Mapping Kickstarter's Success Formula

April 27th 2018

Overview

- Why learn a Winning Formula?
- Exploring Kickstarter
- How to measure a “Successful” Campaign
- Three Regression Models
- Which Features Matter

What is Kickstarter?

Purpose

- To allow small-scale creators to fund their work through donations from members of their community

Finances

- $3 Billion pledged
- 150K successful projects
- 14 Million Backers

Community

- Vast majority of campaigns raised less than 20K
- No money for analytics, advertising, videography, market research, etc.

The Goal:

Demystify the crowdfunding formula to help small scale creators and entrepreneurs.

Exploring Kickstarter:

By Emma Cohen & Miles Pepper
First created

FinalStraw, the world's first collapsible, reusable straw

Clean, compact, and totally badass: For anyone who wants to rid their lives—and the planet—of single-use plastic.

$616,954
pledged of $12,500 goal

13,183
backers

22
days to go

Back this project

Remind me

All or nothing. This project will only be funded if it reaches its goal by Sat, May 19 2018 12:12 AM PDT.

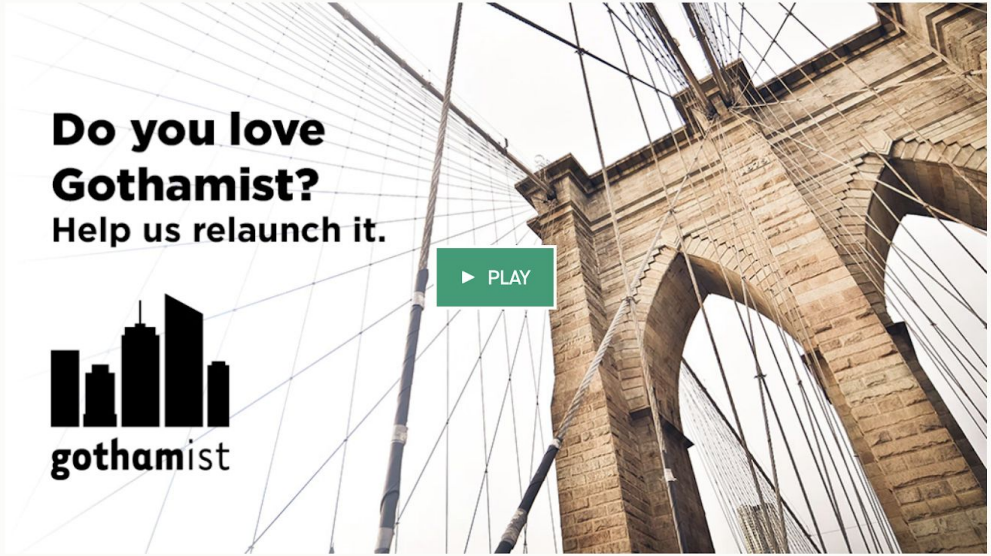
Project We Love · Product Design · Los Angeles, CA

Campaign · FAQ 32 · Updates 4 · Comments 93 · Community

By Gothamist
First created

Bring Back Gothamist

Gothamist has joined New York institution WNYC, now they need you to step up and bring the site back to its former glory.



Project We Love · Journalism · New York, NY

$142,367
pledged of $100,000 goal

2,054
backers

8
days to go

Back this project

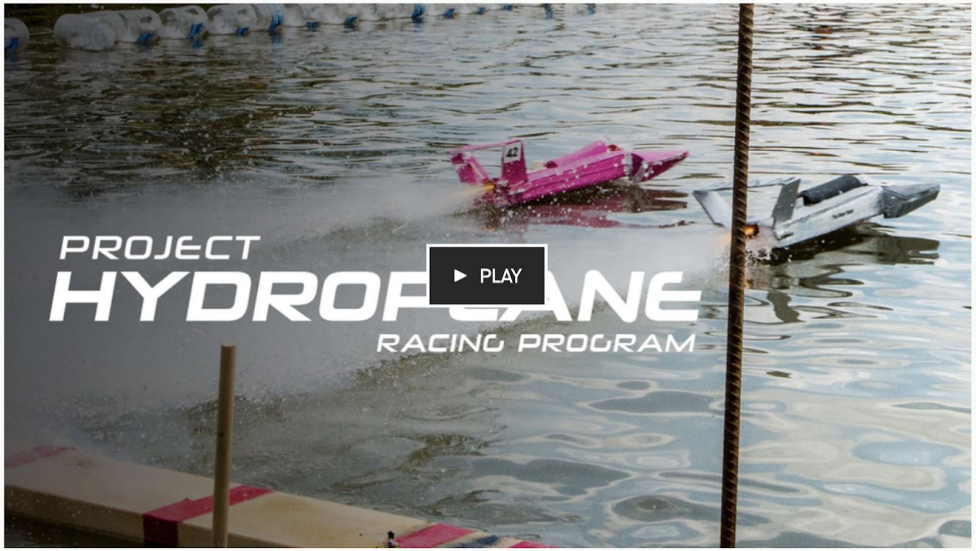
Remind me

All or nothing. This project will only be funded if it reaches its goal by Fri, May 4 2018 8:59 PM PDT.

By Thad Greer
First created

Project Hydroplane Racing Program - STEM Learning Made Fun

Project-based STEM curriculum that fully engages otherwise bored students and prepares them for the real world!



Project We Love Product Design Louisville, KY

$7,348
pledged of $55,000 goal

18
backers

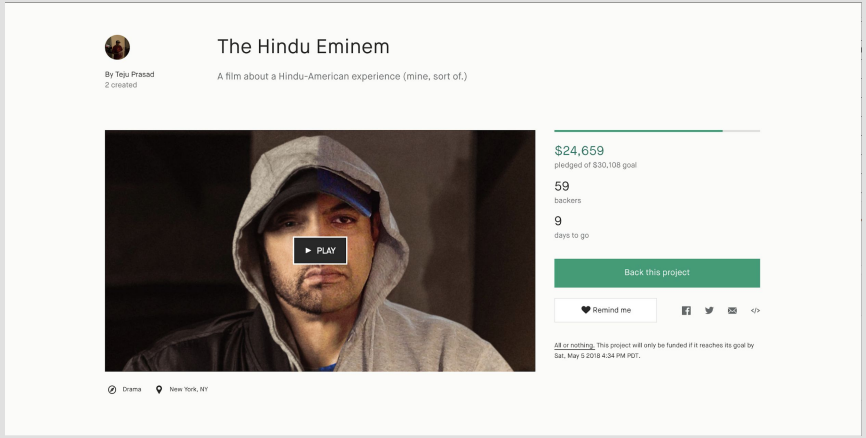
12
days to go

Back this project

Remind me

All or nothing. This project will only be funded if it reaches its goal by Wed, May 9 2018 7:30 AM PDT.

Campaign FAQ 6 Updates 0 Comments 0 Community



https://www.kickstarter.com/projects/1475731975/the-hindu-eminem?ref=discovery&ref=discovery

Pledge $1,000 or more

Pre-Release Screening (NY, NY)

Invitation to Pre-Release Screening (NY, NY)

ESTIMATED DELIVERY
Jul 2019

5 backers

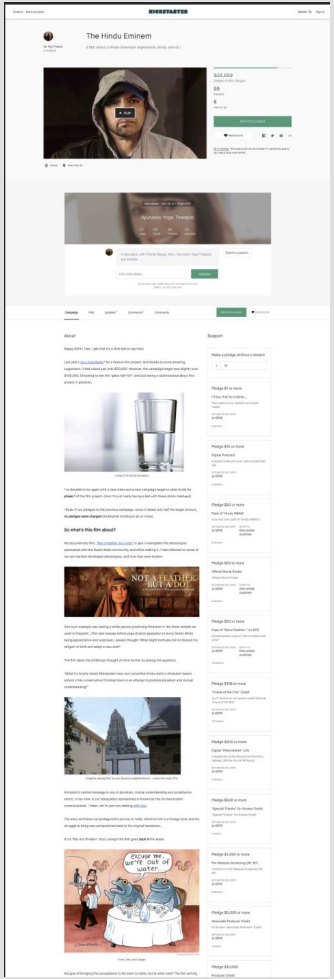
Pledge $5,000 or more

Associate Producer Credit

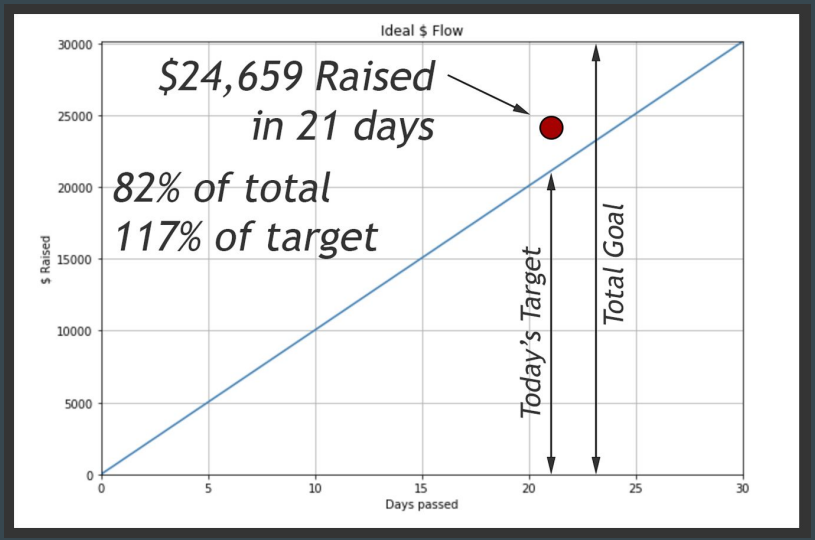
On Screen "Associate Producer" Credit

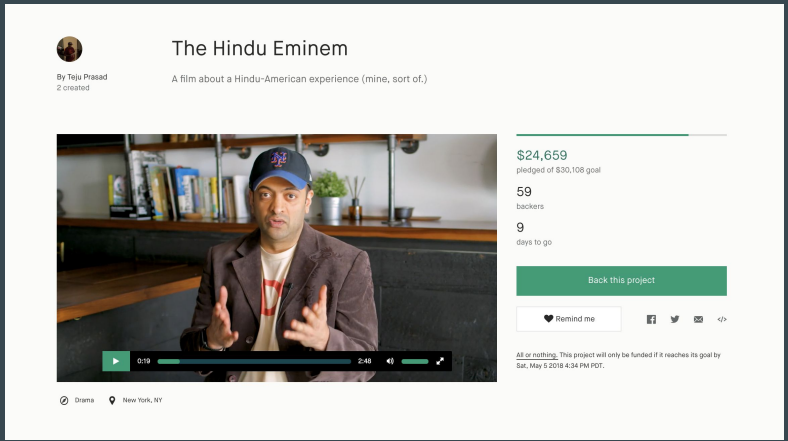
ESTIMATED DELIVERY
Jul 2019

1 backer



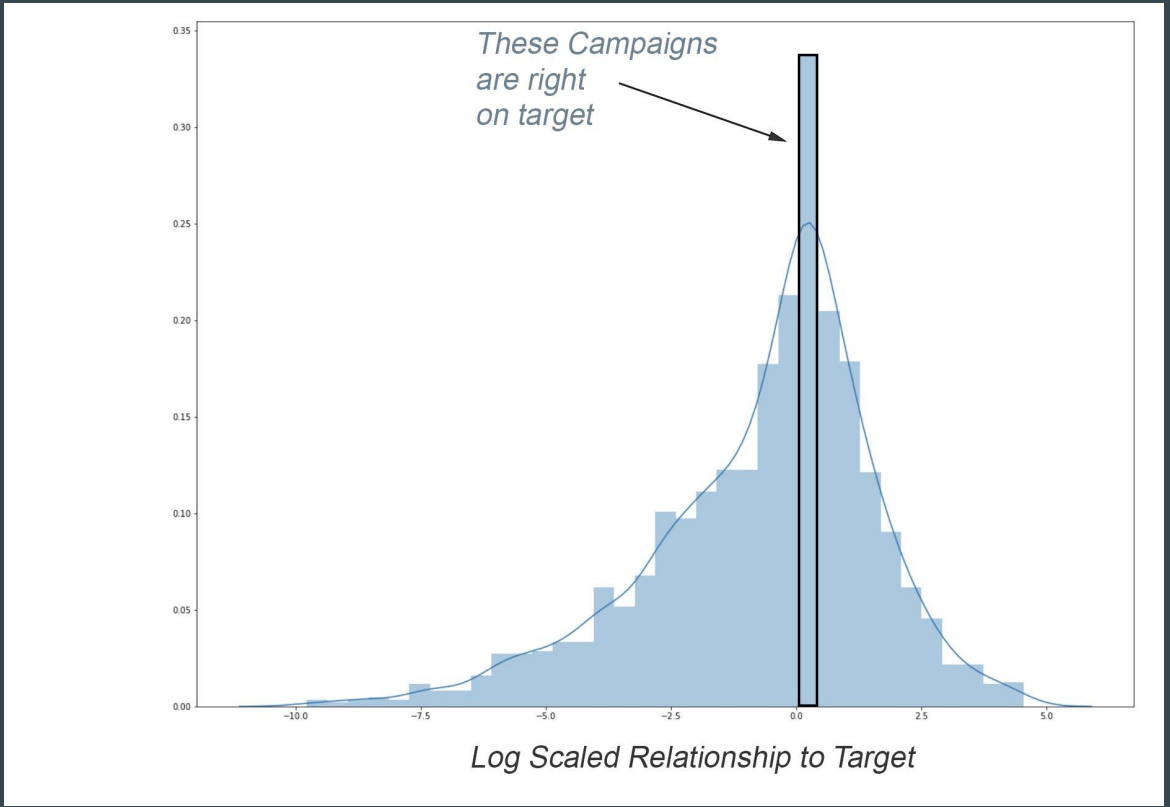
Measuring “Success”





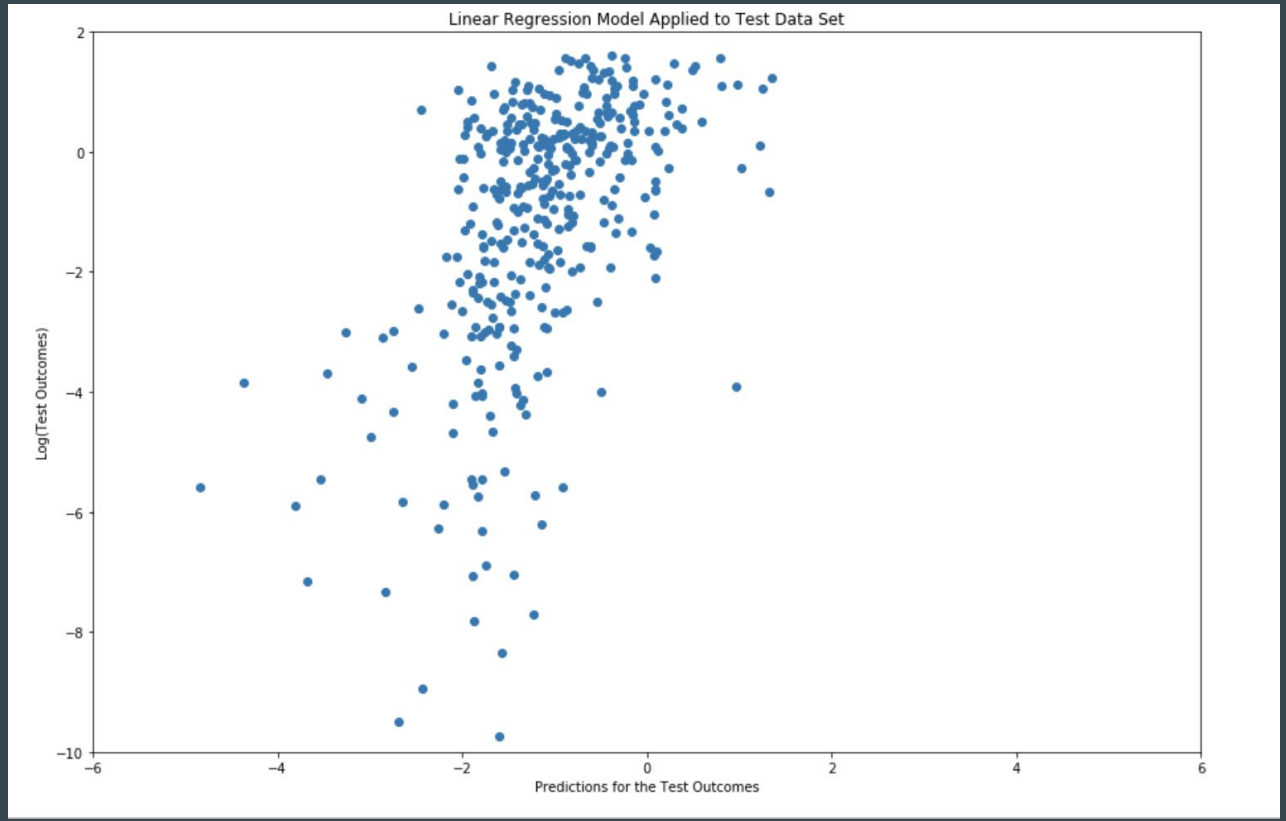
- Kickstarter does not allow search of failed campaigns
- We look at ongoing campaigns
- We estimate their success compared to a straight path toward their target
- This project is at 1.17
- 1.0 means “on track”

Distribution of ‘y’



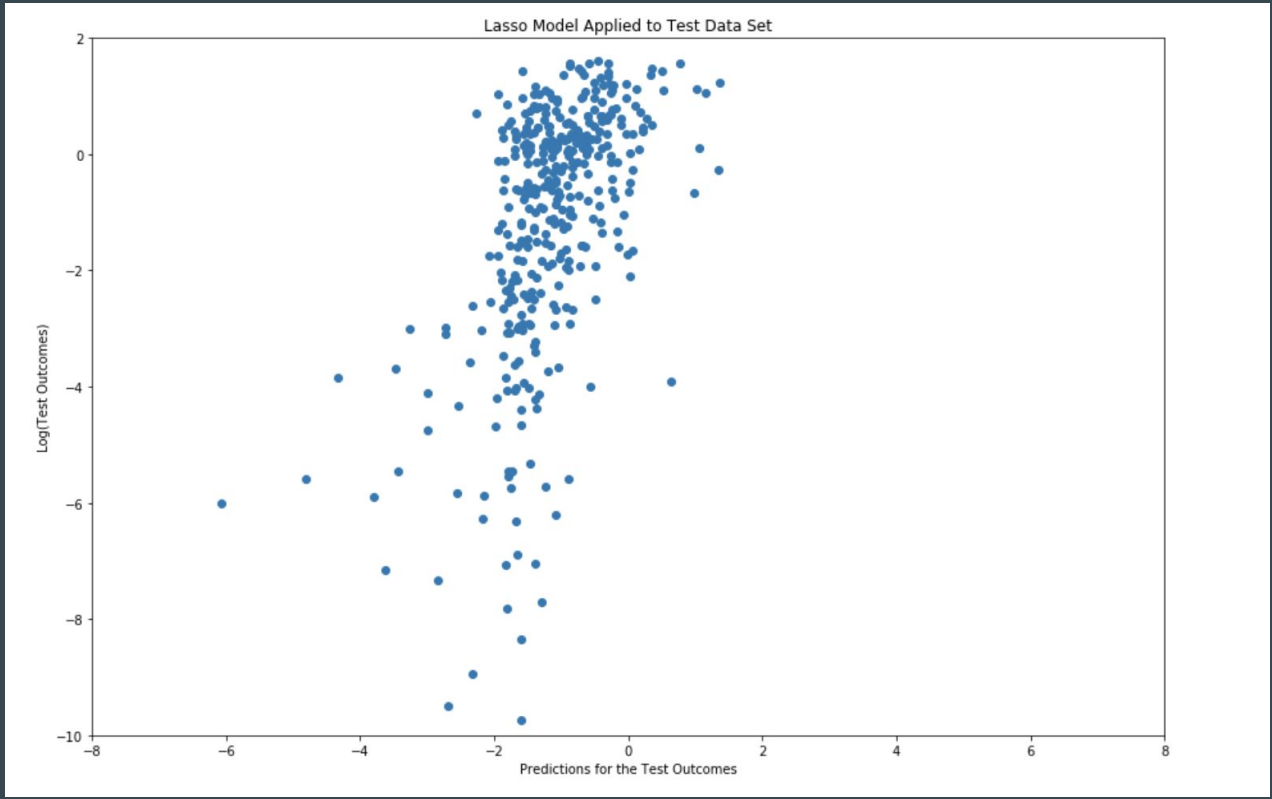
Three Models

Ordinary Least Squares Model



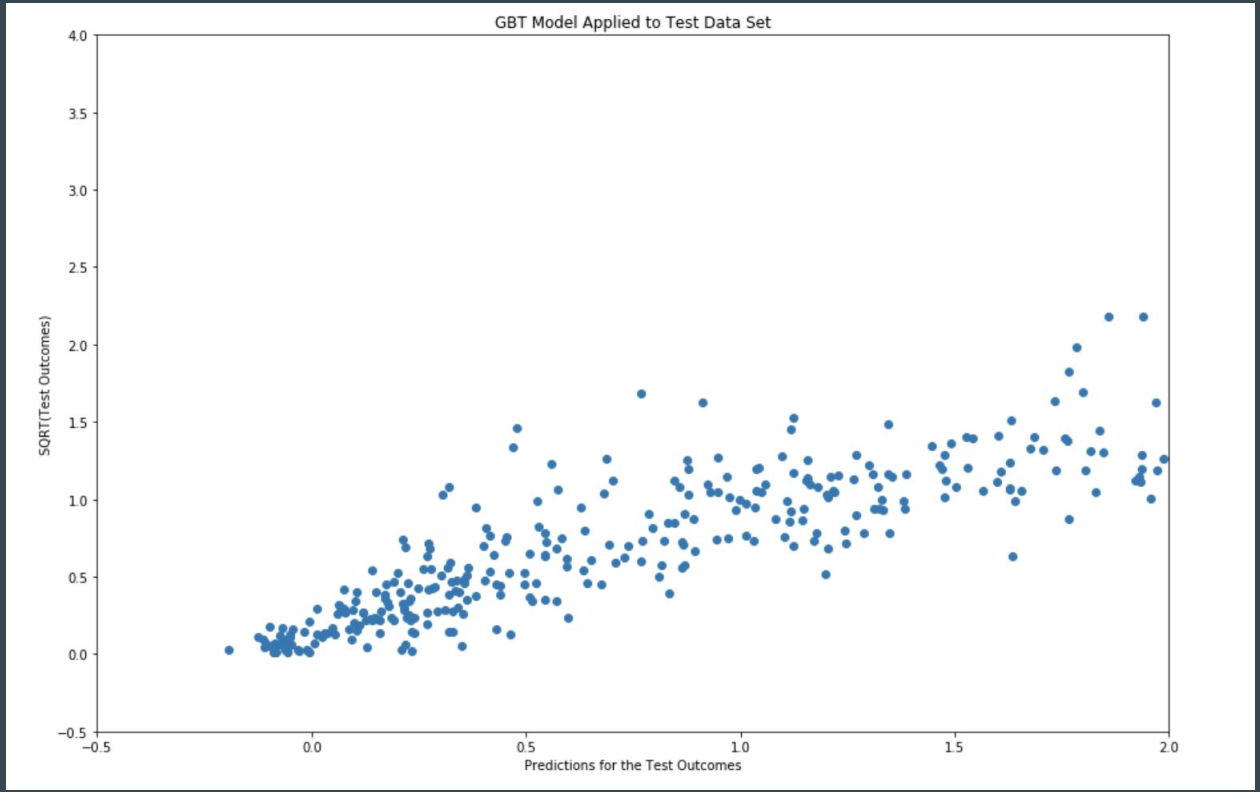
Adjusted RMSE = 0.96248723734544

Polynomial Features and Lasso



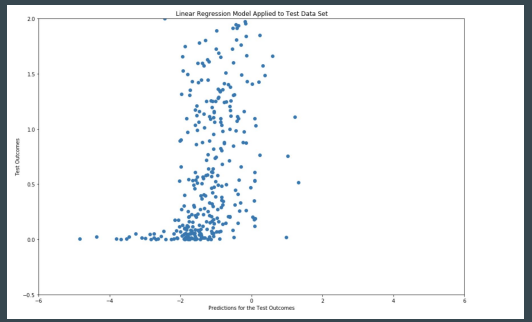
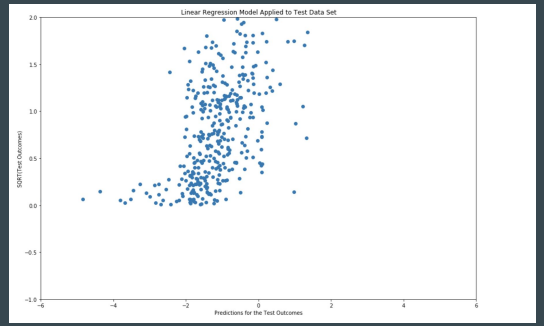
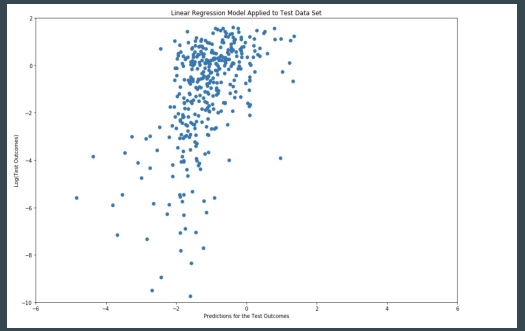
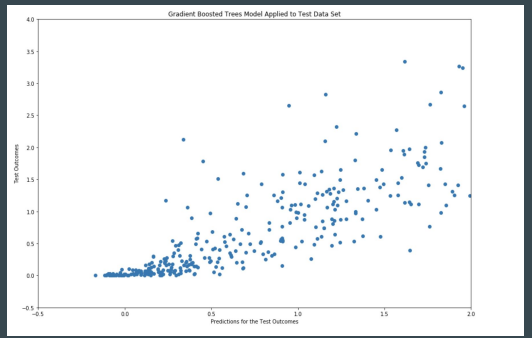
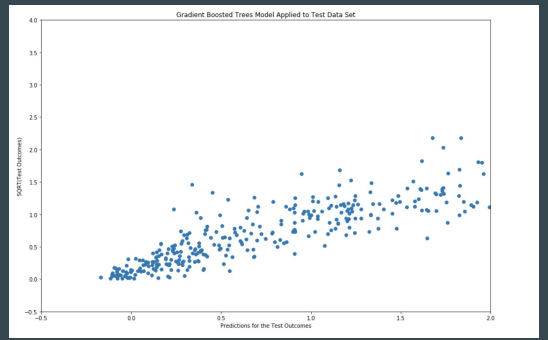
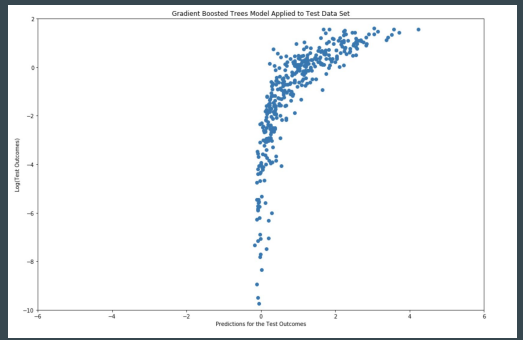
Adjusted RMSE = 1.0268887720267705

Gradient Boosted Trees



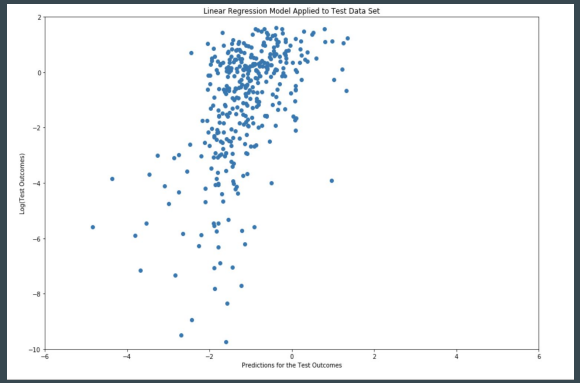
Adjusted RMSE = 0.735480310060743

Appendix: Response Variable Scaling

	(y)	Sqrt(y)	Log(y)
Ordinary Least Squares			
Gradient Boosted Trees			

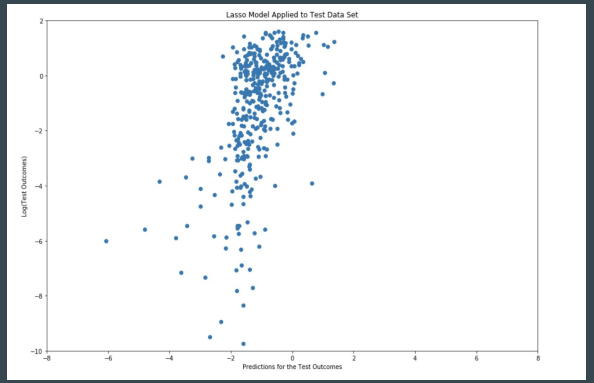
Data Summary:

Ordinary Least Squares



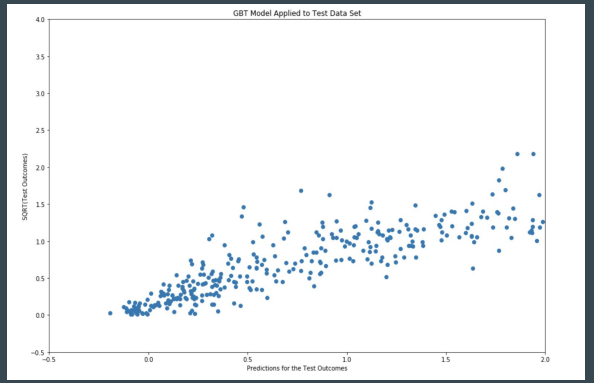
R^2 Max = .418
Adjusted RMSE = 0.96248723734544

L1 Norm (Lasso)



R^2 Max = .275
Adjusted RMSE = 1.0268887720267705

Gradient Boosted Trees



R^2 Max = .586
Adjusted RMSE = 0.735480310060743

Which Features Contain the Variance?

Strong Effect

- Total Backers
- Total Goal
- Number of Images
- Total Text
- Updates

Weak Effect

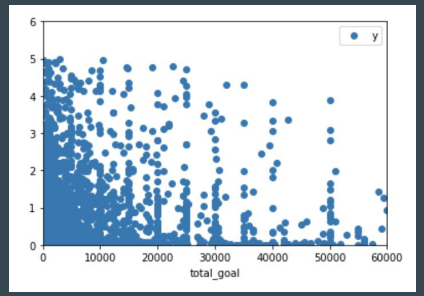
- Number of Tiers
- Length of Title
- Length of Campaign

Surprises

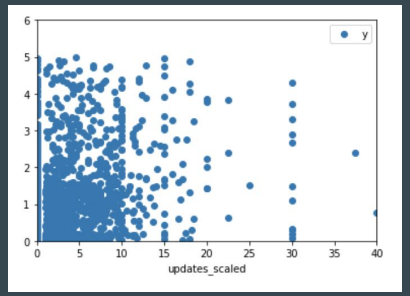
- Video has no effect at all!?

Plots of Largest Predictors

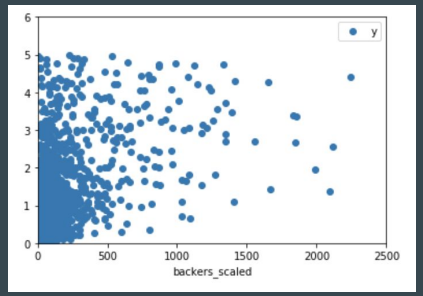
Goal Size



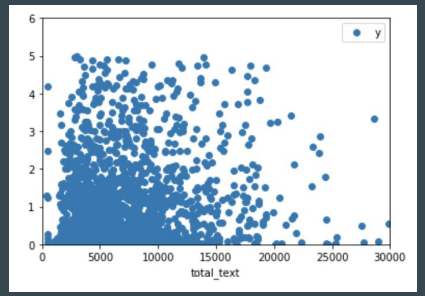
Updates



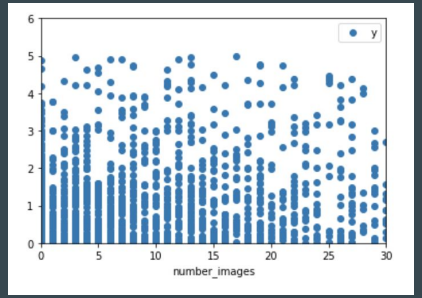
Number of Backers



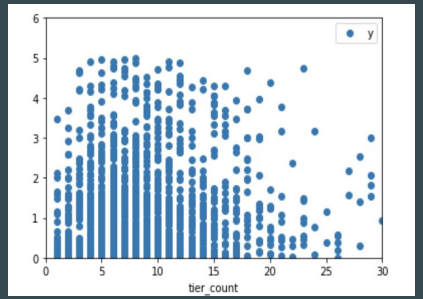
Total Text



Number of Images



Tier Count



R^2 = 0.032

R^2 = 0.050

Conclusions:

- Features themselves are only lightly predictive
- Not improved by looking to simple feature transformations
- Gradient boosted trees model does better but it’s hard to say why at this time
- Get the numerics right but they are clearly not the whole story so focusing on them to the exclusion of your content is a huge mistake.

Future Directions (Data):

- Stretch Goal Size ($)
- Video Length (seconds)
- Tier Dollar Amounts ($)
- Backers per Tier (int)
- Price Point of Lowest Tier ($)
- Prior Successful Campaigns (int)
- Prior Unsuccessful Campaigns (int)
- Video hosted on Youtube vs. Other (boolean)
- Number of Followers of Campaign Creator
- Time of Year (date)
- Video View Count (int)
- Pageviews (int)
- Mentions Burning Man (boolean)
- Category

Future Directions (Modeling):

- Huber loss to give less weight to outliers

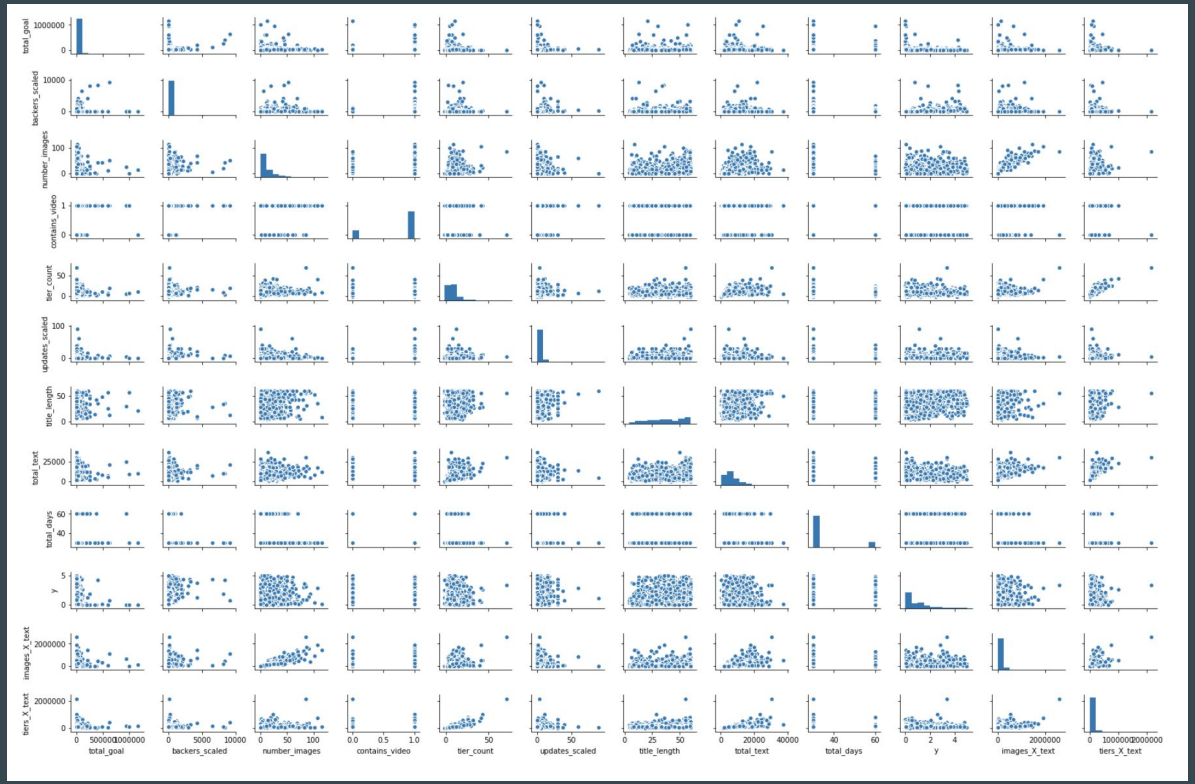
- Motivate the use of the Gradient Boosted Trees model

- Check the assumptions for linear regression
  - Regression is linear in parameters & correctly specified
  - Residuals ( ) should be normally distributed with zero mean
  - The error term must have constant variance
  - Assumption 4: Errors are uncorrelated across observations
  - No independent variable is a perfect linear function of any other independent variable (no perfect multi-collinearity)

Fin!

(Questions)?

Appendix: Feature Pair Plots



Appendix: Correlation Matrix

	total_goal	backers_scaled	number_images	contains_video	tier_count	updates_scaled	title_length	total_text	total_days
total_goal	1.000000	0.262742	0.097397	0.063295	0.035208	0.001725	-0.007649	0.153234	0.050043
backers_scaled	0.262742	1.000000	0.263059	0.110473	0.137613	0.228440	0.005444	0.222911	-0.053239
number_images	0.097397	0.263059	1.000000	0.173836	0.297501	0.209192	0.130526	0.445501	-0.030529
contains_video	0.063295	0.110473	0.173836	1.000000	0.150150	0.129097	0.047185	0.193876	-0.017285
tier_count	0.035208	0.137613	0.297501	0.150150	1.000000	0.180116	0.064274	0.560014	-0.052054
updates_scaled	0.001725	0.228440	0.209192	0.129097	0.180116	1.000000	0.070169	0.223417	-0.070860
title_length	-0.007649	0.005444	0.130526	0.047185	0.064274	0.070169	1.000000	0.153803	-0.006166
total_text	0.153234	0.222911	0.445501	0.193876	0.560014	0.223417	0.153803	1.000000	-0.015411
total_days	0.050043	-0.053239	-0.030529	-0.017285	-0.052054	-0.070860	-0.006166	-0.015411	1.000000

Appendix: OLS Residual Plot