

Manjeel Baral
Data Science
Fellow

Kickstarter[®] Project

funded
with

**KICK
STARTER**

Data Science Workflow



Define
Problem



Gather Data



Explore
Data



Build the
model



Evaluate the
model



Answer the
Problem



Define Problem

- The primary goal of my capstone project is to predict whether the project is successful given the explanatory variables.

I obtained
Kickstarter
project data from
Kaggle with 12
features

Gather Data

SHAPE OF DATA:
12 COLUMNS
WITH 323729
OBSERVATIONS

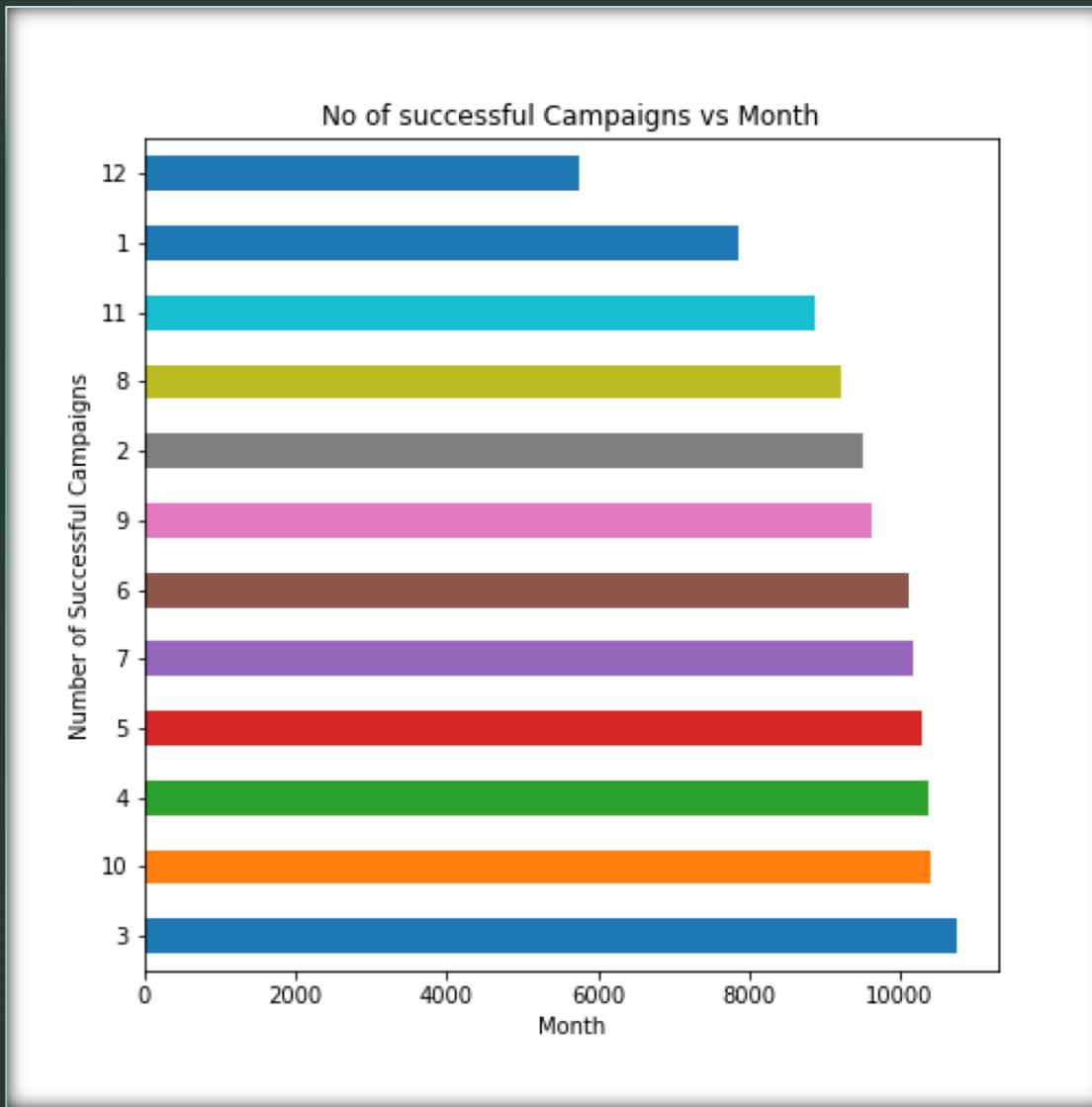
Contd..

MISSING
VALUES: 3790

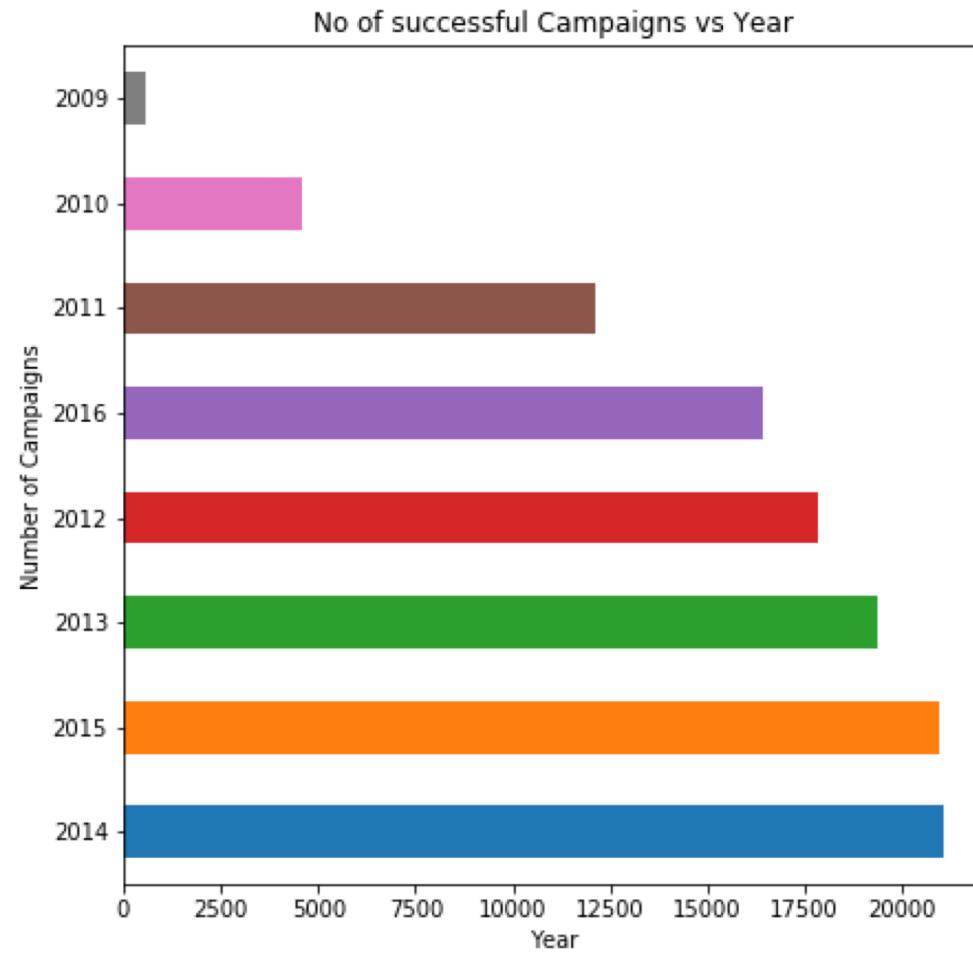
Explore Data

variables	Description
ID	Internal Kickstarter ID
name	name of project
category	category
main_category	category of campaign
currency	currency used to support
deadline	deadline for crowdfunding
goal	Goal for project to be successful
launched	date launched
pledged	amount pledged by crowd
state	Current condition the project is in
backers	number of backers
country	country pledged from
usd pledged	amount of money pledged in USD

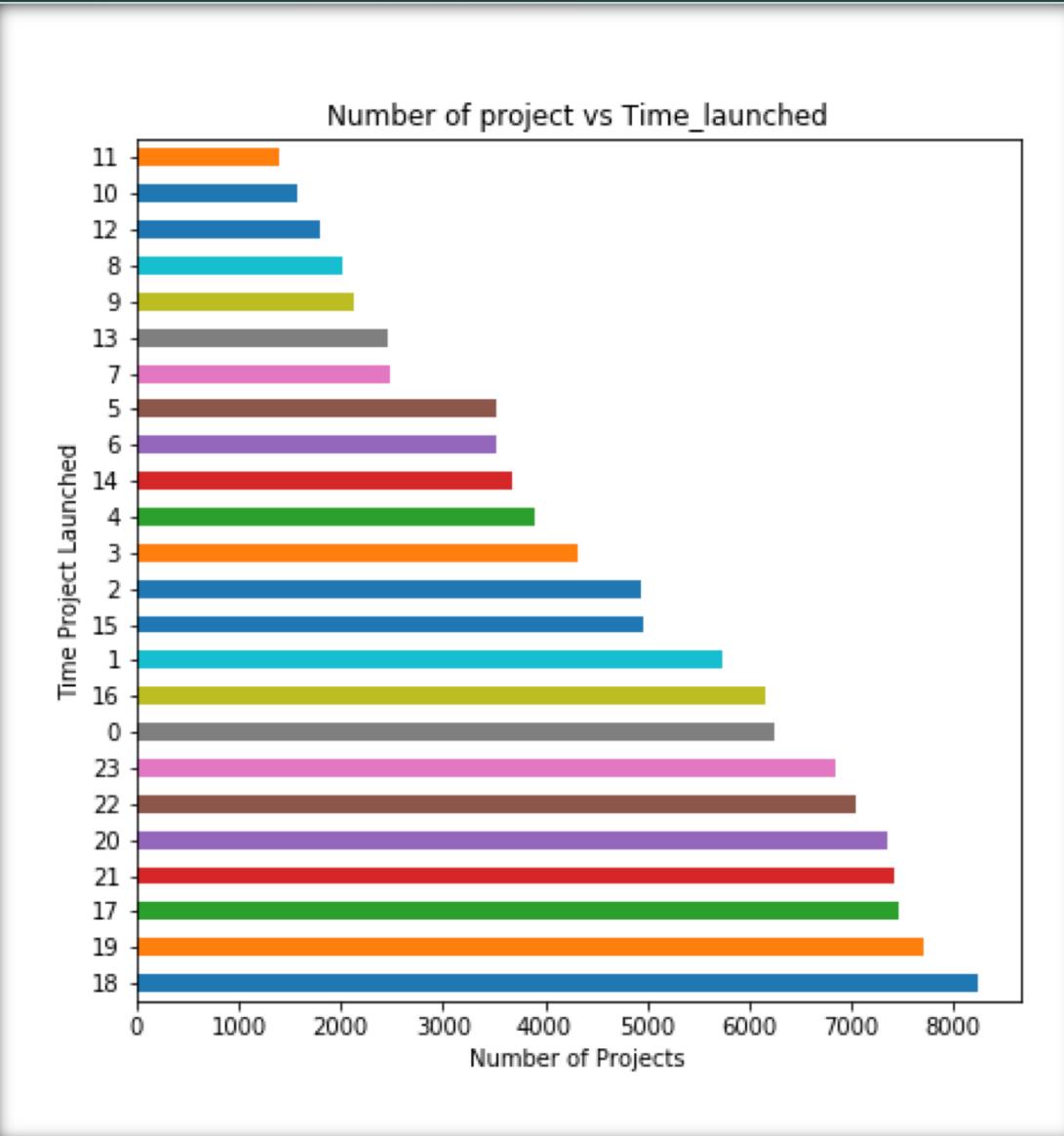
Contd..



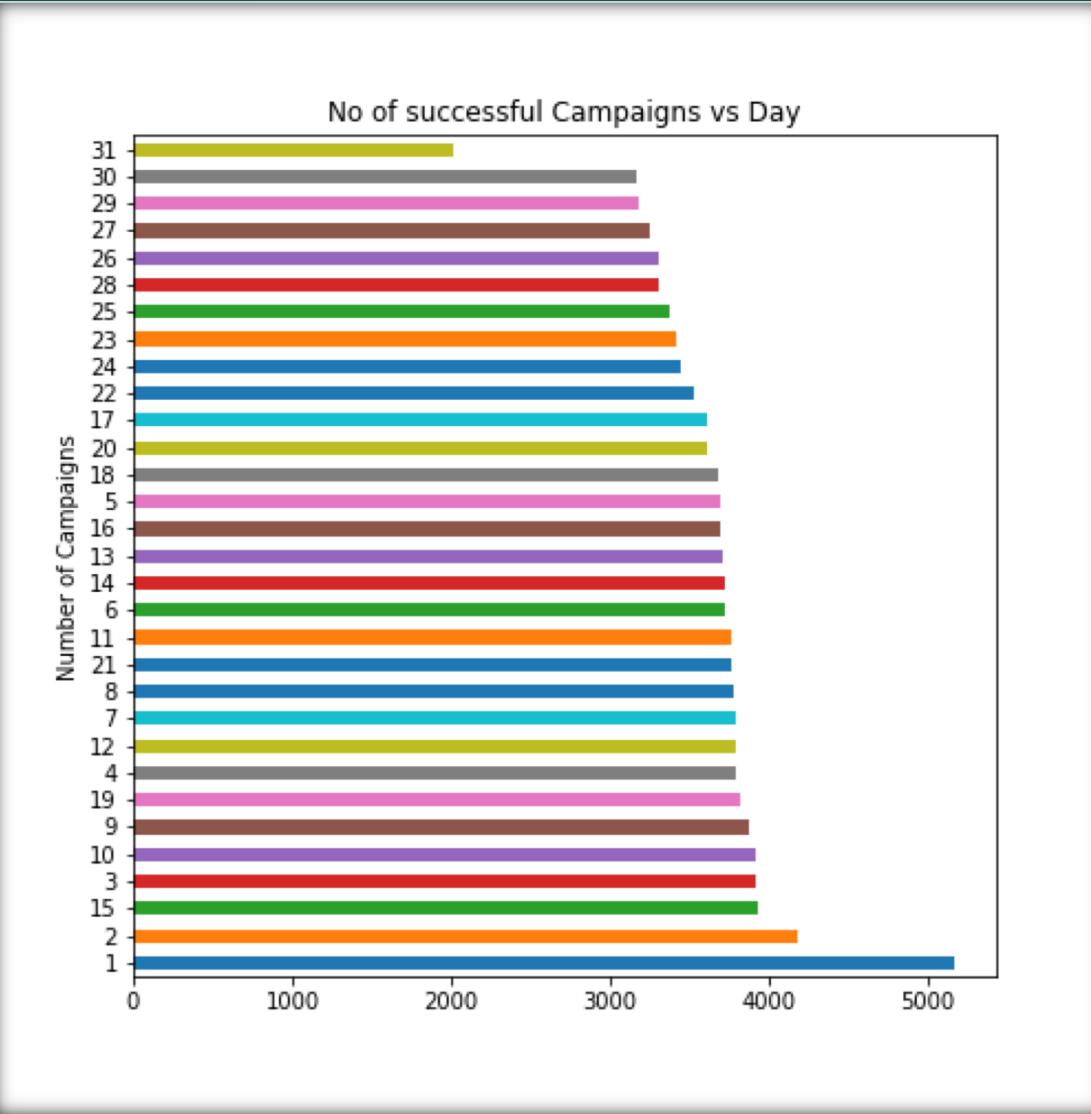
Contd..



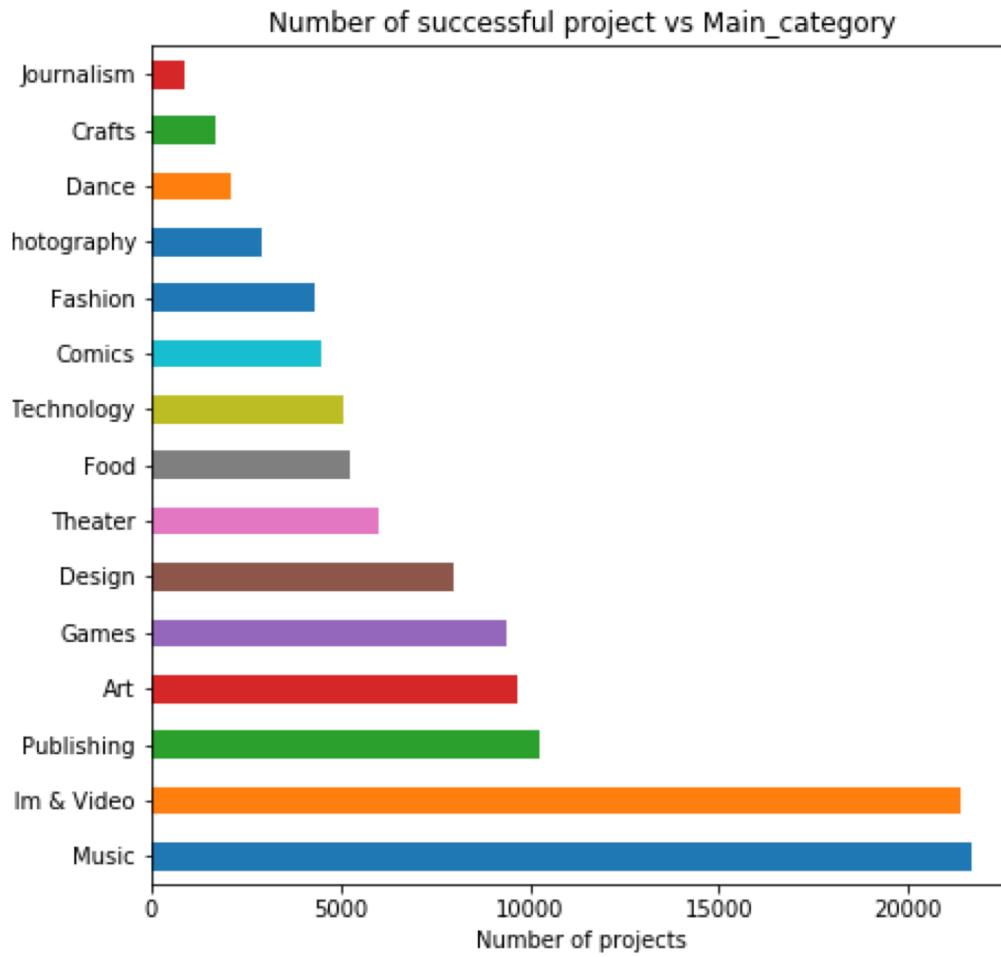
Contd...



Contd..

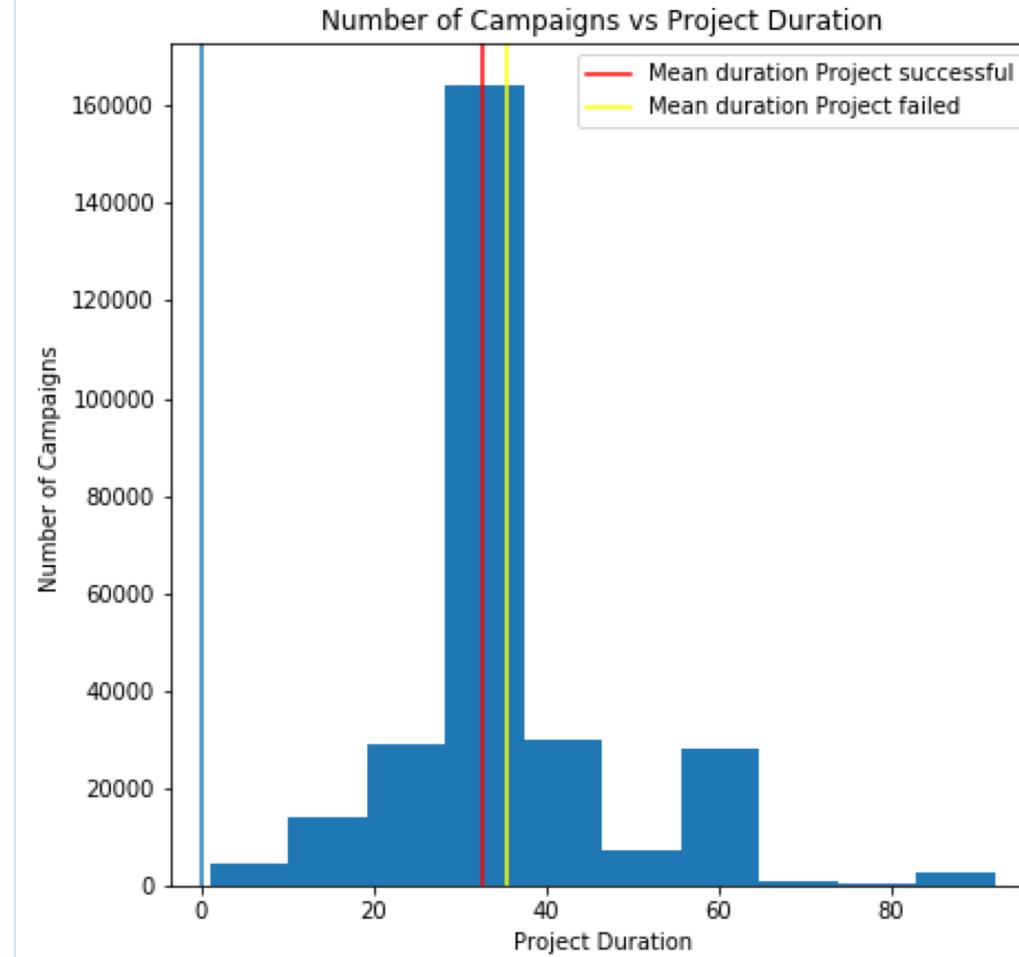


Contd..



Contd..

- Mean duration of failed campaigns 35 days
- Mean duration of successful campaigns 32 days



Build Model



SINCE THE GOAL OF
MY PROJECT IS
MODEL
PREDICTABILITY, I
HAVE LEVERAGED
THE CONCEPT OF
ENSEMBLE METHODS
AND USED GRADIENT
BOOSTING
CLASSIFIER TO FIT
THE DATA.



MODEL WAS FITTED
WITH THE TRAINING
SET AND EVALUATED
ON TESTING SET.



THE TRAIN
ACCURACY SCORE
WAS 0.93



THE TEST SCORE
ACCURACY WAS 0.93



CROSS VALIDATED
SCORE WAS 0.93
(MEAN)

Evaluate Model

True Negatives: 39194
False Positives: 2835
False Negatives: 2123
True Positives: 26121

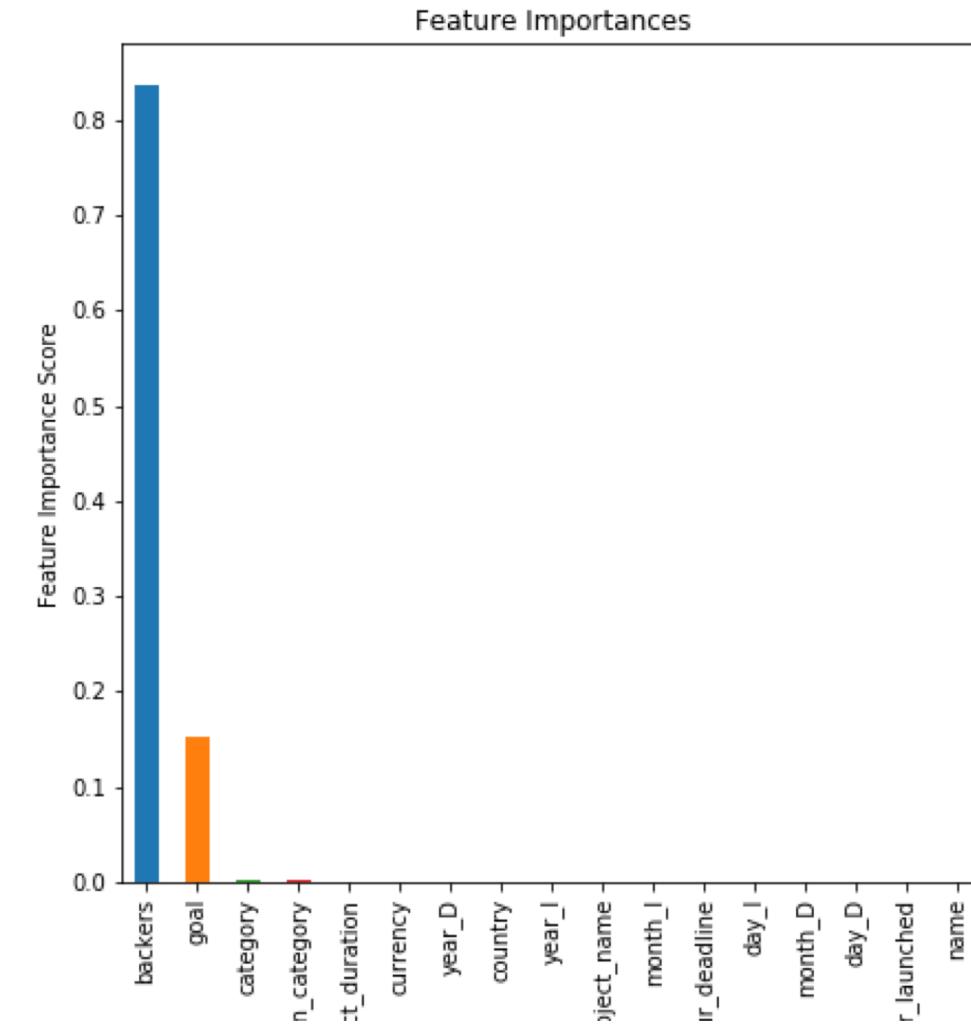
- True positive are project we correctly predict to be successful (True Positive).
- True negative are project we correctly predict to be failed (True negative)
- False positives are project we incorrectly predict to be successful (False positive).
- False negatives are people we incorrectly predict to be failed (False negative).

Evaluate contd..

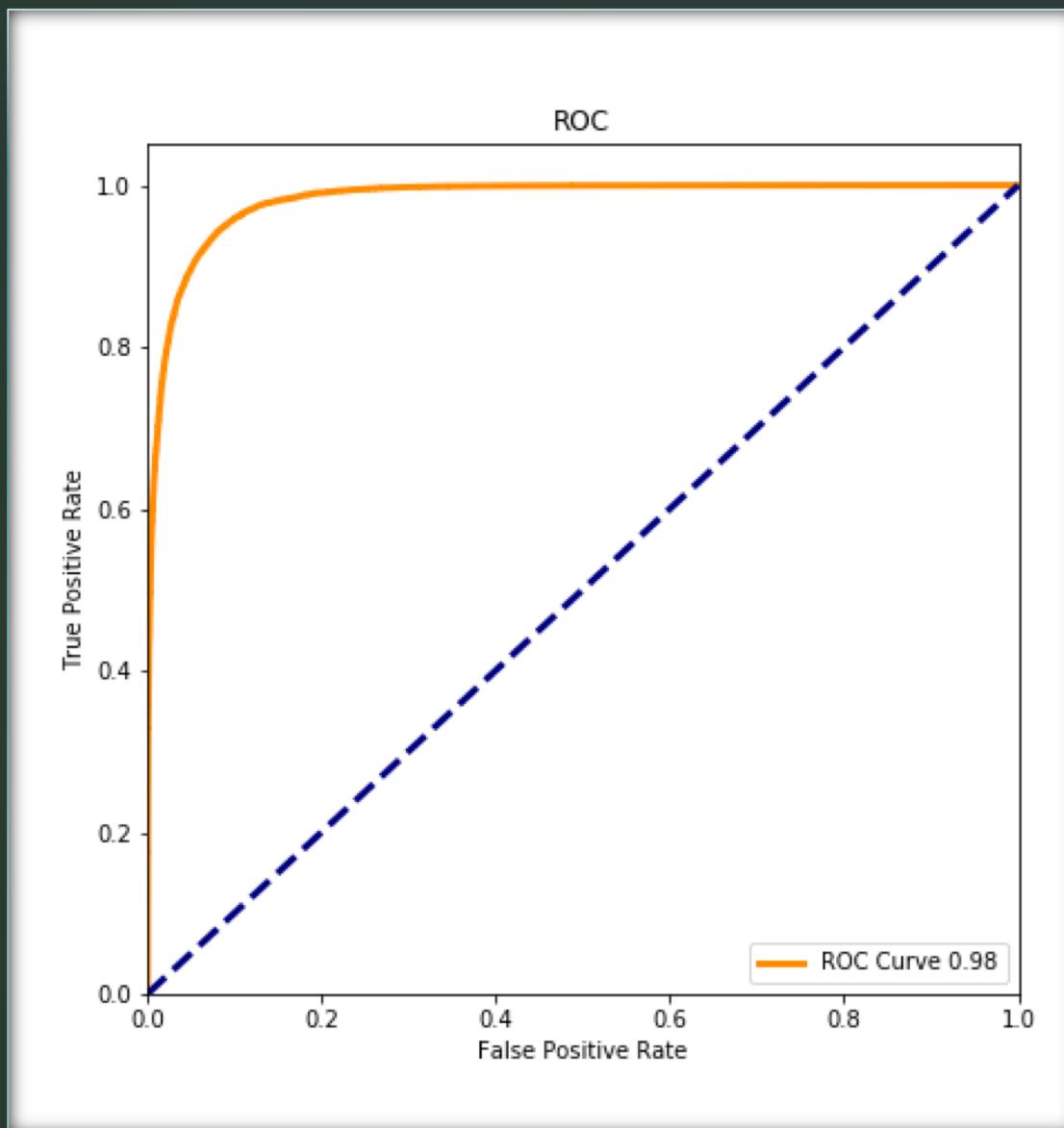
```
Accuracy: 0.9294465868825865  
F1 score: 0.9133216783216783  
Recall: 0.9248335929754993  
Precision: 0.9020928305014505
```

```
classification report:  
precision    recall   f1-score   support  
0            0.95     0.93     0.94     42029  
1            0.90     0.92     0.91     28244
```

Evaluate contd..



Evaluate contd..



Conclusion

- Algorithm has given the most important features as backers and it makes sense as Kickstarter has adopted All-or-nothing funding meaning that no one will be charged for a pledge towards a project unless it reaches its funding goal and to make project successful it needs backers to support and share about the project (word-of-mouth)

A large, stylized, three-dimensional green logo consisting of the letters 'Q' and '&' followed by 'A'. The letters are thick and have a slight shadow, giving them a 3D appearance. The logo is set against a white background.

This Photo by Unknown Author is licensed under [CC BY-NC](#)