

Project Luther Write-up - Kickstarter Board Games

Goal:

Predict success of projects in the “Tabletop Games” category on Kickstarter, and infer ways for a project creator to increase likelihood of success.

Data / Scraping:

Collected project URLs by iterating through Kickstarter explore index.

For each URL, scraped data from:

- Project Page
 - Total \$ raised, total backers
 - Fundraising goal, funding period length / start / end
 - Number of reward tiers, min / median / max reward price,
 - Most backed reward price (usually corresponds to price of 1 copy of the game)
- FAQ / Update / Community sub-pages
 - Number of FAQ, Updates, community comments
 - Number of US backers, first time vs. existing backers
- Creator profile
 - Creator # projects created, # projects backed

Regression analysis

EDA / Cleaning

- Removed rows with null values
- Histogram of target variables (Total \$ raised, total backers)
 - Removed inactive projects with minimal funding, extreme values ($>3x$ standard dev)
- Looked at pair plots, correlation, p-values for regression on all inputs
 - Created scaled % features for US backers and first-time vs. existing backer to remove correlation with total backers
 - Identified subsets of features to add / remove based on correlation and p-values

Regression

- Wrote a function which takes a model object, input array (Xs), and target series (Ys), then runs crossfold validation on the inputs using the model, prints out diagnostic charts and model scores
- Ran validation regressions on 80% of data:
 - Single input
 - Top 4 correlated inputs
 - Other multi-input combinations identified in EDA
 - All inputs
 - Inputs + 2nd degree polynomial terms (interaction, squares, both)
 - Inputs + 3rd degree polynomial terms (interaction, squares & cubes)
 - Lasso CV on 2nd degree polynomial terms

Testing

- Tested best model (2nd degree interaction terms) on 20% holdout data
- Looked at linear regression with all inputs for inference purposes