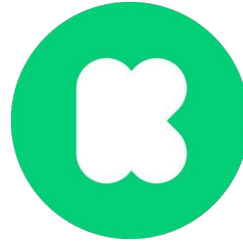




WILL IT SUCCEED ON **KICKSTARTER?**

By Benjamin Karlsberg

How Kickstarter Works



- ▶ “Crowdfunding” Platform
- ▶ Set a pledge goal
- ▶ Give title, description, rewards, video/picture
- ▶ That’s it! Low barrier to entry
- ▶ Funding only received if pledge goal reached
 - ▷ “All-or-Nothing” Condition

Requires a balance of pledge goal and expected funding needed. Failing to deliver can result in legal recourse. (Risky to underfund)

OBJECTIVE

- ▶ Create a model that predicts a project's probability of success, pre-launch
- ▶ Value: Informing realistic expectation of funding

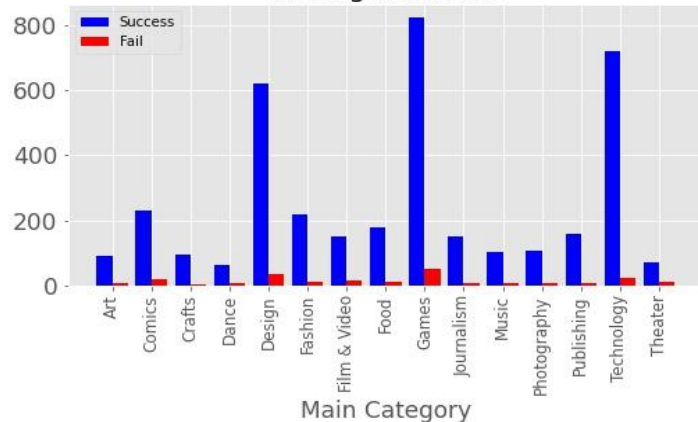
Challenge: Most data sourced from Kickstarter is from during or after project launch (leakage). Not much information available from pre-launch projects.

DATA INFORMATION

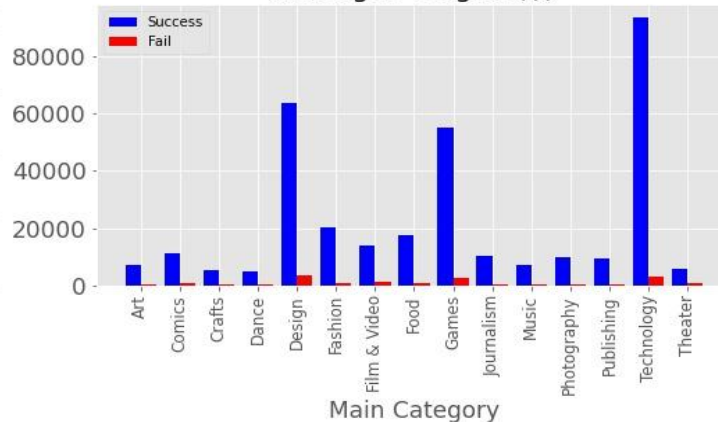
- ▶ Sourced from Kaggle and supplemented with Harvard relational database
- ▶ >35,000 Total Projects between 2009 and 2018
- ▶ Pre-launch features available:
 - ▷ Pledge goal, Category, Sub-category, Currency, Country, Campaign Duration, Reward Count, Description (“Blurb”)
- ▶ Class Imbalance:
 - ▷ 37% Successes, 63% Failures
- ▶ Test-Train Split: 80% Train, 20% Test

EDA Success vs Fail

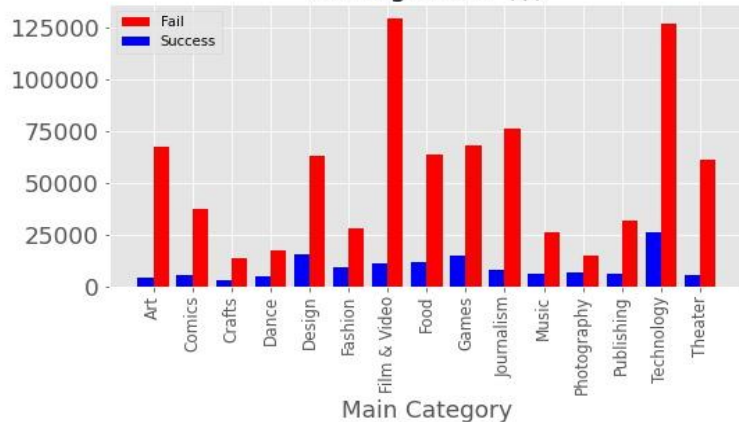
Average Backers



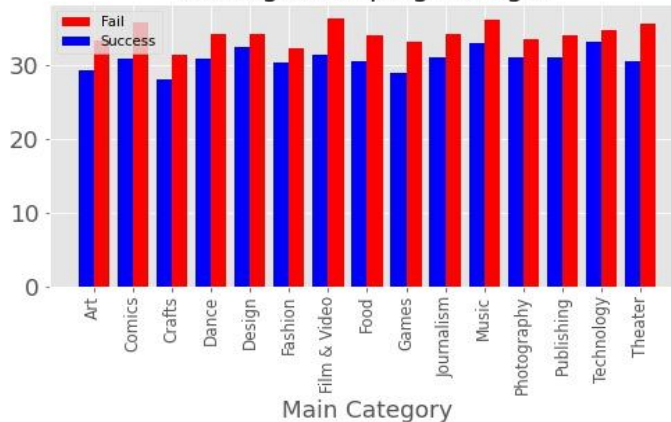
Average Pledged (\$)



Average Goal (\$)



Average Campaign Length

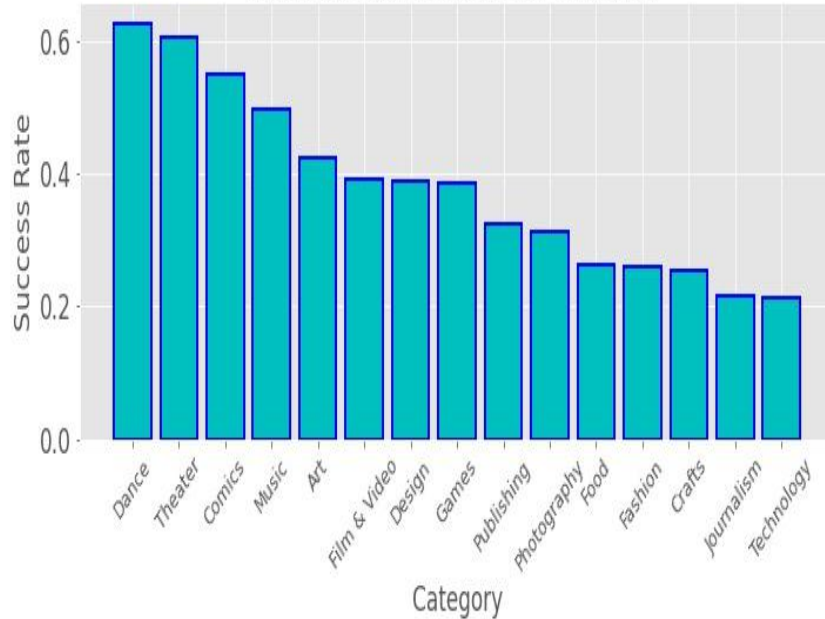


Successes
Went Viral

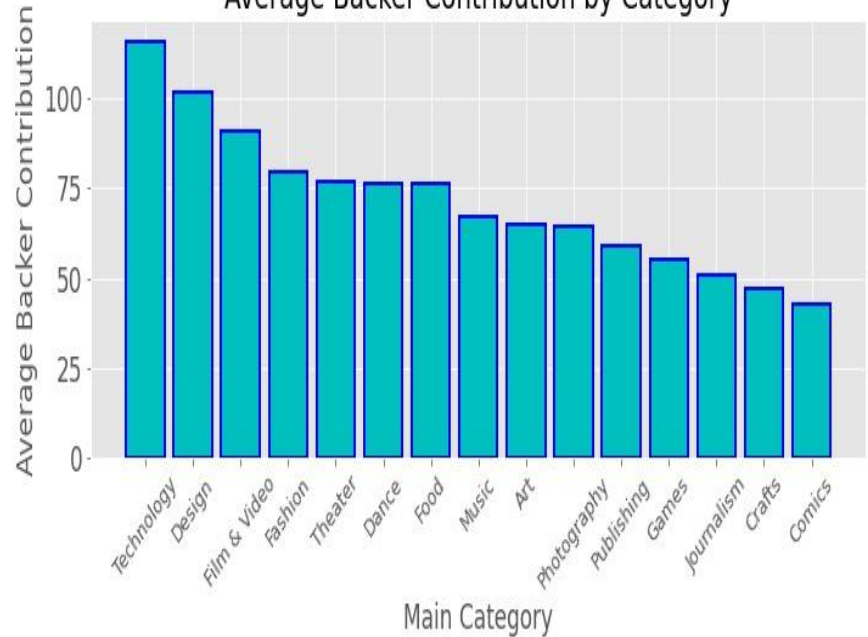
Failures Had
Much Higher
Goals

CATEGORY EDA

Average Success Rate by Category



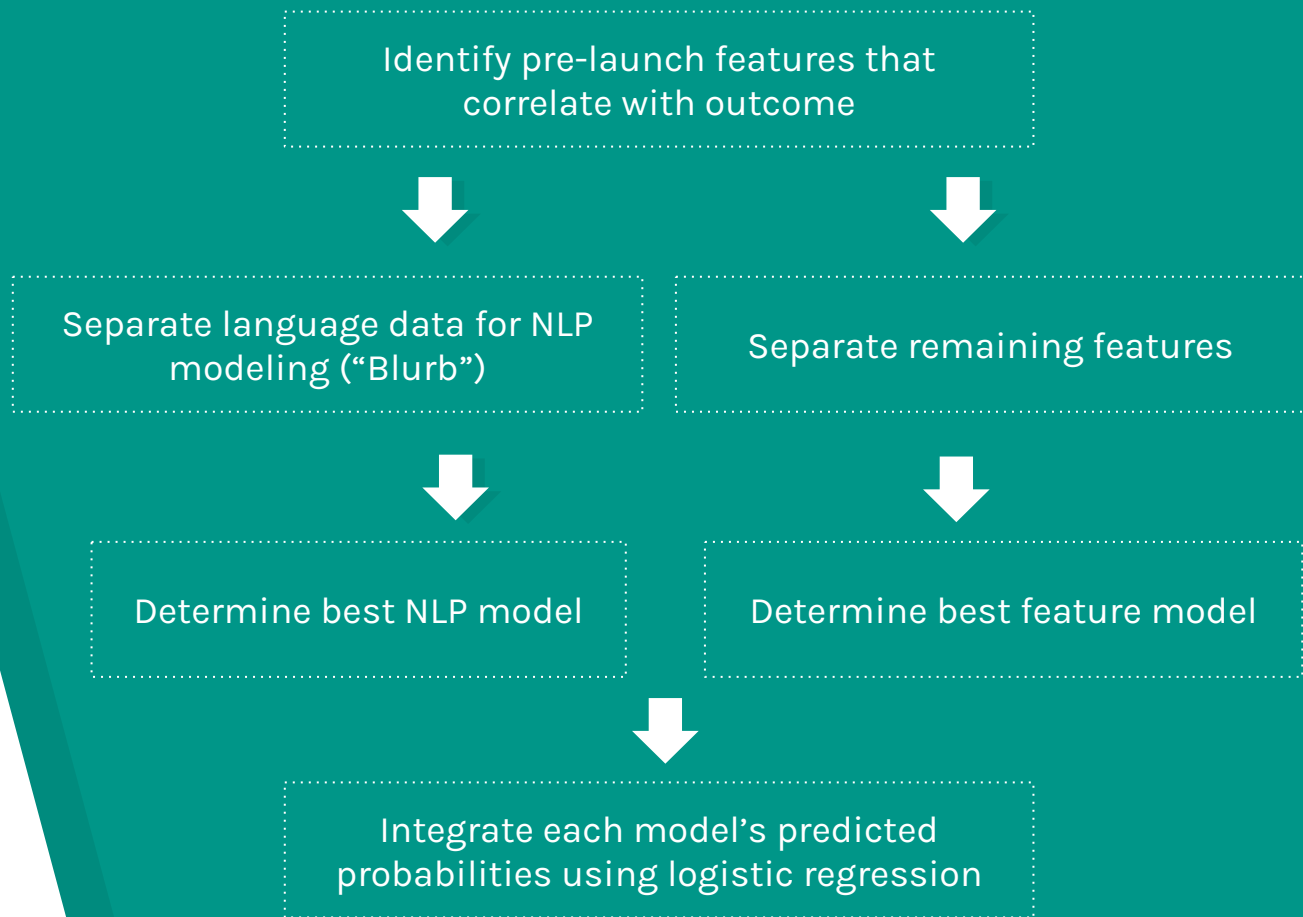
Average Backer Contribution by Category



- ▶ Dance/Theater have the highest success rates
- ▶ Technology gets the largest donations per backer



MODELING STRATEGY



Model Results (Cross Validation)

Features (63% Baseline Accuracy)

	Logistic Regression	Random Forest Classifier	Adaboost	Gradient Boosted Classifier	Neural Net
Accuracy	.694	.700	.722	.730	.715
Precision	.638	.607	.652	.656	.599
Recall	.422	.565	.552	.584	.718

Top 3 Features:

Pledge Goal


Reward Count

Campaign Duration

NLP ("Blurb") (63% Baseline Accuracy)

	Logistic Regression	Random Forest Classifier	Naive Bayes	Gradient Boosted Classifier
Accuracy	.683	.662	.671	.647
Precision	.618	.679	.599	.669
Recall	.401	.185	.367	.115

COMBINED MODEL TRAINING RESULTS

	Baseline	Feature Gradient Boosted Classifier	NLP Logistic Regression	Logistic Combined
Accuracy	0.63	.730	.683	 .762

COMBINED MODEL TESTING RESULTS

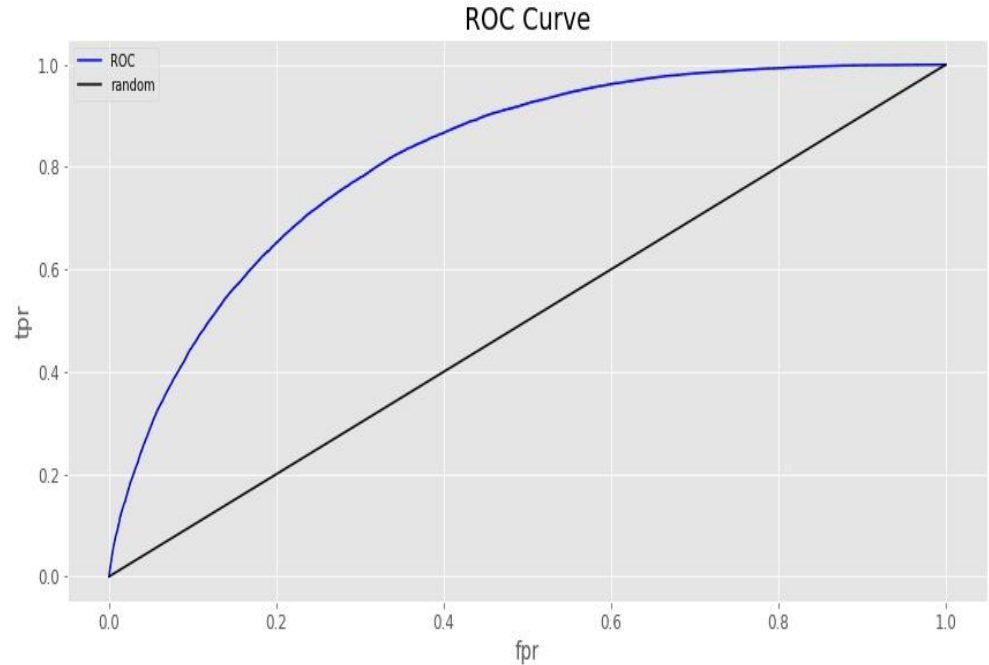
	Logistic Combined
Accuracy	.745
Precision	.680
Recall	.601

COMBINED MODEL TESTING

Confusion Matrix (Positive = class fail)

~70,000 Test Data	Actual Positive	Actual Negative
Predicted Positive	TP = 36822	FP = 10599
Predicted Negative	FN = 7508	TN = 15946

AUC Score	.728
-----------	------





Takeaway **CONCEPTS**

Setting pledge goal too low might cost you out-of-pocket expenses to deliver on promises

Setting it too high will net \$0 in funding, but can always be re-launched

Knowing if your project will succeed or not before launch can help decide how much risk you are willing to take



THANK YOU!

Any questions?