Università della Calabria - Master Degree in Artificial Intelligence and Computer Science

Data Analytics (Data Warehouse and Visualization) course

Introduction

Steam is a video game digital distribution service and storefront by Valve. It was launched as a software client in September 2003 as a way for Valve to provide automatic updates for their games and expanded to distributing and offering third-party game publishers' titles in late 2005. Steam offers various features. Like digital rights management (DRM), game server matchmaking and anti-cheat measures and social networking and games streaming services. It provides the user with automatic game updating, saved game cloud synchronization and community features such as friends messaging, in-game chat and a community market. Offering all of these services, Steam is the store of choice for third-party developers to distribute their video games, in fact the Steam platform is the largest digital distribution platform for PC gaming, estimate around 75% of the market share.

Sommary

Source **3**
goal **4**
Preliminary phase **5**
Relational schema 6
Attribute tree **7**
Edited tree 8
Fact schema **8**
Snowflake schema 9
Analisys sheets and Dashboard with Tableau **10**
Conclusions **23**

Source

Using data gathered from the Steam Store and SteamSpy APIs, this dataset provides information about various aspects of games on the store, such as its genre and the estimated number of owners. Gathered around May 2019, it contains most games of the store released prior to that date. Unrelased titles were removed as well as many non-games like software. The steam.csv file collects unique row about 27033 video games. There are 18 columns described in the following table:

ATTRIBUTE NAME	DESCRIPTION	TYPE
appid	Unique identifier for each title	numerical
name	Game title	String
release_date	Release date in format DD/MM/YYYY	Date
english	Lenguage support: 1 if is in English	Boolean
developer	Nam(s) of developer(s)	String
publisher	Nam(s) of publisher(s)	String
platforms	List of supported platforms	String
required_age	Minimum required age according to PEGI UK	Numerical
categories	List of game categories	string
genres	List of game genres	String
steamspy_tags	List of game steamspy tags	string

achivements	Number of in-game achievements	String
positive_ratings	Numer of positive ratings, from steam spy	numerical
negative_ratings	Numer of negative ratings, from steam spy	Numerical
average_playtime	Average user playtime, from steam spy	Numerical
median_playtime	median user playtime, from steam spy	Numerical
owners	Estimated number of owners. Range like 20000 - 50000	String
Price	Current full price of title in GBP	numerical

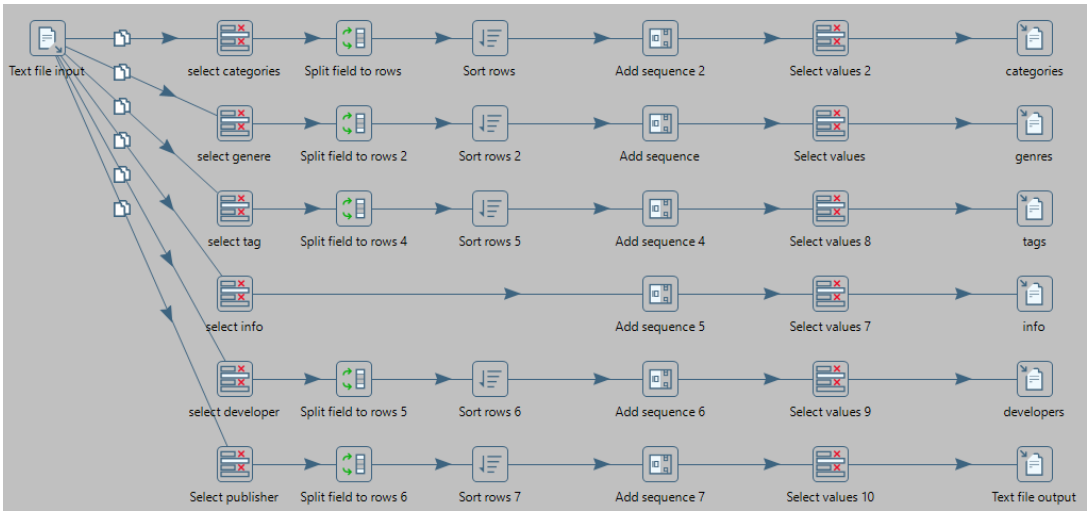
Goal

The main goal is to analyze genres and categories trends over years and months to help third-party developers strategically release their products over time. Also, we want to analyze for each category and genre the ratings of developers, publishers and video games.

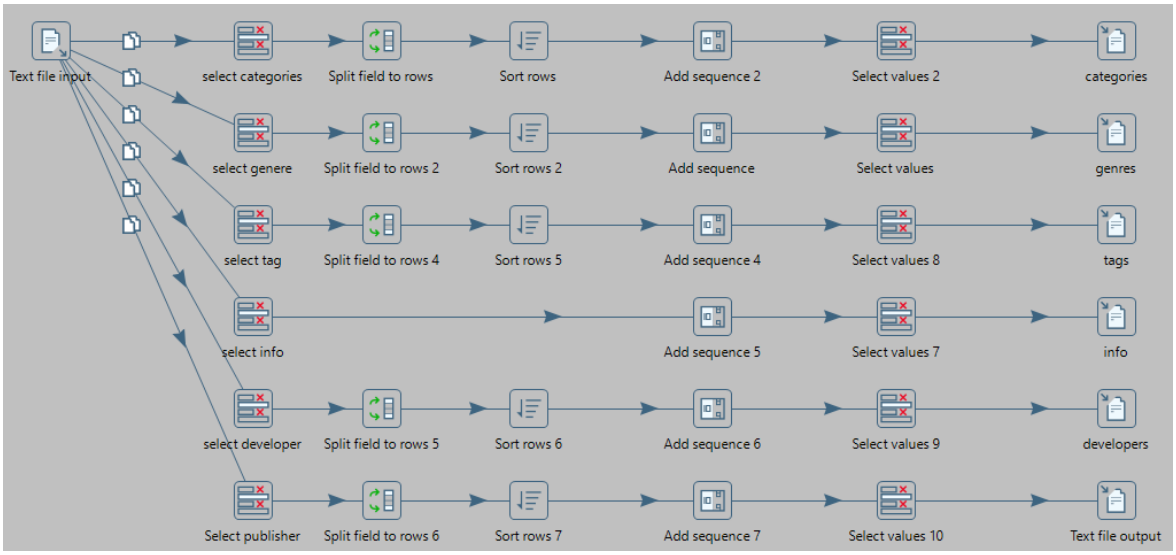
Preliminary phase

An important preliminary phase in necessary before proceeding with the presentation of a relational scheme. The attributes genres, categories, steampsy_tags, developer and publisher can strore multiple values in the same row delimitated with a semicolon. For this problem we create new “bridge” tables that link the main tables to a new table containing all the unique values.

Penthao was used to perform these data transformations. The following screen show the steps performed for extracting all the single value of the attributes.



Then we create the “bridge” tables:



Abridge table is characterized by a weight for each value, that’s important because some value could be more important than others. Pandas was used to create the scrip weights.py, the weights are calculates like 1/n, where n is the number of genres of a single games, for example.

Relational schema

info (appid, name, release_date, positive_ratings, negative_ratings)

Categories (ID_Categories, Categories)

Genres (ID_Genres, Genres)

Tags (ID_Tags, Tags)

Developers (ID_Developer, Developers)

Publishers (ID_Publisher, Publishers)

AppCat (ID, appid: info, ID_Categories: Categories, weights)

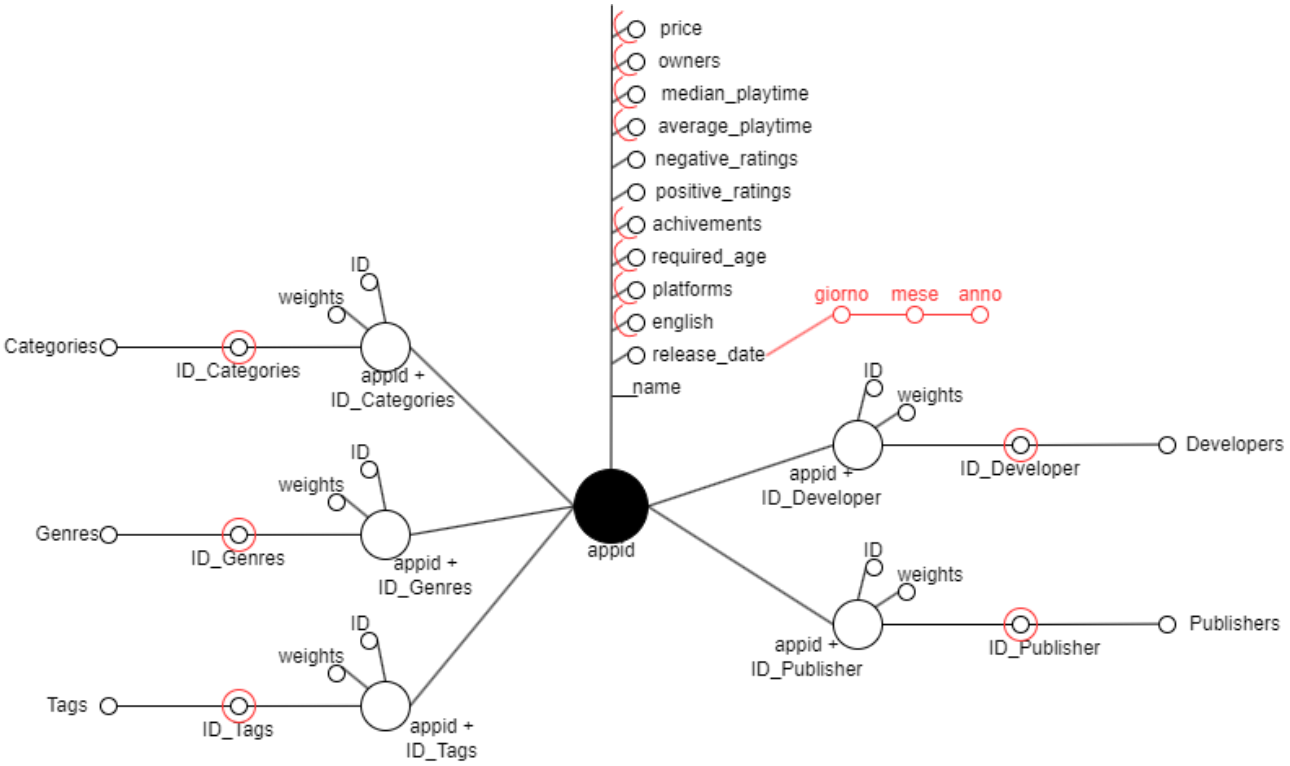
AppGen (ID, appid: info, ID_Genres: Genres, weights)

AppTags (ID, appid: info, ID_Tags: Tags, weights)

AppDev (ID, appid: info, ID_Developer: Developers, weights)

AppPub (ID, appid: info, ID_Publisher: Publishers, weights)

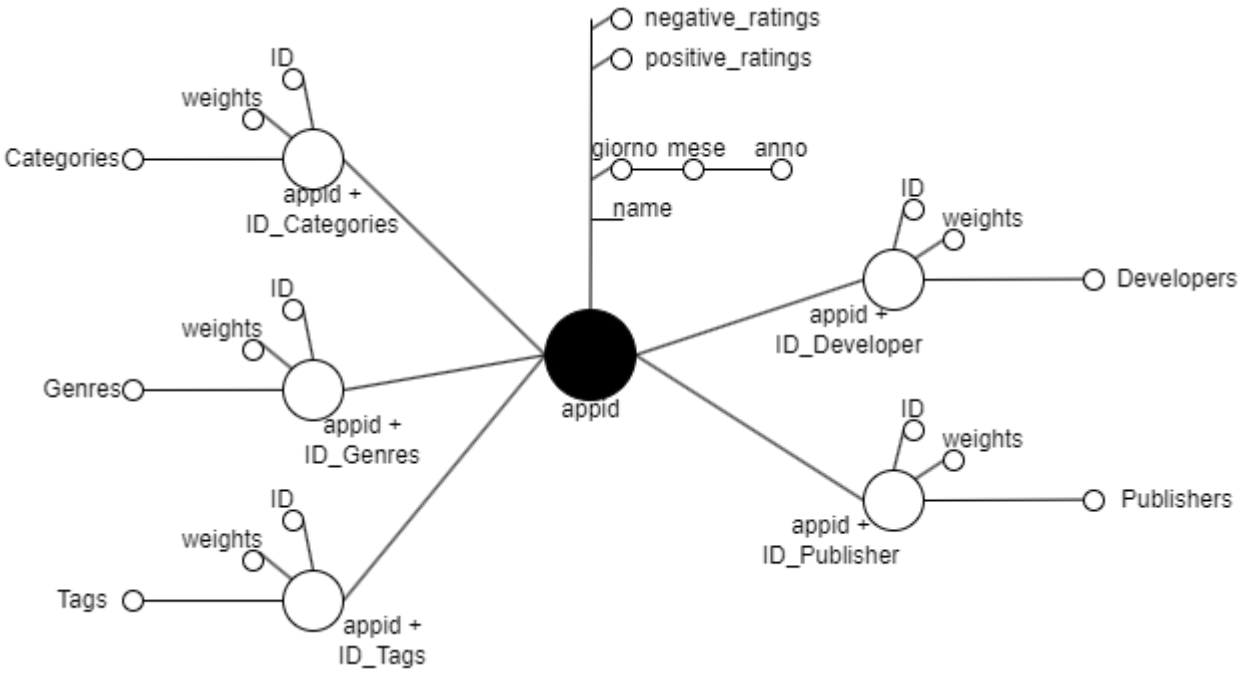
Attribute Tree



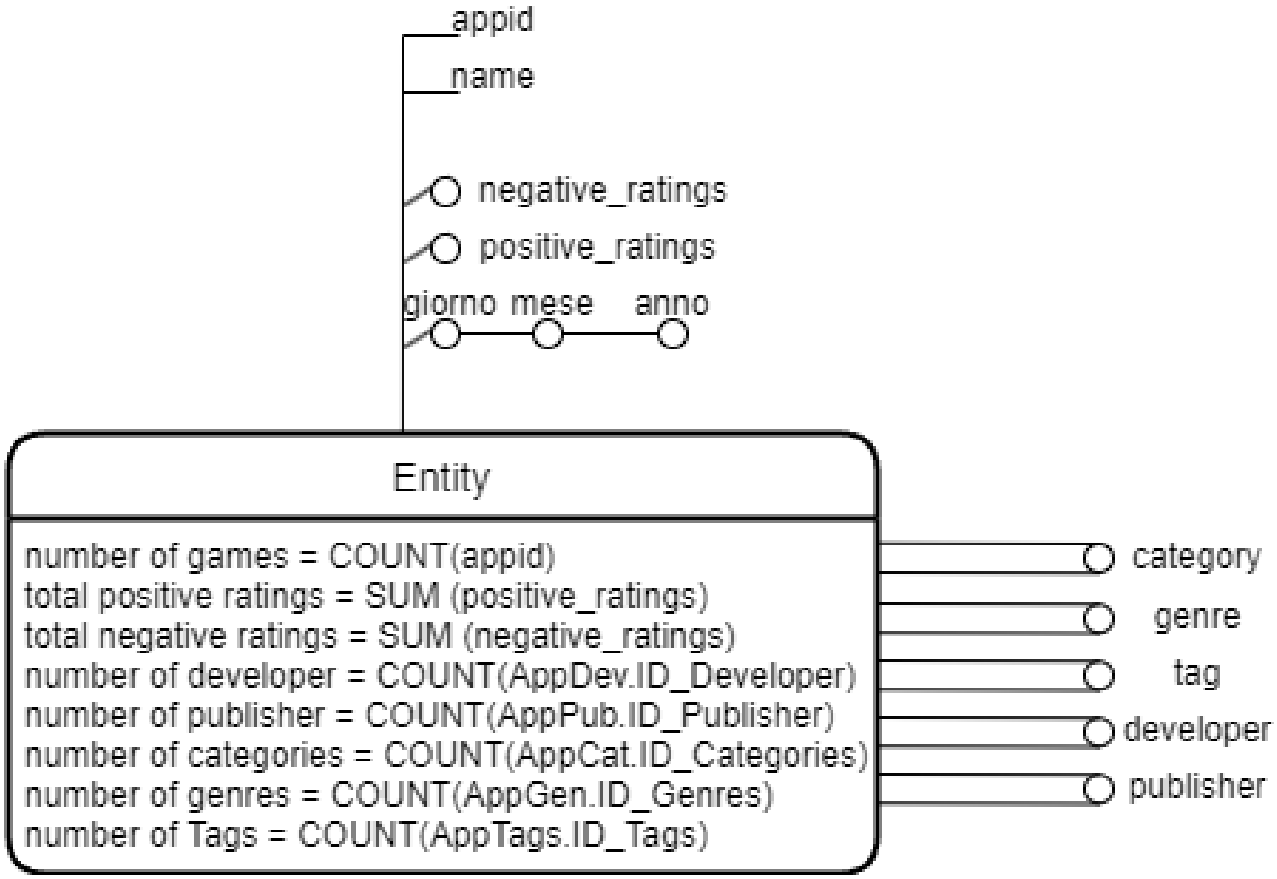
According with the goal explained previously, some attributes are useless for our analysis:

- English, because most of the games support this language. (if this attribute was used foe all the language it could be useful)
- Price, we are not interested in a business analysis.
- Owners, same reason as price.
- Achievements, we can use already the tag “achievement” to know if a game has or not some achivments.
- Required_age, most value are 0 that means games are free age or unranked. So we decided to remove this attribute.
- Platforms, the database a little bit old. The analysis could be not very realistic because steam has created his own OS based on linux in these years.
- Average_playtime and median_playtime because we decided to analyze games and genres based on ratings.

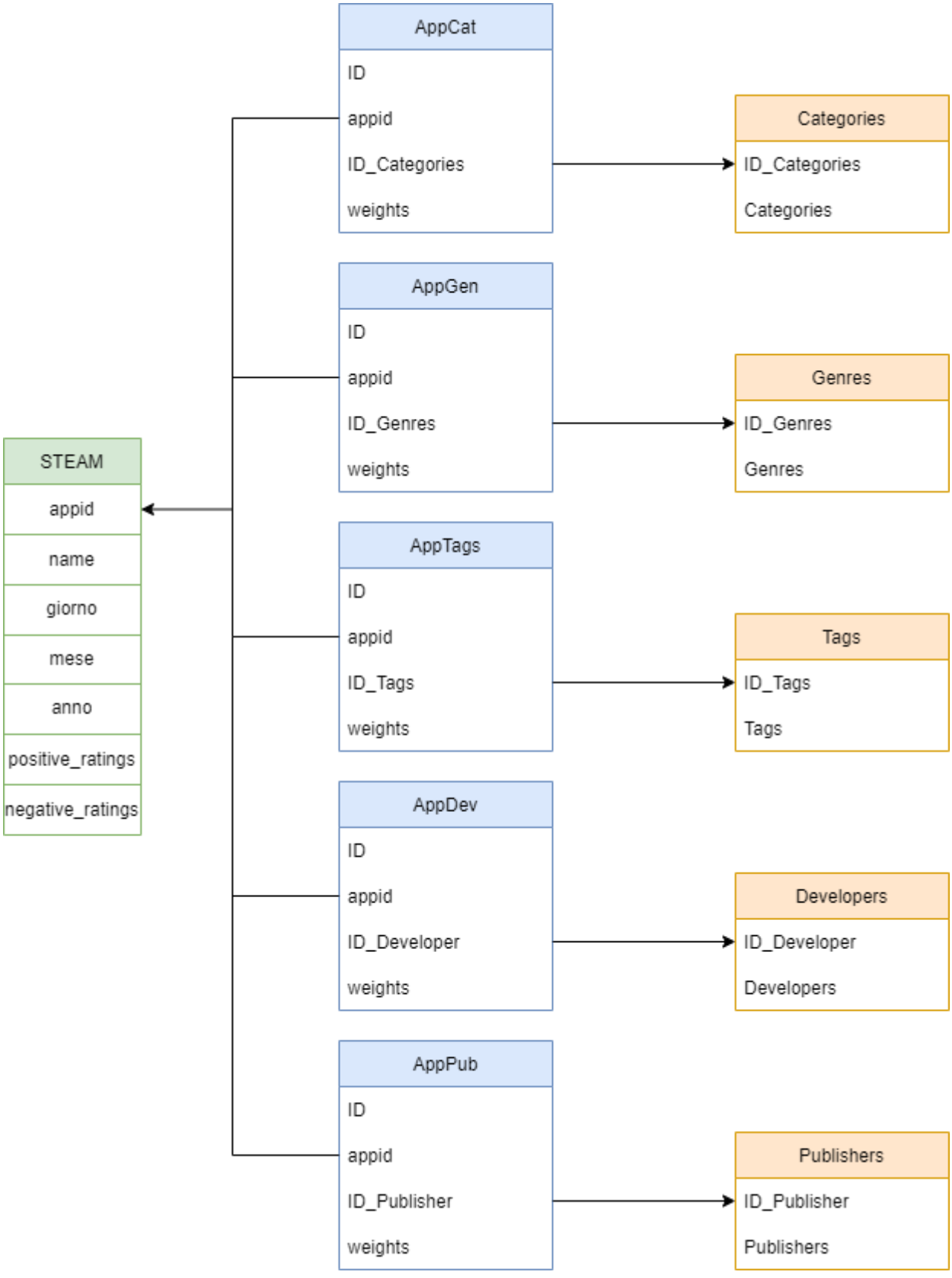
Edited Tree



Fact Schema

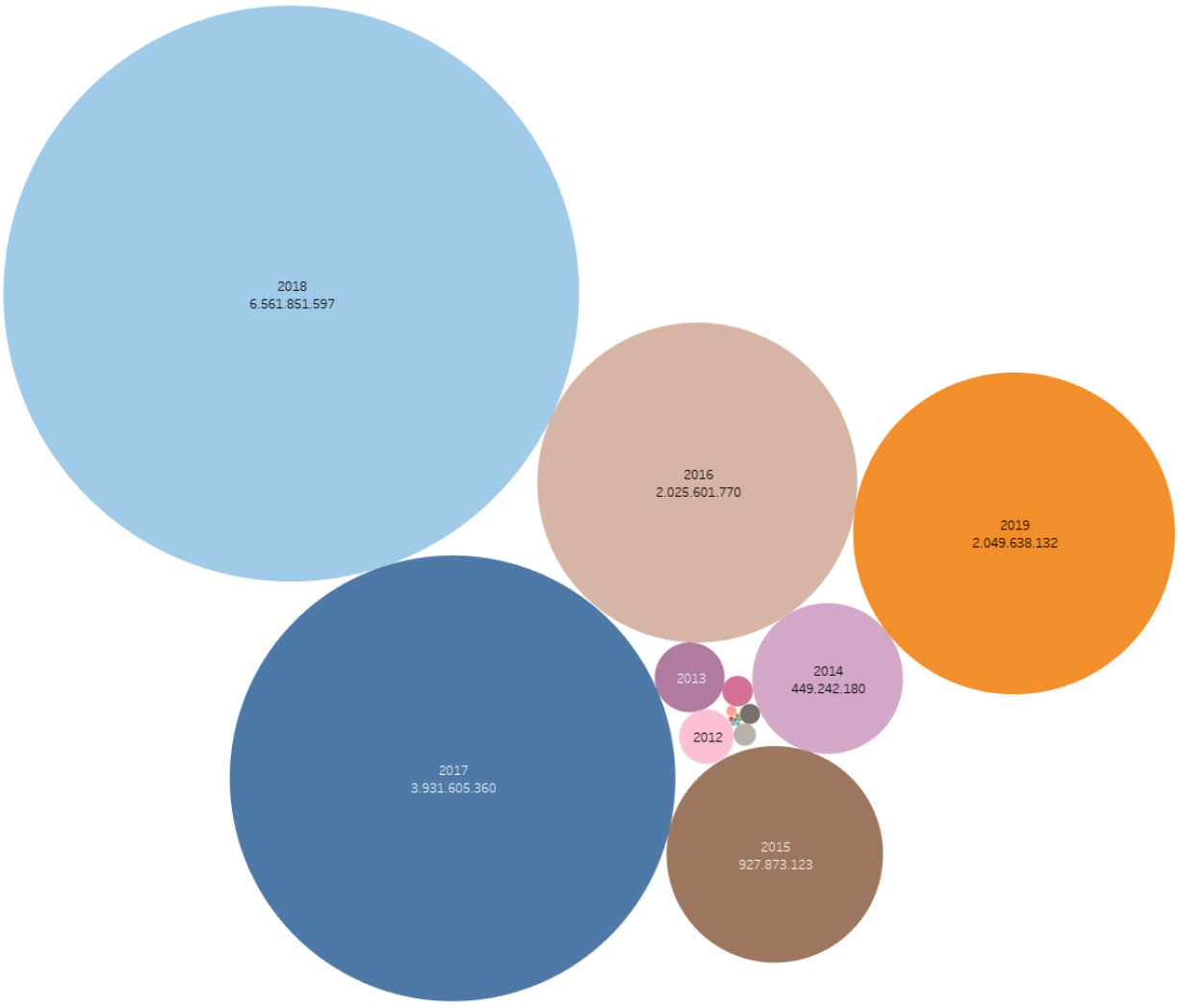


Snowflake Schema



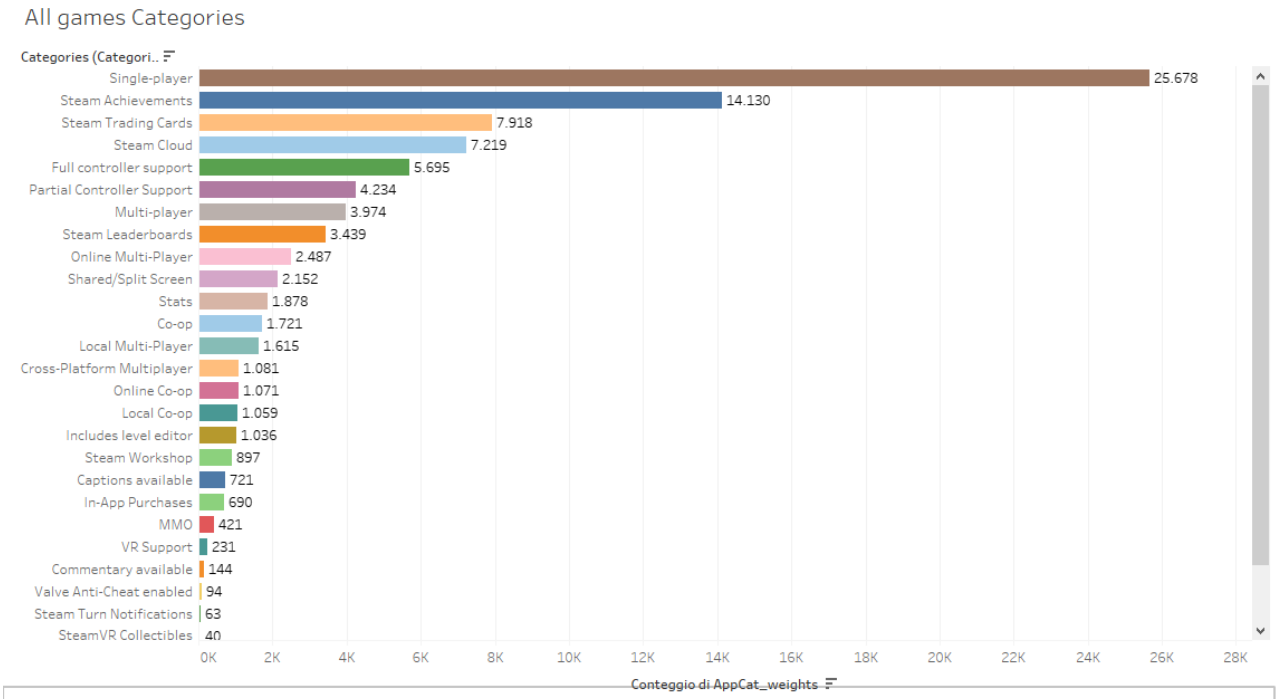
Analysis sheets and Dashboards with Tableau

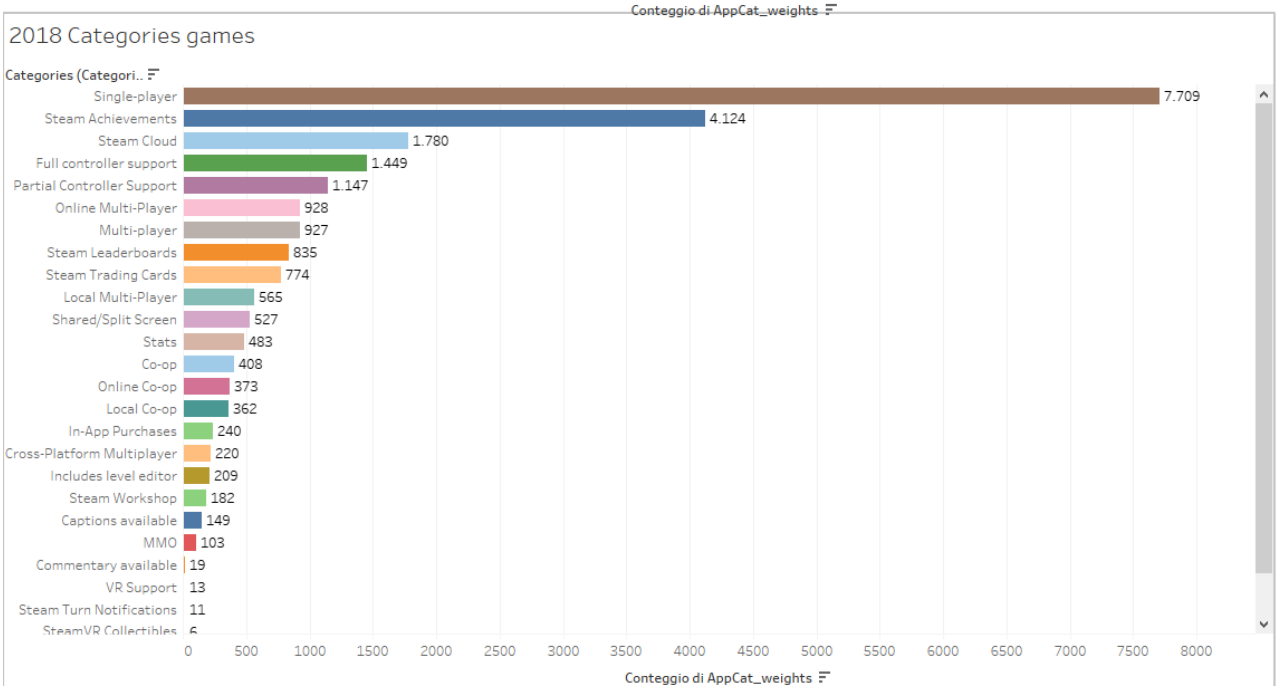
A first look was taken to the numbers of games released on steam every year. As can be observed the number of games increases every year. This is a clear picture of how the world of gaming has exploded in recent years. This dataset has information up to mid-2019, that's why 2019 is smaller than 2018. So, the 2018 was used as reference year for the analysis.



This comparison between “categories of all time” and 2018 games categories shows that:

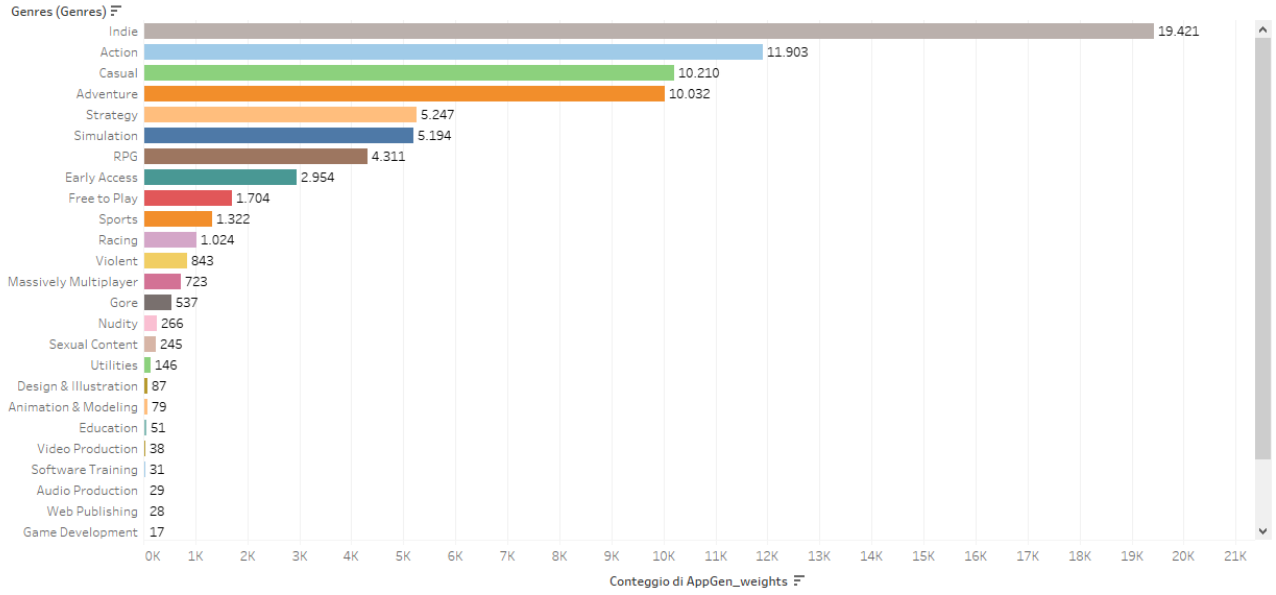
1. Single-player games and achievements are most used. So, single player with achievements is a good combo for the users.
2. The collecting card steam service is going down.
3. Multiplayer games are getting popularity.



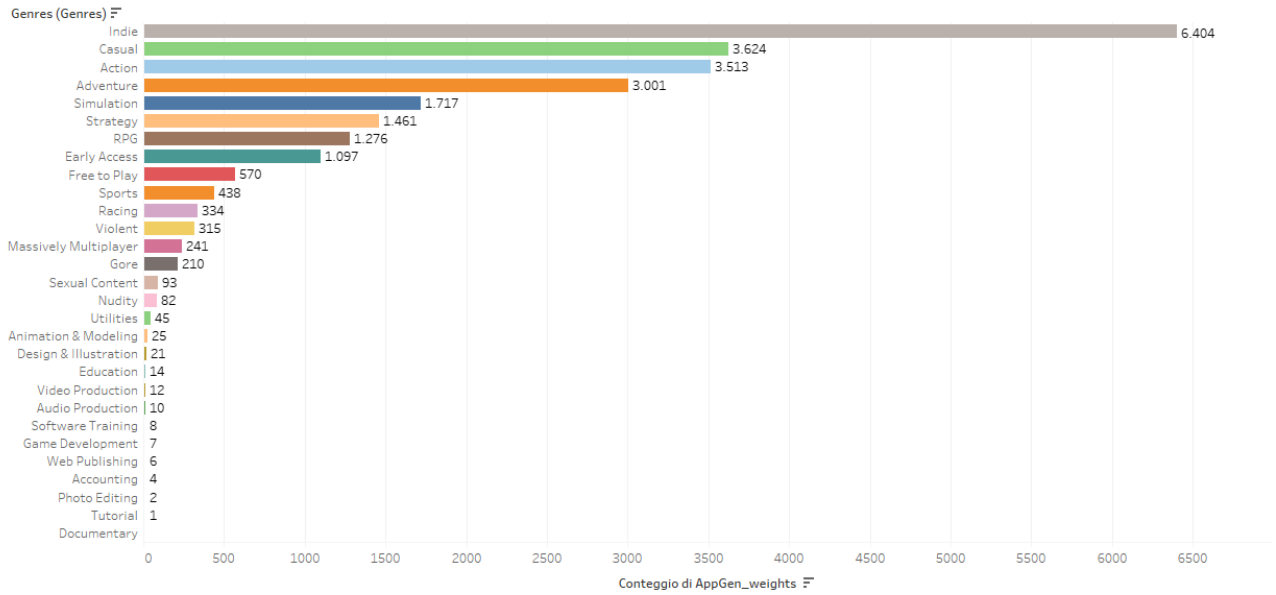


This is 2 sheets are very similar, we can say that they are the same picture. However, is possible to make a consideration on indie games. It's a most used genre of all time and still it is, that's because steam is very friendly with indie software house. In the past steam allowed to all user to create their own games and upload them on "steam greelight", unfortunately this service was closed years ago.

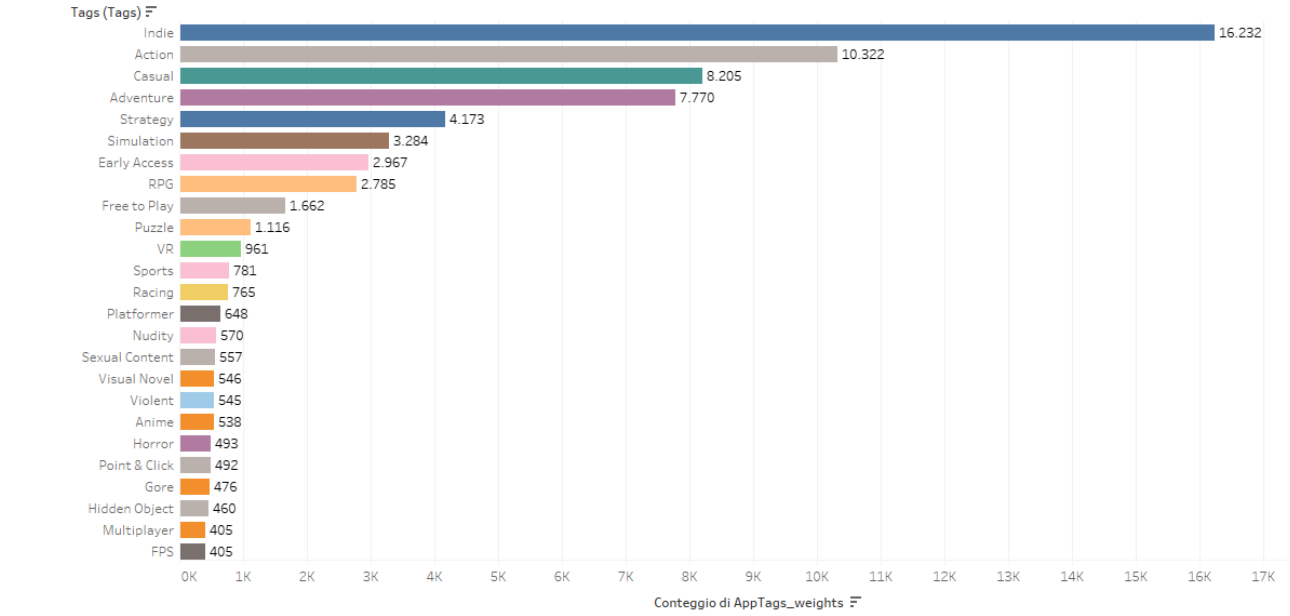
All games Genres



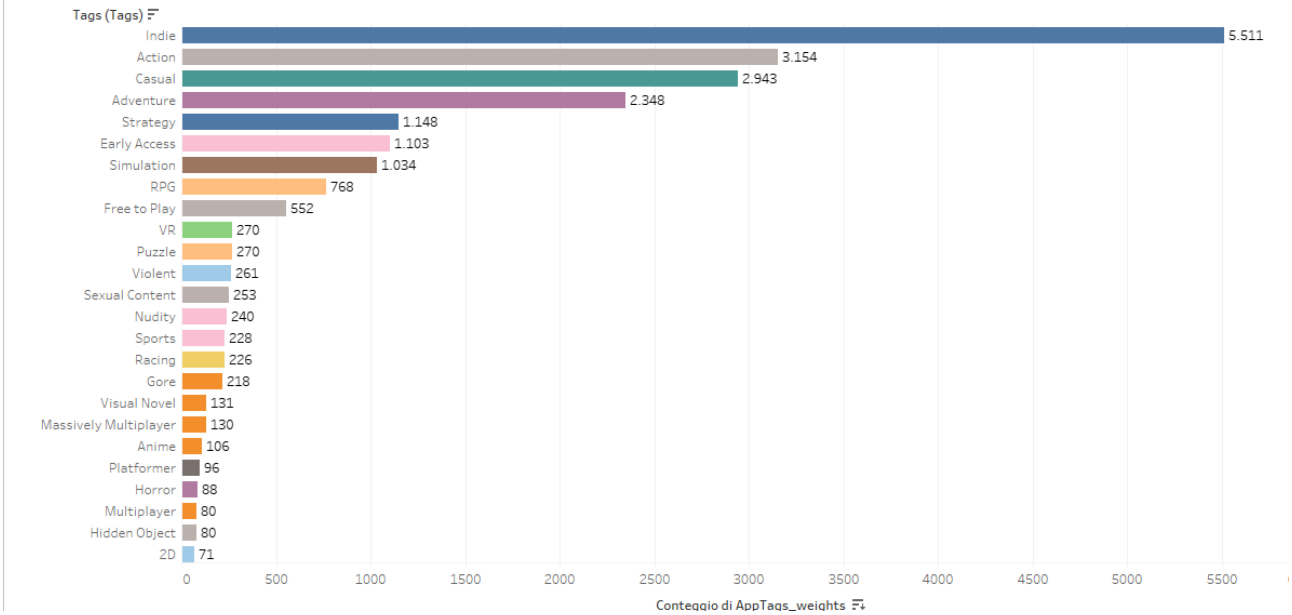
2018 Genres games

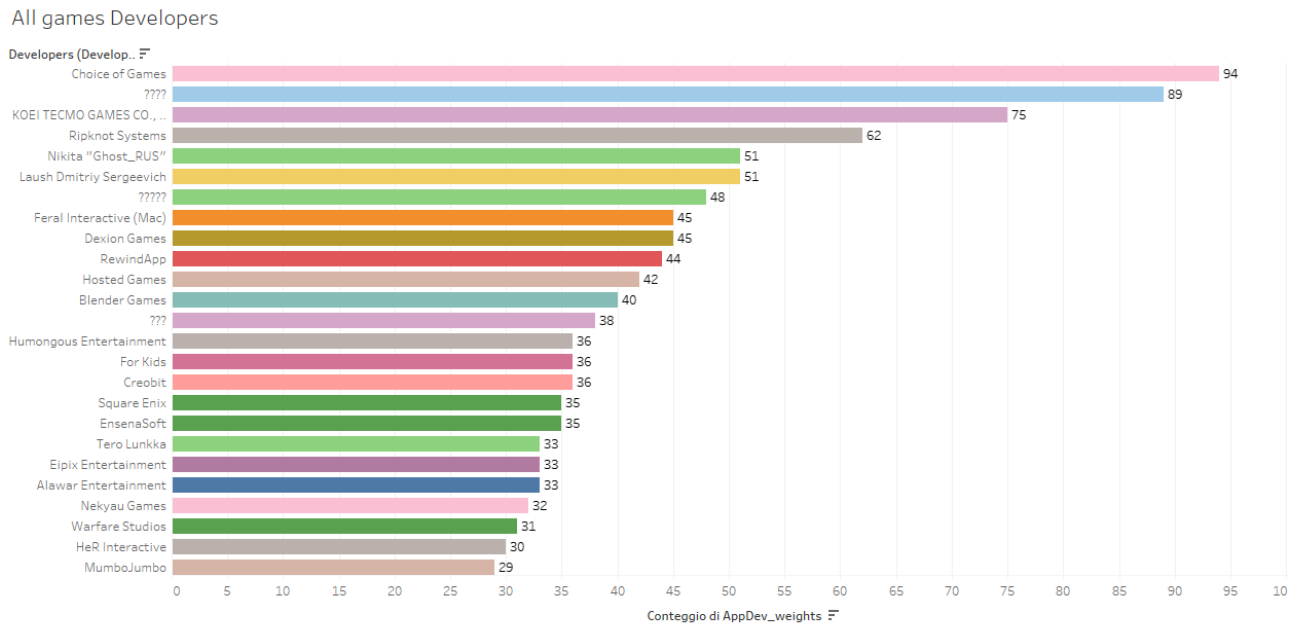


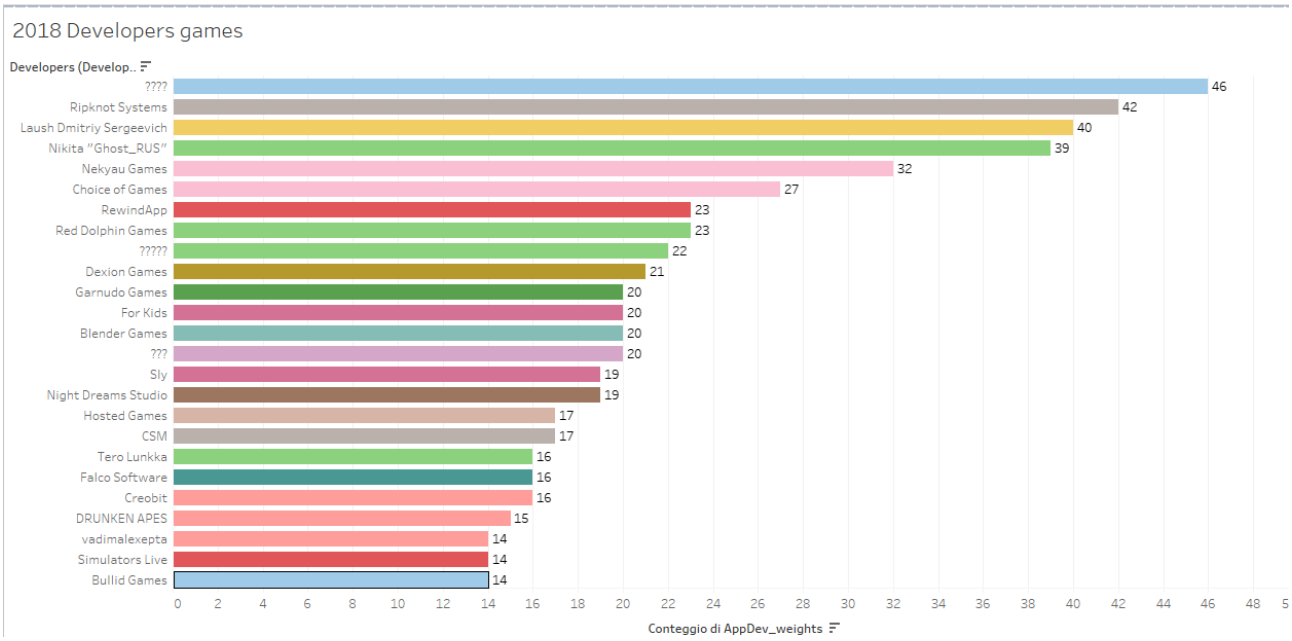
All games Tags



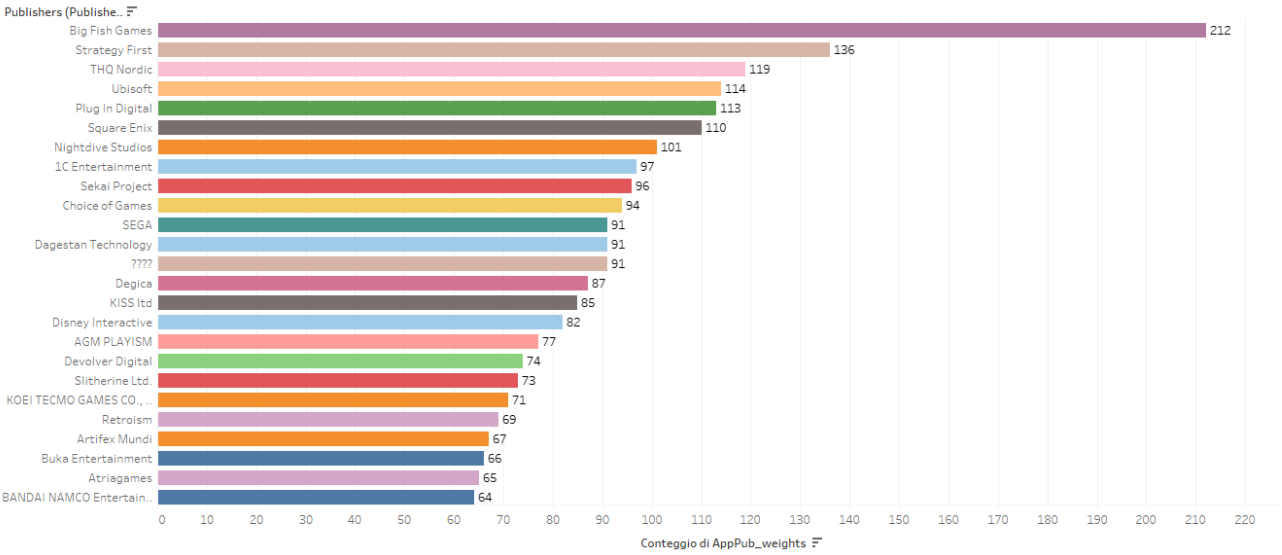
2018 Tags games



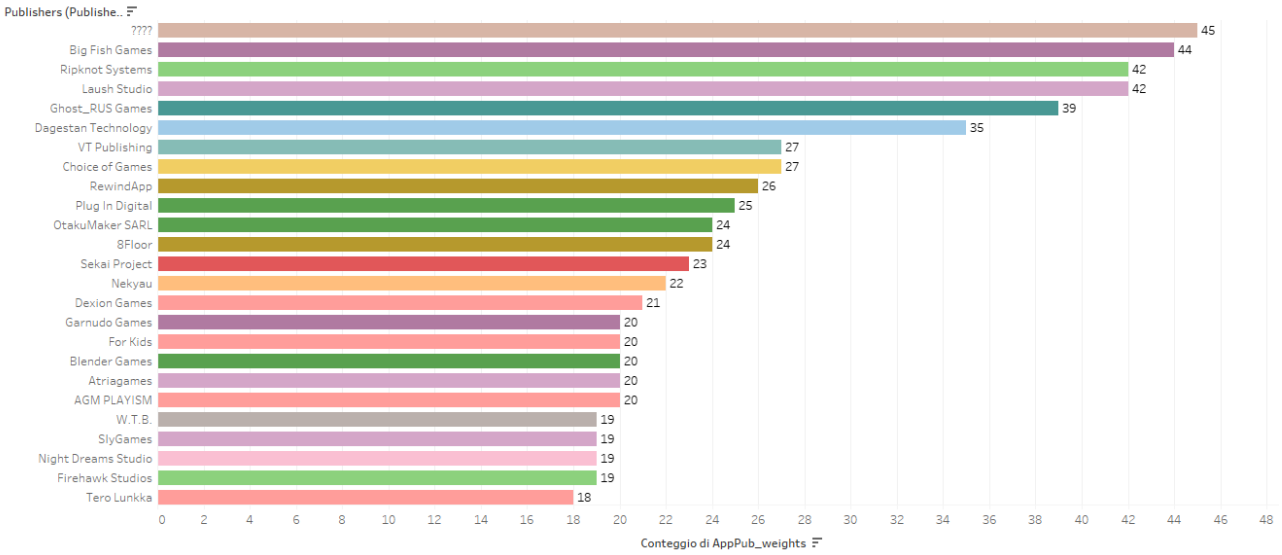

All games Developers
Developers (Develop..
Choice of Games 94
???? 89
KOEI TECMO GAMES CO., .. 75
Ripknot Systems 62
Nikita "Ghost_RUS" 51
Laush Dmitriy Sergeevich 51
????? 48
Feral Interactive (Mac) 45
Dexion Games 45
RewindApp 44
Hosted Games 42
Blender Games 40
??? 38
Humongous Entertainment 36
For Kids 36
Creobit 36
Square Enix 35
EnsenaSoft 35
Tero Lunkka 33
Eipix Entertainment 33
Alawar Entertainment 33
Nekyau Games 32
Warfare Studios 31
HeR Interactive 30
MumboJumbo 29
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 10
Conteggio di AppDev_weights


2018 Developers games
Developers (Develop..
???? 46
Ripknot Systems 42
Laush Dmitriy Sergeevich 40
Nikita "Ghost_RUS" 39
Nekyau Games 32
Choice of Games 27
RewindApp 23
Red Dolphin Games 23
????? 22
Dexion Games 21
Garnudo Games 20
For Kids 20
Blender Games 20
??? 20
Sly 19
Night Dreams Studio 19
Hosted Games 17
CSM 17
Tero Lunkka 16
Falco Software 16
Creobit 16
DRUNKEN APES 15
vadimalexepta 14
Simulators Live 14
Bullid Games 14
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 5
Conteggio di AppDev_weights

All games Publishers

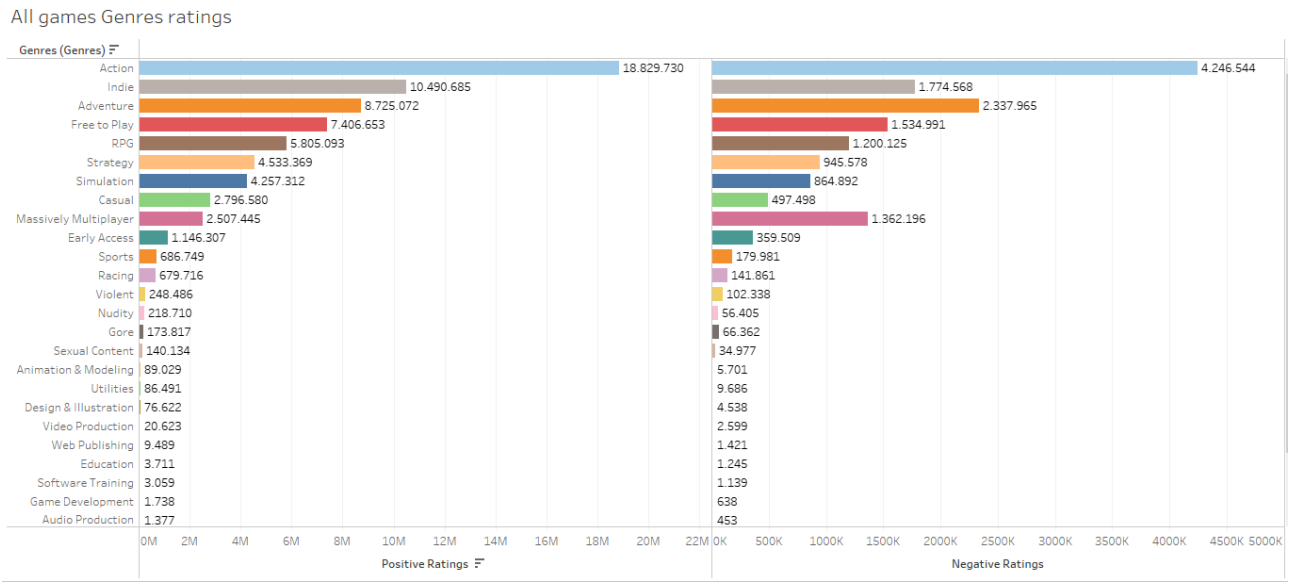


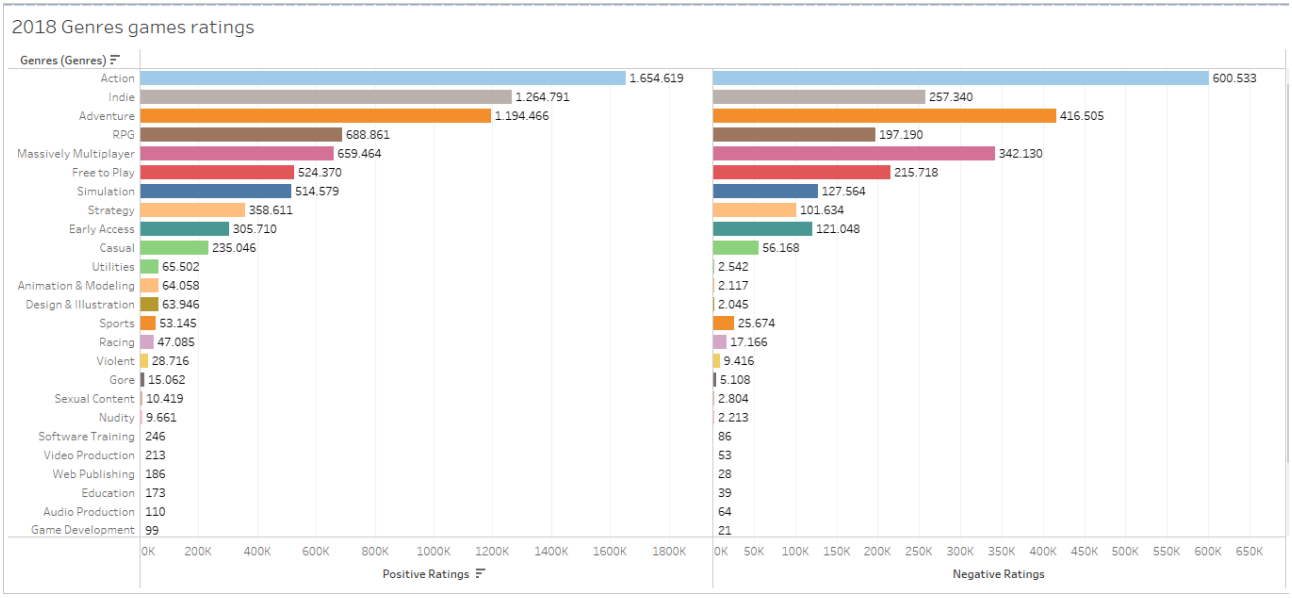
2018 Publishers games



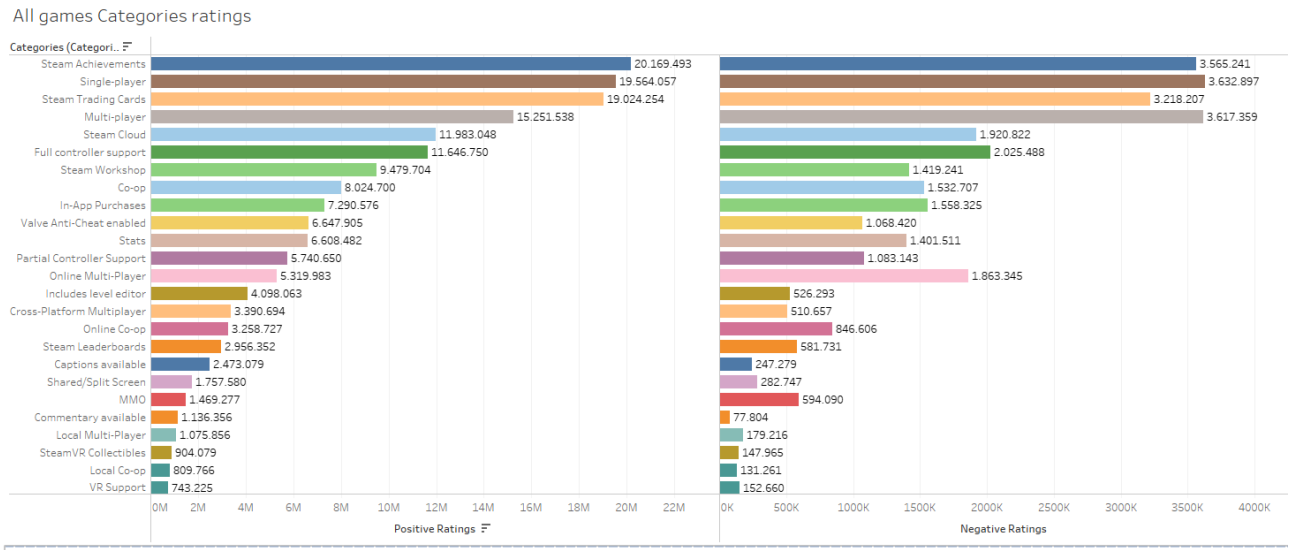
Positive and negative ratings were used to see how much players liked or disliked a game with a certain category or genre and compare it with 2018.

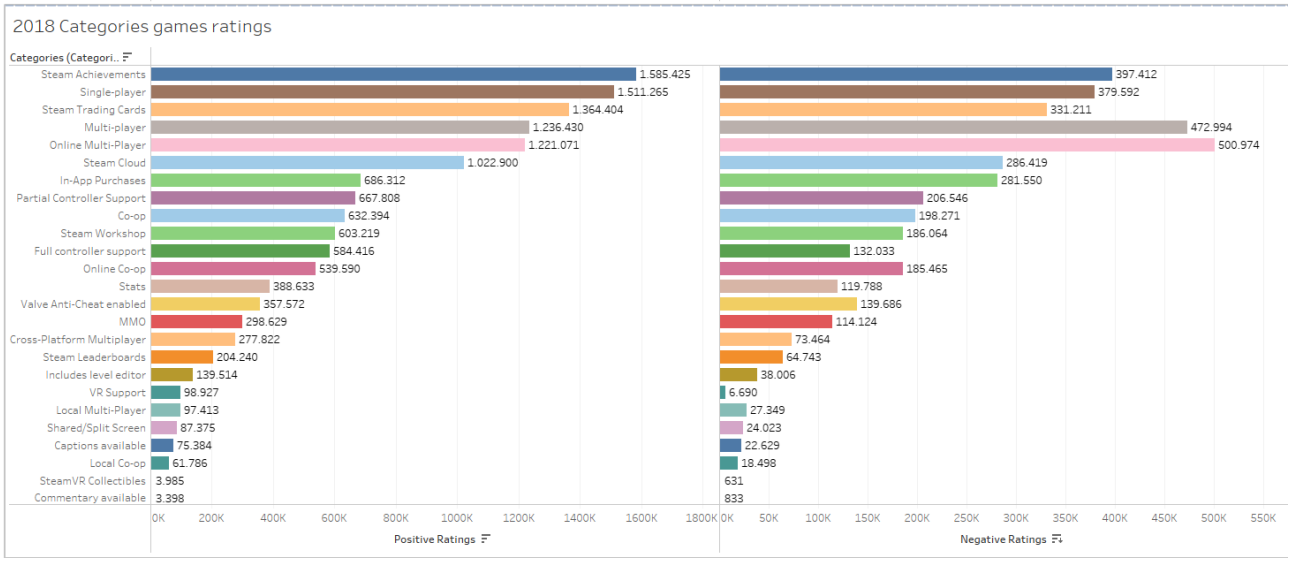
In this case is possible to see that Action, indie and adventure are the top appreciated genres of games, but in 2018 multiplayer and massively multiplayer are growing.





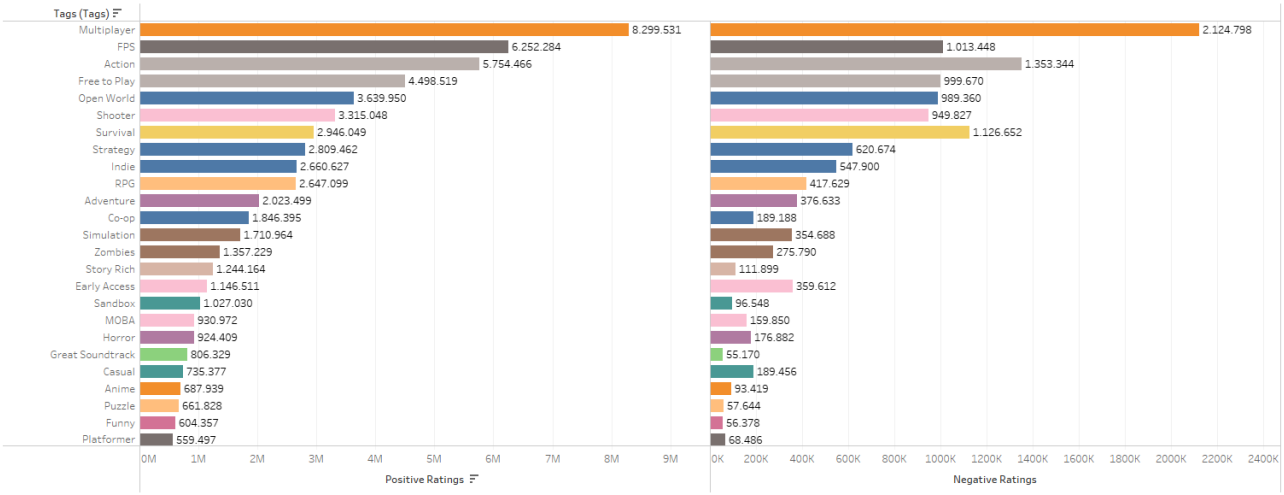
Same situation as Genres sheets. The first three categories are always top appreciated with multiplayer games are growing.



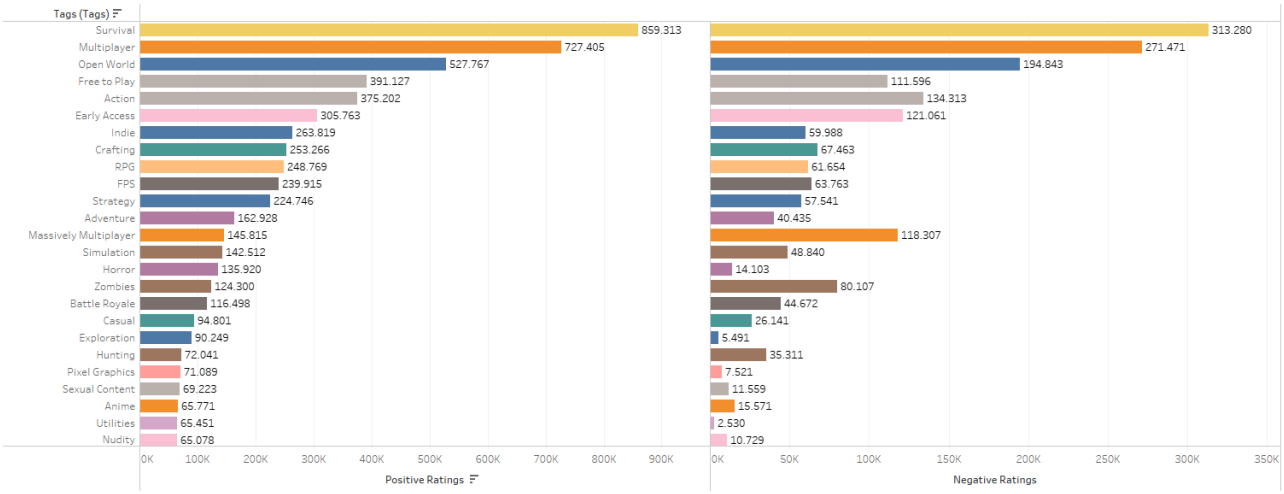


Different situation in Tags with multiplayer, FPS, action and free to play in top tier but in 2018 is possible to see that survival and open world are in top. Maybe in 2018 multiplayer, survival, open world games were the most played and appreciated.

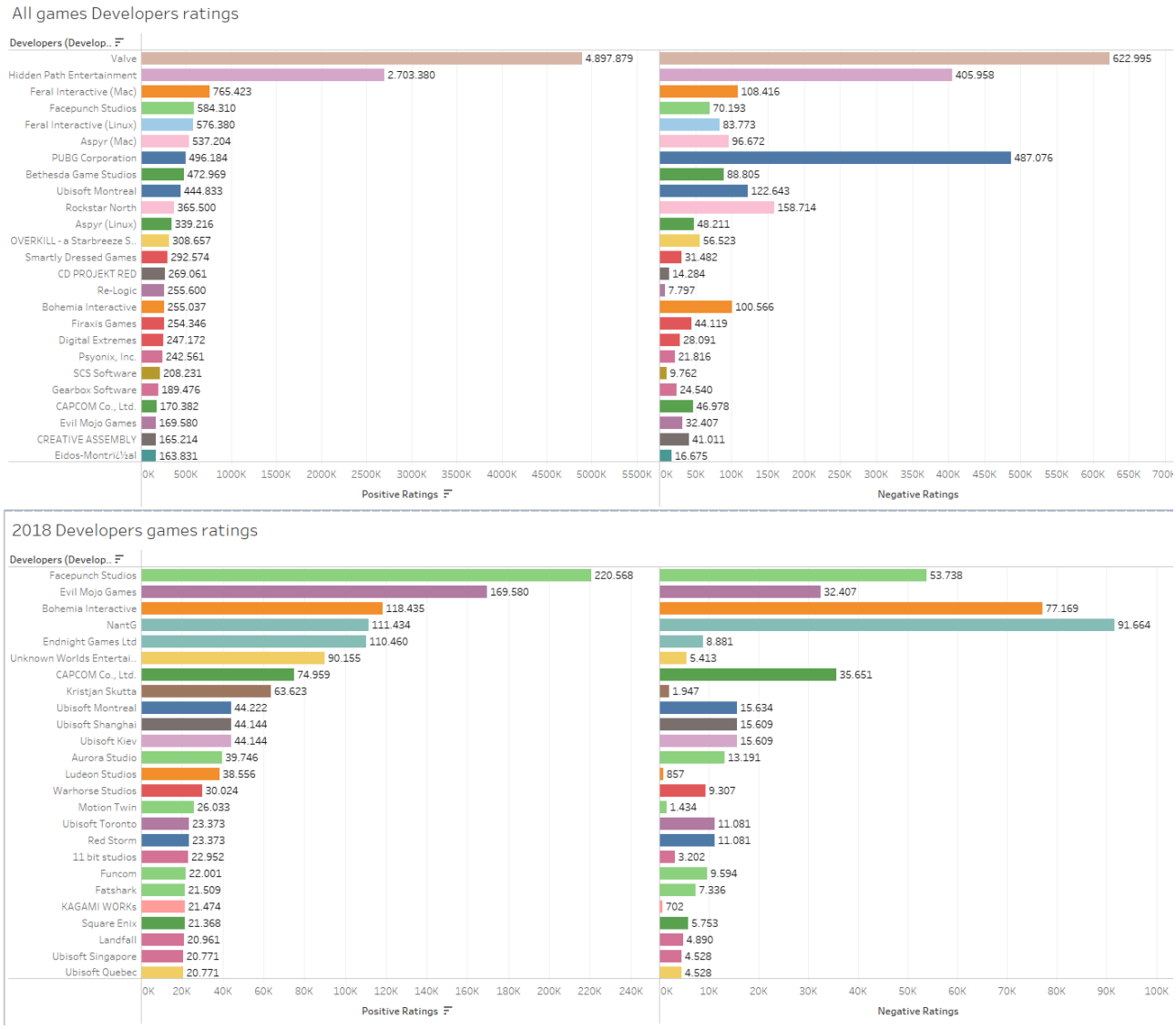
All games Tags ratings



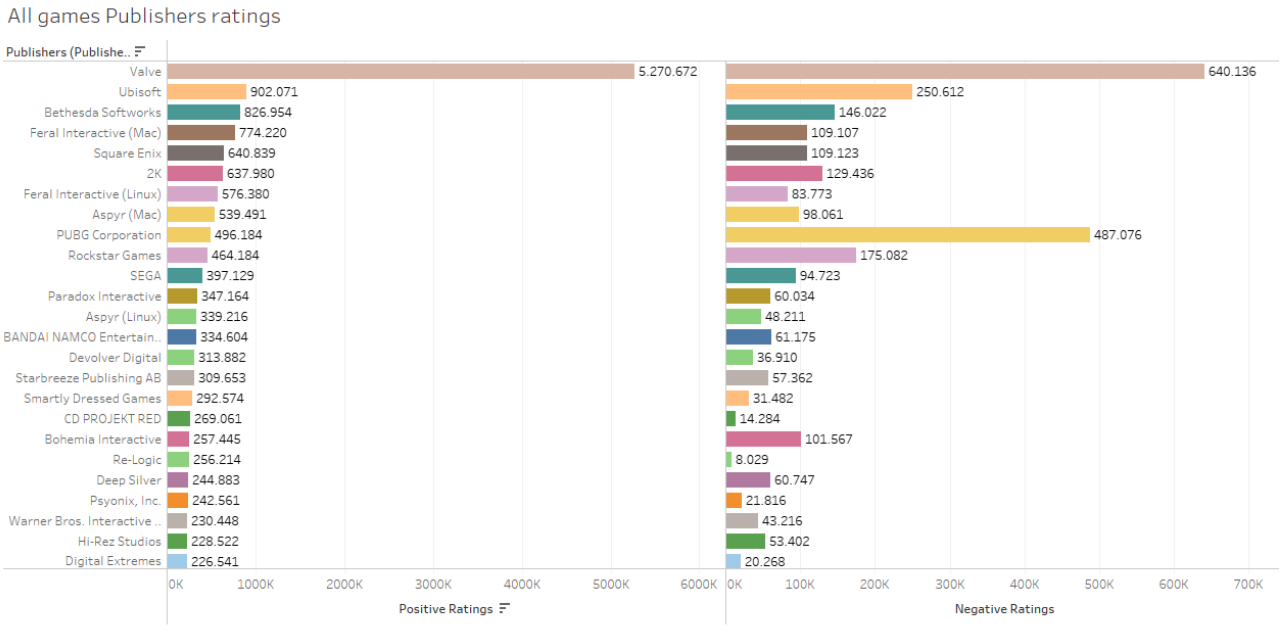
2018 Tags games ratings

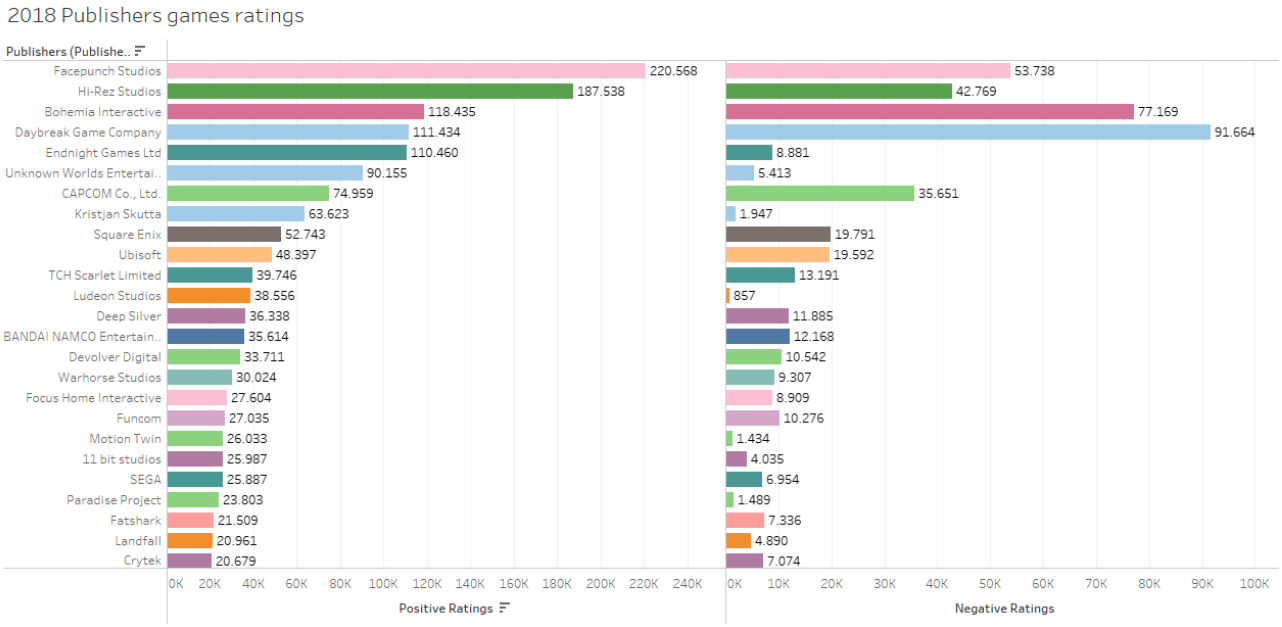


This sheet shows that games developed by valve are the most appreciated. It's not difficult to understand why, games like CS: GO or team fortress 2 have big community of players that still play. Pugb Corporation is very hated from the players.



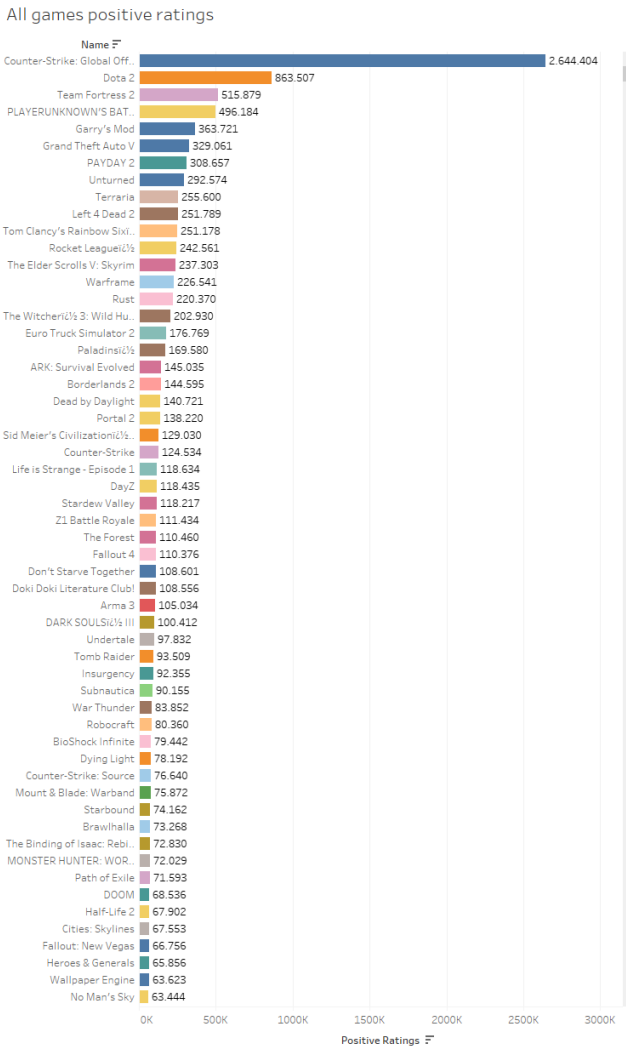
Same situation as before.

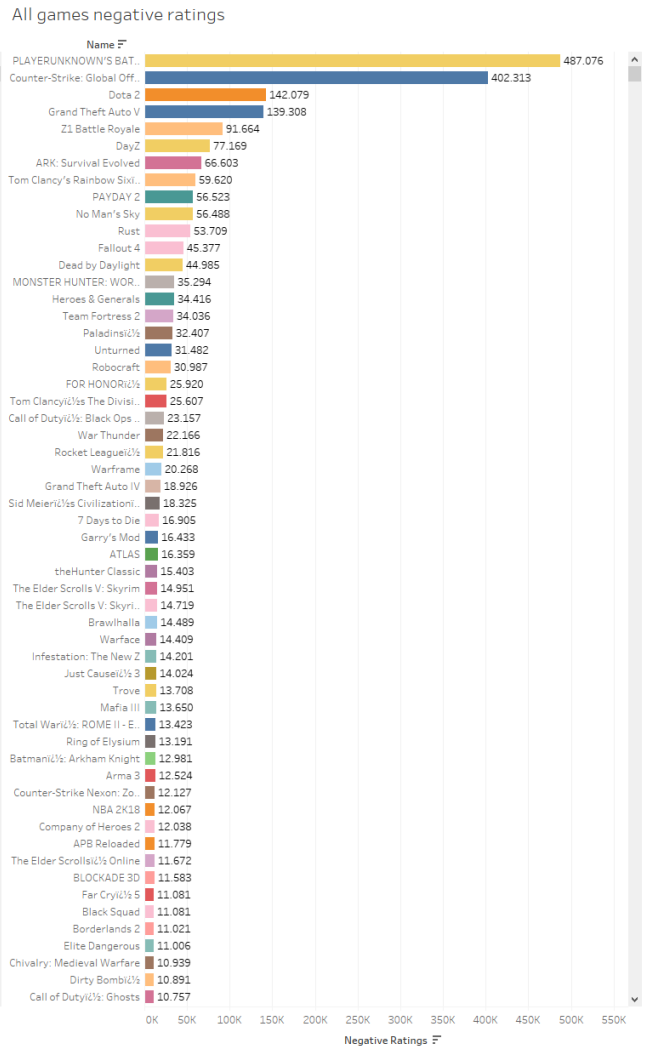




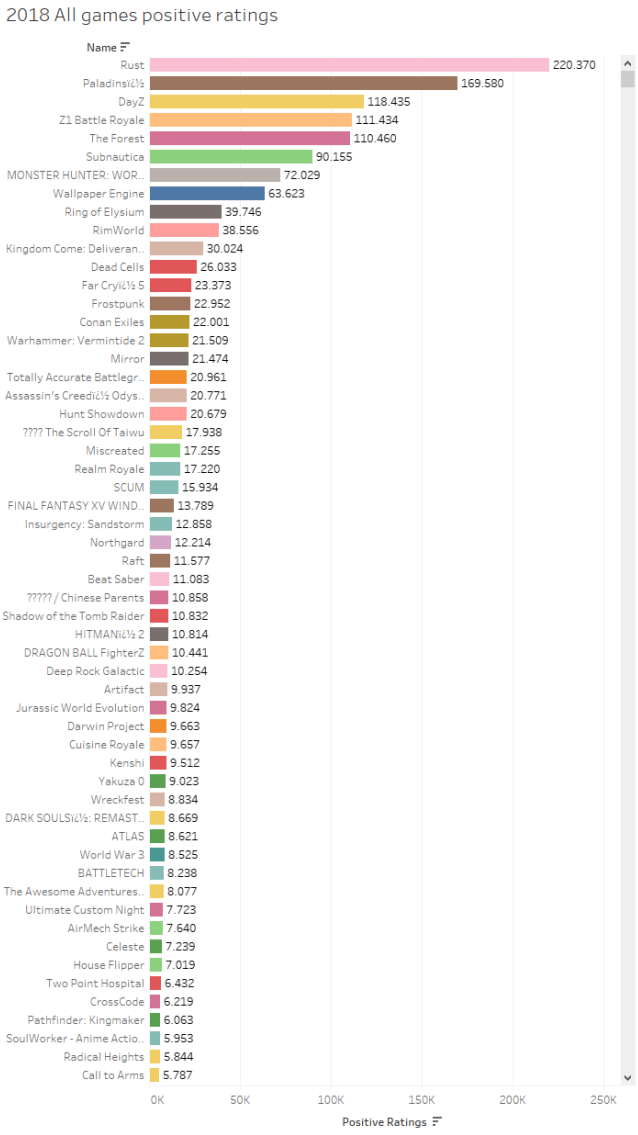
The most popular games of all time are all valve games: CS: GO, Dota 2 and team Fortress 2.

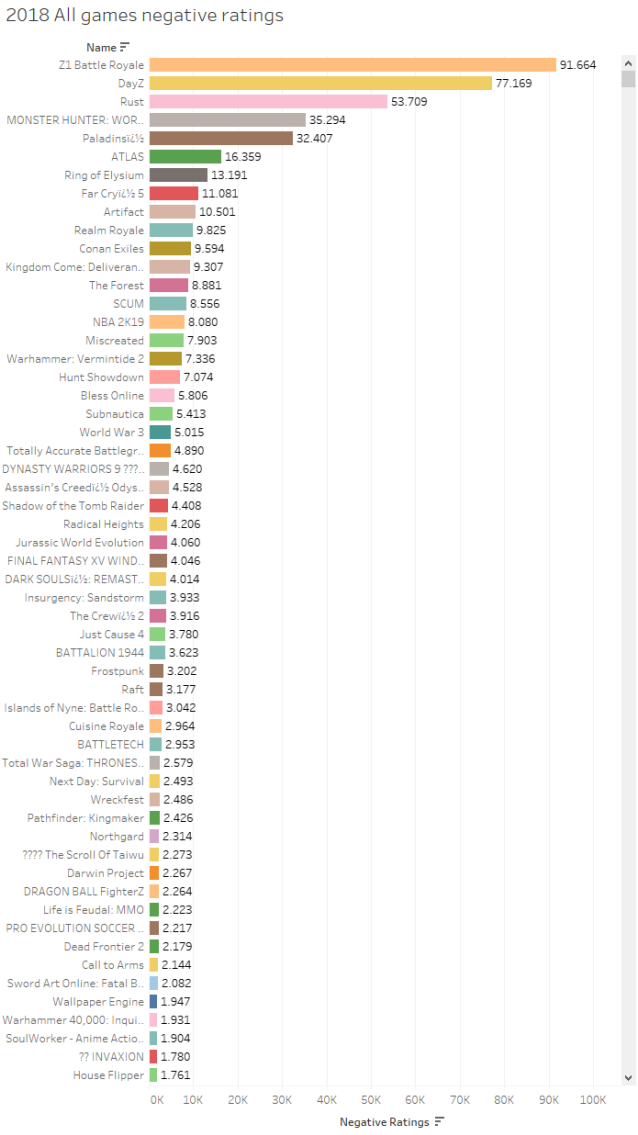
The most unliked games are PUBG and CS: GO, but cs:go has a lot of ratings so is normal that have a lot of negative ratings too.





In 2018 the most popular games are survival like “rust” and “the forest” and battle royal like “Dayz” and “Z1 battle royal” also open world are appreciated like “monster hunter” and “sub nautica”.





Conclusions

- In recent years was possible to see that the most popular games have many different genres and tangs. Means that games have become very complex and bigger then the past. The best way to develop a good game is to experiment a mix different genre of games, doing so is possible to reach different kind of player.
- Considering the trend of genres of the 2018 (last complete year of the dataset), we notice that the Indie games are the most developed. This is an important result that highlights the great growth of the video games sector. Not too many years ago this was very difficult, because the small developers did not have the technologies suitable for development, they were not free to use. So, this is good news, because large companies can take note of these results and maybe finance small future projects, as is already happening.
- Steam is a good starting point where publish your own games, it's the most used client for pc games this makes it possible for your games to be noticed by many players and the ratings and reviews can help expand your user base even further.
- Valve is no longer publish new games, that means Steam is no longer the Valve client for games, but it became THE client for games. Valve is targeting services like cloud saving, chat and call, dedicated server, but the most important is the contribution that valve has made for Gaming on linux with their own OS.