



Predicting Success in Kickstarter Campaigns

Kelli Peluso
Capstone 1
Springboard Data Science Career Track



Table of Contents

- Background
 - The Data
 - The Problem
 - The Approach
 - The Goal
- Data Wrangling
 - Duplicate Values
 - Feature Manipulation
- Data Storytelling
 - Campaign Names
 - Campaign Categories
 - Campaign Goals
 - Campaign Duration
 - Campaign Backers
- Modeling & Statistics
 - Correlations
 - Random Forest Classifier
 - Logistic Regression
- Conclusion
 - Recommendations

Background



Kickstarter Campaigns

The Data: obtained from [Kaggle](#), this dataset includes 378,661 rows of data. Each Kickstarter campaign is represented by one row of data spanning 15 features.

The Problem: The success of Kickstarter campaigns benefits both the companies and creators who launch campaigns, and Kickstarter itself.

The Approach: An analysis of both failed and successful Kickstarter campaigns will address potential features that are predictive of the campaign outcome of future campaigns.

The Goal: Provide Kickstarter with the ability to make data-driven, impactful decisions in regards to: the services they offer, fees that they charge, and future campaign guidelines and recommendations

Data Wrangling



Data Wrangling

- Duplicate Values: The data obtained was fairly clean, and did not require the removal of any duplicate campaigns.
- Null Values: Two features had null values - 'name' (4) and 'usd_pledged' (3797). The 4 null values from 'name' were dropped, and the entire 'usd_pledged' feature was dropped, as there is an additional feature (usd_pledged_real) that is complete and more reliable.
- Feature Manipulation:
 - `campaign_state > binary_state`
 - Removal of 'live', 'undefined', 'suspended',
 - Creation of binary feature for campaign outcomes (target classifier)
 - `name_len, name_cl, main_category_encode, goal_binned, pledge_binned, backers_binned, campaign_duration, camp_days`

Data Storytelling

Campaign Name



All Campaigns

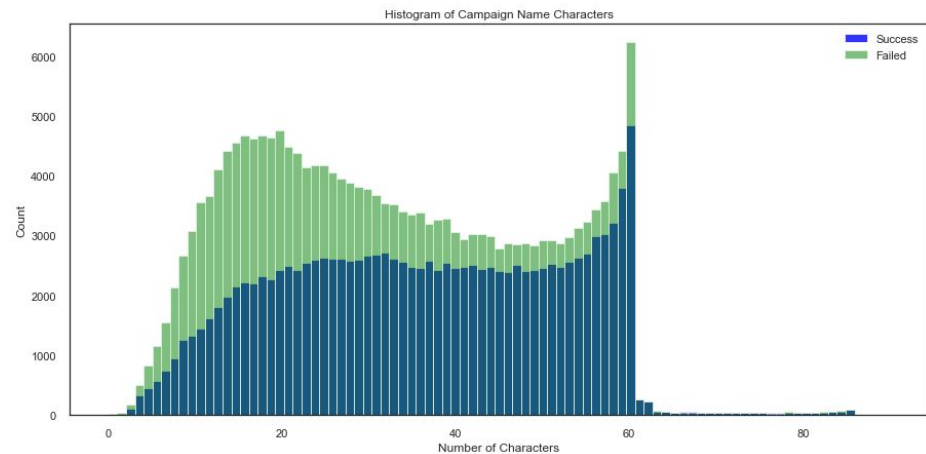
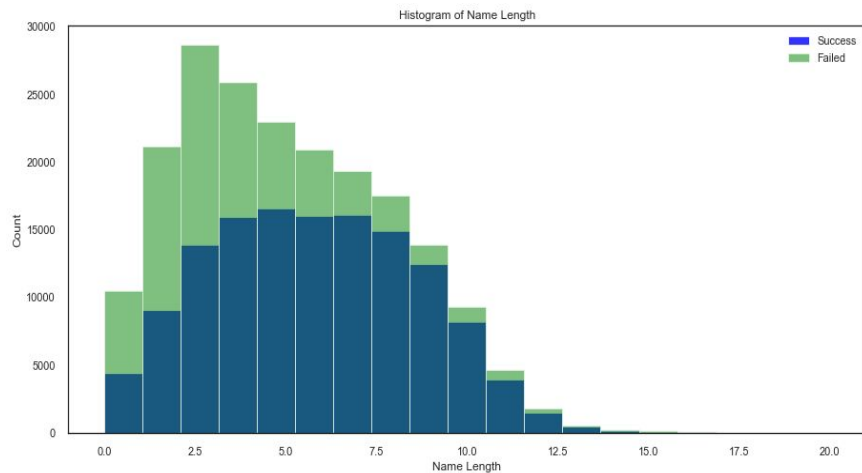


Successful Campaigns

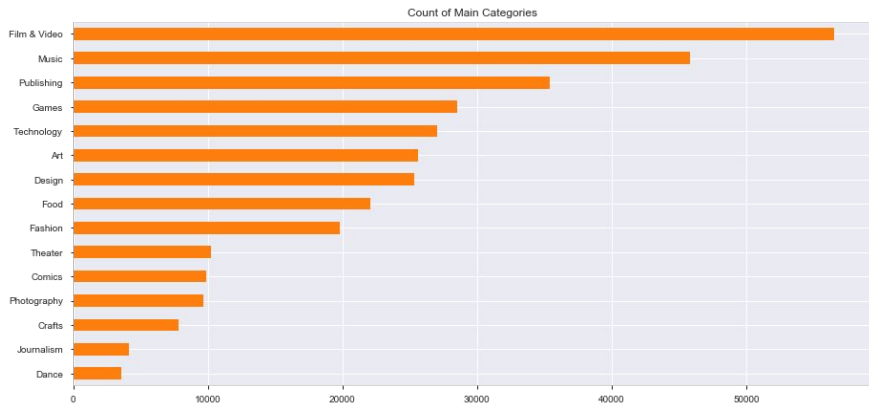


Failed Campaigns

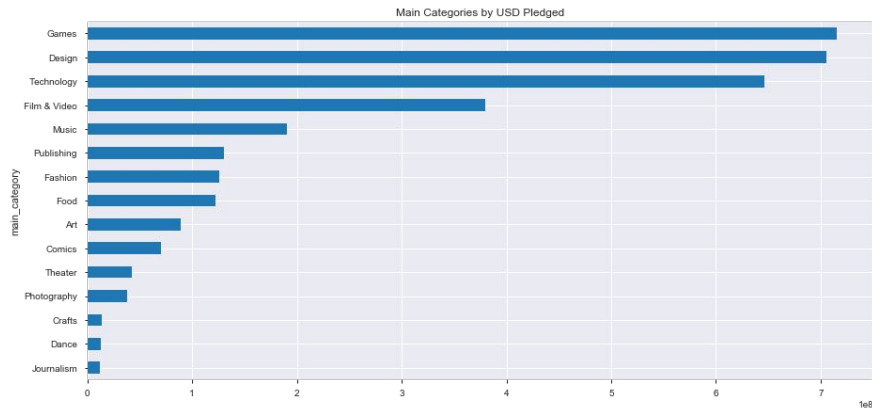
Campaign Names



Main Categories

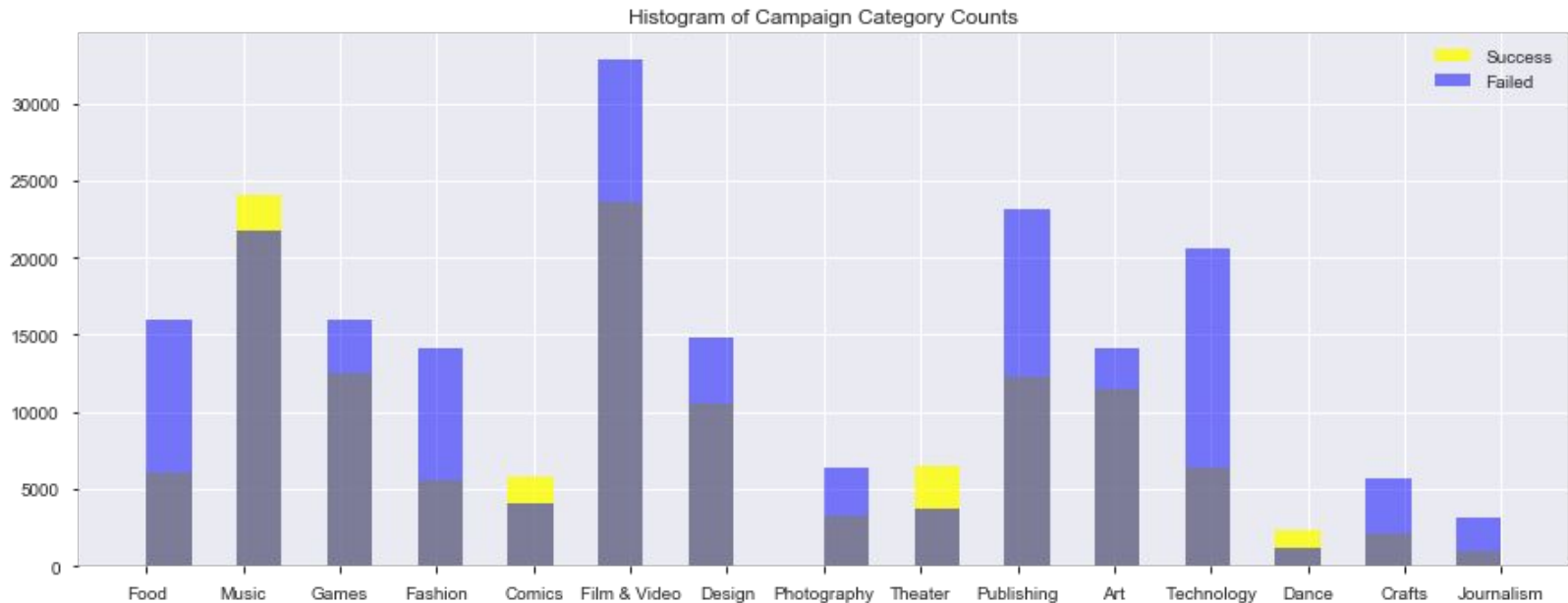


Main Categories by Count

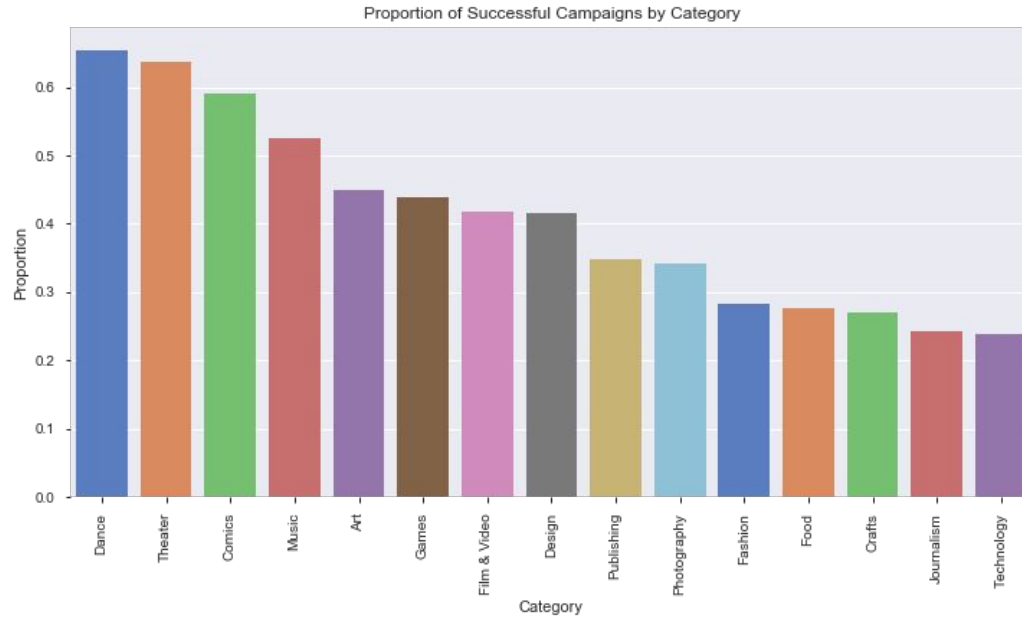


Main Categories by USD Pledged

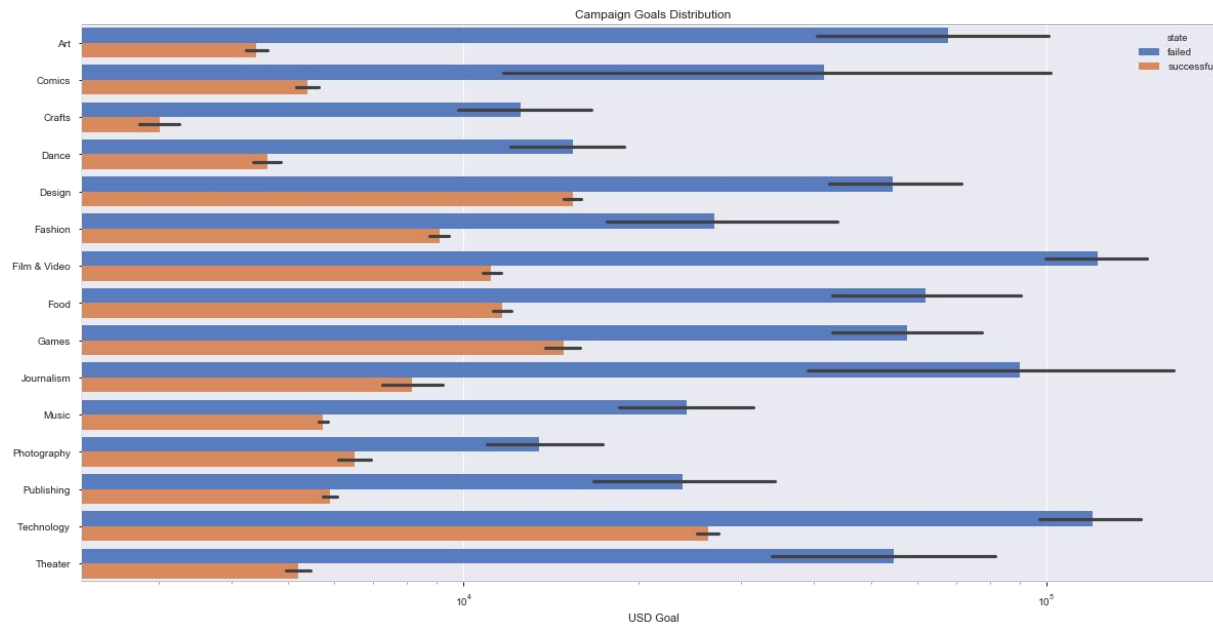
Main Categories



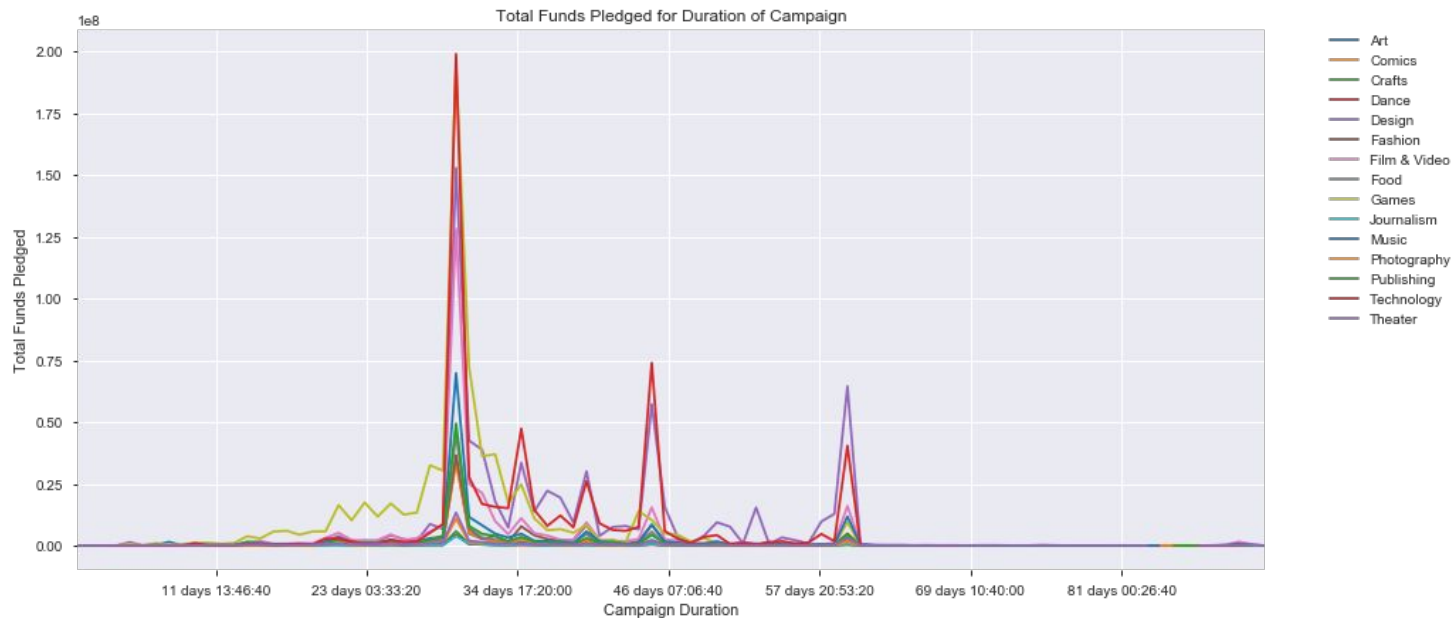
Main Categories



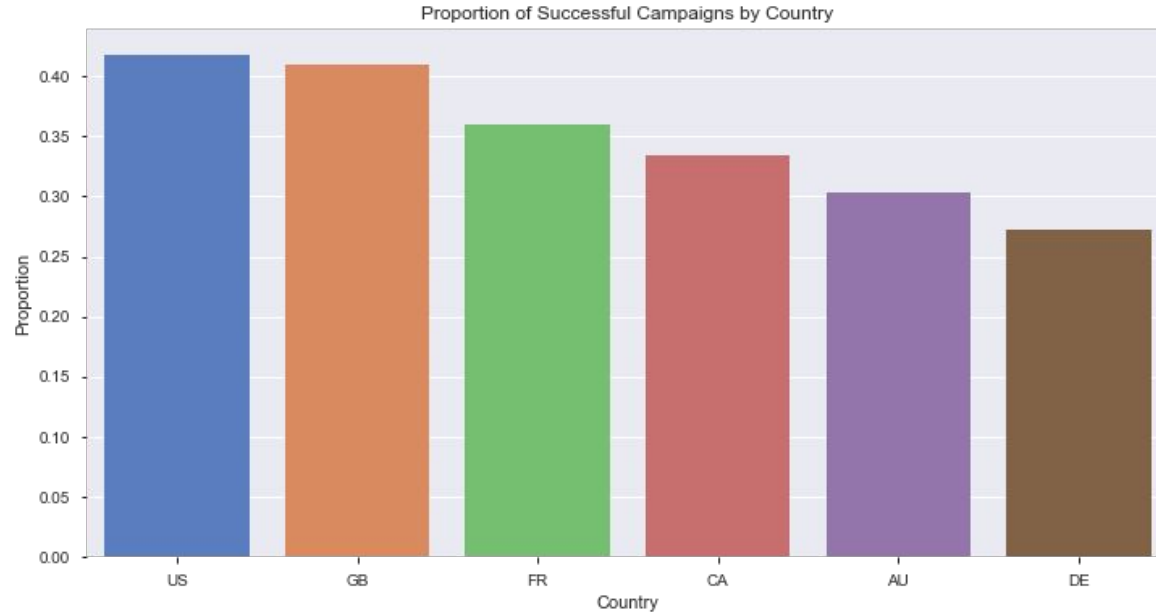
Campaign Goals



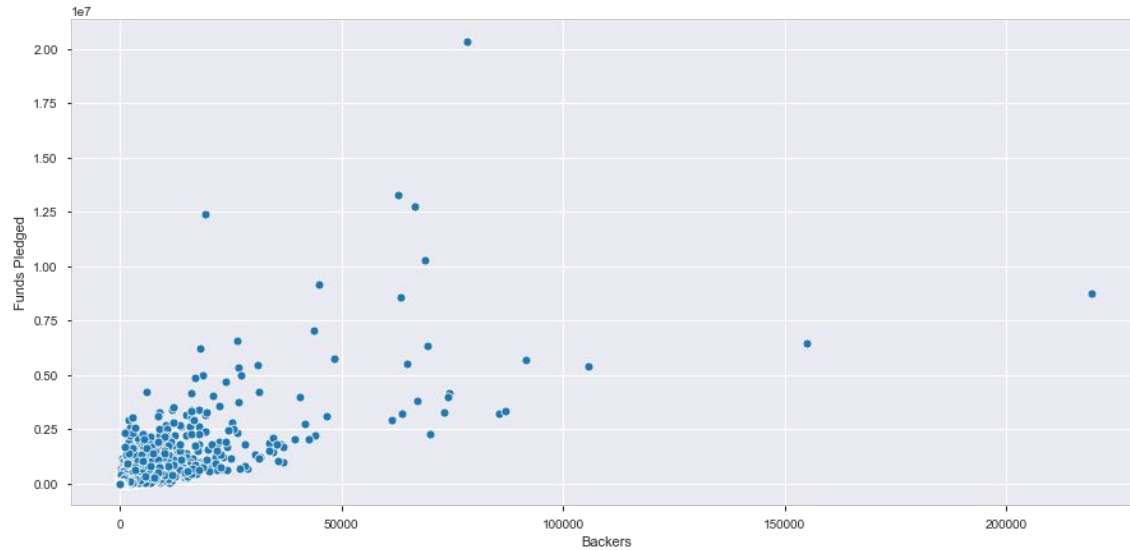
Campaign Duration



Campaign Countries of Origin



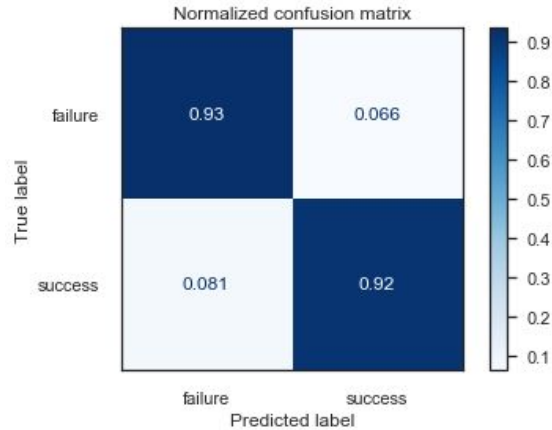
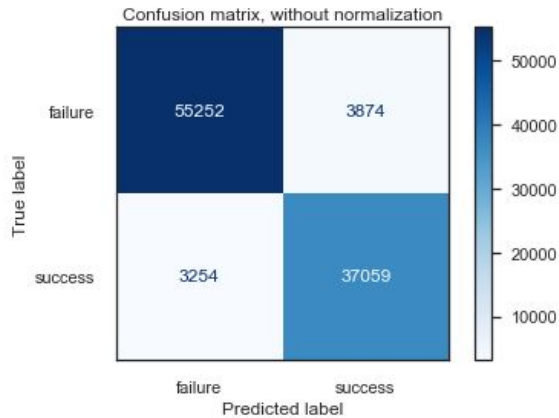
Campaign Backers



Relationship between Funds Pledged and Campaign Backers

Modeling & Statistics

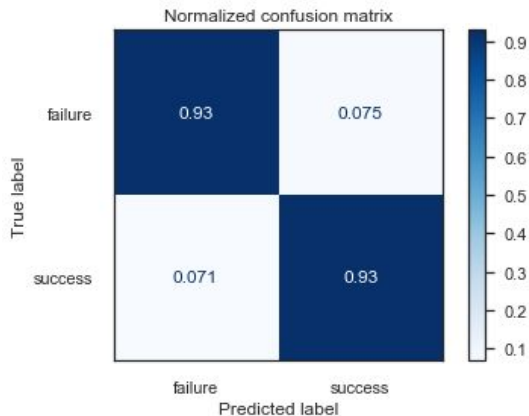
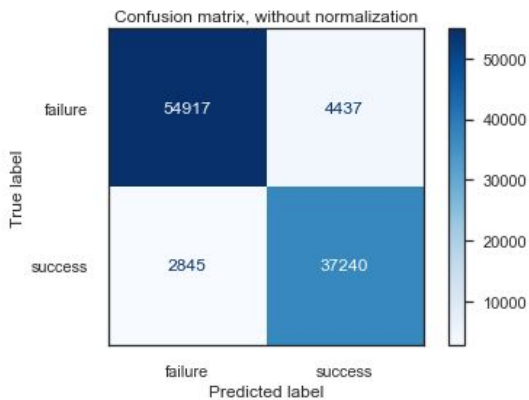
Random Forest Classifier



Metric	Result
Accuracy Score	92.87%
Precision Score	90.38%
Recall Score	92.11%
F1 Score	91.24%
ROC_AUC	92.75%

Default Parameters

Random Forest Classifier

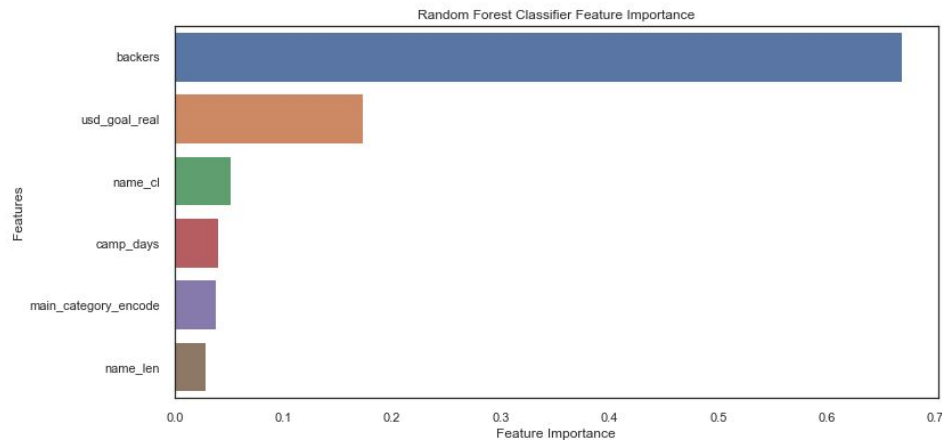


Metric	Result
Accuracy Score	92.68%
Precision Score	89.35%
Recall Score	92.90%
F1 Score	91.09%
ROC_AUC	92.71%

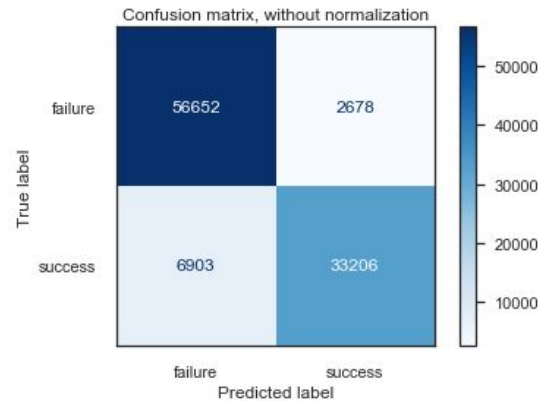
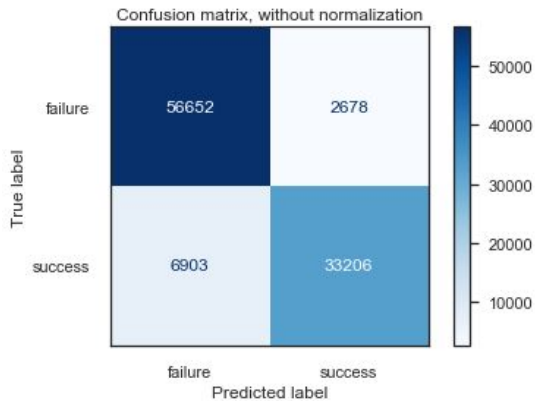
Optimized Parameters

Random Forest Classifier

Metric	Default	Optimized	+/-
Accuracy Score	92.87%	92.68%	-0.21%
Precision Score	90.38%	89.35%	-1.14%
Recall Score	92.11%	92.90%	0.86%
F1 Score	91.24%	91.09%	-0.16%
ROC_AUC	92.75%	92.71%	-0.04%

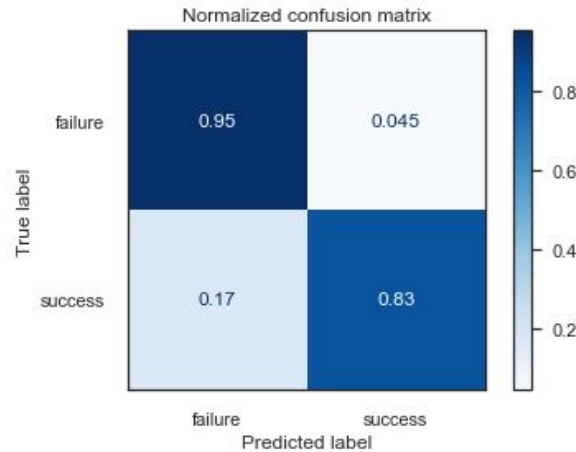
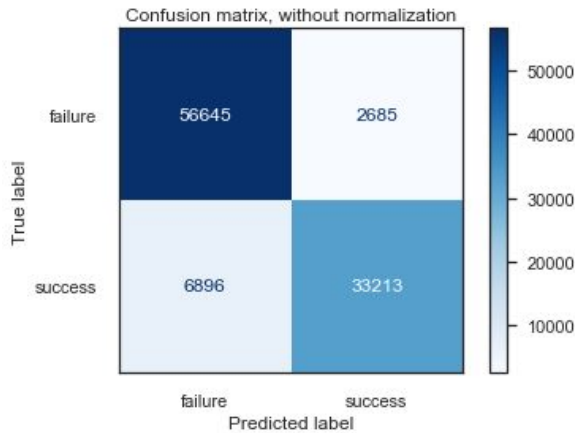


Logistic Regression



Metric	Result
Accuracy Score	90.36%
Precision Score	92.54%
Recall Score	82.79%
F1 Score	38.26%
ROC_AUC	89.14%

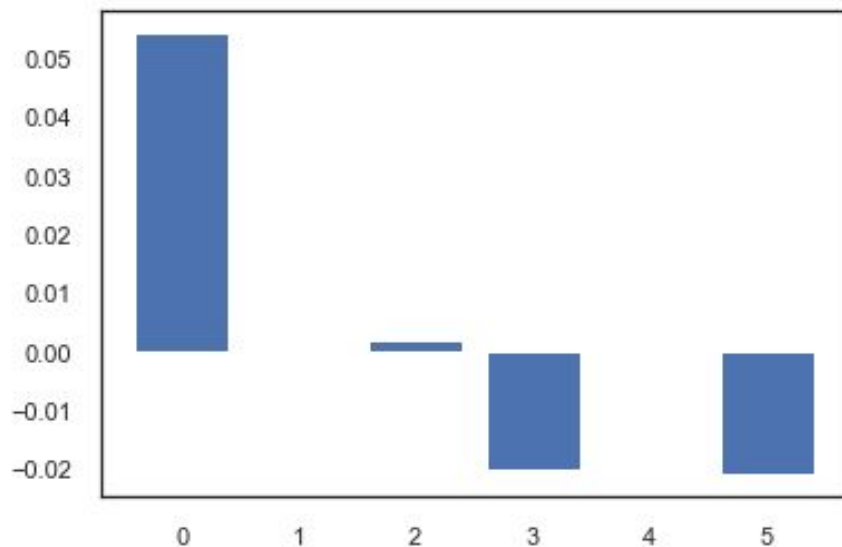
Logistic Regression



Metric	Result
Accuracy Score	90.36%
Precision Score	92.52%
Recall Score	82.81%
F1 Score	38.26%
ROC_AUC	89.14%

Logistic Regression

Feature Importance



Feature	Importance
0 - backers	0.05413
1 - name_cl	-0.0003
2 - name_len	0.00182
3 - camp_days	0.00182
4 - usd_goal_real	-0.00023
5 - main_category_encode	-0.02089

Conclusions



Thoughts and Recommendations

After examining the data, and fine-tuning the random forest classifier, I believe that I can make several recommendations to Kickstarter. These recommendations should assist in their development of campaign services, campaign recommendations and guidelines, and their fee structure.

1. Campaigns with smaller, less ambitious goals tend to be more successful
2. The number of campaign backers will help determine whether or not a campaign will succeed or fail
3. Main categories of campaigns do matter, with some seeing much higher rates of success (Music, Comics, Theater, Dance) than others (Film & Video, Food, Technology, Publishing)