## Predicting Success of Steam Games

## Task

Investigate how well can be post-release success of games predicted from basic information about games (available before release)

No data regarding previews, early reviews, user activity on social networks...

## Steam

Featured liem
$\square$ Sotware - Hardware -
Videos
News

For You


Largest platform for selling PC games

Since 2004
Over 10,000 games in total

Sign in to add this item to your wishlist, follow it, or mark it as not interested

| Buy Rocket League | A |  | 2 | Single-player |
| :---: | :---: | :---: | :---: | :---: |
|  | 19,99€ | Add to Cart | 8 | Mulit-player |
|  |  |  | 8 | Cross-Platiorm Mulitiayer |
| Buy Rocket League 4-Pack | A |  | 28 | Co-op |
|  | 54,996 | Add to Cart | 8 | Local Co-op |

## Steam Charts

 Rocket LeagueTracks how many players are in game every hour

Monthly statistics
Since July 2012
Not very accurate measure of success (?)
(however, Steam Charts is the best source with sufficient history)
Zoom 48h 7 d 1m $3 \mathrm{~m} \quad 6 \mathrm{~m}$ 1y All


Compare with others

Month
Last 30 Days
November 2015
October 2015
September 2015
August 2015
July 2015
June 2015

Avg. Players
$16,942.3$
$16,276.4$
16,709.3
$17,439.9$
19,144.3
10,789.6

Gain
$+665.9$
-432.9
-730.6
$-1,704.5$
$+8,3547$
$+10,787.7$

From Nov 29, 2015 To Dec 6, 2015

| \% Gain | Peak Players |
| :---: | ---: |
| $4.09 \%$ | 37,300 |
| $2.59 \%$ | 37,175 |
| $4.19 \%$ | 32,204 |
| $8.90 \%$ | 34,391 |
| $7.43 \%$ | 35,686 |
| $51.27 \%$ | 29,027 |
| - | 25 |

## SteamSpy

Tracks number of owners

Since April, 2015
No access to history
Manually collected data for about 1,400 games

Still not a good measure of success?

## steamspy

Rocket League


Developer: Psyonix, Inc. Publisher: Psyonix, Inc.
Genre: Action, Indie, Racing, Sports
Languages: English, French, Italian, German, Spanish, Dutch, Portuguese, Japanese, Korean, Russian, Tukkish
Tags: Multiplayer (1476), Racing (1430), Soccer (1229), Sports (1142), Competitive (974) Team-Based (943), Football ( (847), Online Co-Op (800), Action (694), Co-op (620), Funny ( 5933 ), Fast-Paced ( 5855 ), Local Multiplayer ( 5344 ), Local Co-Op ( 530 ), Great Soundtrack (519), Split Screen (502), 4 Player Local (473), Singleplayer (467), Indie (362), Casual (343)

Category: Single-player, Multi-player, Co-op, Shared/Split Screen, Cross-Platorm Multiplayer, Steam Achievements, Full controller support, Steam Trading Cards, Steam cloud, Stats
Release date: Jul 7,201 rice: $\$ 19.99$
Score rank: $87 \%$ Userscore: $93 \%$ Old userscore: $93 \%$ Metascore: $86 \%$ Wners: 4,327,767 $\pm 53,25102016-04-21$
ayers in telast 2 weeks: 1,239,904 $\pm 28,629$ (28.65\%)
Players total: $4,153,044 \pm 52,178$ (95.96\%)
YouTube stats: $3,998,144$ views and 28,875 comments for videos uploaded last week. over 50 new videos uploaded yesterday
Playtime in the last 2 weeks: 08:32 (average) 0:23 (median)
Playtime total: $62: 51$ (average) $18: 20$ (median)

OWNERS CHANGES AUDIENCE (2 WEEKS) AUDIENCE TOTAL CCU (DAILY) CCU (HOURLY) TWITCH (DAILY)
twitch (hourin) youtube reviews more like this related ceocraphy playtime (total)
PLAYTIME (2 WEEKS)
Owners data:


## Data

About 4,700 games from July, 2012 - July, 2016
Completely omitted Early Access titles and Free-to-play titles
Info such as genres, price, release date, descriptions, languages, features

+ Thumbnail and screenshots of every game

Full image, rPlotter

## Screenshots

K-means used to extract 5 most dominant colors


First screenshot and thumbnail


Reduced image, frequent



Reduced image, rPlotter


Reduced image, frequent


Reduced image, frequent


## Subsetting

Pick only games whose publisher has a history of at least 2 games
Covers 30-40 \% games

## Notable attributes

Previous games: how many, max and min players of any game, Gini index
Developer, Publisher (limited to max 53, rest is "other")
text descriptions
Price
HW requirements: disk space, RAM, GPU
Languages
Number of colors on thumbnail and first screenshot + dominant thumbnail color

## Results

Training, validation, test set ( $60 \%, 20 \%, 20 \%$ )

| Classification (10 classes) | Random Forest | Regression (values 0-15) | SVM |
| :--- | :--- | :--- | :--- |
|  |  | Correlation coefficient | 0.80 |
|  |  | MAE | 1.24 |
|  |  | RMSE | 1.67 |
| Accuracy | $46 \% / 44 \%$ | within +-1 from actual | $53 \% / 52 \%$ |
| within +-1 from actual | $87 \% / 84 \%$ | within +-2 from actual | $82 \% / 77 \%$ |



