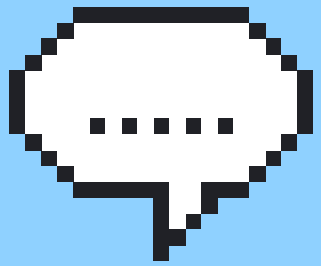
To Be or Not To Be

A Positive-Reviewed Game

By Mihaela Catan

Introduction

If you try and think off the top of your head about the best video games, only a few will come to mind. If you stop and remember all the great games you feel nostalgic about – old or new – you'll find there's a lot more than you first thought!



The Dataset

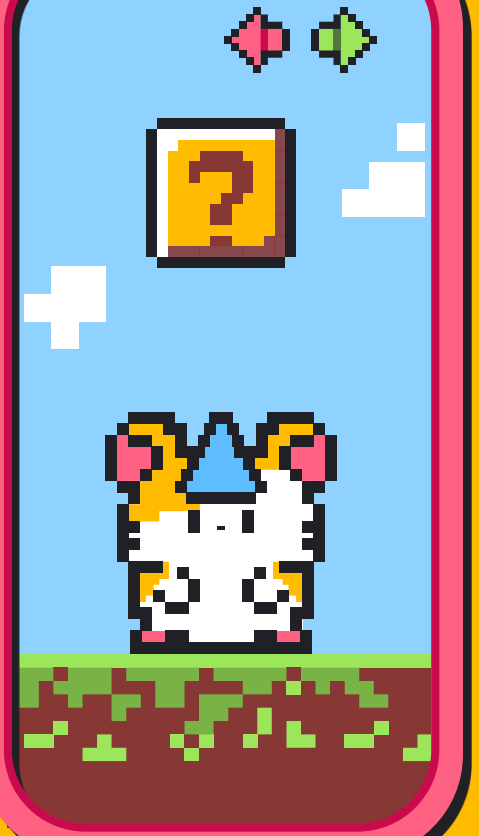
Games						
app_id	title	release_date	genres	categories	developer	publisher
publisher	original_price	discouted_price	dlc-available	age_rating	content_descriptor	about_description
win_support	mac_support	linux_support	awards	overall_review	overall_review_count	recent_review
recent_review_%	recent_review_count	overall_review_%				

numerical/bool object

Target

overall_review_%

A continuous data column



Let's Go

Questions are allowed during presentation

In [3]:
```
data = pd.read_csv('/content/drive/MyDrive/Exam_01_07_2024_4/data.csv')
data.head()
```

Out[3]:

	app_id	title	release_date	genres	categories	developer	publisher	original_price	discount_percentage	discour
0	730	Counter-Strike 2	21 Aug, 2012	Action, Free to Play	Cross-Platform Multiplayer, Steam Trading Card...	Valve	Valve	NaN	NaN	
1	570	Dota 2	9 Jul, 2013	Action, Strategy, Free to Play	Steam Trading Cards, Steam Workshop, SteamVR C...	Valve	Valve	NaN	NaN	
2	2215430	Ghost of Tsushima DIRECTOR'S CUT	16 May, 2024	Action, Adventure	Single-player, Online Co-op, Steam Achievement...	Sucker Punch Productions	PlayStation PC LLC	NaN	NaN	
3	1245620	ELDEN RING	24 Feb, 2022	Action, RPG	Single-player, Online PvP, Online Co-op, Steam...	FromSoftware Inc.	FromSoftware Inc.	NaN	NaN	
4	1085660	Destiny 2	1 Oct, 2019	Action, Adventure, Free to Play	Single-player, Online PvP, Online Co-op, Steam...	Bungie	Bungie	NaN	NaN	

5 rows × 24 columns

Note

There were no duplicates, but...

app_id	0	content_descriptor	40122
title	0	about_description	138
release_date	57	win_support	0
genres	87	mac_support	0
categories	45	linux_support	0
developer	190	awards	0
publisher	211	overall_review	2477
original_price	37638	overall_review_%	2477
discount_percentage	37638	overall_review_count	2477
discounted_price	240	recent_review	36994
dlc_available	0	recent_review_%	36994
age_rating	0	recent_review_count	36994

How Did I Get Rid off NaNs?

01 LLM Models

02 Drop Rows

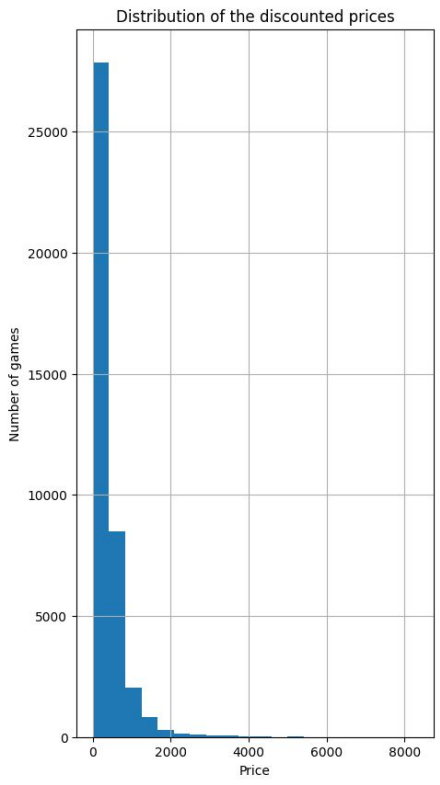
03 Drop Columns

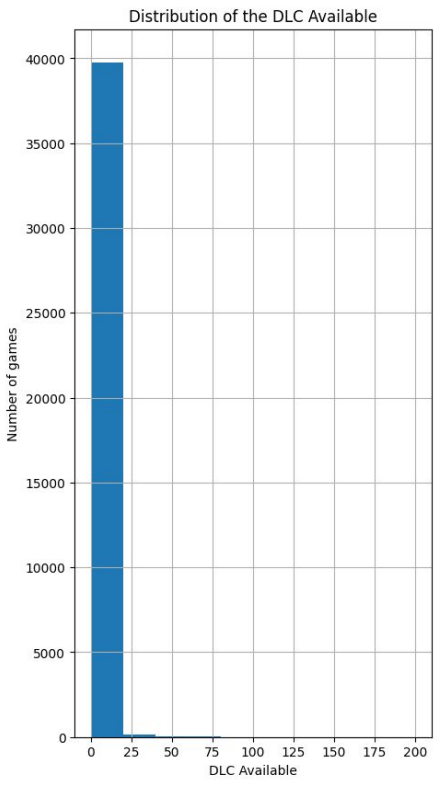
04 Statistics

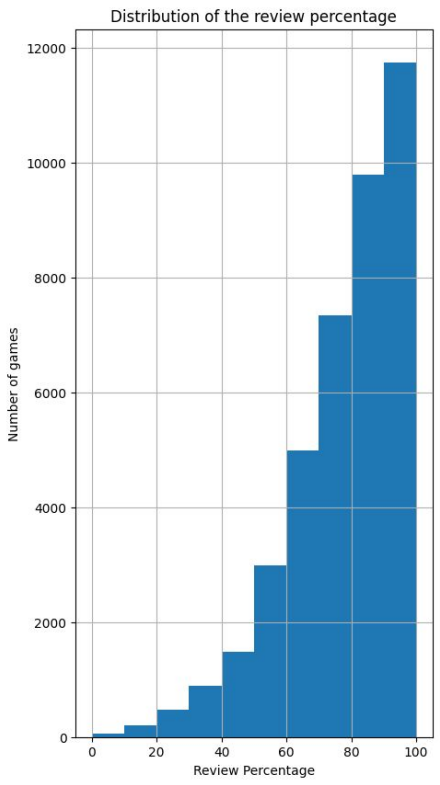
Data at Glance

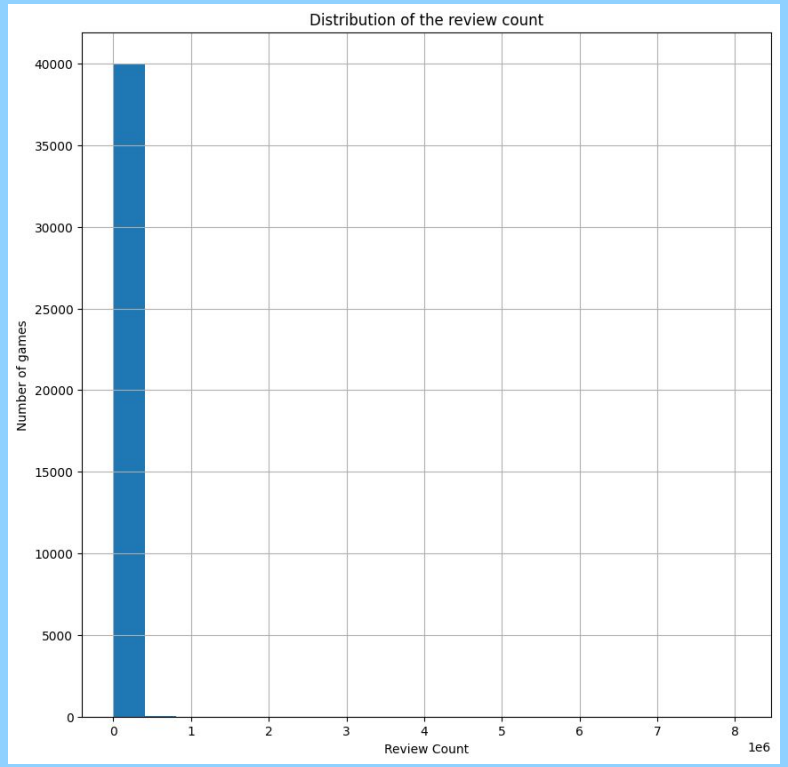
```
data.describe()
```

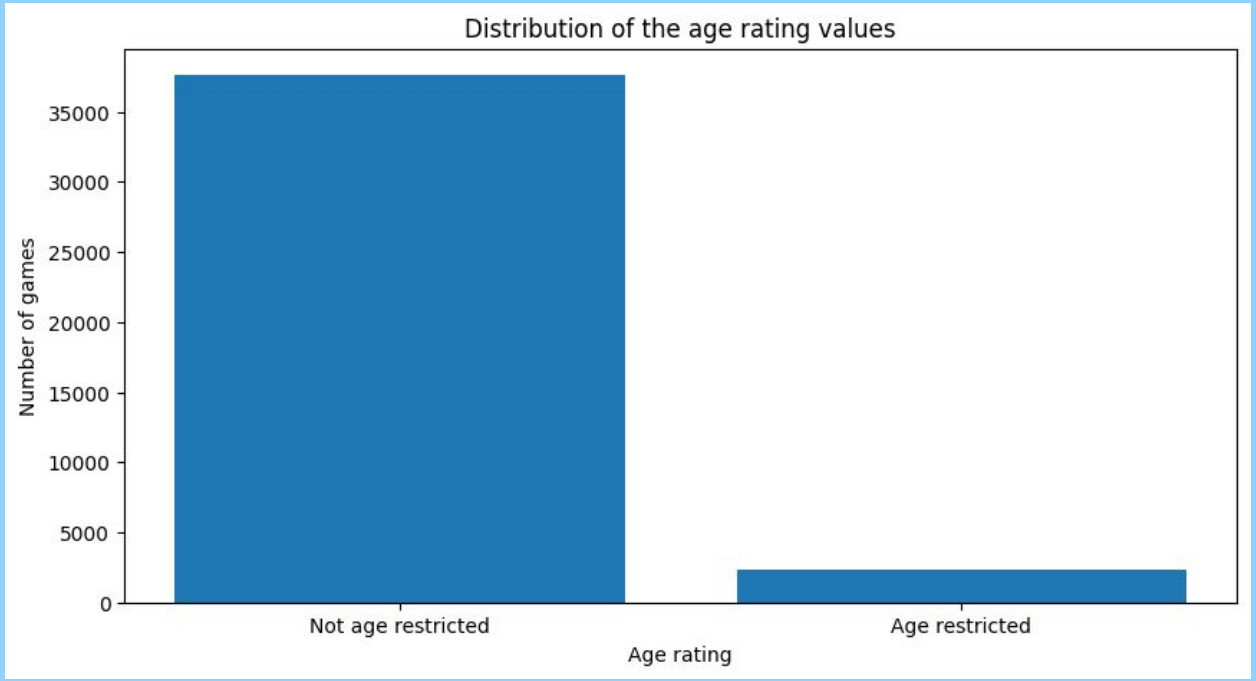
	discounted_price	dlc_available	age_rating	awards	overall_review_%	overall_review_count
count	39952.000000	39952.000000	39952.000000	39952.000000	39952.000000	3.995200e+04
mean	355.262165	0.758385	0.058270	0.325115	77.170730	2.500934e+03
std	450.308165	4.814032	0.234256	1.290559	17.666591	4.892032e+04
min	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000e+01
25%	59.000000	0.000000	0.000000	0.000000	68.000000	2.100000e+01
50%	250.000000	0.000000	0.000000	0.000000	81.000000	5.900000e+01
75%	479.000000	0.000000	0.000000	0.000000	91.000000	2.810000e+02
max	8325.000000	200.000000	1.000000	41.000000	100.000000	8.062218e+06

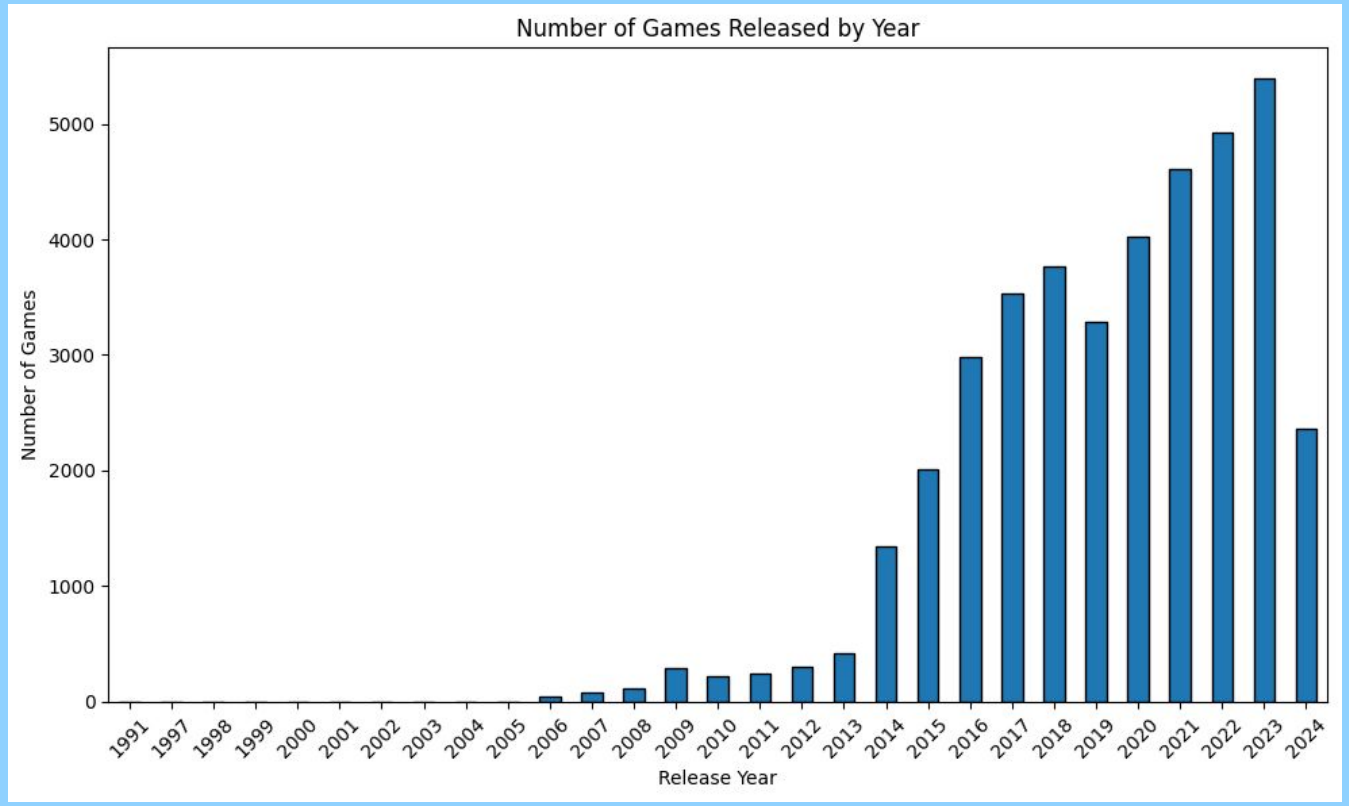

Distribution of the discounted prices
Number of games
Price
0
5000
10000
15000
20000
25000
0
2000
4000
6000
8000

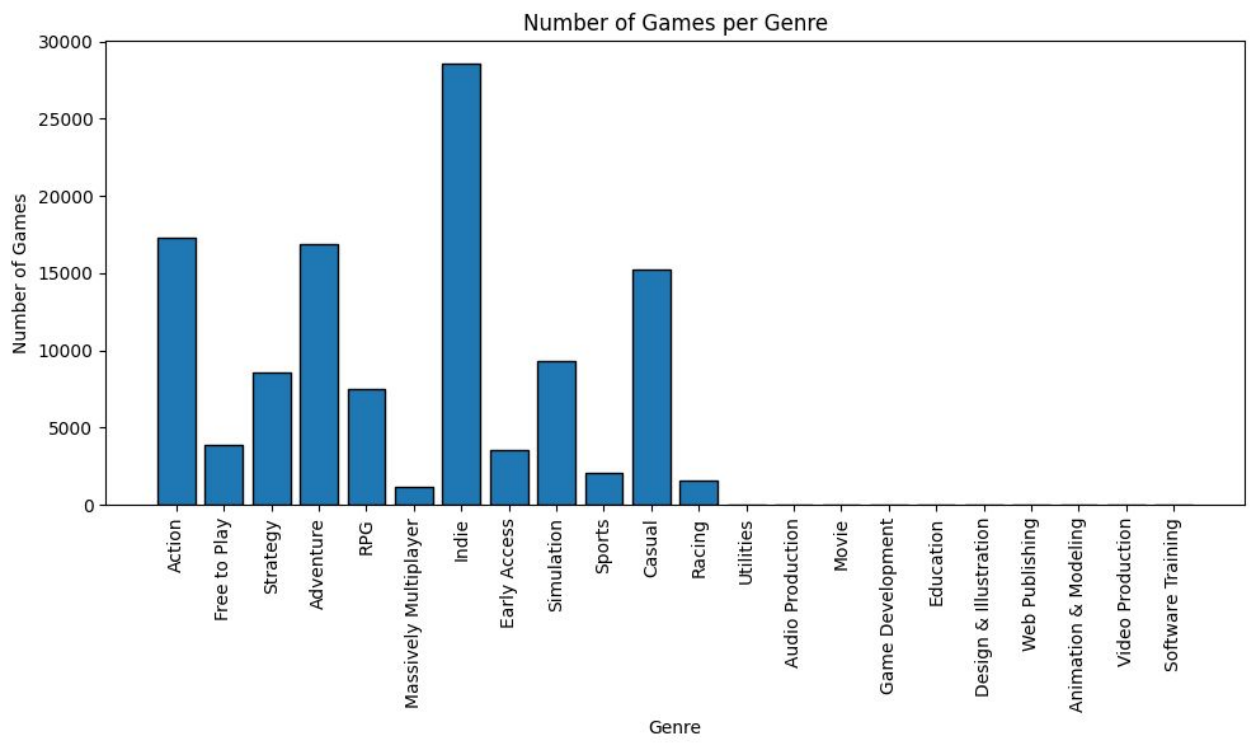

Distribution of the DLC Available
Number of games
DLC Available
0
5000
10000
15000
20000
25000
30000
35000
40000
0
25
50
75
100
125
150
175
200

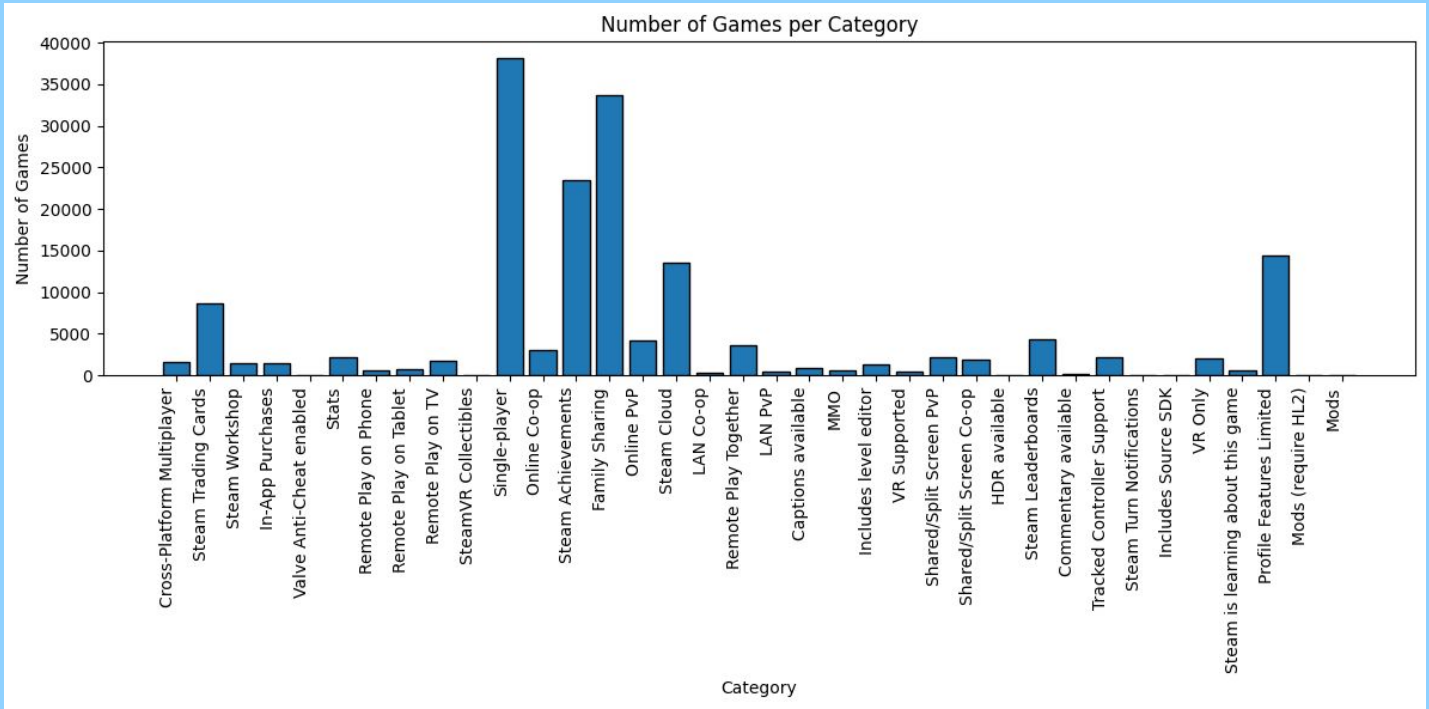

Distribution of the review percentage
Number of games
Review Percentage
0
2000
4000
6000
8000
10000
12000
0
20
40
60
80
100


Distribution of the review count
Number of games
Review Count
1e6
0
5000
10000
15000
20000
25000
30000
35000
40000
0
1
2
3
4
5
6
7
8


Distribution of the age rating values
Number of games
35000
30000
25000
20000
15000
10000
5000
0
Not age restricted
Age restricted
Age rating


Number of Games Released by Year
Number of Games
0
1000
2000
3000
4000
5000
1991
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
Release Year


Number of Games per Genre
Number of Games
30000
25000
20000
15000
10000
5000
0
Action
Free to Play
Strategy
Adventure
RPG
Massively Multiplayer
Indie
Early Access
Simulation
Sports
Casual
Racing
Utilities
Audio Production
Movie
Game Development
Education
Design & Illustration
Web Publishing
Animation & Modeling
Video Production
Software Training
Genre

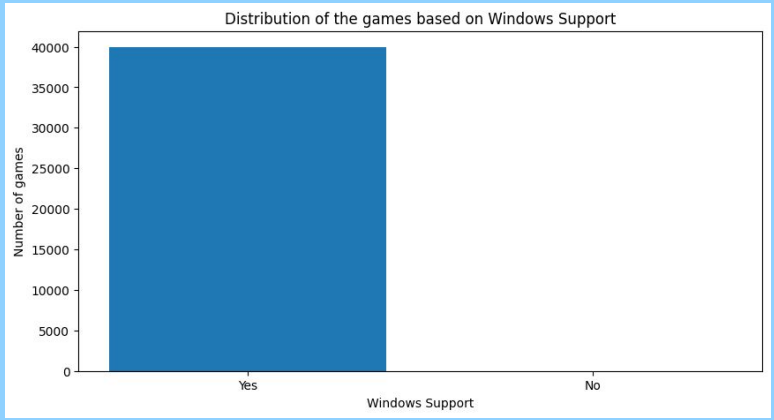

Number of Games per Category
Number of Games
40000
35000
30000
25000
20000
15000
10000
5000
0
Cross-Platform Multiplayer
Steam Trading Cards
Steam Workshop
In-App Purchases
Valve Anti-Cheat enabled
Stats
Remote Play on Phone
Remote Play on Tablet
Remote Play on TV
SteamVR Collectibles
Single-player
Online Co-op
Steam Achievements
Family Sharing
Online PvP
Steam Cloud
LAN Co-op
Remote Play Together
LAN PvP
Captions available
MMO
Includes level editor
VR Supported
Shared/Split Screen PvP
Shared/Split Screen Co-op
HDR available
Steam Leaderboards
Commentary available
Tracked Controller Support
Steam Turn Notifications
Includes Source SDK
VR Only
Steam is learning about this game
Profile Features Limited
Mods (require HL2)
Mods
Category

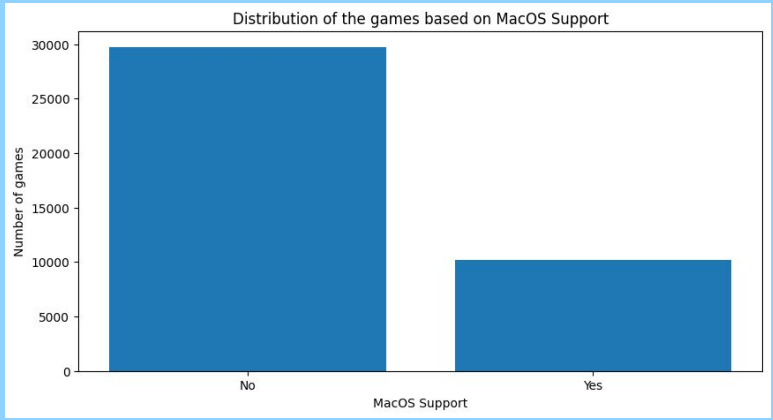
```
developers = data['developer'].value_counts()
developers
```

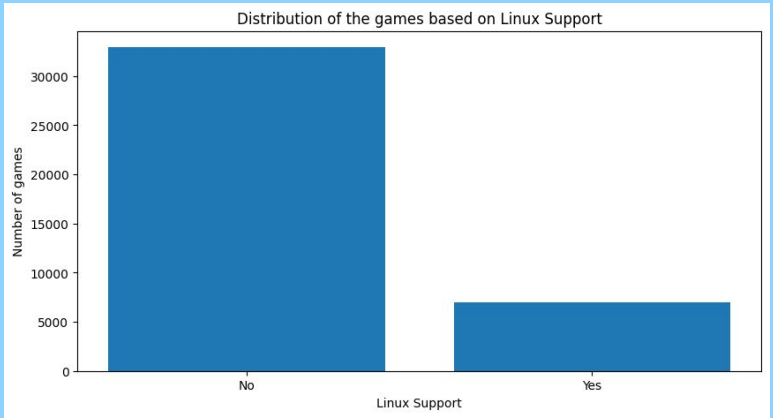
```
developer
Choice of Games              122
Laush Dmitriy Sergeevich     103
Sokpop Collective             87
Square Enix                   80
Dnovel                        74
                            ...
Thaddeus Psarris               1
Rutger van Dijk                1
ALFINA GAMERS                  1
Morgopolis Studios             1
Stoneworks Games               1
Name: count, Length: 23910, dtype: int64
```

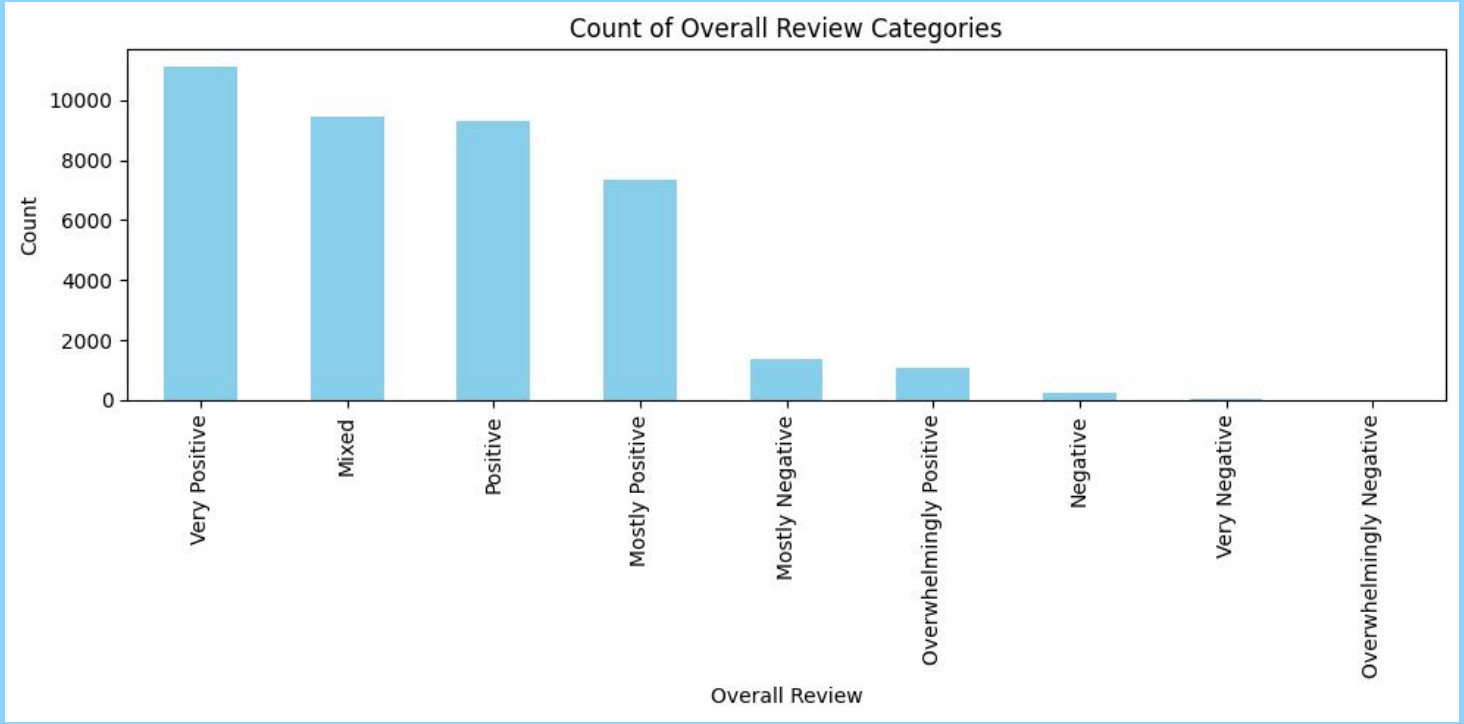
```
publishers = data['publisher'].value_counts()
publishers
```

```
publisher
Big Fish Games             164
Conglomerate 5             148
Kagura Games               135
Sekai Project              132
THQ Nordic                 131
                          ...
SmartOliveGames LTD          1
GKStudio                     1
Heavy Spectrum Limited       1
Pixel Gangsta Studio         1
Stoneworks Games             1
Name: count, Length: 19913, dtype: int64
```

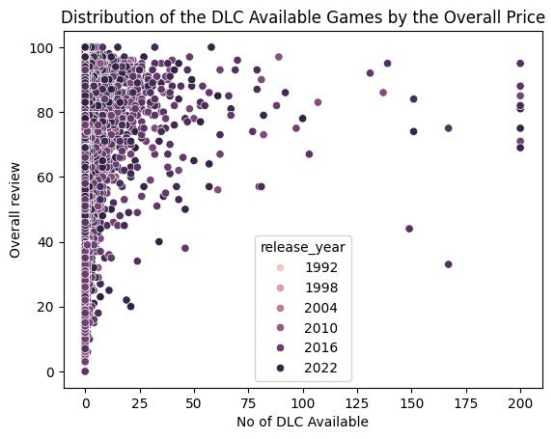

Distribution of the games based on Windows Support
Number of games
40000
35000
30000
25000
20000
15000
10000
5000
0
Yes
No
Windows Support

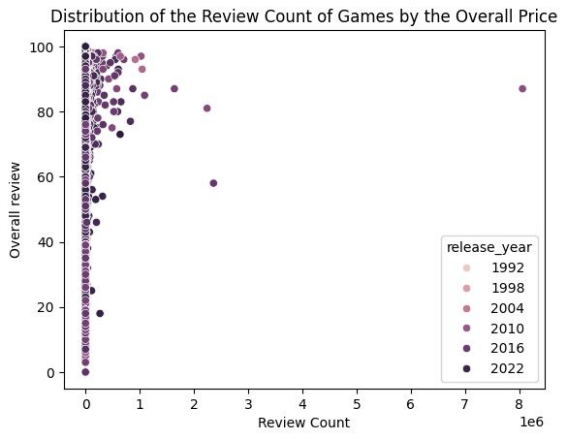

Distribution of the games based on MacOS Support
Number of games
30000
25000
20000
15000
10000
5000
0
No
Yes
MacOS Support

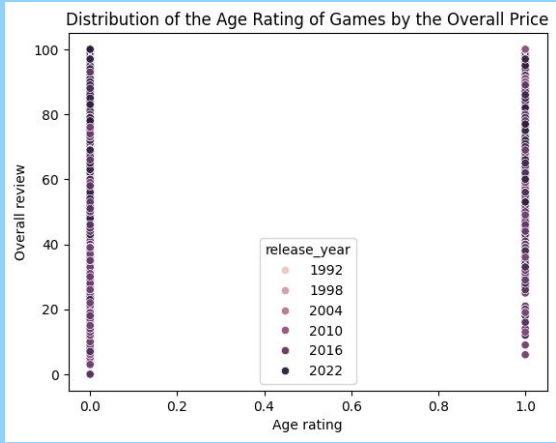

Distribution of the games based on Linux Support
Number of games
30000
25000
20000
15000
10000
5000
0
No
Yes
Linux Support

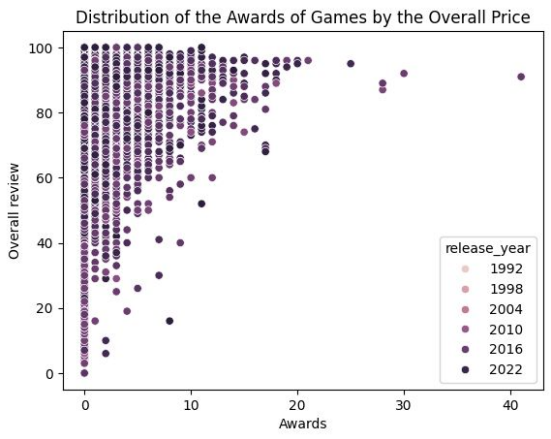

Count of Overall Review Categories
Count
10000
8000
6000
4000
2000
0
Very Positive
Mixed
Positive
Mostly Positive
Mostly Negative
Overwhelmingly Positive
Negative
Very Negative
Overwhelmingly Negative
Overall Review

Distribution of the Discounted Price Games by the Overall Price
Overall review
Discounted Price
release_year
1992
1998
2004
2010
2016
2022

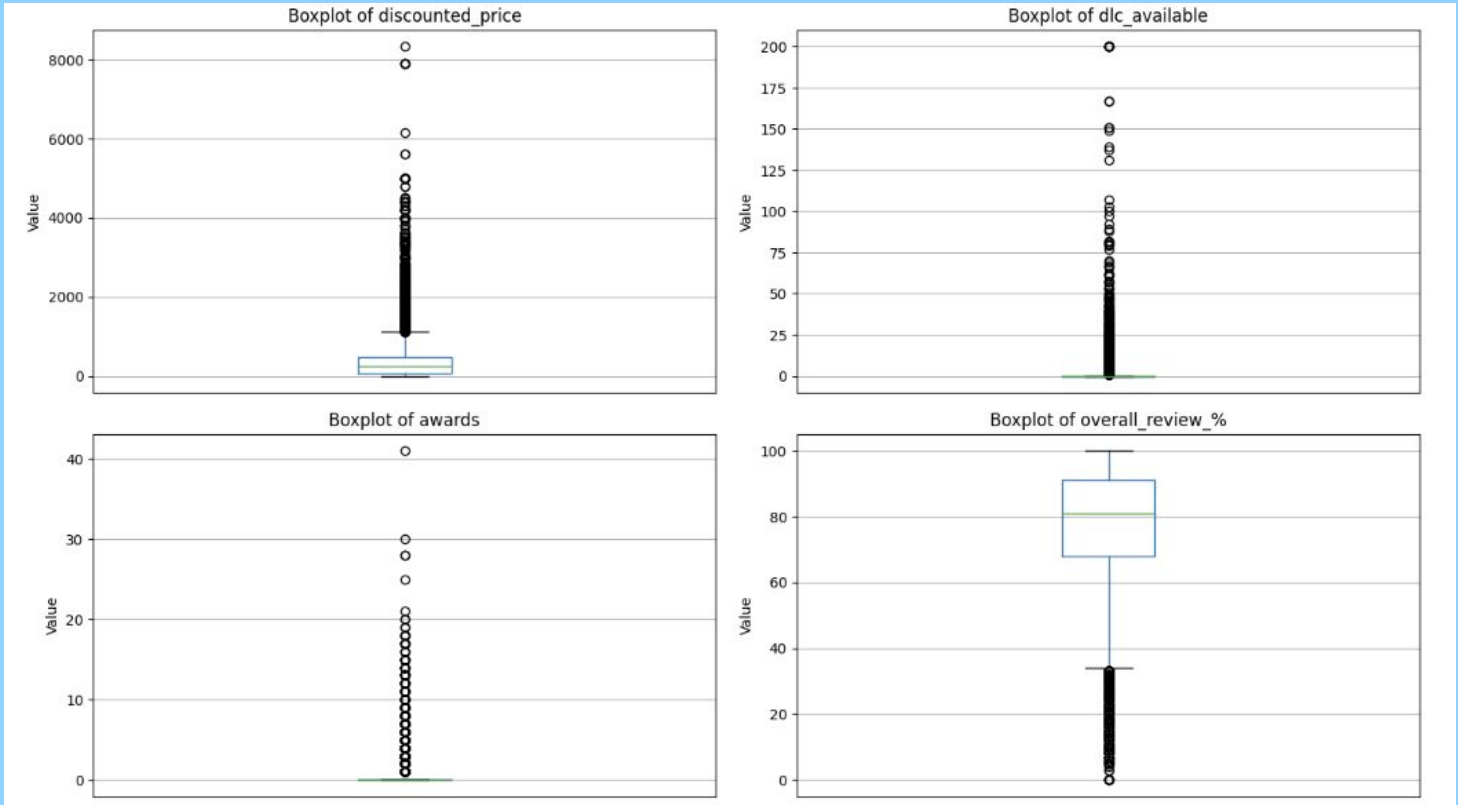

Distribution of the DLC Available Games by the Overall Price
Overall review
No of DLC Available
release_year
1992
1998
2004
2010
2016
2022


Distribution of the Review Count of Games by the Overall Price
Overall review
Review Count
1e6
release_year
1992
1998
2004
2010
2016
2022


Distribution of the Age Rating of Games by the Overall Price
Overall review
Age rating
release_year
1992
1998
2004
2010
2016
2022


Distribution of the Awards of Games by the Overall Price
Overall review
Awards
release_year
1992
1998
2004
2010
2016
2022

Outliers...


Boxplot of discounted_price
Value
8000
6000
4000
2000
0
Boxplot of dlc_available
Value
200
175
150
125
100
75
50
25
0
Boxplot of awards
Value
40
30
20
10
0
Boxplot of overall_review_%
Value
100
80
60
40
20
0

Categorical Columns?

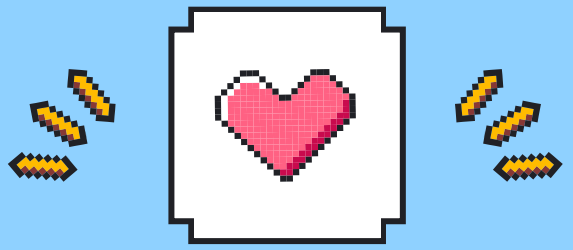
New Columns

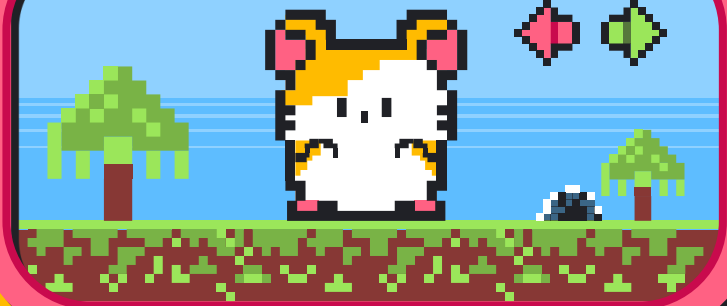
One-Hot Encoding

Convert to Numerical

Model Creation

Model	MAE	MSE	R^2
Linear Regression	4.33	29.78	0.90
Decision Tree R	5.32	51.82	0.83
Random Forest R	4.23	30.35	0.90
Gradient Boosting Regressor	4.22	29.62	0.91
Neural Network	77.21	6274.01	-19.12





Conclusion

Thanks

Questions?