



SUCCESS IS NO ACCIDENT

Predicting Kickstarter Campaign Success

Presented by: Crystal Huang
Metis Bootcamp Project IV







\$ \$ \$



KICKSTARTER

Powerful crowd-funding platform

201,498

Successfully funded projects

\$ 5,810,323,737

Total dollars pledged to Kickstarter projects



All or nothing. This project will only be funded if it reaches its goal by Wed, June 2 2021 2:59 AM EDT.

Challenges:

- **All-or-nothing** Funding model
 - **No** one is charged for a **pledge** towards a project **unless** the funding **goal is reached**
- **Creators** want to succeed
 - If campaign fails, **time and money wasted** for campaign creators
- **Backers** don't want to miss out on an amazing successful campaign





Goal:
Build a classification model to
predict Kickstarter campaign
success



TOOLS

Data Source

Datasets obtained from a scraping website and complied by aggregating monthly data from January –April 2021

870,114 rows,
38 columns

Data Storage & Access



SQLAlchemy

Data Manipulation



pandas



NumPy

Plot graphs



matplotlib

seaborn

Modeling & Testing



**scikit
learn**

App



Flask
web development,
one drop at a time

METHODOLOGY

Metric Selection



- **ROC/AUC** curve for model comparison
- **F1 score** –
 - Precision - Minimize **False Positives**
(not predict too many success that will turn out to be failure)
 - Recall - Minimize **False Negatives**
(capture as many success as possible)

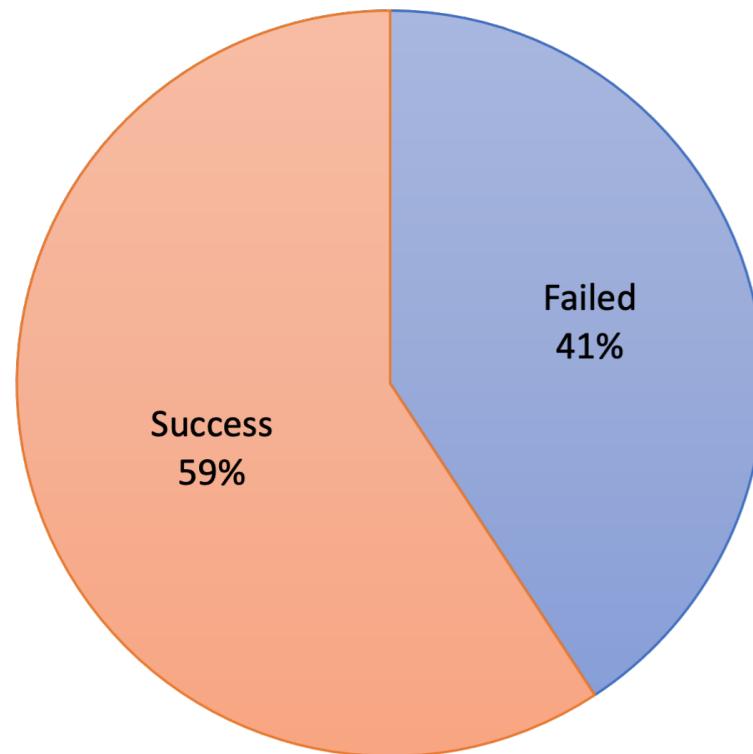


METHODOLOGY

EDA & Feature Engineering

Target
Success Outcome

Percentage of Campaign Outcome



Not too imbalanced



METHODOLOGY

Features

- ~~Backers, pledges, etc.~~
- Goal (in USD)
- Campaign Duration
- Preparation Duration
- **Getting Featured**
- Based in US
- Category (dummies)
- Length of Description

EDA & Feature Engineering



Watermelons Are Not Strawberries

A photobook about inner transformation and the pursuit of self-awareness that comes from navigating the challenges of parenting.

Sandra Bacchi
First created · 7 backed More

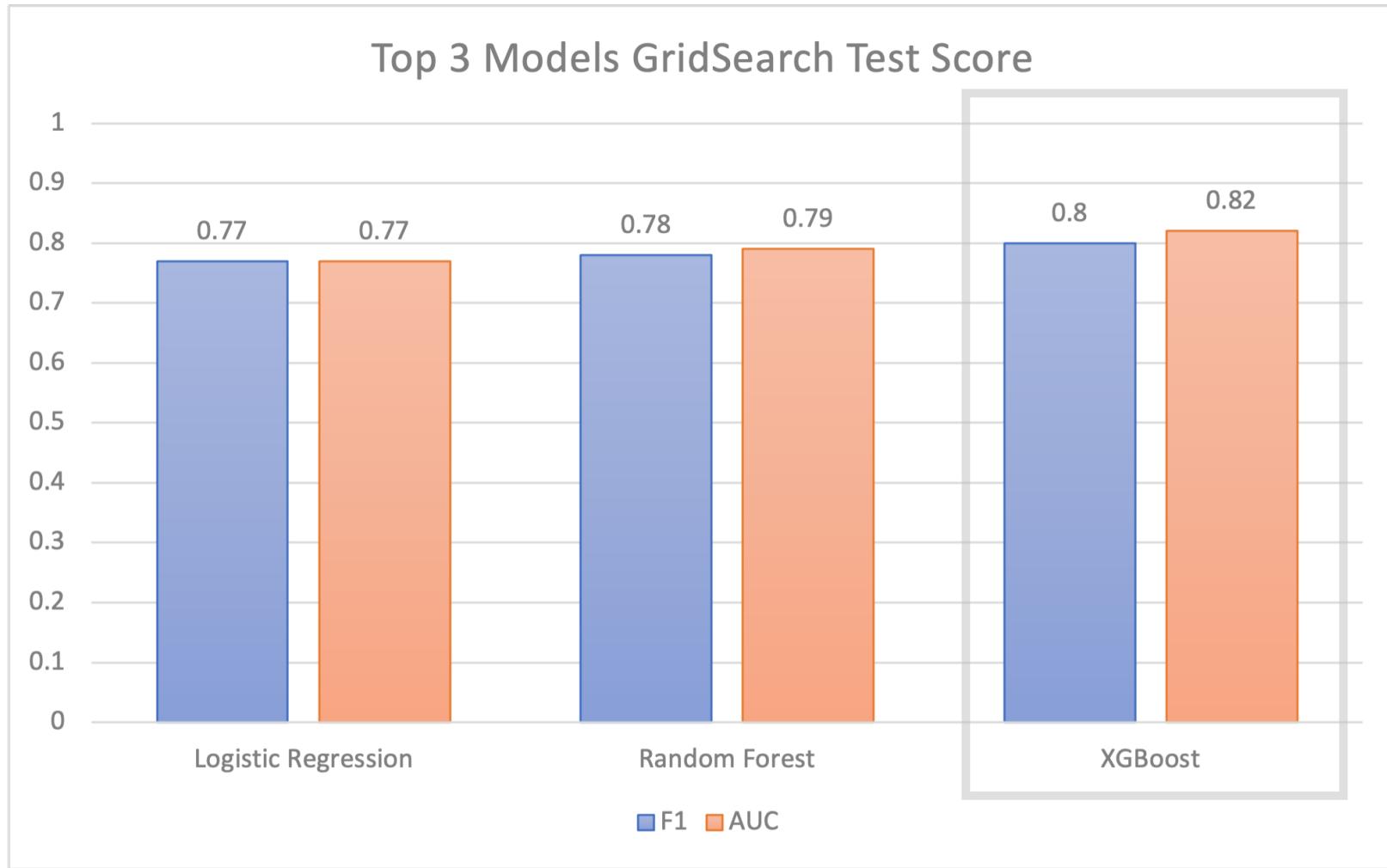
\$7,082 pledged of \$14,000 87 backers 23 days to go

Project We Love Photobooks Pittsburgh, P

Final Dataset:
189,162 data points
20 features

METHODOLOGY

Models Evaluation & Selection



RESULTS

Final Model Performance on Test Set

Final Model: XGBoost

	Validation	Test
F1	0.81	0.80
AUC	0.82	0.82

	Baseline Model	Final Model (XGBoost)
F1	0.75	0.80
AUC	0.65	0.82

▲ 0.05

▲ 0.17

RESULTS

Final Model Performance on Test Set

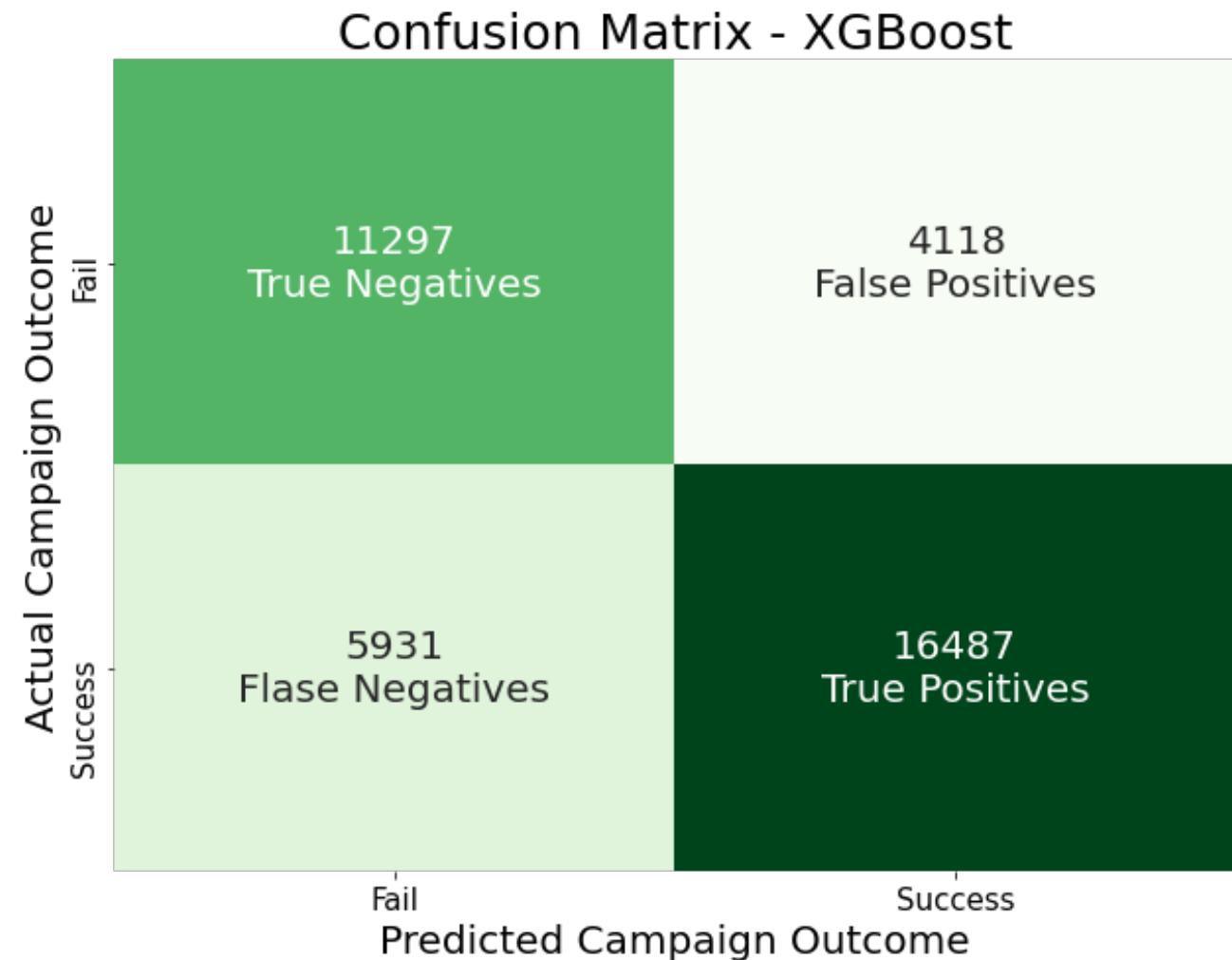
Final Model: XGBoost

	Validation	Test
F1	0.81	0.80
AUC	0.82	0.82

	Baseline Model	Final Model (XGBoost)
F1	0.75	0.80
AUC	0.65	0.82

▲ 0.05

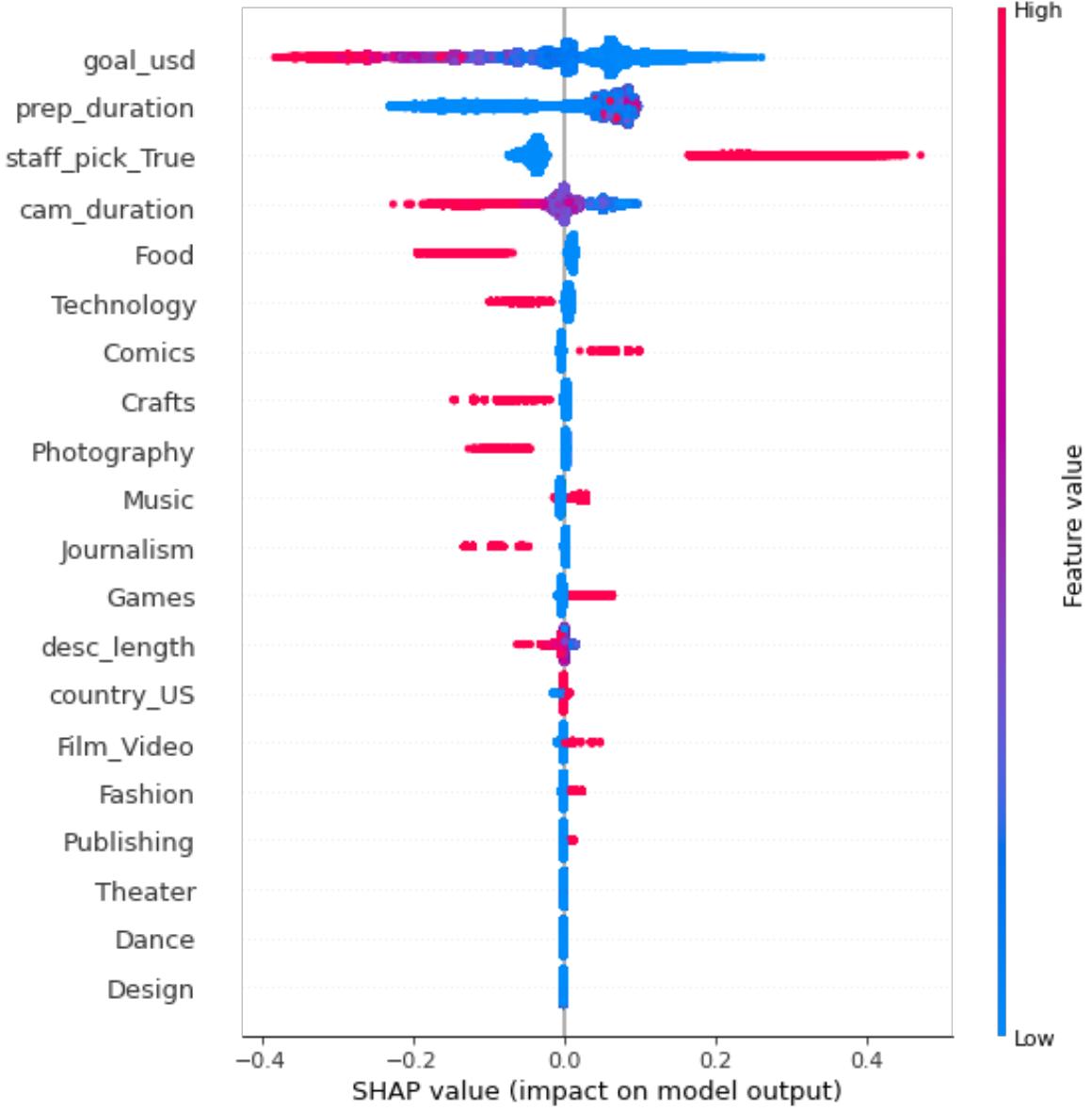
▲ 0.17



RESULTS

Feature Importance

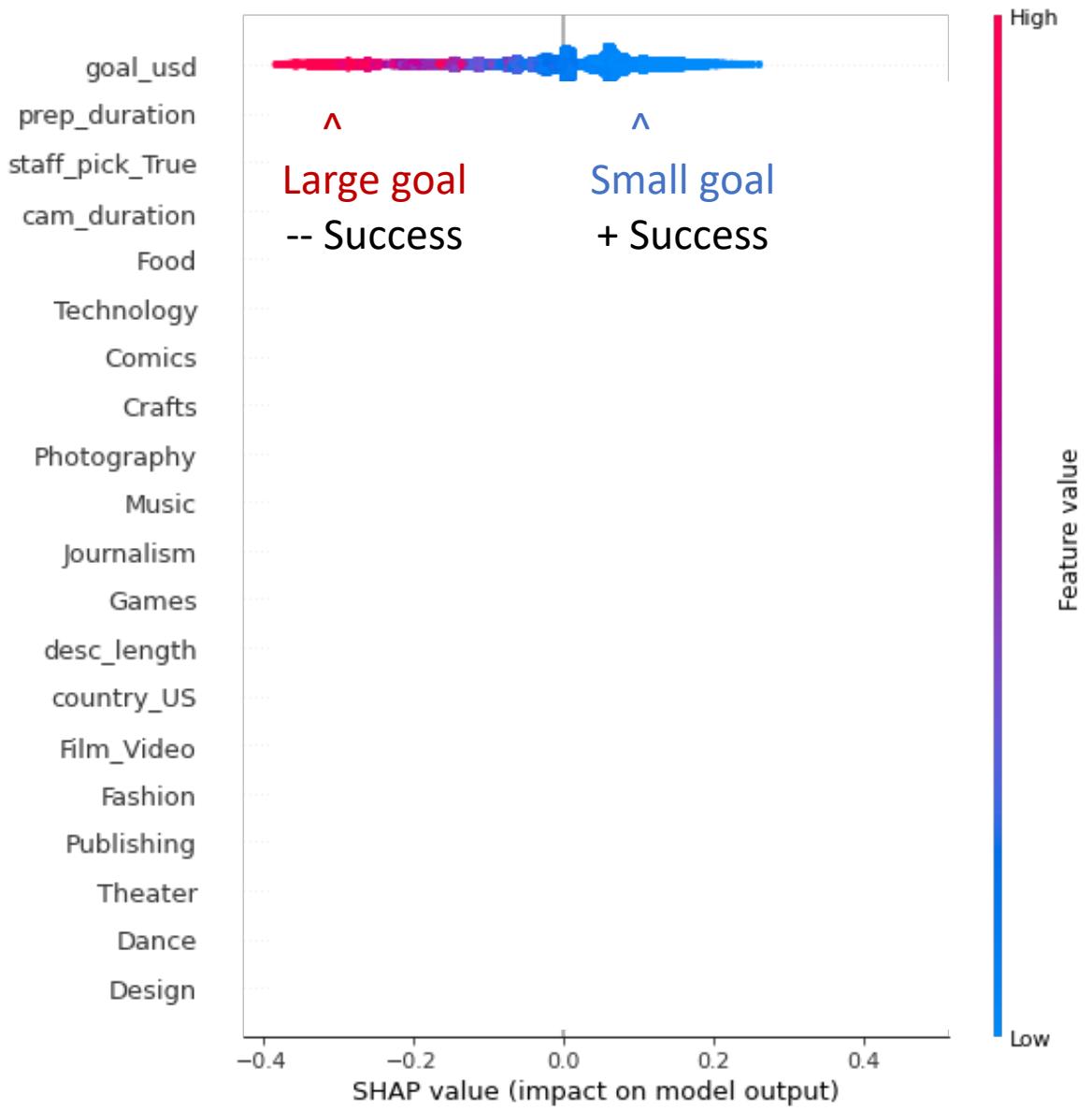
- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters



RESULTS

Feature Importance

- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters

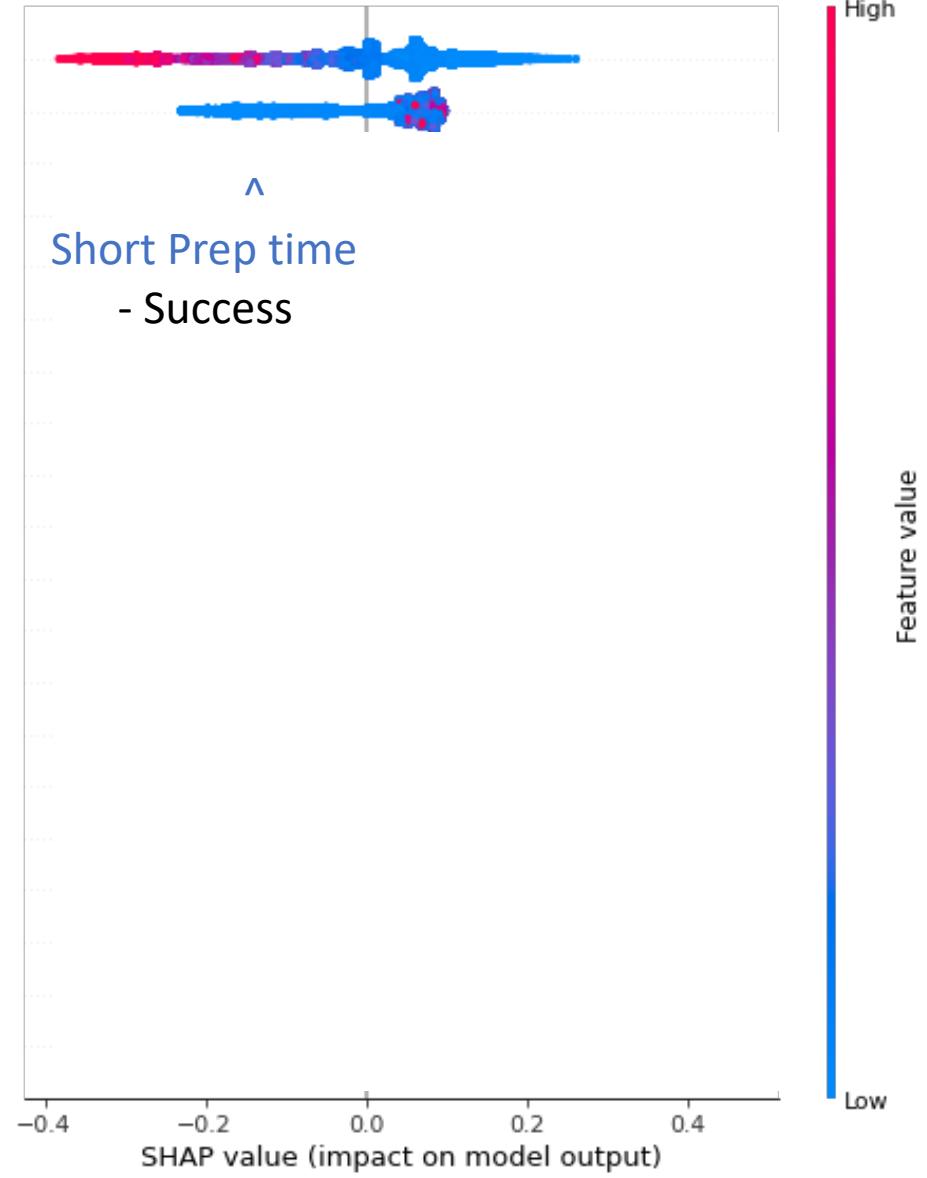


RESULTS

Feature Importance

- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters

goal_usd
prep_duration
staff_pick_True
cam_duration
Food
Technology
Comics
Crafts
Photography
Music
Journalism
Games
desc_length
country_US
Film_Video
Fashion
Publishing
Theater
Dance
Design

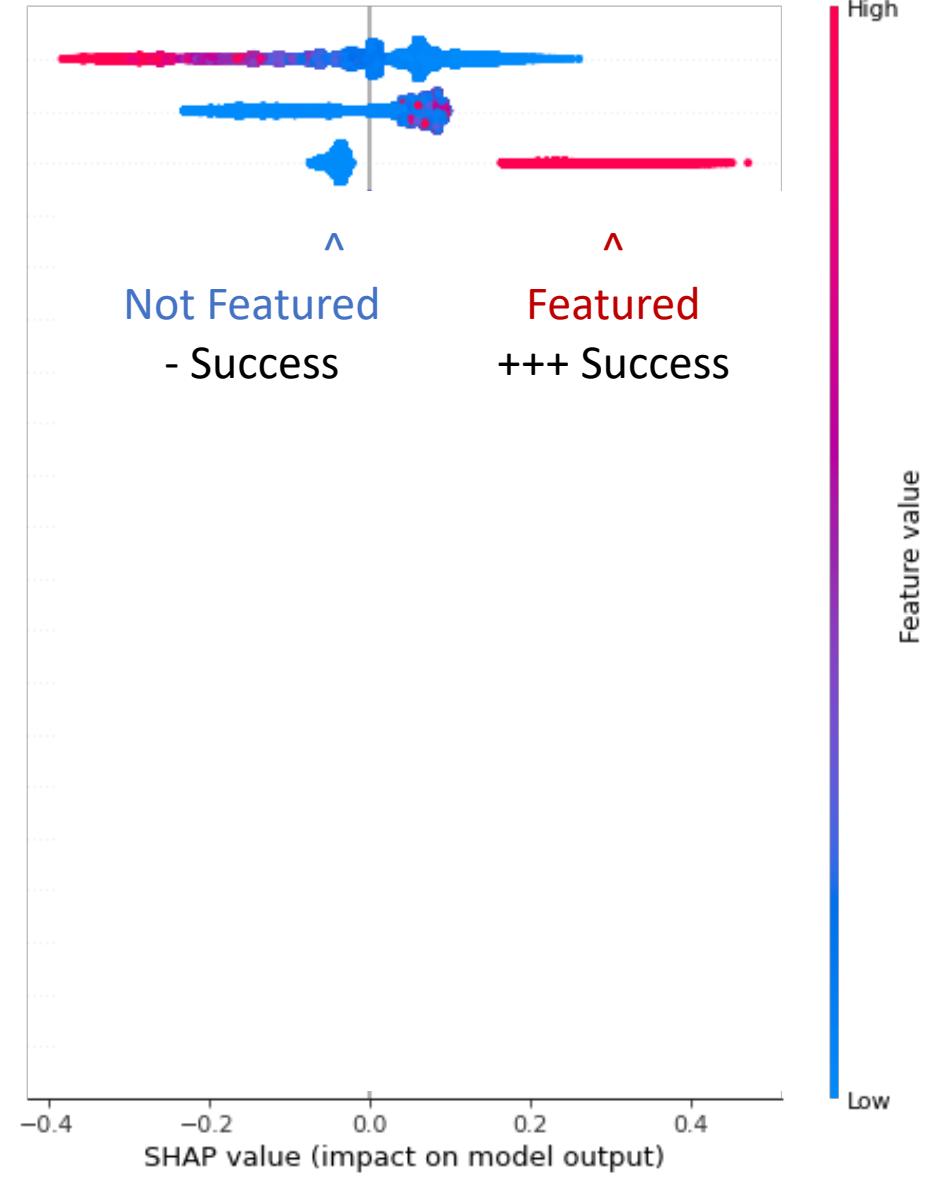


RESULTS

Feature Importance

- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters

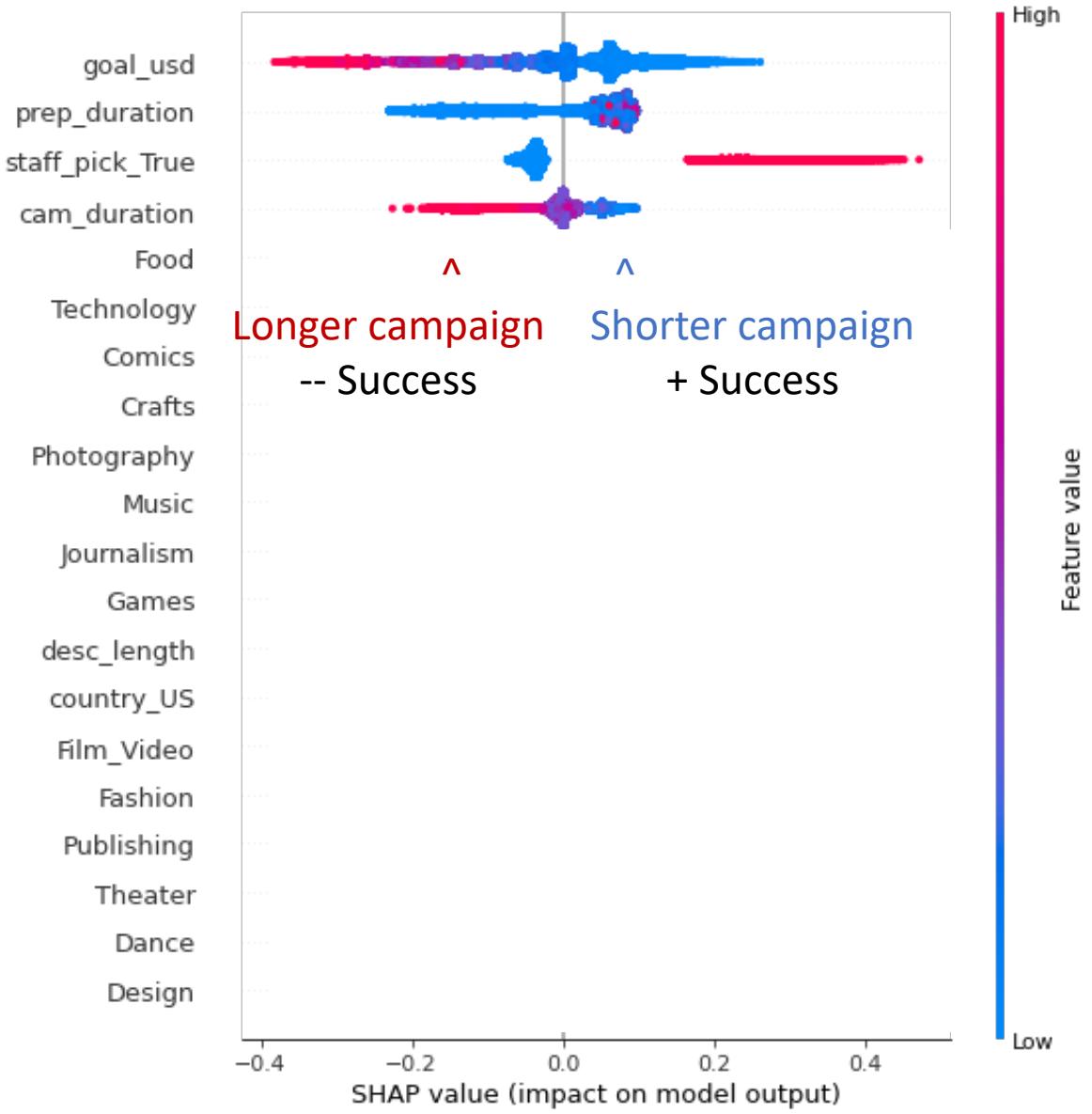
goal_usd
prep_duration
staff_pick_True
cam_duration
Food
Technology
Comics
Crafts
Photography
Music
Journalism
Games
desc_length
country_US
Film_Video
Fashion
Publishing
Theater
Dance
Design



RESULTS

Feature Importance

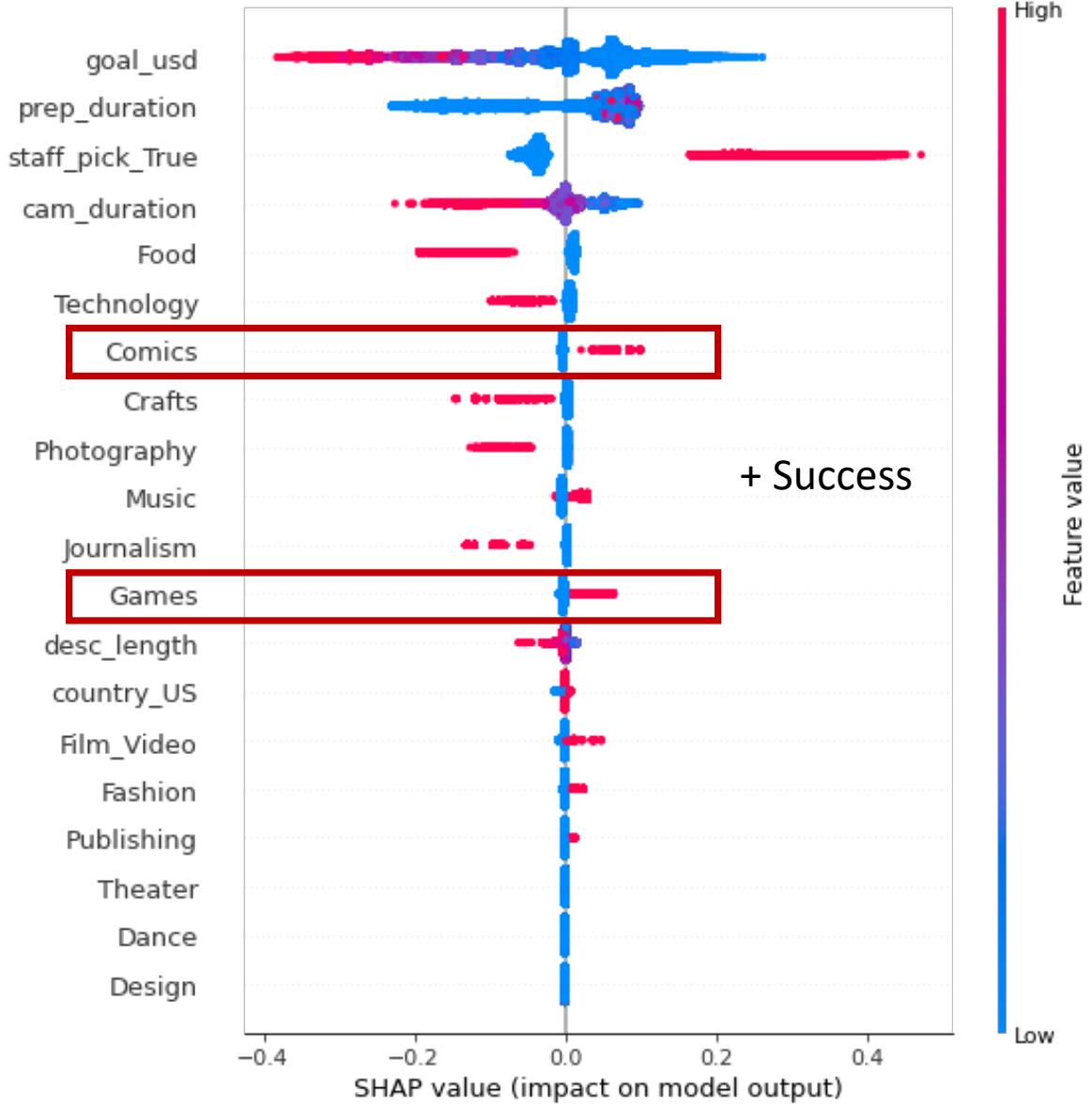
- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters



RESULTS

Feature Importance

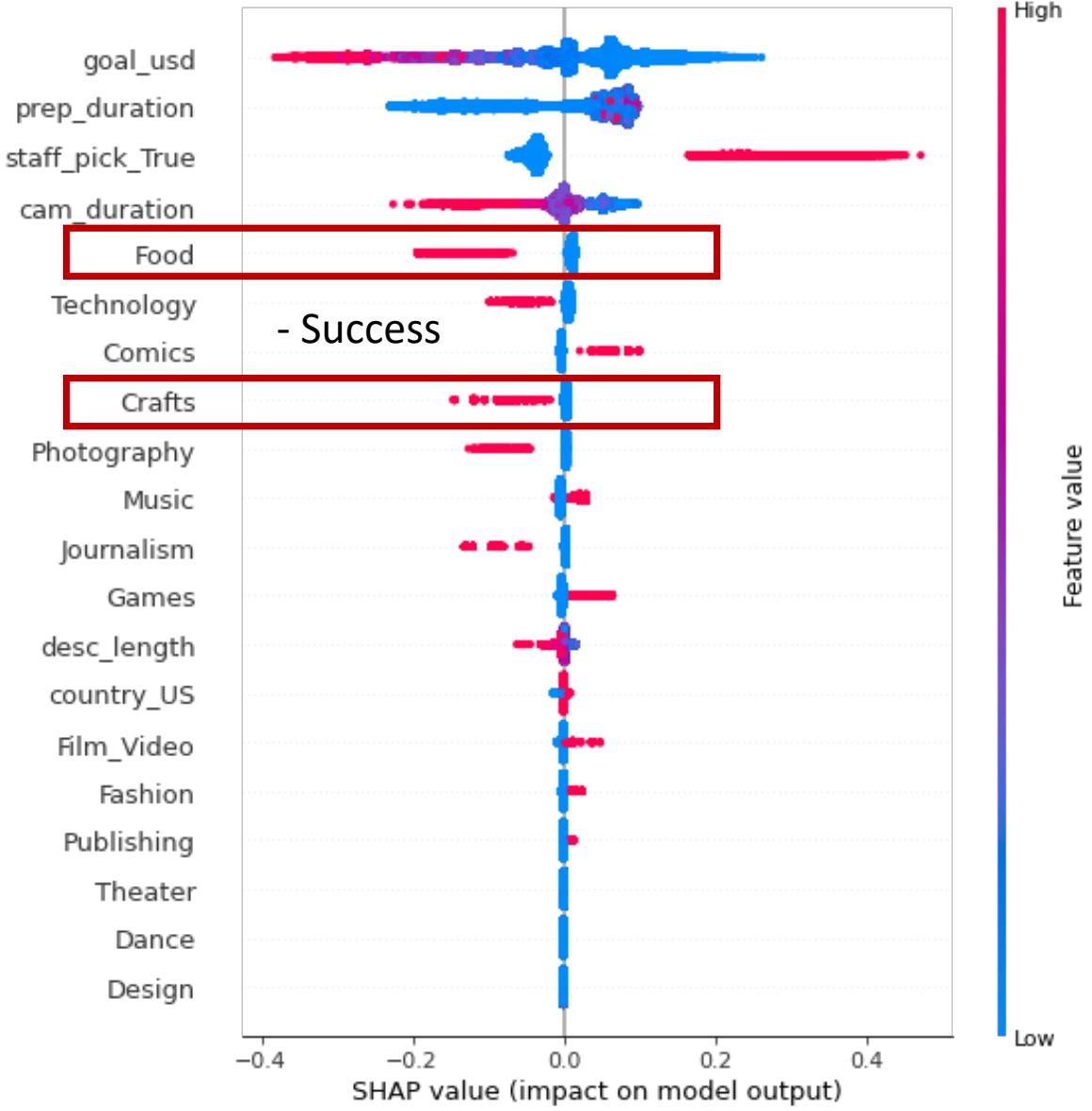
- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters



RESULTS

Feature Importance

- Low goal can benefit the campaign
- Shorter prep time don't do well
- Featured on Kickstarter improves chance of success
- Shorter campaigns do better
- Choice of category matters



APPLICATION

Use Case 1

Kickstarter Prediction

127.0.0.1:5000

Kickstarter Success Prediction

Let's bring your project to life!

Tell me about your campaign:

What is your goal (in USD?)

Is it based in the US? Yes

What is your campaign duration (in days)?

What is your preparation duration (in days)?

What is your campaign description length in words(word count)?

Is this a staff pick? Yes

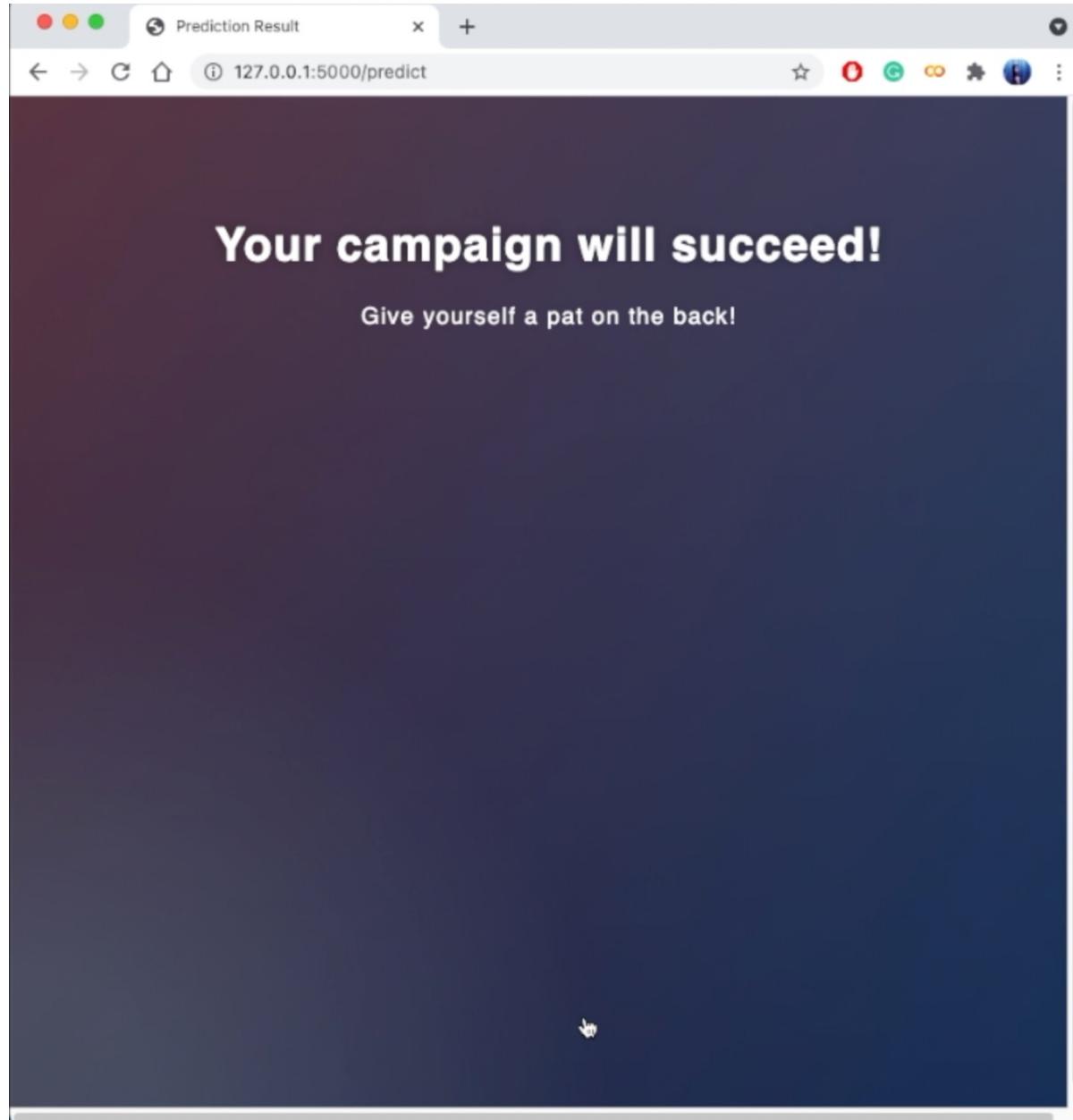
Choose your category: Comics

Predict



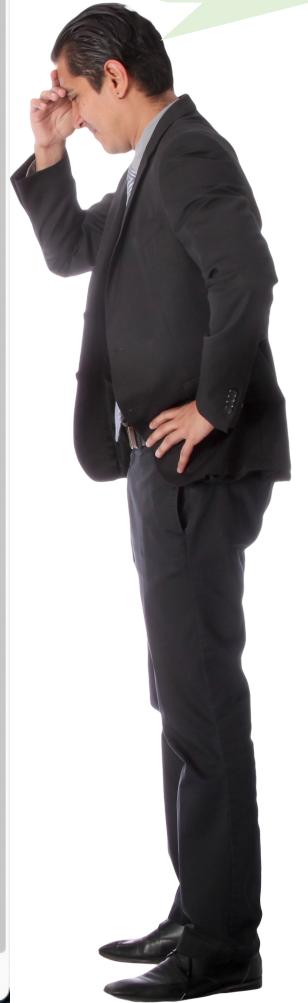
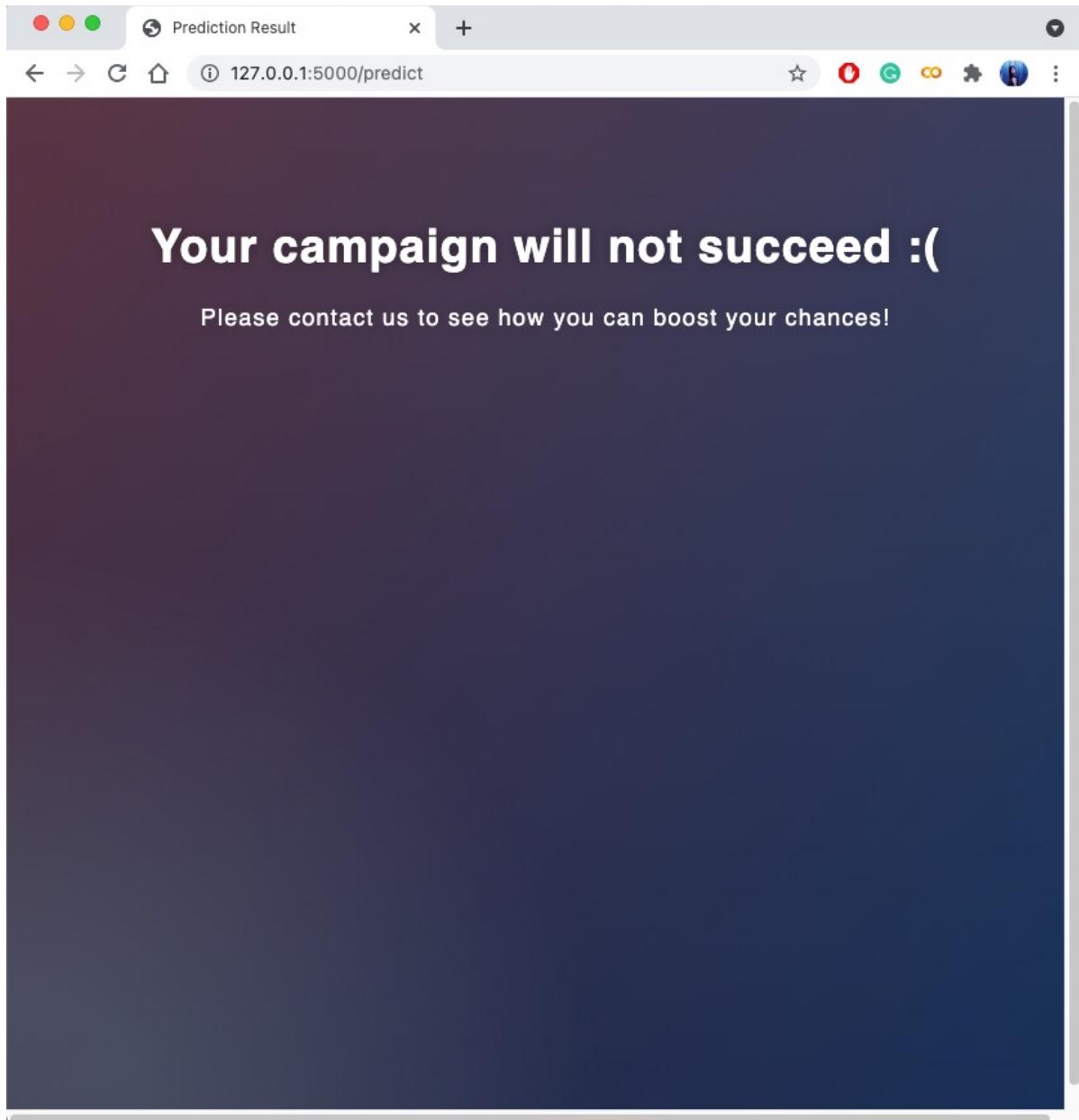
APPLICATION

Use Case 1



APPLICATION

Use Case 2



APPLICATION

Use Case 2

Let's Boost Your Chances!

- Set small goal** that fits the scope of the campaign
- Increase campaign prep time**
- Don't stretch** the campaign duration **too long**
- Get featured** by Kickstarter if possible
- Choose category** carefully



FUTURE WORK

1

Build more robust model

More data, features (Amount of pledge giveaways, etc.)

2

Deploy Flask app on Heroku

Full functionality and visual appealing

3

What will the future bring?

Explore how quickly campaigns are fulfilled



THANK YOU

Questions?

A photograph of a person's hands holding a glowing incandescent lightbulb. The bulb is illuminated, casting a warm glow. The background is blurred with a gradient from teal at the top to orange at the bottom. A large, semi-transparent dark grey rectangular box covers the center of the image, containing the text.

CRYSTAL-CTRL



CRYSTALHUANG-DS

Reference

- Interpretable Machine Learning with XGBoost

<https://towardsdatascience.com/interpretable-machine-learning-with-xgboost-9ec80d148d27>

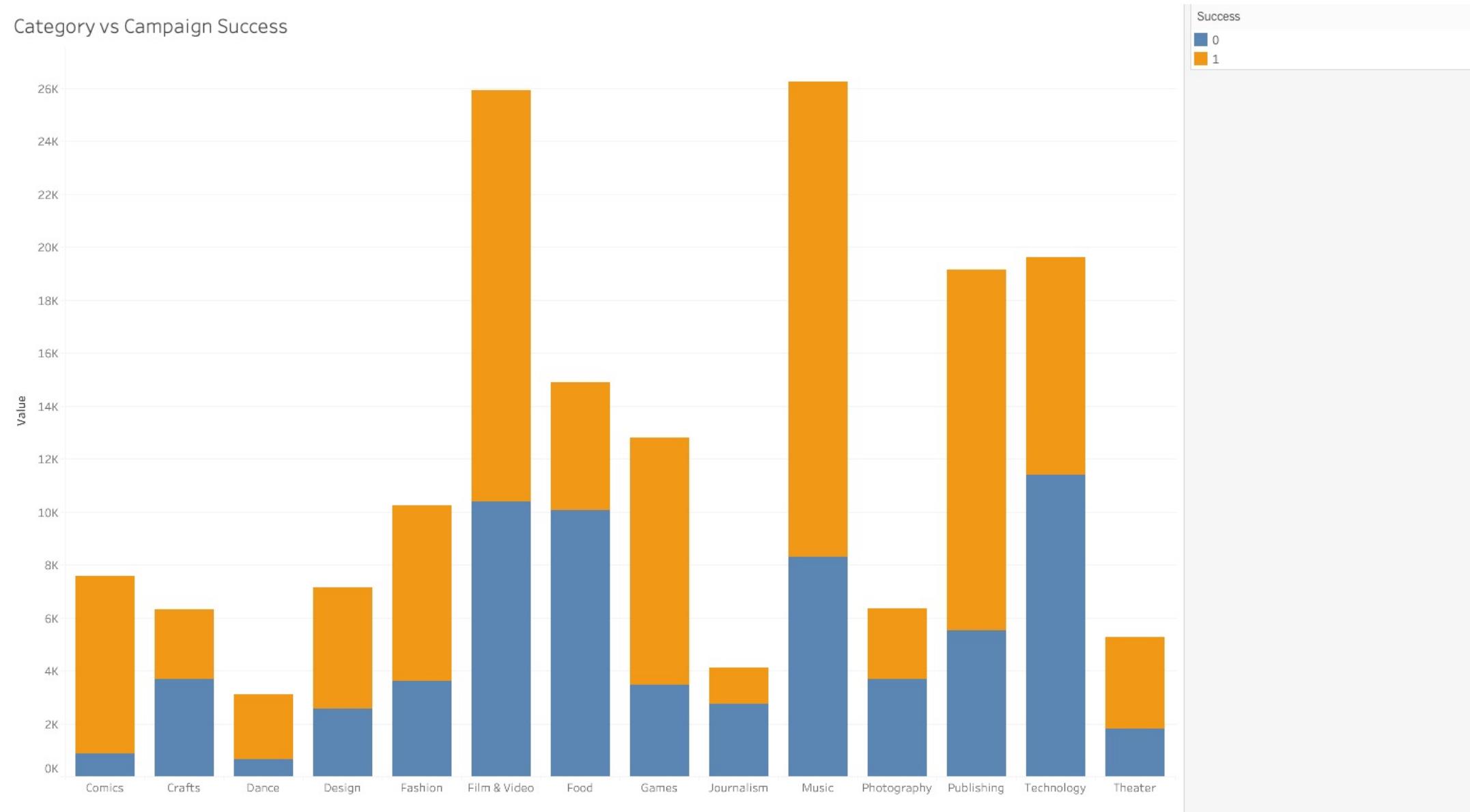
Appendix 1

Classification Project Workflow



Appendix 2

Category vs Campaign Success



Appendix 3

Baseline Model

Baseline Model

Logistic Regression
with Regularization

- F1 : 0.75
- AUC: 0.65

Add more
Features



Feature
engineering

Model 2.0

Logistic Regression
with Regularization

- F1 : 0.77
- AUC: 0.77

Appendix 4

Model Comparison

