Modeling Game Review Data

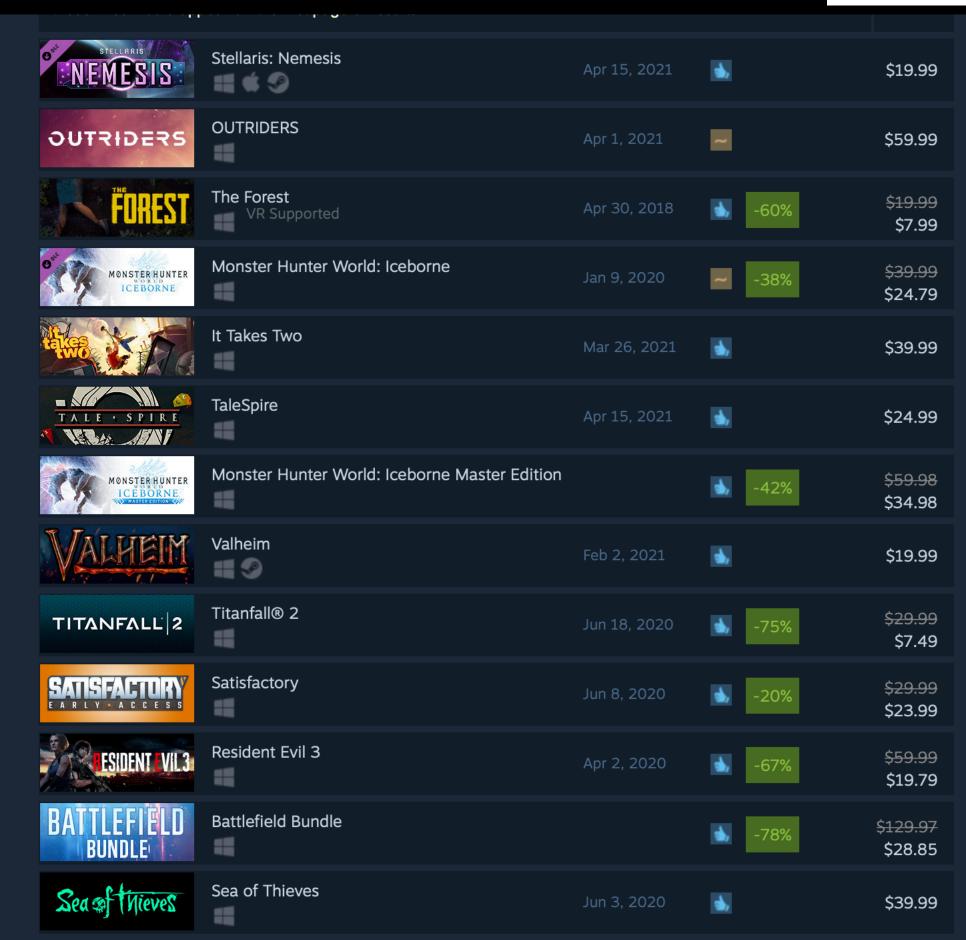# Methodology
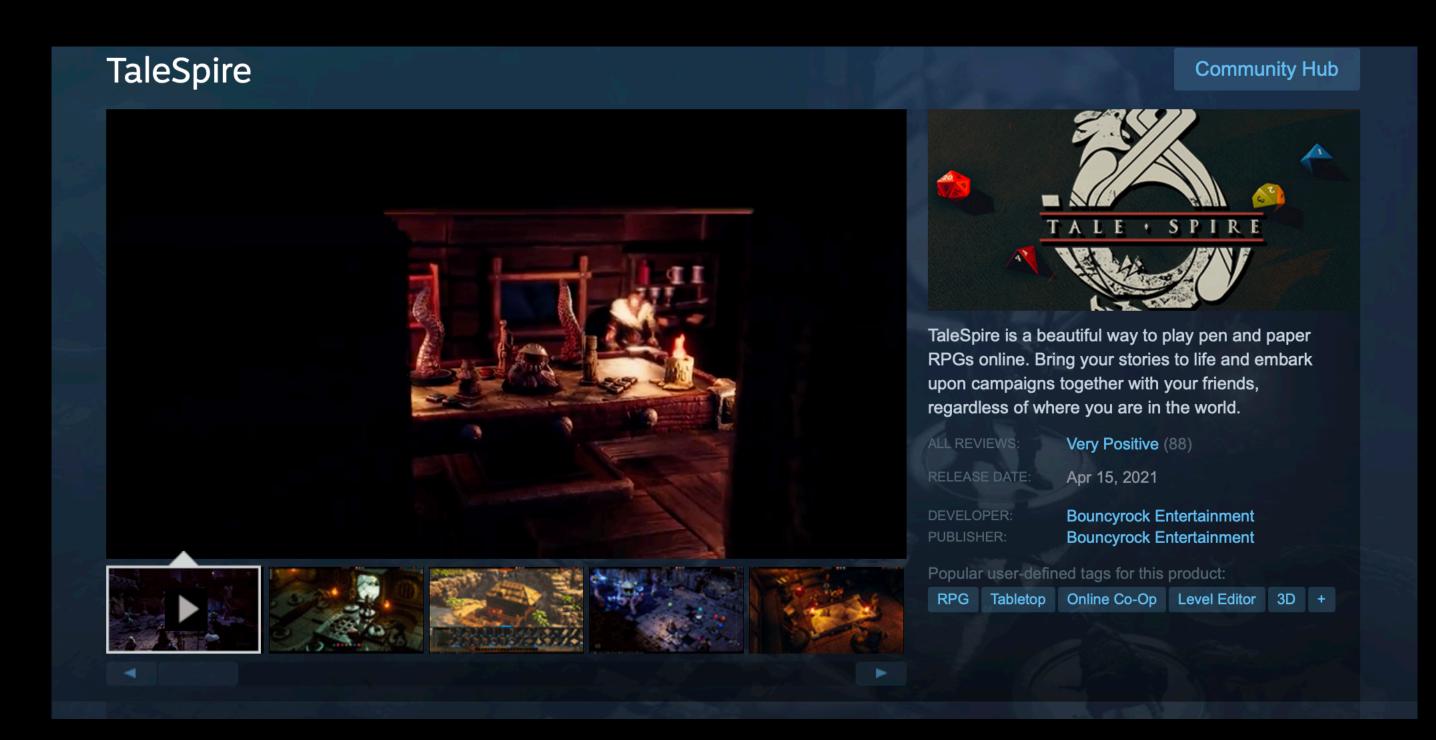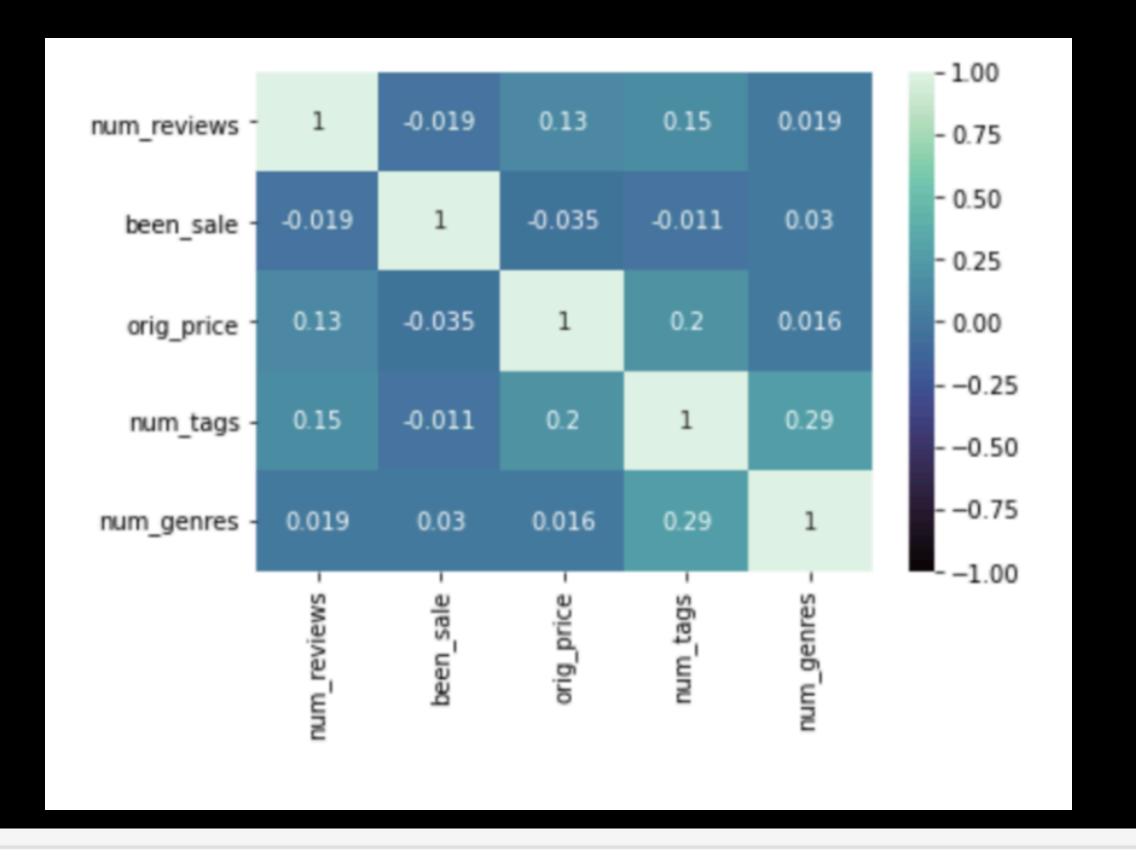

- Features were scraped from over 16k game pages

Baseline Modeling .03


```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5528 entries, 0 to 5527
Data columns (total 6 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   num_reviews   5528 non-null    int64
 1   release_date  5526 non-null    object
 2   been_sale     5528 non-null    int64
 3   orig_price    5528 non-null    float64
 4   num_tags      5528 non-null    float64
 5   num_genres    5528 non-null    float64
dtypes: float64(3), int64(2), object(1)
memory usage: 259.2+ KB
```

```
0.0328302298223444
[-1668.08354042    312.50037055    859.3763758
-8474.274915866785
```

-initial r^2

-coeficients for each column

-intercept

Feature Engineering
82 Columns

- 6 numerical columns

- Several categorical

- Final: 82 columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5528 entries, 0 to 5527
Data columns (total 15 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Unnamed: 0      5528 non-null    int64
 1   title           5528 non-null    object
 2   num_reviews     5528 non-null    int64
 3   release_date    5526 non-null    object
 4   orig_price      5528 non-null    float64
 5   been_sale       5528 non-null    int64
 6   review_rating   5528 non-null    object
 7   rating_val      5521 non-null    float64
 8   num_languages   2102 non-null    float64
 9   tags            5528 non-null    object
 10  num_tags        5528 non-null    float64
 11  genres          5528 non-null    object
 12  num_genres      5528 non-null    float64
 13  publisher       5523 non-null    object
 14  top_pubs        5528 non-null    object
dtypes: float64(5), int64(3), object(7)
memory usage: 647.9+ KB
```

Lasso Regularization

- Price

- Rating

- Publishers

```
('orig_price', 4023.861405248766),
('been_sale', -290.42170773673683),
('rating_val', 699.7324947475248),
('num_tags', 4793.755372488246),
('num_genres', 0.0),
('year_released', -3473.0458848682997),
('languages_filled', 6724.944309134146),
('BANDAI NAMCO Entertainment', -774.8231135935904),
('Bethesda Softworks', 1010.0328039938552),
('Devolver Digital', 297.7402707251717),
('Disney', -576.6962798739536),
('Electronic Arts', -578.0312769914183),
('Kagura Games', -179.16318825029904),
('Nightdive Studios', -242.07395354529697),
('Paradox Interactive', 759.2700199485986),
('SEGA', -487.7257302877862),
('Square Enix', -685.1406895023088),
('THQ Nordic', -840.3422726283607),
('Ubisoft', -831.05304326482),
('Xbox Game Studios', 229.22680947826834),
('other', 0.0),
```

Scoring

- PolynomialFeatures

0.6314130761771726

- Final scoring

0.07617142187529358