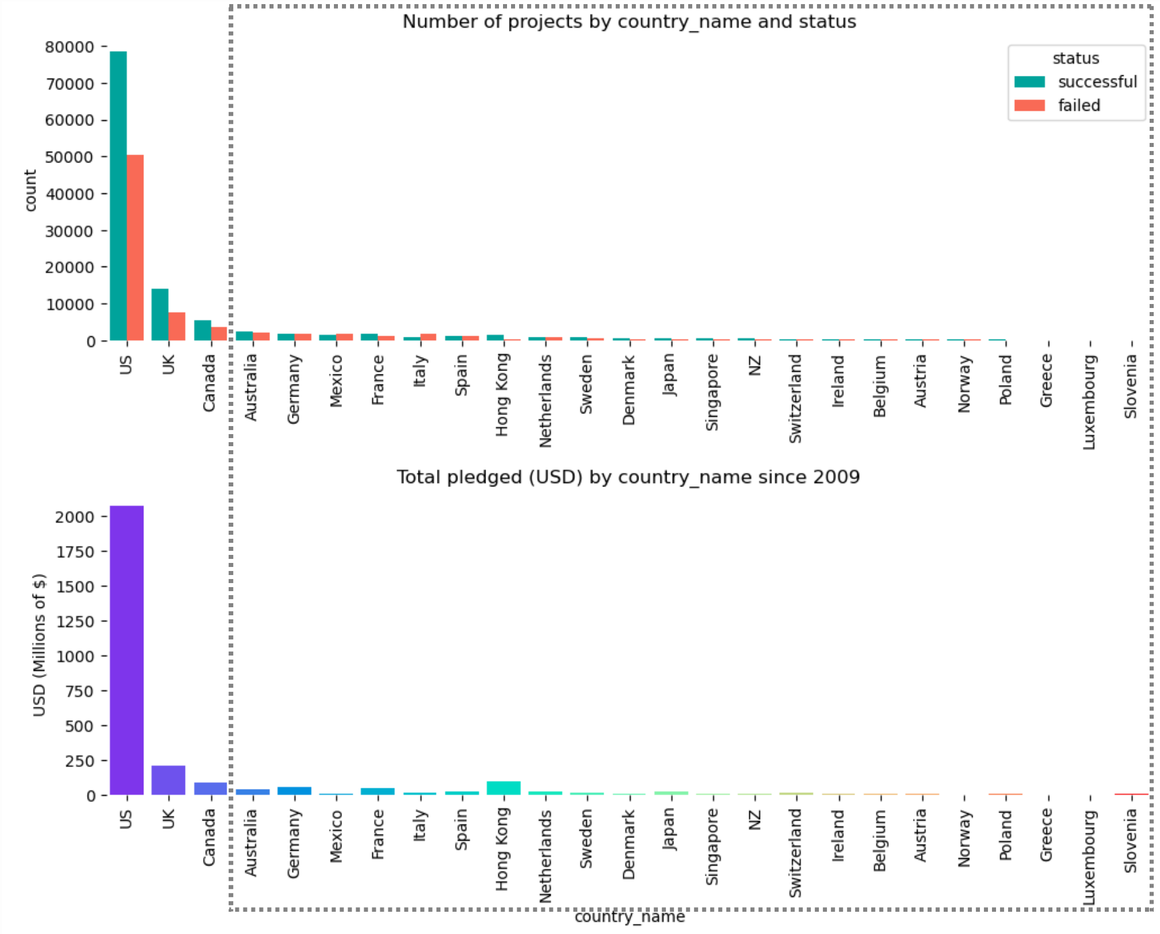


- Most popular - film & video, music, art
- Least popular - dance, journalism, theater

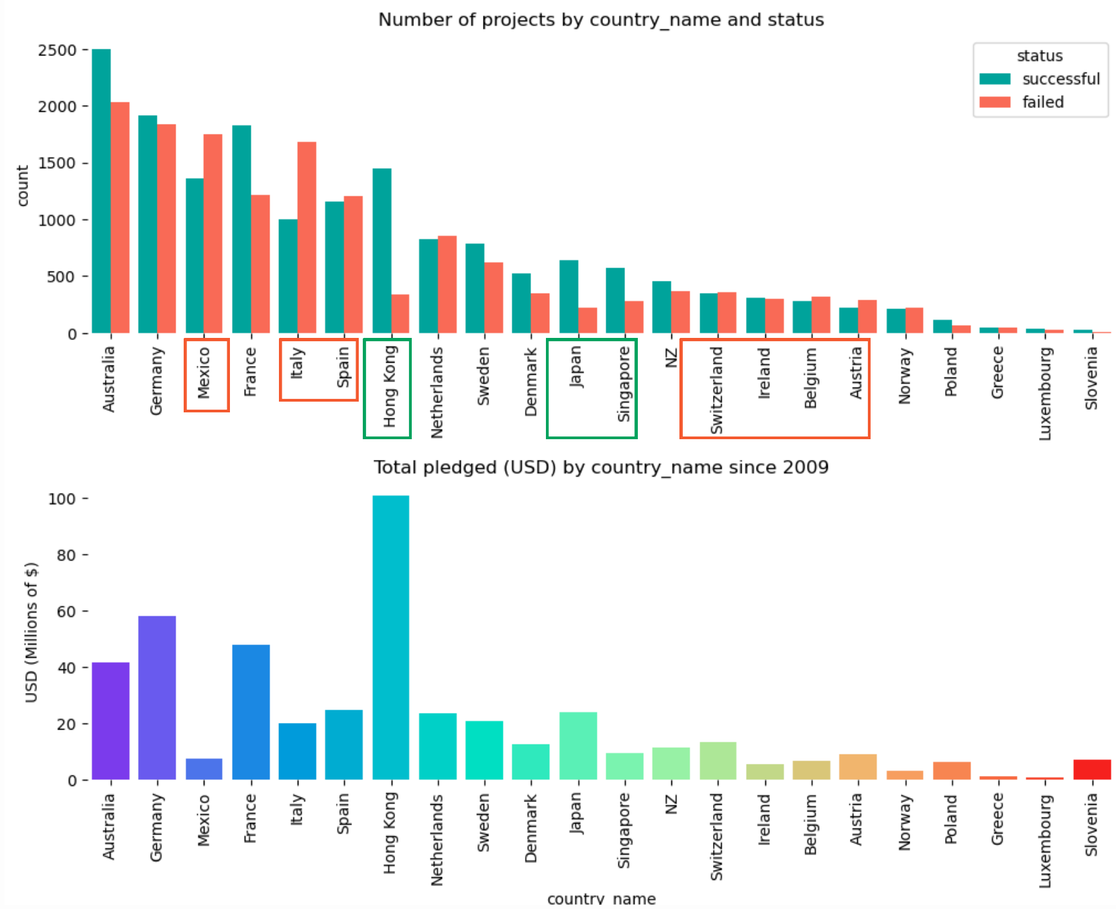
- Comics - highest % of successful projects
- Technology, food, crafts, photography, journalism - higher % of failed projects

- **Technology** attracted the most funding (~$1B), then **games** (~$400M), **film & video** (~$300M)

BY LOCATION



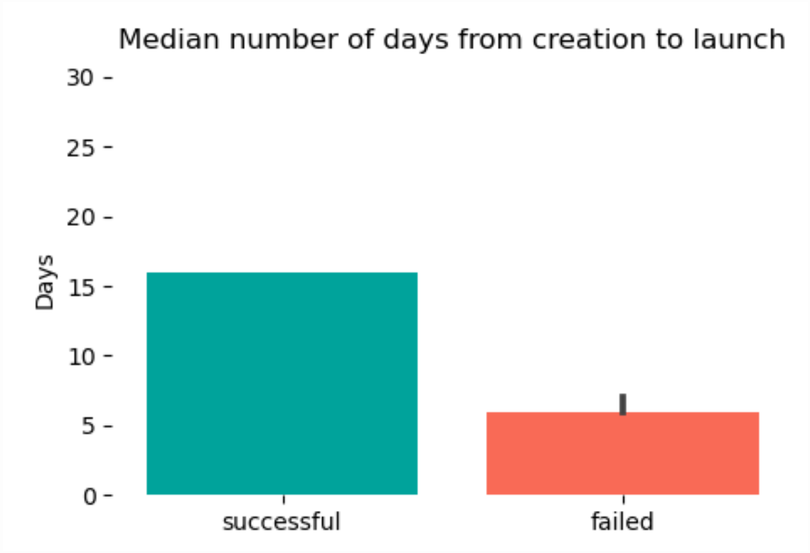
Exclude US, UK, Canada

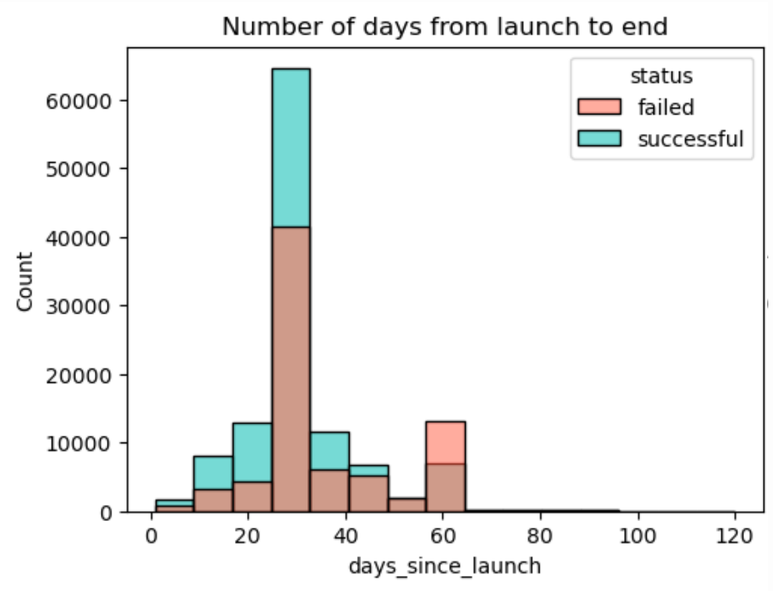


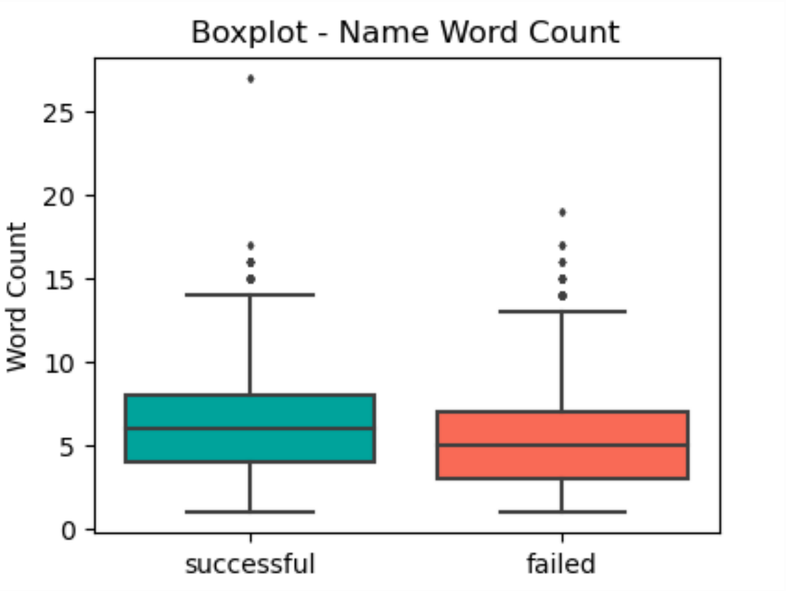
Most projects have been based in US, raising a total of over $2B in funding.

Projects based in **Asia** - high proportion of success
Projects based in **Mexico/Europe** - higher proportion of failure

BY DURATION & NAME







Successful projects take ~10 days more from creation to launch

Most projects are live for 30 days. As duration gets longer, projects tend to fail

Successful projects tend to have slightly longer project names.

MODELLING OVERVIEW

1

DATA PRE-PROCESSING

- Define predictors
  - Exclude future data
- Define target (Status)
- Feature engineering

- Project duration,
- Combined category
- Word count, polarity

2

CHOOSE MODELS

- Logistic regression
- Gaussian Naive Bayes
- Random Forest
- Adaboost
- XGBoost

3

MODELLING
BASELINE

Train models with default settings. Evaluate baseline performance.

4

HYPERPARAMETER TUNING

Find optimal hyperparameters for each model.

5

MODELLING
HYPERPARAMETER TUNED

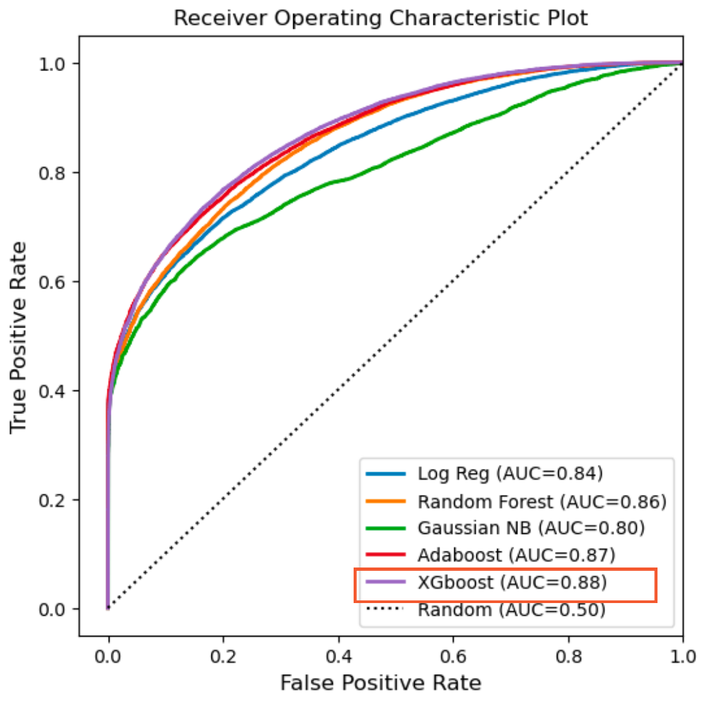
Train models with selected hyperparameters. Evaluate against baseline.

EVALUATION

- **Accuracy** - main evaluation metric as target is well balanced
- **Precision** - number of projects **falsely** labelled as successful (false positive)

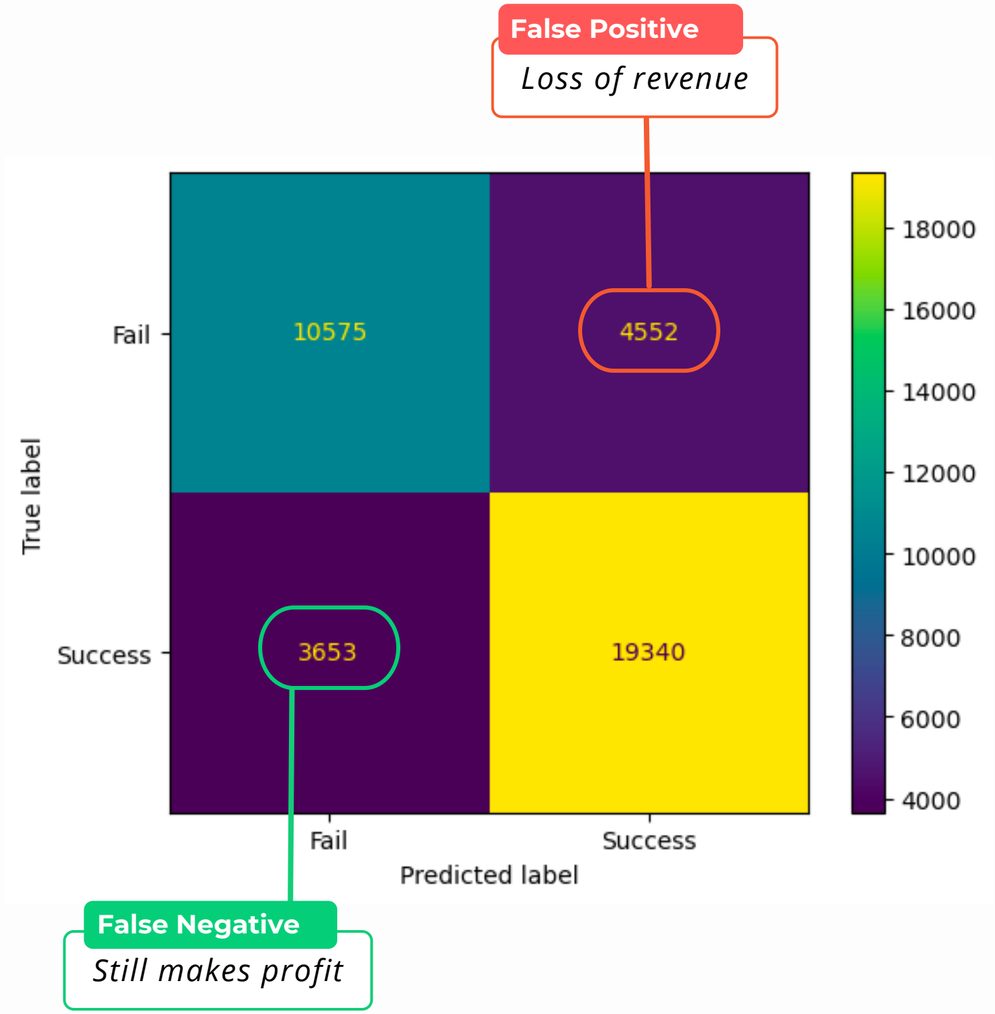
	Logistic Regression	Gaussian Naive Bayes	Random Forest*	Adaboost	XGBoost
Accuracy (Baseline)	75%	63%	78%	76%	**78%**
Accuracy (Tuned)	75%	70%	78%	78%	**79%**
Precision (Baseline)	81%	99%	78%	80%	**82%**
Precision (Tuned)	81%	92%	77%	81%	**81%**

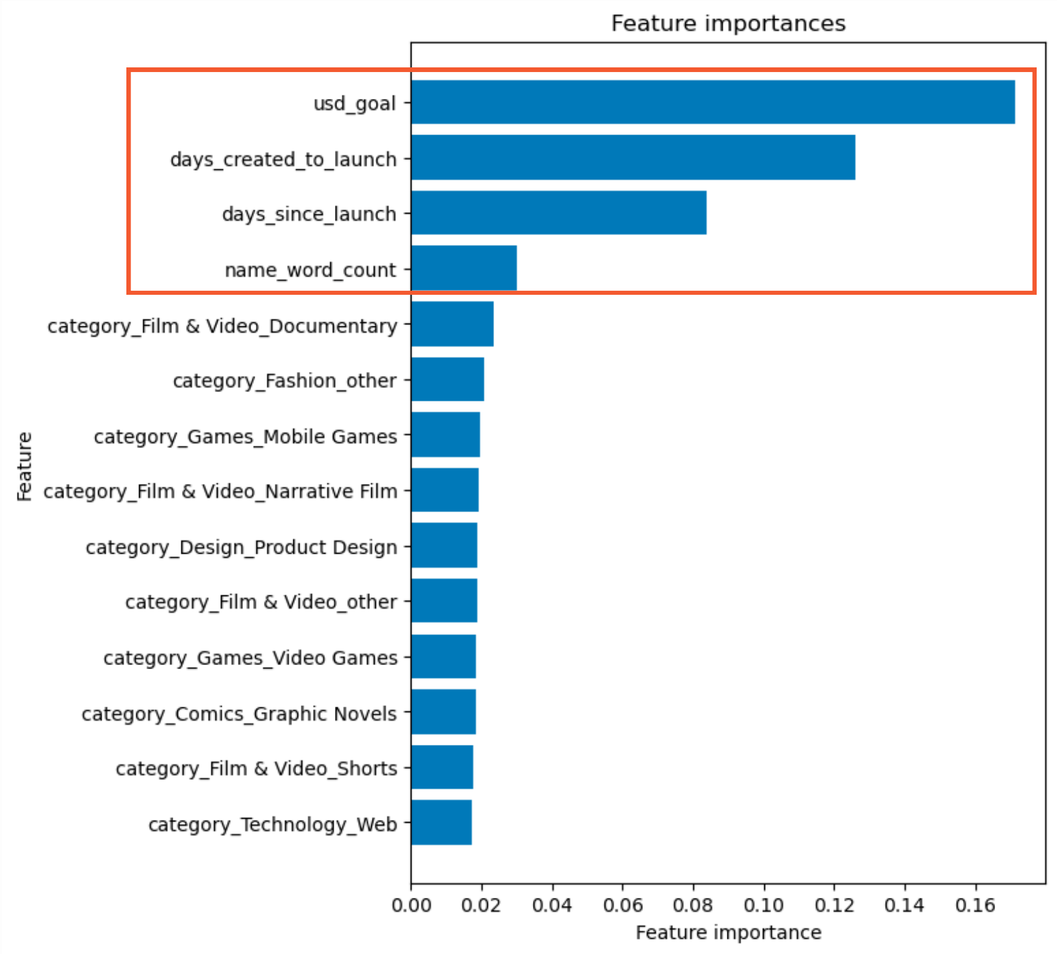
** Signs of overfitting present in baseline*



Considering accuracy, precision and AUC scores, **XGBoost** has been chosen as the best model.

EVALUATION





Funding goal, project duration and project name word count are important features.

CONCLUSION

- Business goal - increase **revenue** by increasing **success rate**
- We were able to predict the outcome of projects with **79%** accuracy
- Identify & promote projects at risk of failure
- Factors driving success rate include
  - Funding goal
  - Project duration
  - Category
  - Location

RECOMMENDATION

Funding Goal

Encourage smaller, realistic goals (average ~$3500)

Project Creation

Taking time between creating project to launch (1~2 weeks).

Duration

Shorter is generally better. Recommend around 30 days.

Category

- Technology attracts significant pledges at lower success rate
- Comics, games, film & video projects have high success rate

Next Steps

- Improve data collection process
- Deploy as web app
- Predict $ pledged

THANK YOU

Q & A