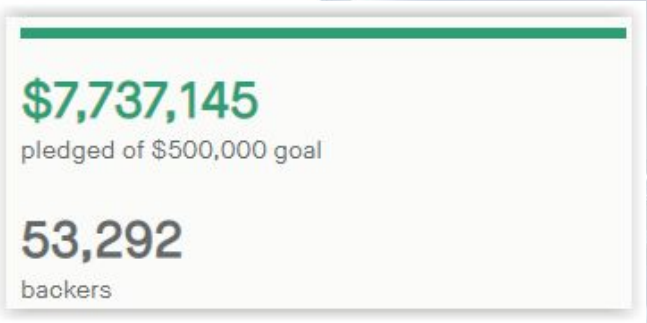
Modeling Funding of Kickstarter Tabletop Games

Luke LaJoie

Introduction



- Many new board games are funded through Kickstarter
- Acquire data from Kickstarter
- Determine features which most affect funding

Methodology

Scrape data from Kickstarter

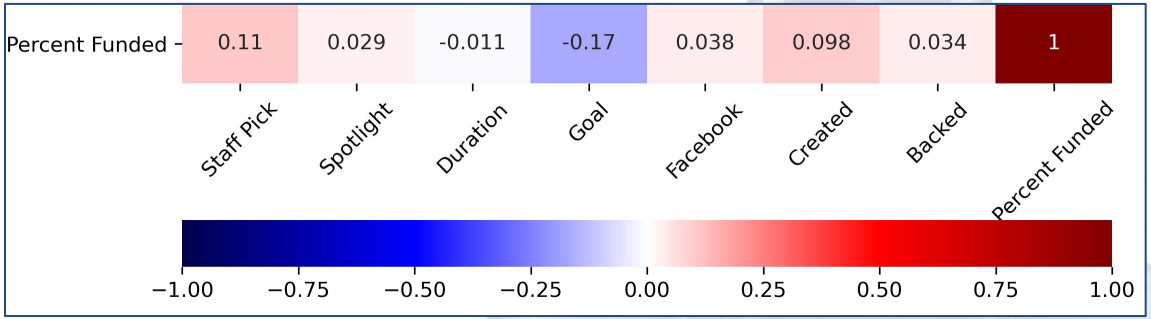
Clean and explore data

Compare regression models

Find features relevant to funding

Methodology

Correlation Heatmap



Metrics

- R^2
- Adj R^2
- RMSE

Models

- OLS
- Ridge
- Lasso

Results

After splitting the data and running the models

- OLS and Ridge had very similar R-Squared: ~0.05
- 2nd degree polynomial mean: negative R-Square by factor of 100
- LassoCV R-Squared ~0.06
    - Data split differently, so not valid comparison

- OLS, Ridge and LassoCV had very similar RMSE
- 2nd degree polynomial

Results

- Added all features shown in heatmap and stat models with OLS

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.6661	0.154	43.374	0.000	6.364	6.968
Staff Pick	0.4962	0.111	4.474	0.000	0.279	0.714
Funding Goal	-0.1336	0.020	-6.585	0.000	-0.173	-0.094
Created Projects	0.1227	0.030	4.073	0.000	0.064	0.182
Duration	0.0076	0.004	2.098	0.036	0.000	0.015

Conclusions

- These models did not account for much variance
- May want to consider factor marketing, advertising, YouTube reviewers, etc.

Future Work

- Get more data
  - More Kickstarter scraping
  - Kicktraq, Boardgamegeek

- Classification may be more relevant to model if projects will fund or not

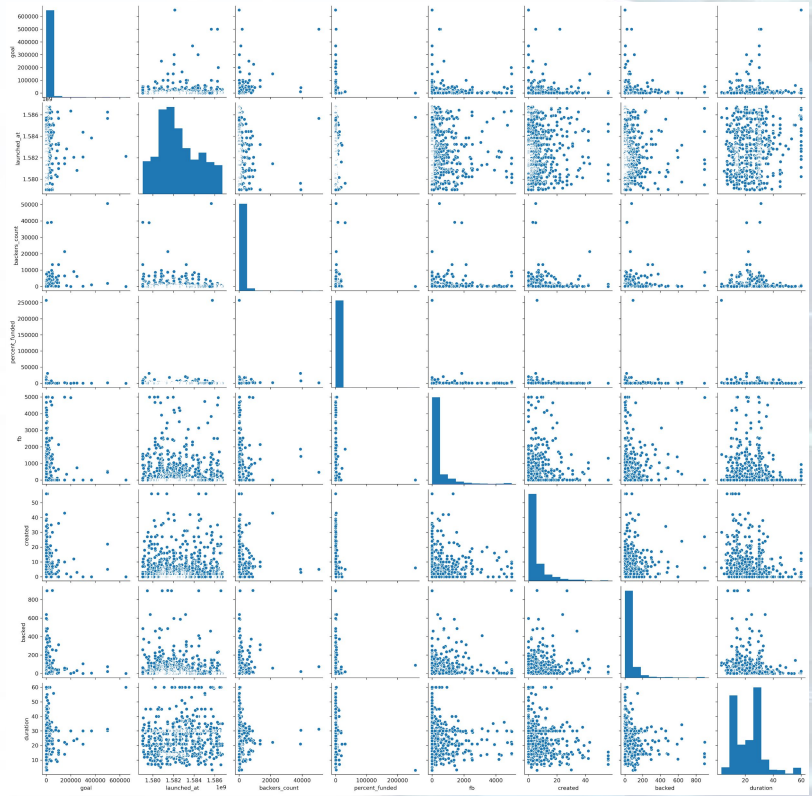
Questions?

Questions?

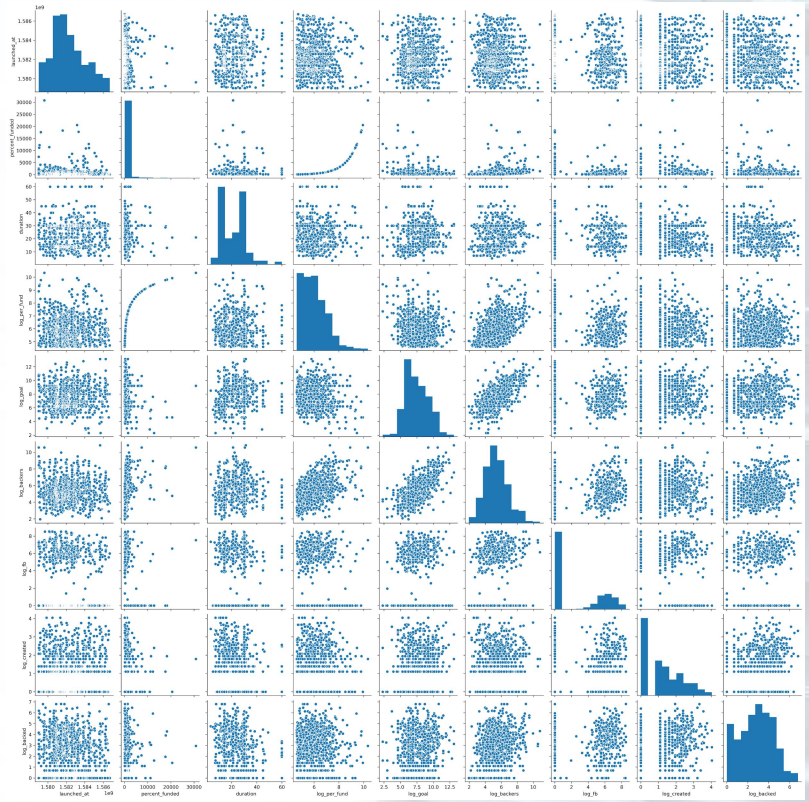
Appendix

1 slide
Recommendations, interesting insights

Before Transformations



After Transformations



OLS Regression

OLS Regression Results

Dep. Variable:	log_per_fund	R-squared:	0.066
Model:	OLS	Adj. R-squared:	0.062
Method:	Least Squares	F-statistic:	15.89
Date:	Fri, 17 Apr 2020	Prob (F-statistic):	1.42e-12
Time:	01:19:46	Log-Likelihood:	-1259.1
No. Observations:	900	AIC:	2528.
Df Residuals:	895	BIC:	2552.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.6661	0.154	43.374	0.000	6.364	6.968
staff_pick[T.True]	0.4962	0.111	4.474	0.000	0.279	0.714
log_goal	-0.1336	0.020	-6.585	0.000	-0.173	-0.094
log_created	0.1227	0.030	4.073	0.000	0.064	0.182
duration	0.0076	0.004	2.098	0.036	0.000	0.015

Omnibus:	109.819	Durbin-Watson:	1.955
Prob(Omnibus):	0.000	Jarque-Bera (JB):	155.248
Skew:	0.892	Prob(JB):	1.94e-34
Kurtosis:	3.980	Cond. No.	120.

CV: Simple, Ridge, Poly(2nd)

Simple regression scores: [0.026579358917850793, 0.042986726079853654, 0.04064433499357778, 0.03350101722456533, 0.02419757543410994]

Ridge scores: [0.02648539845169773, 0.042966486374552915, 0.04063672561083442, 0.03381675445031196, 0.02499222870346618]

Degree 2 polynomial: [0.11601087366630859, 0.09066726237683354, 0.049376268356580266, -0.014727048686371534, -0.006935456404602469]

Simple mean cv r^2: 0.034 +- 0.007

Ridge mean cv r^2: 0.034 +- 0.007

Degree 2 polynomial mean cv R^2: 0.047 +- 0.052

Simple mean cv rmse: 0.997 +- 0.045

Ridge mean cv rmse: 0.997 +- 0.045

Degree 2 polynomial mean cv rmse: 0.990 +- 0.047

LassoCV

```
Out[278]: [('staff_pick', 0.09426745031501062),
           ('is_starrable', -0.0318220984852419),
           ('spotlight', -0.0),
           ('duration', 0.0716453234341529),
           ('log_goal', -0.2037579781868695),
           ('log_fb', 0.0),
           ('log_created', 0.167554641533311),
           ('log_backed', -0.0)]
```

```
In [279]: # Make predictions on the test set using the new model
          test_set_pred = lasso_model.predict(X_te)
```

```
In [280]: # Find the MAE and R^2 on the test set using this model
          mae(y_test, test_set_pred)
```

```
Out[280]: 0.8478544186745868
```

```
In [281]: lasso_CV_rmse = np.sqrt(mean_squared_error(y_test, test_set_pred))
          lasso_CV_rmse
```

```
Out[281]: 1.0729815039311457
```

```
In [282]: lasso_CV_r2 = r2_score(y_test, test_set_pred)
          lasso_CV_r2
```

```
Out[282]: 0.04107750228775209
```

Modeling Funding of Kickstarter Tabletop Games

Luke LaJoie

Introduction

- Many new board games are funded through Kickstarter
- Acquire data from Kickstarter
- Determine features which most affect funding