

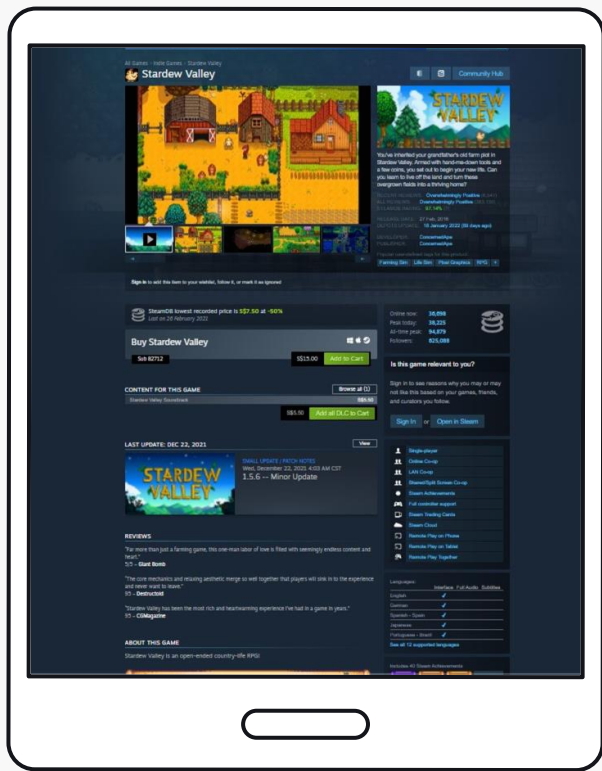


SC1015

Mini- Project: Steam Store Analysis



SC18 Group 10:
Babu Rishe, Wong
Ri Hong, Xing Mian



Intro- duction

<https://store.steampowered.com/>



STEAM®



This is a slide structure based on a multi-purpose presentation

You can delete this slide when you're done editing the presentation

Fonts	To view this template correctly in PowerPoint, download and install the fonts we used
Used and alternative resources	An assortment of graphic resources that are suitable for use in this presentation
Thanks slide	You must keep it so that proper credits for our design are given
Colors	All the colors used in this presentation
Infographic resources	These can be used in the template, and their size and color can be edited
Customizable icons	They are sorted by theme so you can use them in all kinds of presentations

For more info:
SLIDESGO | SLIDESGO SCHOOL | FAQs

You can visit our sister projects:
FREEPIK | FLATICON | STORYSET | WEPIK | VIDFY

Motivation



Table of contents

1/

**Data
Preparation
& EDA**

2/

**Machine
Learning**

3/

**Natural
Language
Processing**

4/

Conclusion

1/

Data Preparation & EDA



Data sources



Steam Store Games
(Clean dataset)
<https://www.kaggle.com/datasets/nikdavis/steam-store-games>



Data Cleaning



Using NumPy and Pandas



Data Preparation



Data Cleaning

Using Numpy and Pandas

1. Remove day and month from 'release_date'
1. Normalizing playtime and ratings
1. Remove delimiters in 'genres', 'categories' and 'steamspy_tags'
1. Calculating 'positive_percentage'
1. Combine rating percentage and total rating number: 'SteamDB_rating', by referring to Steam Database's rating algorithm
1. Define positively rated games as those 'SteamDB_rating' > Q3 quartile of overall 'SteamDB_rating', which will be 'tier'=1; Otherwise 'tier'=0;

	release_date	platforms	categories		genres	steamspy_tags	positive_ratings	negative_ratings	average_playtime	median_playtime	
0	2000-11-01	windows,mac,linux			Multi-player,Online Multi-Player,Local Multi-Player,Valve Anti-Cheat enabled	Action	Action,FPS,Multiplayer	124534	3339	17612	317
1	1999-04-01	windows,mac,linux			Multi-player,Online Multi-Player,Local Multi-Player,Valve Anti-Cheat enabled	Action	Action,FPS,Multiplayer	3318	633	277	62
2	2003-05-01	windows,mac,linux			Multi-player,Valve Anti-Cheat enabled	Action	FPS,World War II,Multiplayer	3416	398	187	34
3	2001-06-01	windows,mac,linux			Multi-player,Online Multi-Player,Local Multi-Player,Valve Anti-Cheat enabled	Action	Action,FPS,Multiplayer	1273	267	258	184
4	1999-11-01	windows,mac,linux			Single-player,Multi-player,Valve Anti-Cheat enabled	Action	FPS>Action,Sci-fi	5250	288	624	415
5	2000-11-01	windows,mac,linux			Multi-player,Online Multi-Player,Valve Anti-Cheat enabled	Action	Action,FPS,Multiplayer	2758	684	175	10
6	1998-11-08	windows,mac,linux			Single-player,Multi-player,Online Multi-Player,Steam Cloud,Valve Anti-Cheat enabled	Action	FPS,Classic>Action	27755	1100	1300	83
7	2004-03-01	windows,mac,linux			Single-player,Multi-player,Valve Anti-Cheat enabled	Action	Action,FPS,Multiplayer	12120	1439	427	43
8	2001-06-01	windows,mac,linux			Single-player	Action	FPS>Action,Sci-fi	3822	420	361	205
9	2004-11-16	windows,mac,linux			Single-player,Steam Achievements,Steam Trading Cards,Captions available,Partial Controller Support,Steam Cloud,Includes Source SDK	Action	FPS>Action,Sci-fi	67902	2419	691	402
10	2004-11-01	windows,mac,linux			Multi-player,Cross-Platform Multiplayer,Steam Achievements,Steam Cloud,Valve Anti-Cheat enabled,Stats,Includes Source SDK	Action	Action,FPS,Multiplayer	76640	3497	6842	400
11	2004-06-01	windows,mac,linux			Single-player	Action	FPS>Action,Sci-fi	3767	1053	190	214
12	2010-07-12	windows,mac,linux			Multi-player,Cross-Platform Multiplayer,Steam Achievements,Steam Trading Cards,Valve Anti-Cheat enabled,Stats,Includes Source SDK	Action	FPS,World War II,Multiplayer	10489	1210	1356	134
13	2004-11-01	windows,mac,linux			Multi-player,Valve Anti-Cheat enabled,Includes Source SDK	Action	Action,FPS,Multiplayer	6020	787	311	32
14	2005-10-27	windows,mac,linux			Single-player,Commentary available	Action	FPS>Action,Singleplayer	5783	1020	46	29
15	2006-05-01	windows,mac,linux			Multi-player,Valve Anti-Cheat enabled	Action	Action,FPS,Multiplayer	1362	473	102	81
16	2006-06-01	windows,mac,linux			Single-player,Steam Achievements,Captions available,Partial Controller Support,Steam Cloud,Stats,Includes Source SDK,Commentary available	Action	FPS>Action,Sci-fi	7908	517	281	184
17	2007-10-10	windows,mac,linux			Single-player,Steam Achievements,Captions available,Partial Controller Support,Includes level editor,Includes Source SDK,Commentary available	Action	Puzzle,First-Person,Singleplayer	51801	1080	288	137
18	2007-10-10	windows,mac,linux			Single-player,Steam Achievements,Captions available,Partial Controller Support,Steam Cloud,Stats,Includes Source SDK,Commentary available	Action	FPS>Action,Sci-fi	13902	696	354	301
19	2007-10-10	windows,mac,linux			Multi-player,Cross-Platform Multiplayer,Steam Achievements,Steam Trading Cards,Captions available,Steam Workshop,In-App Purchases,Partial Controller Support,Valve Anti-Cheat enabled,Stats,Includes level editor,Commentary available	Action,Free to Play	Free to Play,Multiplayer,FPS	515879	34036	8495	623

Before
and after
cleaning

1.

'release_date'
2.

Normalizing
3.

Remove delimiters
4.

'positive_percentage'
5.

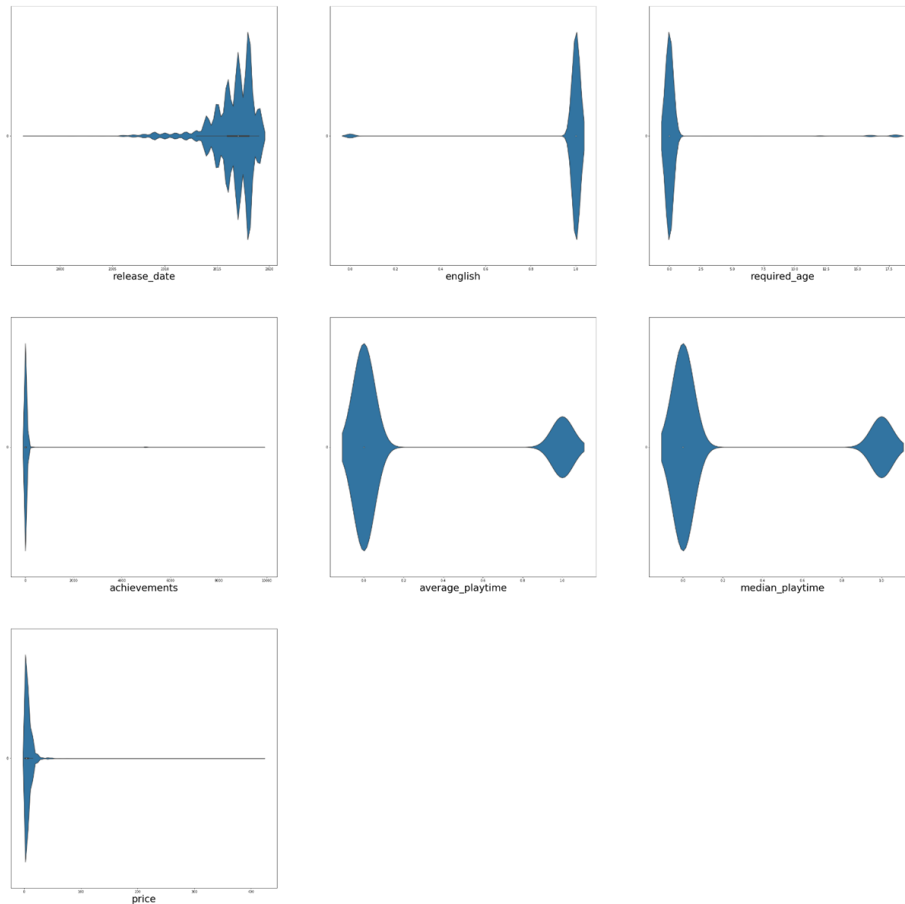
'steamDB_rating'
6.

'tier'

	release_date	platforms	categories		genres	steamspy_tags	positive_ratings	negative_ratings	average_playtime	median_playtime	positive_percentage	steamdb_rating	ser
0	2000	windows,mac,linux			[Multi-player,Online Multi-Player,Local Multi-Player,Valve Anti-Cheat enabled]	[Action,FPS,Multiplayer]	124534	3339	1.000000	1.000000	0.973688	96.013967	1
1	1999	windows,mac,linux			[Multi-player,Online Multi-Player,Local Multi-Player,Valve Anti-Cheat enabled]	[Action,FPS,Multiplayer]	3318	633	1.000000	1.000000	0.839787	81.170349	1
2	2003	windows,mac,linux			[Multi-player,Valve Anti-Cheat enabled]	[Action,FPS,World War II,Multiplayer]	3416	398	1.000000	1.000000	0.895648	86.259763	1
3	2001	windows,mac,linux			[Multi-player,Online Multi-Player,Local Multi-Player,Valve Anti-Cheat enabled]	[Action,FPS,Multiplayer]	1273	267	1.000000	1.000000	0.826623	79.672853	1
4	1999	windows,mac,linux			[Single-player,Multi-player,Valve Anti-Cheat enabled]	[Action,FPS>Action,Sci-fi]	5250	288	1.000000	1.000000	0.947996	91.404024	1
5	2000	windows,mac,linux			[Multi-player,Online Multi-Player,Valve Anti-Cheat enabled]	[Action,FPS,Multiplayer]	2758	684	1.000000	1.000000	0.801278	77.532197	1
6	1998	windows,mac,linux			[Single-player,Multi-player,Online Multi-Player,Steam Cloud,Valve Anti-Cheat enabled]	[Action,FPS,Classic>Action]	27755	1100	1.000000	1.000000	0.961878	94.069559	1
7	2004	windows,mac,linux			[Single-player,Multi-player,Valve Anti-Cheat enabled]	[Action,FPS,Multiplayer]	12120	1439	1.000000	1.000000	0.893871	87.141068	1
8	2001	windows,mac,linux			[Single-player]	[Action,FPS>Action,Sci-fi]	3822	420	1.000000	1.000000	0.900990	86.854902	1
9	2004	windows,mac,linux			[Single-player,Steam Achievements,Steam Trading Cards,Captions available,Partial Controller Support,Steam Cloud,Includes Source SDK]	[Action,FPS>Action,Sci-fi]	67902	2419	1.000000	1.000000	0.963601	94.942376	1
10	2004	windows,mac,linux			[Multi-player,Cross-Platform Multiplayer,Steam Achievements,Steam Cloud,Valve Anti-Cheat enabled,Stats,Includes Source SDK]	[Action,FPS,Multiplayer]	76640	3497	1.000000	1.000000	0.966360	94.111794	1
11	2004	windows,mac,linux			[Single-player]	[Action,FPS>Action,Sci-fi]	3767	1053	1.000000	1.000000	0.781535	75.961741	1
12	2010	windows,mac,linux			[Multi-player,Cross-Platform Multiplayer,Steam Achievements,Steam Trading Cards,Valve Anti-Cheat enabled,Stats,Includes Source SDK]	[Action,FPS,World War II,Multiplayer]	10489	1210	1.000000	1.000000	0.896072	87.293078	1
13	2004	windows,mac,linux			[Multi-player,Valve Anti-Cheat enabled,Includes Source SDK]	[Action,FPS,Multiplayer]	6020	787	1.000000	1.000000	0.894384	85.741185	1
14	2005	windows,mac,linux			[Single-player,Commentary available]	[Action,FPS>Action,Singleplayer]	5783	1020	1.000000	1.000000	0.850056	82.549795	1
15	2006	windows,mac,linux			[Multi-player,Valve Anti-Cheat enabled]	[Action,FPS,Multiplayer]	1362	473	1.000000	1.000000	0.742324	71.701625	0
16	2006	windows,mac,linux			[Single-player,Steam Achievements,Captions available,Partial Controller Support,Steam Cloud,Stats,Includes Source SDK,Commentary available]	[Action,FPS>Action,Sci-fi]	7908	517	1.000000	1.000000	0.938838	90.391869	1
17	2007	windows,mac,linux			[Single-player,Steam Achievements,Captions available,Partial Controller Support,Includes level editor,Includes Source SDK,Commentary available]	[Action,Puzzle,First-Person,Singleplayer]	51801	1080	1.000000	1.000000	0.979577	95.142160	1
18	2007	windows,mac,linux			[Single-player,Steam Achievements,Captions available,Partial Controller Support,Steam Cloud,Stats,Includes Source SDK,Commentary available]	[Action,FPS>Action,Sci-fi]	13902	696	1.000000	1.000000	0.920252	92.709546	1
19	2007	windows,mac,linux			[Multi-player,Cross-Platform Multiplayer,Steam Achievements,Steam Trading Cards,Captions available,Steam Workshop,In-App Purchases,Partial Controller Support,Valve Anti-Cheat enabled,Stats,Includes level editor,Commentary available]	[Action,Free to Play,Free to Play,Multiplayer,FPS]	515879	34036	1.000000	1.000000	0.938107	92.991214	1

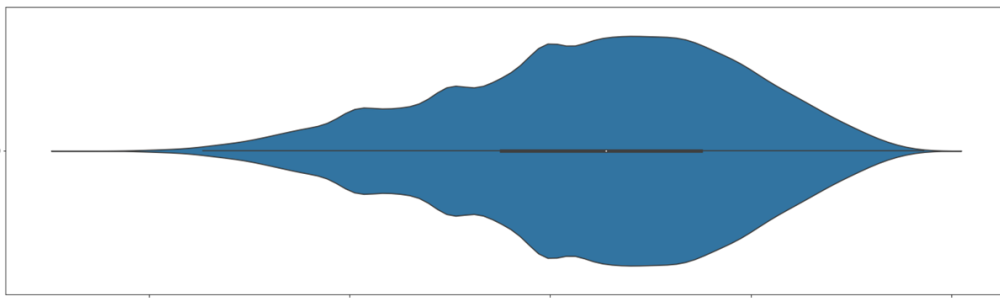
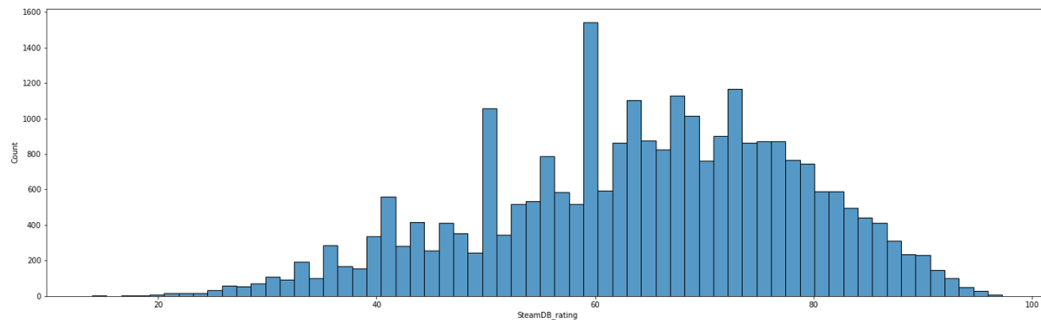
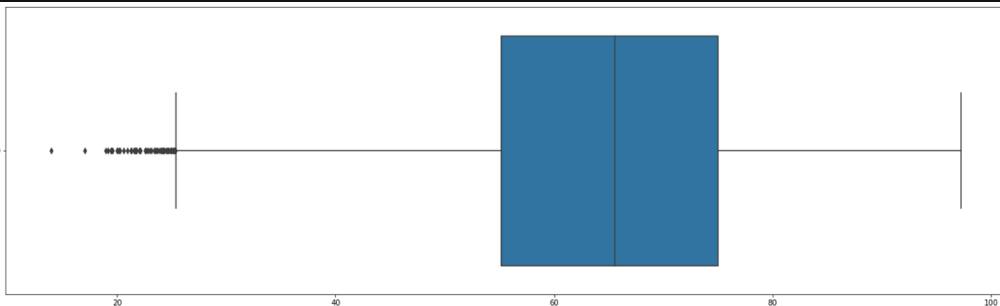
Univariate Exploratory Data Analysis

On numerical data



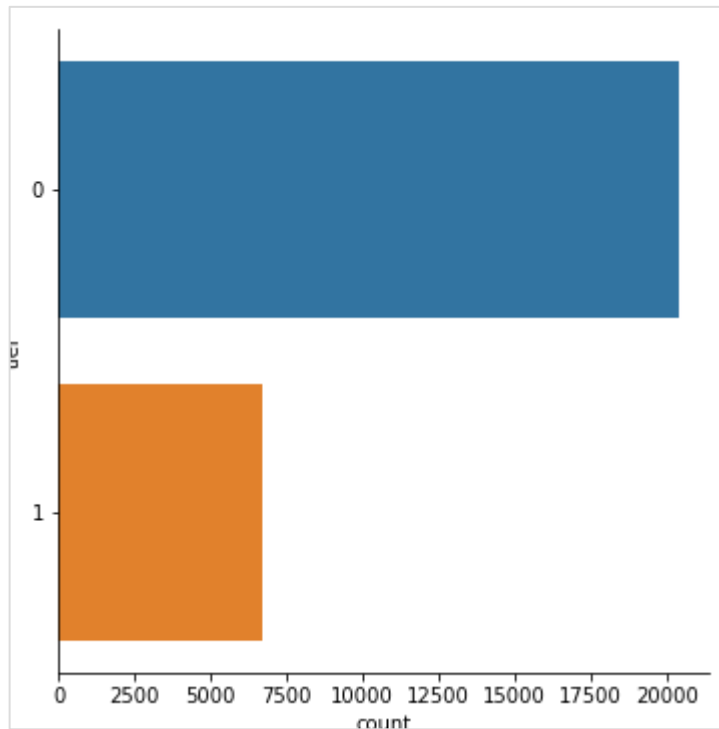
Univariate Exploratory Data Analysis

‘SteamDB rating’



Univariate Exploratory Data Analysis

'tier'



0	20369
1	6706

Univariate Exploratory Data Analysis

word cloud for
unique genres

#

indie, action and casual games
have the largest frequency

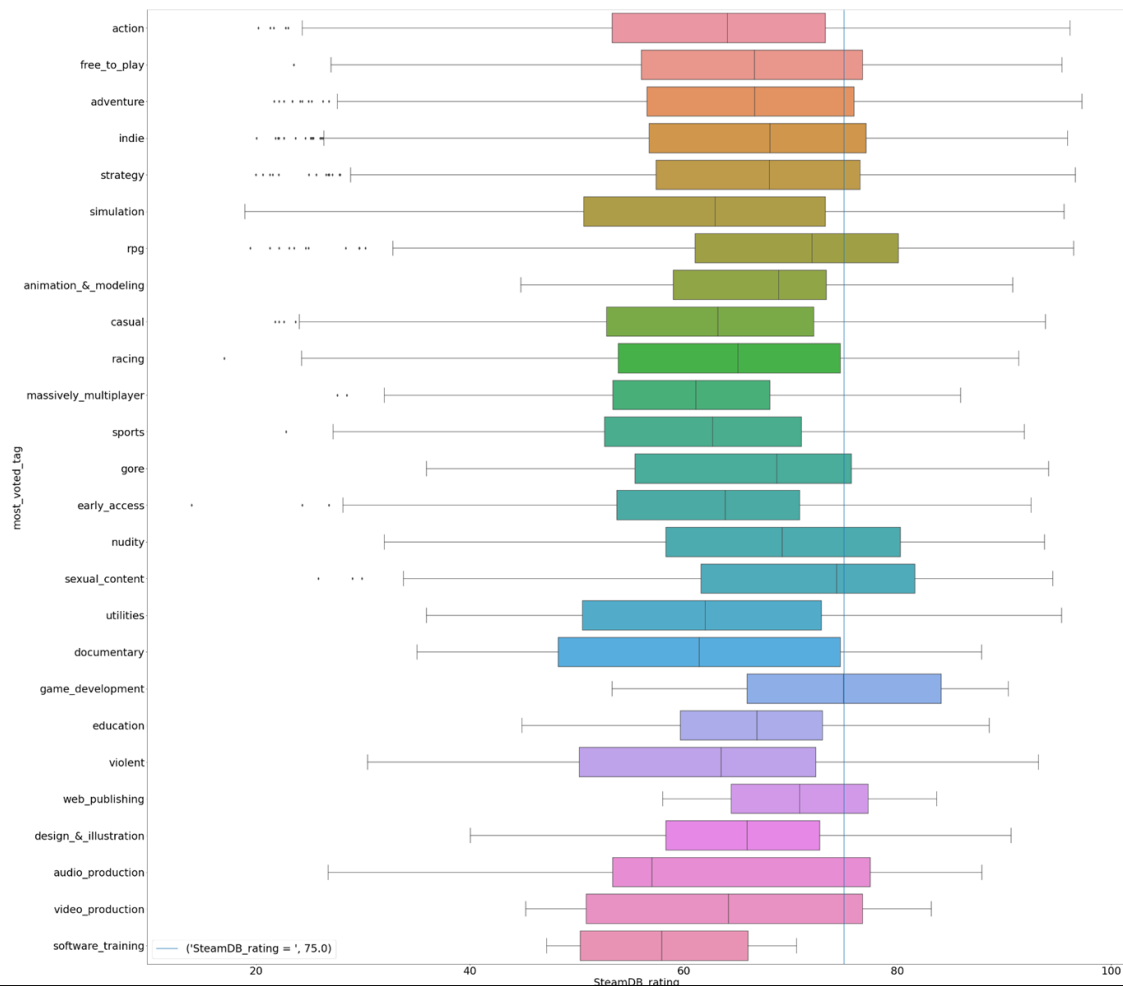


Bivariate Exploratory Data Analysis

'most_voted_tag'
VS
'SteamDB_rating'

#

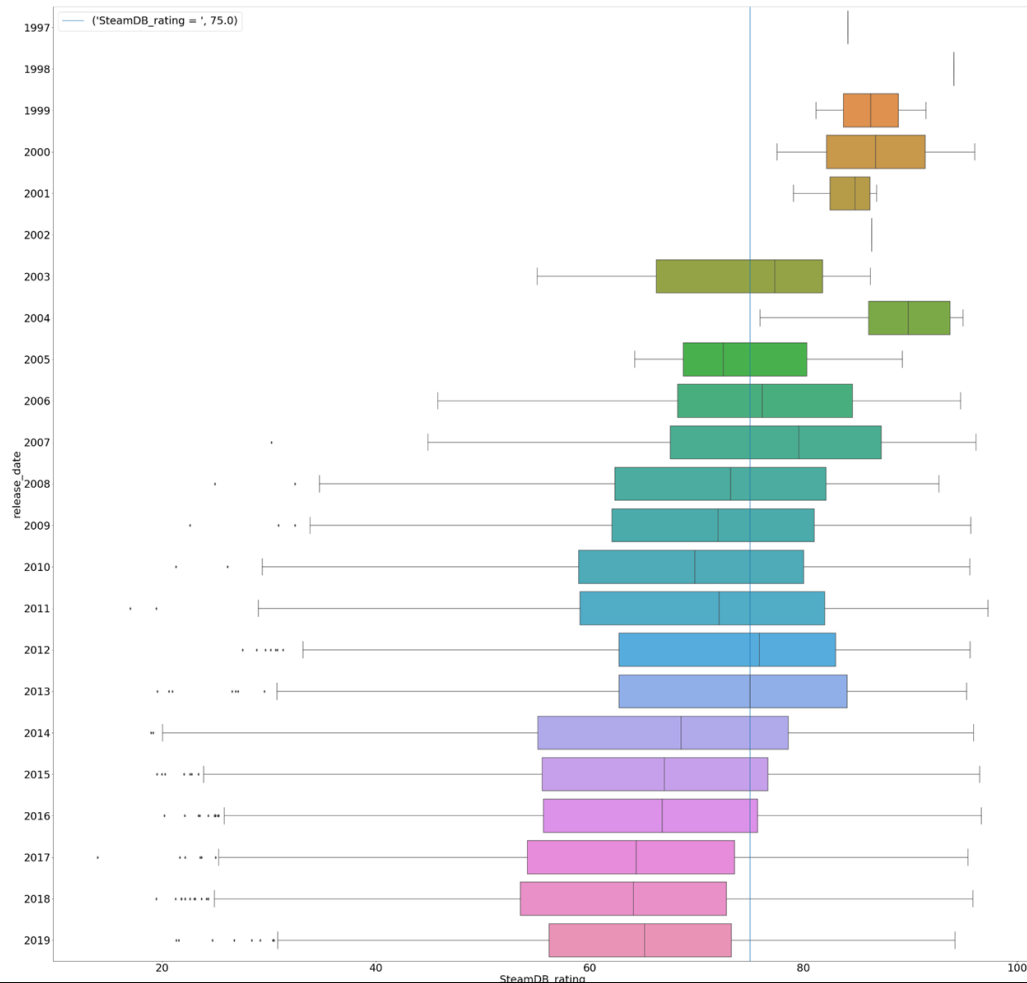
Lowest average rating:
audio_production
Highest average rating:
rpg, game_development
and sexual_content



Bivariate Exploratory Data Analysis

'release_date'
VS
'SteamDB_rating'

The earlier the year of release,
the higher the rating



Bivariate Exploratory Data Analysis

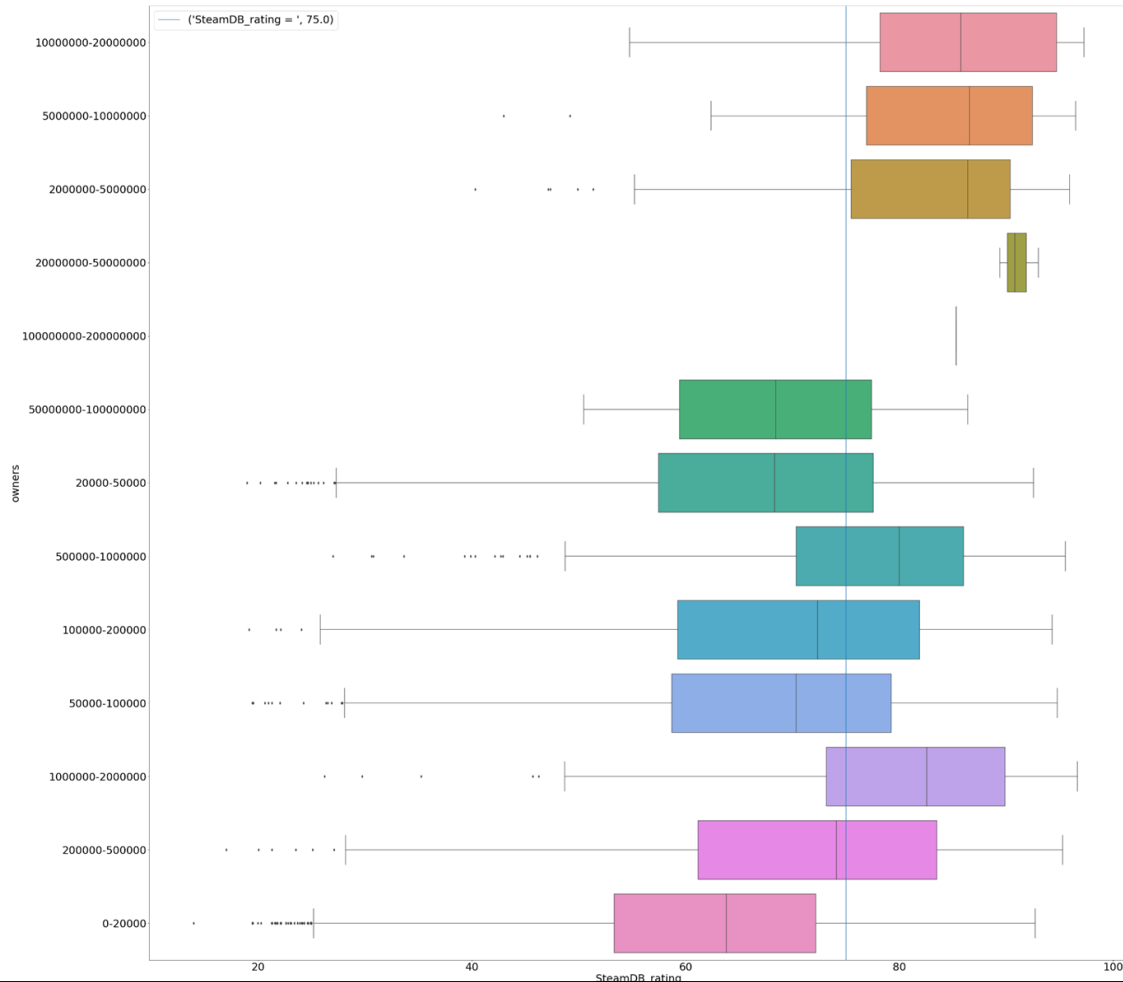
'owners'

VS

'SteamDB_rating'

#

The more the owners,
the higher the rating



Bivariate Exploratory Data Analysis

'developers'
vs
'SteamDB_rating'

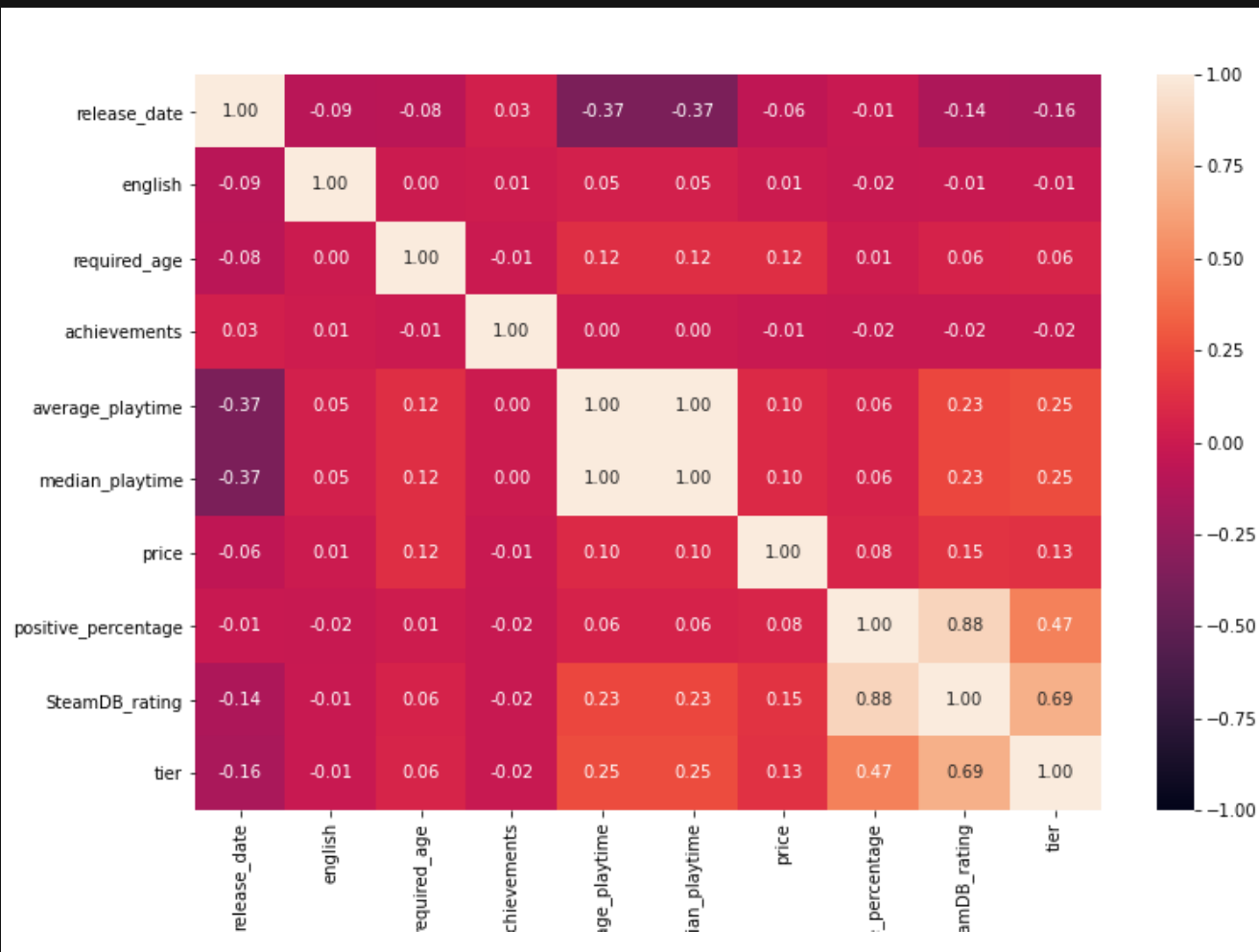
top 10 rated developers

SteamDB_rating	
publisher	
Wube Software LTD.	96.630243
Ludeon Studios	95.847566
ConcernedApe	95.535628
Team Salvato	95.387628
Kristjan Skutta	95.361847
Gears for Breakfast	95.268788
Square Enix;Feral interactive (Mac);Feral Interactive (Linux)	94.988123
Ysbryd Games;AGM PLAYISM	94.897869
Holy Priest	94.844861
Steel Crate Games	94.511332

Multivariate Exploratory Data Analysis

On
numerical
data

'release_date' has a relatively
strong correlation with
playtime;
while playtime
has a relatively strong
correlation with tier and
'SteamDB_rating':



2/

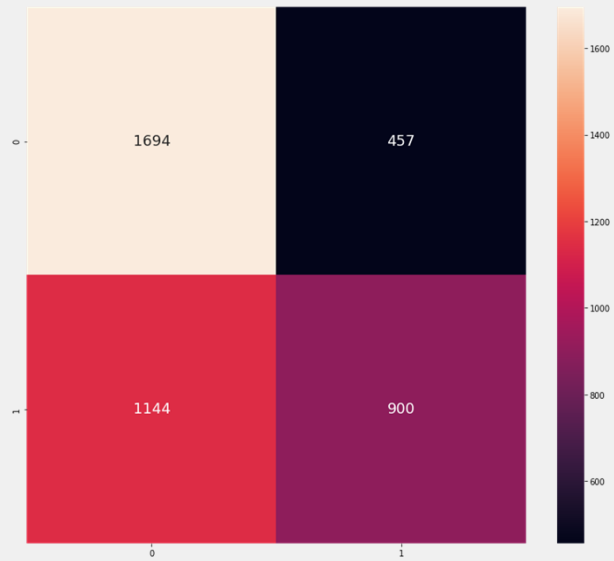
Machine learning



Data Extraction + Encoding data

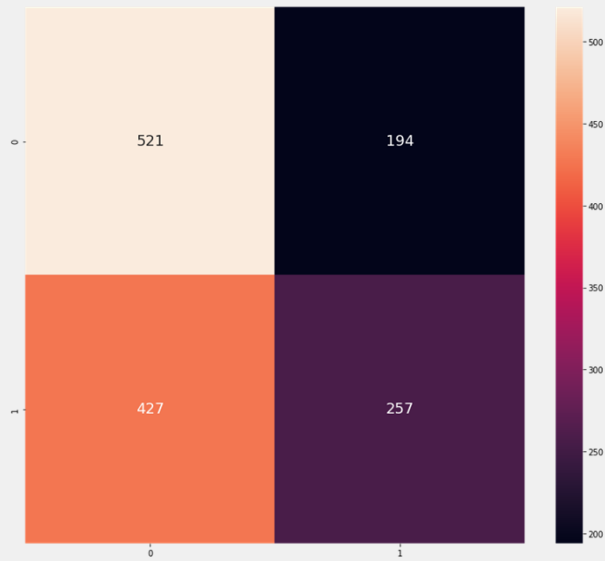
1. Extraction
 - a. Rows with >50 ratings
 - b. >0 Average playtime
1. Label Encoding
 - a. developer
 - b. Most_voted_tag
 - Combine tags with <50 games into “others”
1. Models used
 - a. Decision Tree
 - b. Random Forest
 - c. Naive Bayes

Decision Tree (median_playtime - tier)



Train data

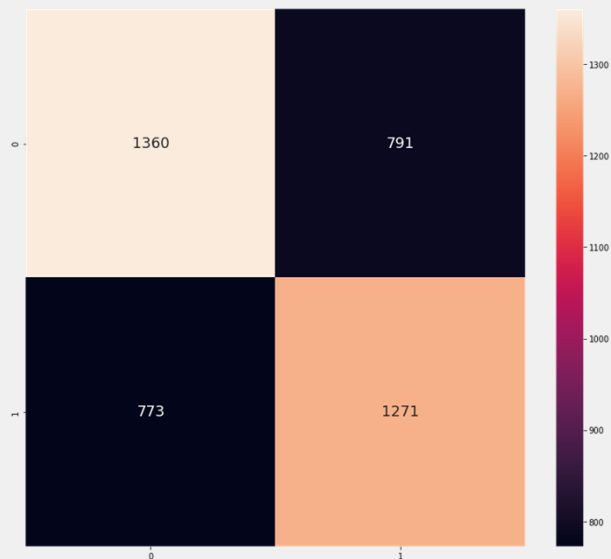
Classification accuracy: 0.6184
False negative rate: 0.5597
False positive rate: 0.2125



Test data

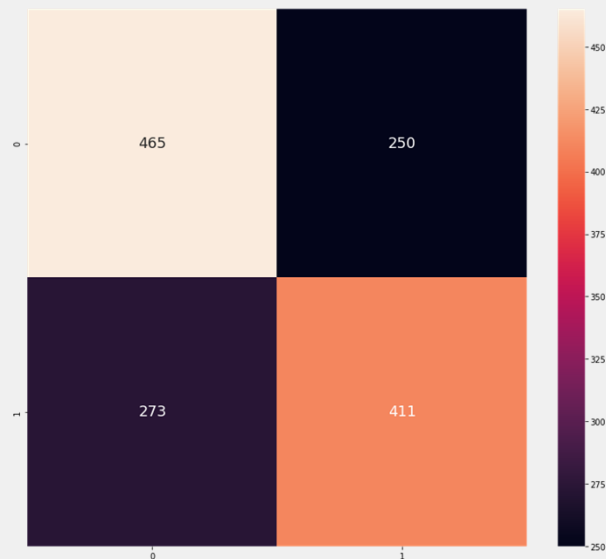
Classification accuracy: 0.5561
False negative rate: 0.6243
False positive rate: 0.2713

Decision Tree (price-tier)



Train data

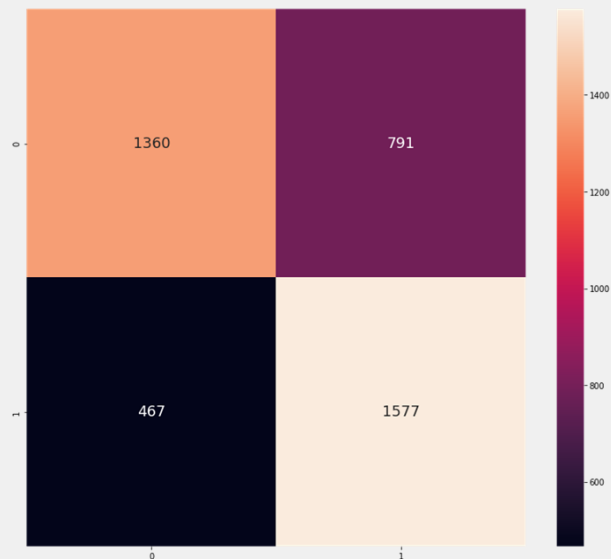
Classification accuracy: 0.6272
False negative rate: 0.3782
False positive rate: 0.3677



Test data

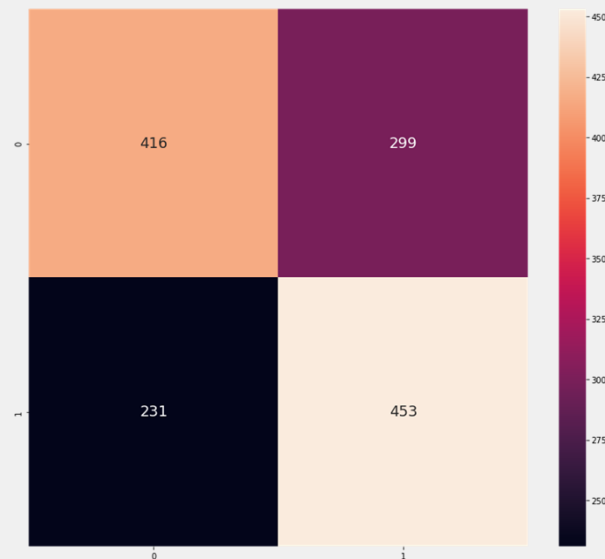
Classification accuracy: 0.6262
False negative rate: 0.3991
False positive rate: 0.3497

Decision Tree (multivariate)



Train data

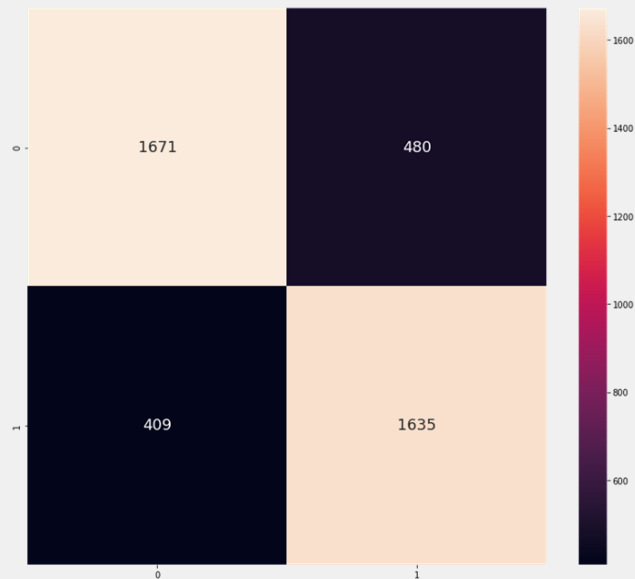
Classification accuracy: 0.7001
False negative rate: 0.2285
False positive rate: 0.3677



Test data

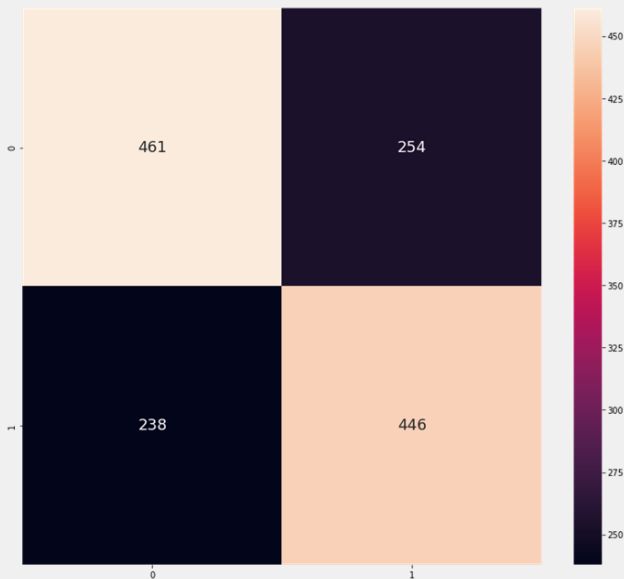
Classification accuracy: 0.6212
False negative rate: 0.3377
False positive rate: 0.4182

Random forest



Train data

Classification accuracy: 0.7881
False negative rate: 0.2001
False positive rate: 0.2232



Test data

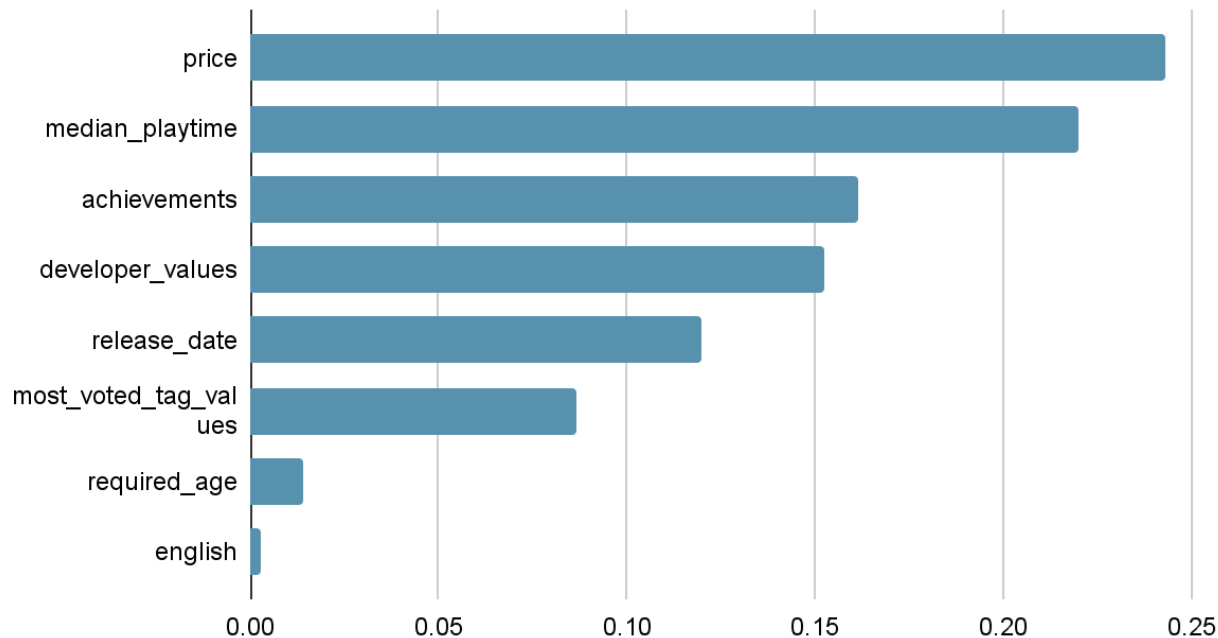
Classification accuracy: 0.6419
False negative rate: 0.3480
False positive rate: 0.3552

Random forest Feature importance

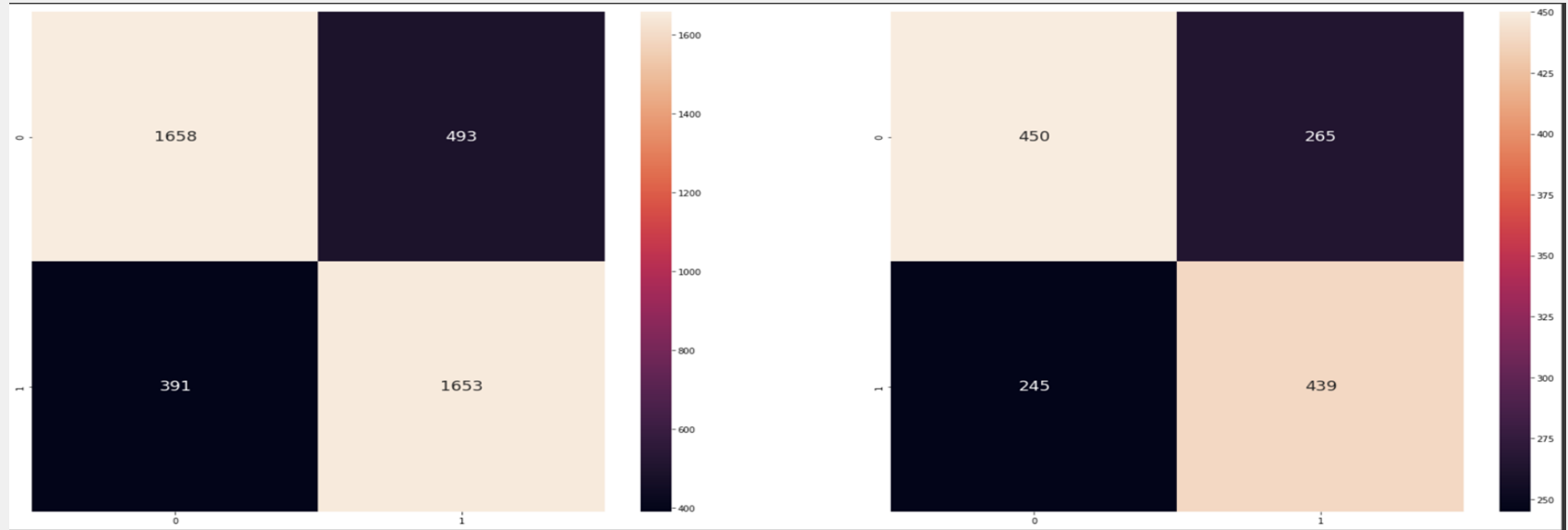
Price & median
playtime are most
important

Required age and
english are
insignificant

Feature importance



Naive Bayes (median_playtime-tier)



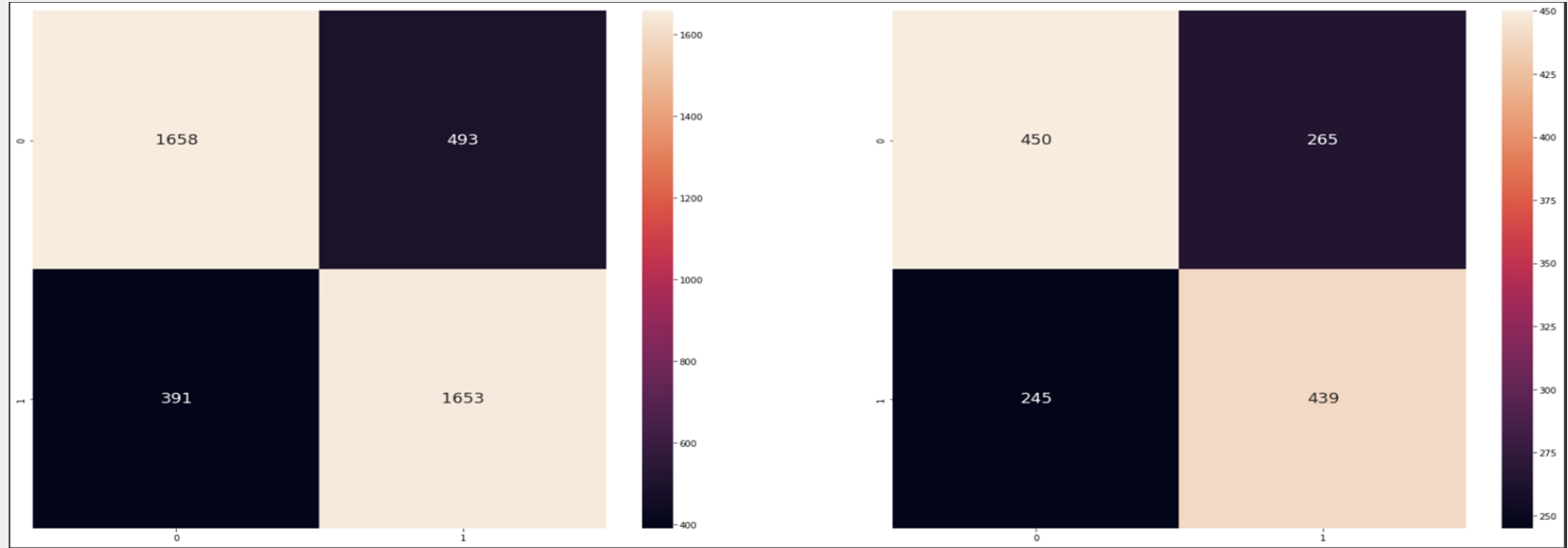
Train data

Classification accuracy: 0.5134
False negative rate: 0.2642
False positive rate: 0.2924

Test data

Classification accuracy: 0.5067
False negative rate: 0.3404
False positive rate: 0.3602

Naive Bayes (price-tier)



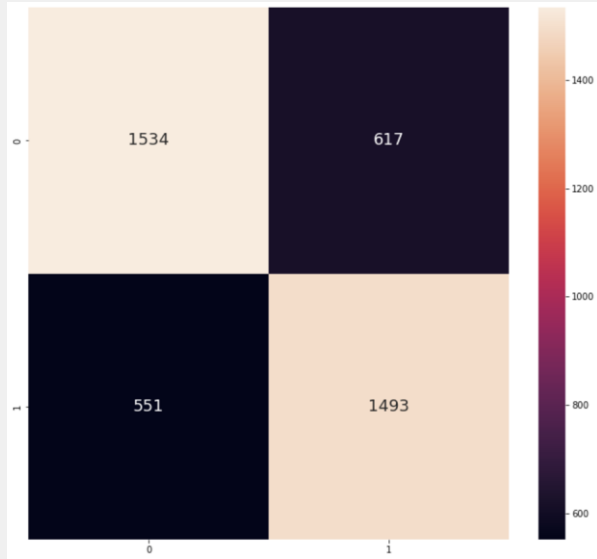
Train data

Classification accuracy: 0.5842
False negative rate: 0.1908
False positive rate: 0.2297

Test data

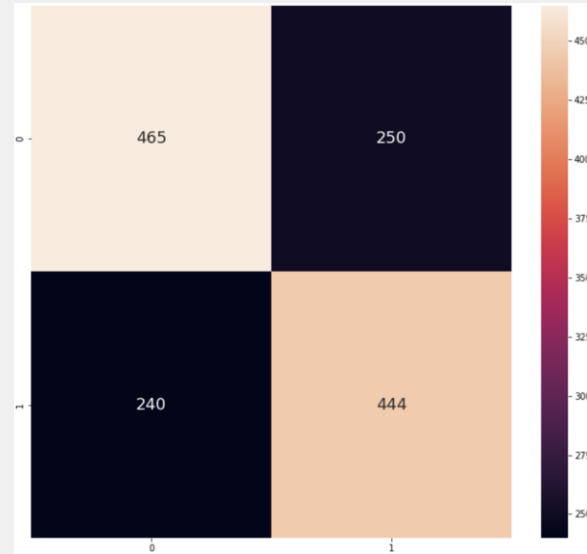
Classification accuracy: 0.5761
False negative rate: 0.3404
False positive rate: 0.3602

Naive Bayes (multivariate)



Train data

Classification accuracy: 0.6021
False negative rate: 0.2642
False positive rate: 0.2924



Test data

Classification accuracy: 0.5983
False negative rate: 0.3404
False positive rate: 0.3602

3/

Sentiment Analysis (NLP)



Data Extraction + Data Preprocessing

1. Extraction
 - a. Reviews for Steam games
 - b. Filter reviews for “Portal 2” only
1. Data Preprocessing
 - Remove Stopwords



Exploring positive words used in reviews for “Portal 2”

1. Drop empty rows of reviews
1. Extract the top 30 words used in the reviews as a whole excluding common english words
1. Append the number of these top positive words found in each review into “positive_review_count” column
1. Created a column for total word count
1. Calculated percentage of positive review words (%) for each review
1. Used VADER (Valence Aware Dictionary and Sentiment Reasoner) to determine the sentiment value in each review

Data Preprocessing

Removing common English words

A.k.a stopwords

Used PorterStemmer module to
Display only the root words in the reviews

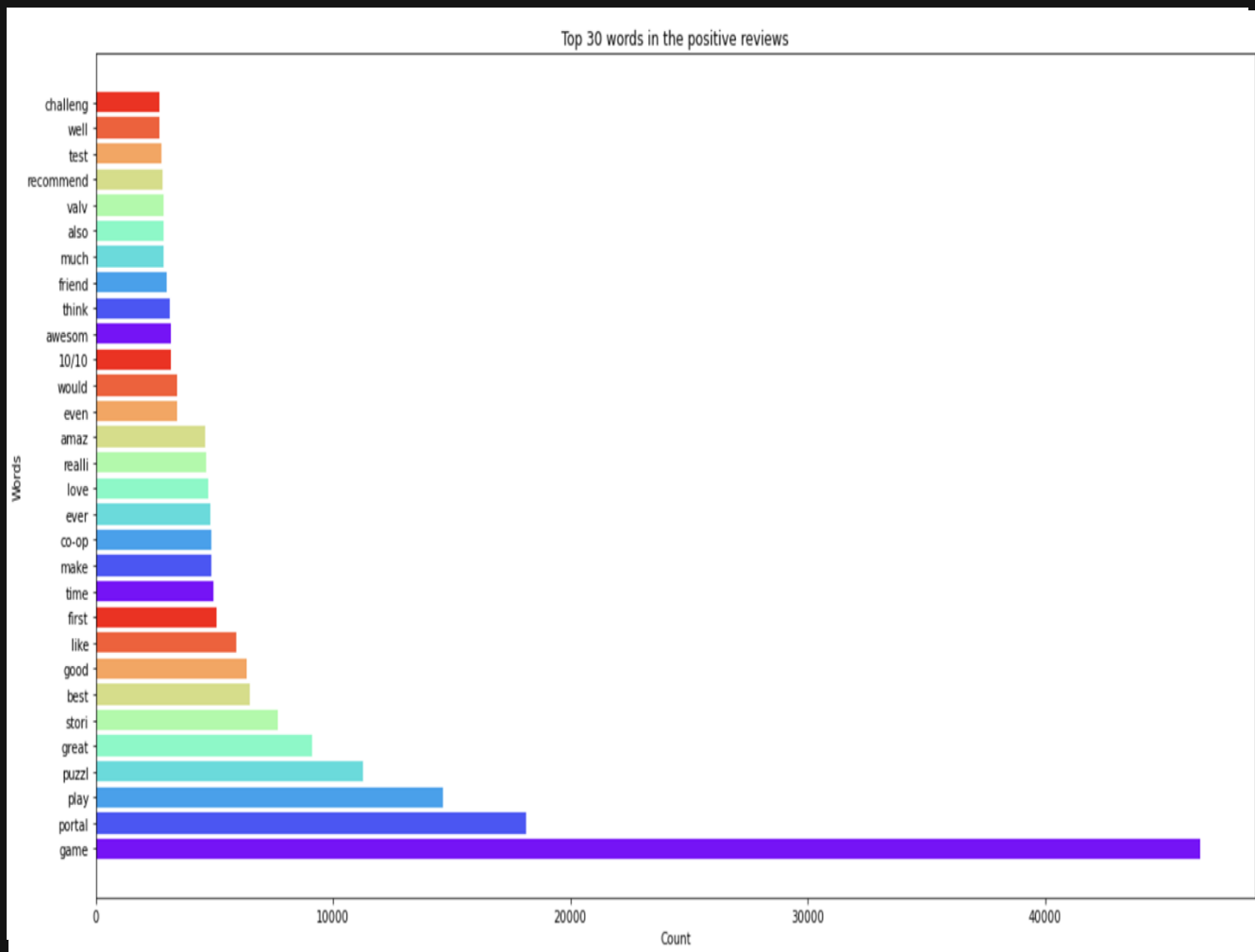
processed_review

best game ever made valv great stori beauti di...
step game step play game step smart step beat ...
pleas valv learn count past
much experi game littl time spend play great g...
brilliant game ever made gameplay humor charac...
...
game sure life
great game play friend mess though singl playe...
veri co-op play friend
great stori love play friend
game call load screen serious wrong load time ...

Top 30 words found in reviews for Portal 2

#

Excluding common English words
Like articles.



Percentage positive word count

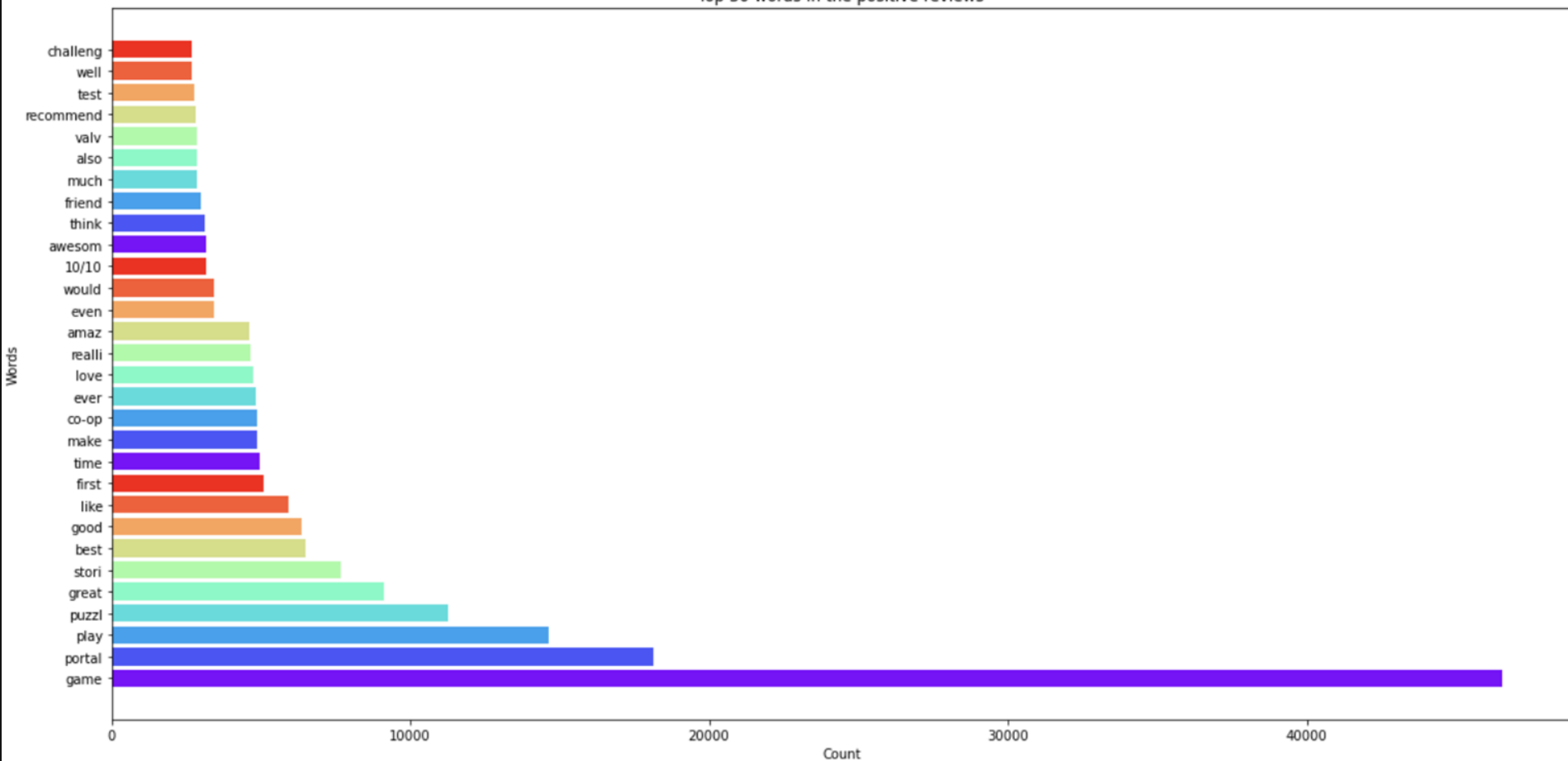
processed_review	positive_review_count	word_count	percentage of positive review words(%)
best game ever made valv great stori beauti di...	12	41	29.268293
step game step play game step smart step beat ...	8	41	19.512195
pleas valv learn count past	1	7	14.285714
much experi game littl time spend play great g...	7	29	24.137931
brilliant game ever made gameplay humor charac...	6	42	14.285714

Detecting polarity within texts using VADER (Valence Aware Dictionary and Sentiment Reasoner)

processed_review	positive_review_count	word_count	percentage of positive review words(%)	Review Polarity Score
best game ever made valv great stori beauti di...	12	41	29.268293	0.9799
step game step play game step smart step beat ...	8	41	19.512195	0.7579
pleas valv learn count past	1	7	14.285714	0.0000
much experi game littl time spend play great g...	7	29	24.137931	0.7579
brilliant game ever made gameplay humor charac...	6	42	14.285714	0.8225

	Unnamed: 0	app_id	app_name	review_text	review_score	review_votes	processed_review
0	6125847	620	Portal 2	The best game ever made by Valve. Great story,...	1	1	best game ever made valv great stori beauti di...
1	6125848	620	Portal 2	Step 1: Get game Step 2: Play game Step 3: Be ...	1	0	step game step play game step smart step beat ...
2	6125849	620	Portal 2	Please Valve, Learn to count past 2	1	0	pleas valv learn count past
3	6125850	620	Portal 2	Don't have much experience in this game, but w...	1	0	much experi game littl time spend play great g...
4	6125851	620	Portal 2	One of the most brilliant games ever made. The...	1	0	brilliant game ever made gameplay humor charac...
...
38919	6164766	620	Portal 2	If you don't have this game I'm not sure what ...	1	0	game sure life
38920	6164767	620	Portal 2	great game to play with friends, and to mess w...	1	0	great game play friend mess though singl playe...
38921	6164768	620	Portal 2	Very fun co-op to play with friends	1	0	veri co-op play friend
38922	6164769	620	Portal 2	Great story, and i love when i play with friends	1	0	great stori love play friend
38923	6164770	620	Portal 2	This game should be called : Loading Screens 2...	-1	0	game call load screen serious wrong load time ...

Top 30 words in the positive reviews





References:

[Introducing Steam Database's new rating algorithm · SteamDB](#)

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik



Thanks !

