Visually Representing Video Game Data

- Data Visualization 2023 -

Project Report

Group 5

University of Southern Denmark - Odense
Software Engineering

Software Engineering
University of Southern Denmark
https://www.sdu.dk/

Title:
Visually representing differences in Video Game data

Theme:
Data Visualization

Project Period:
Fall Semester 2023

Project Group:
5

Participant(s):
Steffen Buhl Krøis
Daniel Vilhelm Kim
Martin Løvborg
Andreas Timm Adriansen

Supervisor(s):
Parisa Niloofar

Page Numbers: 15

Date of Completion:
December 27, 2023

Abstract:

Video games have been a fun and creative entertainment source for many years. The sheer growth of game sales, as well as other encapsulating facts about video games like the popularity of different genres and the earning of publishers, goes unnoticed in the daily life of single individuals. In this paper we explore the visualization of the data, from the years 1980-2016 of game sales, game rating, genres and more throughout several years. This is accomplished by building a Python Dash Dashboard to improve the quality of the visual data display. The dashboard clearly displays a growth in video game sales globally each year. Along with that, our finding also shows that the most highly regarded game, based on the feedback from the users, is Need for Speed: Most Wanted. 11 different graphs have been created to display the data in a meaningful and insightful way.
Through the use of Python Dash, a detailed and interactive visualization of exciting video game statistics have been displayed, and many personal curiosities satisfied.

Contents

Chapter 1

Background and Motivation

Video games have been a ubiquitous and captivating form of entertainment for several decades, captivating the imaginations of millions around the globe. From the early days of pixelated adventures to the immersive and visually stunning worlds of contemporary gaming, the evolution of video games has mirrored technological advancements and cultural shifts. Despite the general public's infatuation with video games, the group have no firsthand experience with the global landscape of video game sales, which made it worthwhile to investigate.

Our enduring fascination with video games is rooted in their ability to transport players to alternate realities, offering a unique blend of narrative, interactivity, and artistic expression. The industry's growth and diversification have been remarkable, with an ever-expanding array of genres and platforms catering to diverse tastes and preferences. However, while the allure of video games is evident, a deeper understanding of the global dynamics, including sales patterns, critical reception, and genre-specific earnings, has been lacking.

The motivation behind this report is to unveil the insights within the vast realm of video game data. As enthusiasts, researchers, and consumers, we have long marveled at the evolving landscape of video games, but we now seek a deeper understanding of the quantitative aspects that define this dynamic industry.

Chapter 2

Project Objective

The objective of this report is to expand knowledge and satisfy curiosity.

The group wishes to learn how the global sales of video games progressed throughout the years. And how the differences in sales differ, based on different regions. The group is particularly interested in the EU and NA regions.

The group is also heavily interested in the sales of different genres of games. The group wants to see how each genre of games, stacks up to each other in sales to learn what type of games the world likes on a global scale.

Furthermore, the group is interested in the most popular games. Popularity is derived from both the user score, but also critics score, so the group would like to see how that differentiates, and to see if there are any clear outliers between the two.

The group is also interested in which of the platforms released by the various companies, have the most game sales on them.

And lastly, also which publishers have sold the most.

Chapter 3

Data

The ”Video Games Sales Dataset” on Kaggle contains data about video game sales across various platforms. It includes information on game names, platforms, year of release, genre, publisher, and sales in North America, Europe, Japan, and other regions. The dataset is useful for analyzing trends in the video game industry, including regional preferences and the popularity of different genres and platforms. The data set contains 3 csv files, one with sales from ps4, one with sales from xboxone and lastly one with all sales data which this project will use. Link to Dataset.

The data file contains a little over 16k entries, of those 11562 has a unique ‘Name’, it has 16 columns of which 7 are of String type, the rest are numerical integer/decimal. The dataset contains the following 10 base columns Name, Platform, Year_of_Release, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, Other_Sales and Global_Sales. On top of that only approximately 60% of the entries contains data for all of the additional columns which are Rating, Developer, User_Score, Critic_Score, User_Count and Critic_score.
Critic_score is an Aggregate score compiled by Metacritic staff, Critic_count is the number of critics used in coming up with the Critic_score. User_score is the score by Metacritic’s subscribers, User_count is the number of users who gave the user_score, Developer is the party responsible for creating the game and finally Rating which is the ESRB ratings.

This dataset features information about video game sales, ratings, and other relevant details. It’s designed for comprehensive analysis of the video game industry’s trends and patterns up to December 22, 2016.

Chapter 4

Visualization

4.1 Requirements

To effectively display the data from the dataset, the group will develop a visualisation dashboard with a design focused on delivering an interactive user experience. This dashboard will allow users to delve into the different aspects of the video game sales data through various interactive graphs.
The dashboard is to be a general overview of the dataset, but to make sure it is sufficient the must have features for the visualisation dashboard include:

- Graphs for visualising regional data:
  - Charts showcasing the relationship between regional sales and time.
- Graphs for visualising data regarding genre:
  - Charts displaying the difference in sales for the different genres.
- Graphs for visualising game popularity and rating:
  - Charts showing the best games by user and critic scores.
- Graphs commenting on platforms and publishers:
  - Charts describing which platforms and publishers came out on top.
- Interactivity:
  - The graphs of the visualisation must be affected by the following choices of the user:
    * Which years do you want to display?
    * Which regions?
    * Which platforms are you interested in?

    * Which publishers?
    * And lastly, which genre of games?
  - The graphs must also be interactable in the sense that extra information will be displayed when the user hovers over them.
  - Animated graphs can also help in visualising the data.

4.2 Design choices

For visualising the regional data, the group has chosen to make the following graphs:

- Line chart displaying yearly sales distribution for each region.
- Bubble chart on a geographic map showing total sales for each region.
- Animated bar chart showing yearly sales distribution for each region.

The line chart is ideal for displaying the yearly sales distribution for each region. Its strength lies in its ability to clearly illustrate trends and changes over time. By connecting the data points with lines, it also becomes easier for users to track rises and falls in sales, observe patterns, and compare different regions simultaneously.
The bubble chart on a geographic map is chosen to show total sales for each region. This type of chart is effective because it allows for the representation of data in terms of location we cannot capture visually elsewhere. The size of the bubbles also visually shows the user the difference in total sales. This type of graph therefore provides a spatial context to the sales data, helping users understand how sales are distributed geographically. Unfortunately, we only have a few regions to gather data from and it would be better if we could narrow it down by country, or perhaps just more regions.
For a dynamic representation, an animated bar chart is used. By animating, a time dimension is added to the standard bar chart, showing how sales distributions change year by year. This animation helps in focusing on a year at a time. It is also effective in telling a story of how sales trends have evolved in each region over time.

The genre visualisations will consist of the following graphs:

- Pie chart showing the percentage of total sales for different genres.
- Horizontal bar chart showing total sales for different genres in numbers.

The pie chart is used to show the percentage of total sales across different game genres. Its strength is in illustrating the market share of each genre in a simple, easy-to-understand format. However, pie charts have limitations, especially when dealing with a large number of categories or when the differences in data points are subtle. They can make it difficult to accurately compare the segments of the chart, especially when angles are close to each

other.
To overcome the limitations of the pie chart, a horizontal bar chart is also used, showing the total sales for different genres in numbers. This format allows for easier comparison between genres as it is simpler to compare lengths than angles.

When visualising the games' popularity, it will be done with:

- A bar chart showing the top ten video games by user score.
- A bar chart showing the top ten video games by user vote count.
- A bar chart showing the top ten video games by critic score.
- A scatter plot showing the relationship between the user scores and the critic scores.

The use of bar charts for showing the top ten video games by user score, user vote count, and critic score is due to the same reasons stated before. Bar charts are effective here as they allow for clean comparisons across different games and different metrics. They clearly highlight which games are leading in each category.
The scatter plot comes in handy for showing how user scores and critic scores relate to each other. It's great for spotting patterns, trends, or even unusual points in the data. Each game is represented as a dot, with its position based on its user and critic scores. This setup makes it easy to see if these scores usually agree or if they often tell a different story.

Lastly, we have the graphs for commenting on the platforms and publishers:

- A bar chart for showing the total sales of each platform.
- A bar chart for showing the top ten publishers based on total sales.

They will be shown as bar charts for the same reasons as mentioned earlier.

Chapter 5

Results

The answer to the groups first objective is shown in the first 3 graphs. These are the ones that shows the sales distribution in each region in millions, the graph types are line, geo scatter and animated bar chart. Two of those(line and animated bar) also showcases the combined global sales of the video games.
The line plot highlights the evolution of region sales over time, from 1980 to 2020. Most notable are the years 2000 and 2003, where after some growth the sales dipped a little before getting a even bigger growth. Peak sales was in 2008, where the global total amount of video game sales was 671.79 million, in which EU sales consisted of 181.14 million and NA sales consisted of 348.69 million.

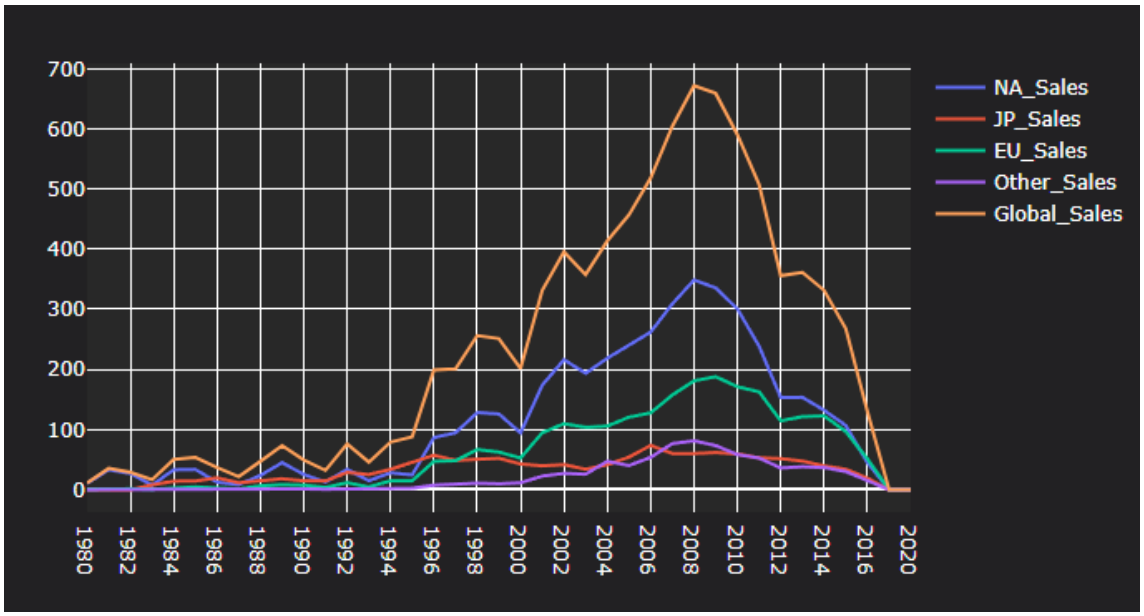

Figure 5.1: Line Plot for region

What the scatter graph(useful for correlation and relation) with a world map in the background spotlights a bubble of each region located in the right place on the map (other sales have been located in the middle mostly south of the globe to represent south America, Africa and Australia). A quick overview tells that NA sales are nearly double the amount of the EU sales, with NA sales being 4335.94 million and EU sales being 2397.29 million.

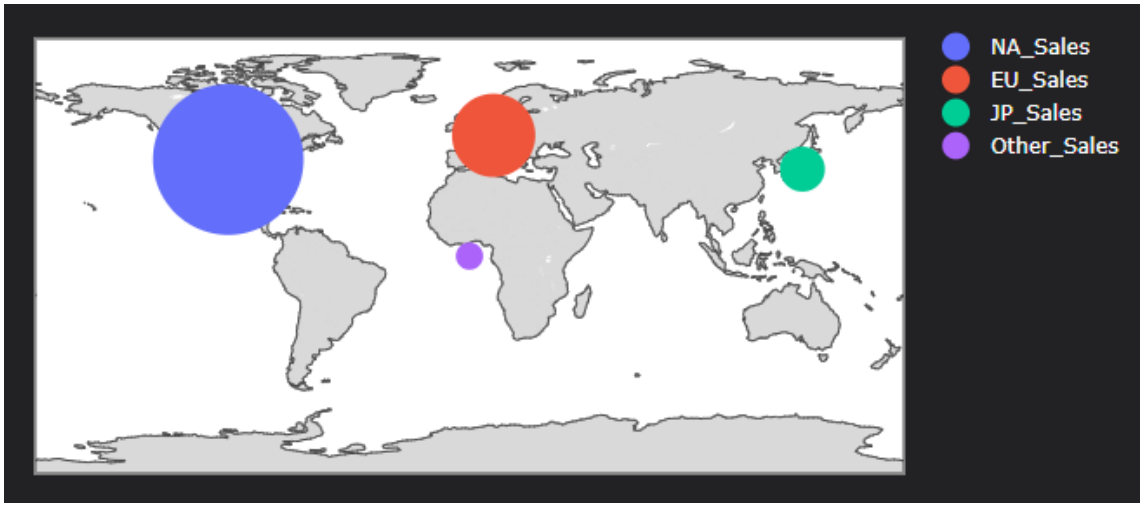

Figure 5.2: Geo Scatter Graph for region

As for the animated bar chart(good for rankings), it visualizes the sales increasing and decreasing each year. This chart inform that the NA region came of to a quick start in video game sales, before the other region, and it never let that lead go. While the EU region for the most part being behind it in second. It is also easier to see that the NA sales consist of around half of the total sales.

The second objective was about the genres of the video games, that was sold. For this objective, the group made a pie chart and a bar chart to answer the question. Both displays are over the distribution of sales by game genre, where as the pie chart present the sales in percentage of the whole video game sales and great for comparisons, meanwhile the bar chart presents the sales numbers in millions.

From the two chart it can be said that action, sport and shooter genre are the most liked in the world with all over 10% and 1000 million of global sales (action: 19.5% & 1717.62 million, sports: 14.9% & 1310.39 million, shooter: 11.8% & 1041.83 million). How they stack up against the least liked genres, which are puzzle, adventure and strategy. These genre are all under 5% and 300 million of global sales (puzzle: 2.73% & 240.33 million, adventure: 2.65% & 233.25 million, strategy 1.96% & 172.69 million).

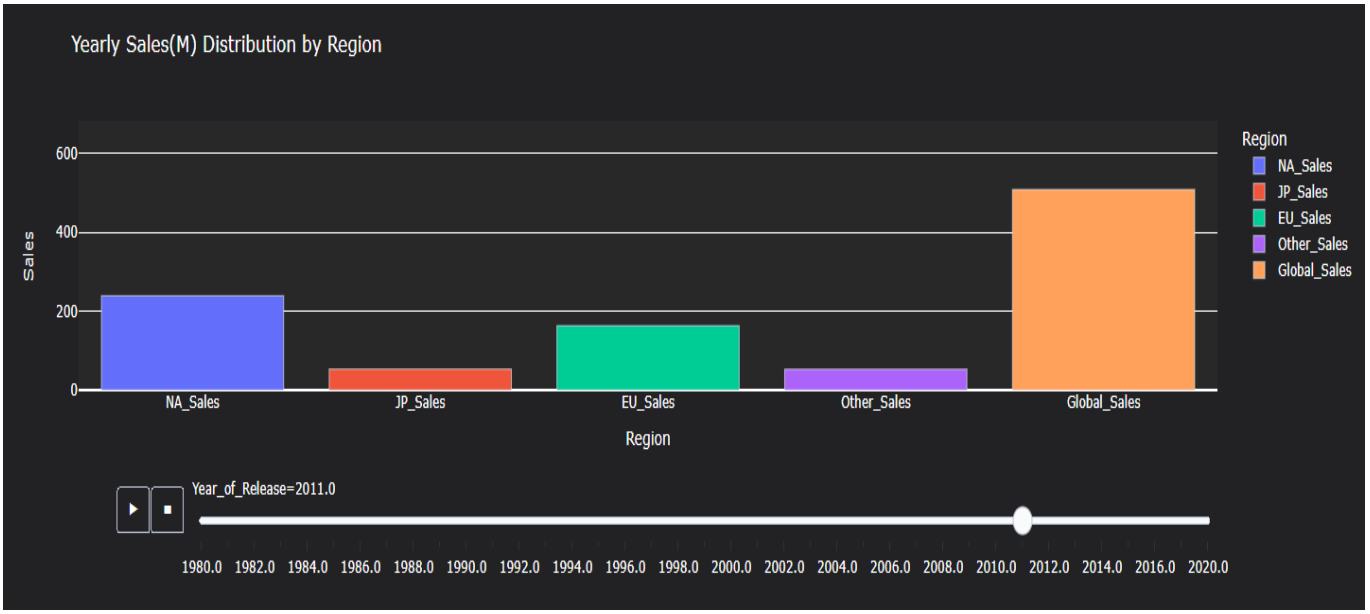

Figure 5.3: Bar Chart for region

Next are the objective around the popularity of the games, here the user score, user vote count and critic score will be looked upon. 3 bar chart have been created for each parameter, with and extra scatter graph that compares critic and user score. The bar chart only shows the top 10 games, but the scatter graph shows all games.
The user score bar chart has "Need for speed: most wanted" in first with a 75.4 score and "Need for speed carbon" at number 10 with a 51.7 score. So quite a differences just between top 10, first place is also rated with somewhat of a big margin compared to second place("Lego star wars II: the original trilogy" at 63.5 score).

For the user count bar chart, first place has a big lead on second and second is also quite a distance away from third. In first are "the witcher 3: wild hunt" with just under 25k votes, second ("call of duty: modern warfare") have just a little over 20k vote counts, third ("the elder scrolls V: skyrim") has around the 15k mark of counts. "Counter-strike: source" in number 10 has 9851 votes, so again in the top 10 there are a big margin. Interestingly none of the top 10 game for the user score and count charts are the same, which means for those games that has most votes have not been voted with a high score or has but also have a good amount of bad scores. It also means that the games that have a high user score have been voted by in comparison few people.

When it comes to the critic score bar chart, there is not that big of a differences between first place ("madden NFL 07" 624 score) and tenth place ("rayman origins" 511 score), like

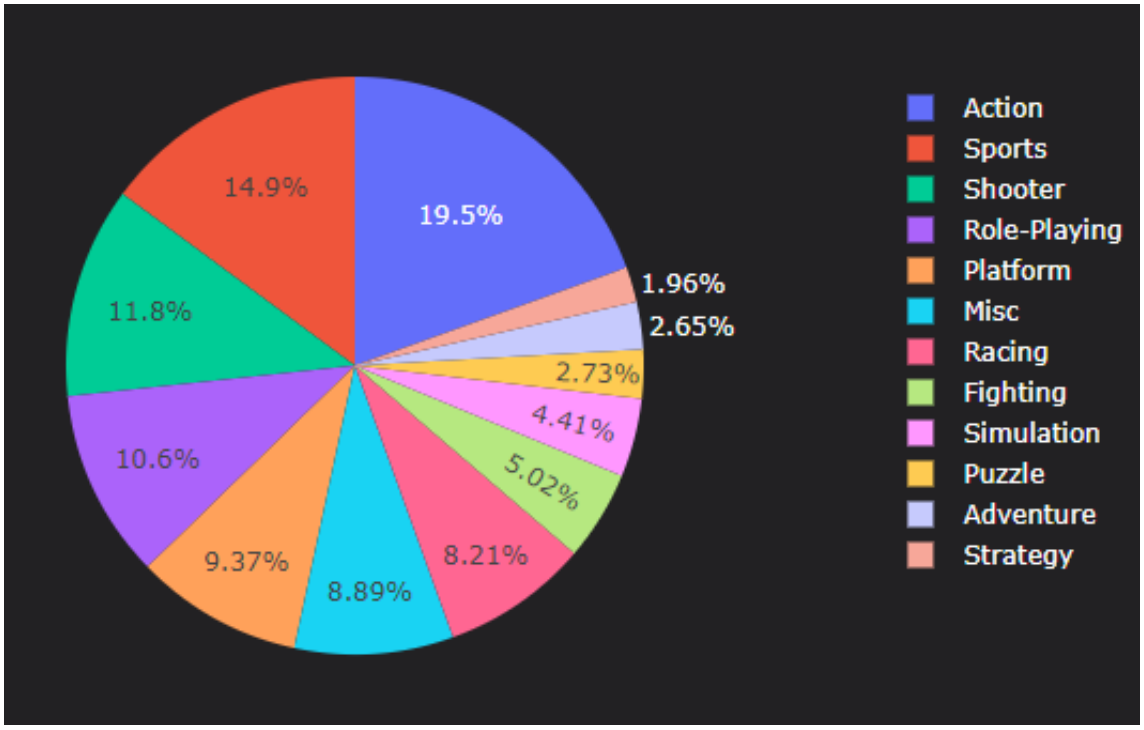

Figure 5.4: Pie Chart for genre

in the user score chart, but then contrast between the 2 can be seen upon. While there are some differences, such as 2 rayman games being in the critics top 10, there are more similarities with the same 6 games (examples: 2 need for speed games) making in both top 10. Although not in the same order, but it can be concluded that both user and critic alike like "Need for speed: most wanted" and "Madden NFL 07". In general the users likes Lego games more than critics, where as critic have a bit more diverse taste in games, however prioritises madden NFL games.

The scatter graph uses filtered data, so the top games from the other visualization might not be the same as this one. It can be seen that an outlier in the bottom is "ride to hell" with 13 in critic score and 1 in user score (disliked both) and one of the top rated game by both is "resident evil 4" with 96 in critic score and 9.4 in user score.

For what platforms have had the most game sales, a bar chart of the distribution on sales by each platform was developed. Here PS2 and X360 lead the top with 2467M and 1922M sales respectively, and at the bottom are the GG and PCFX platform with 0.08M and 0.06M sales in that order.

Lastly the objective that surround the publisher, and on which one have sold most.

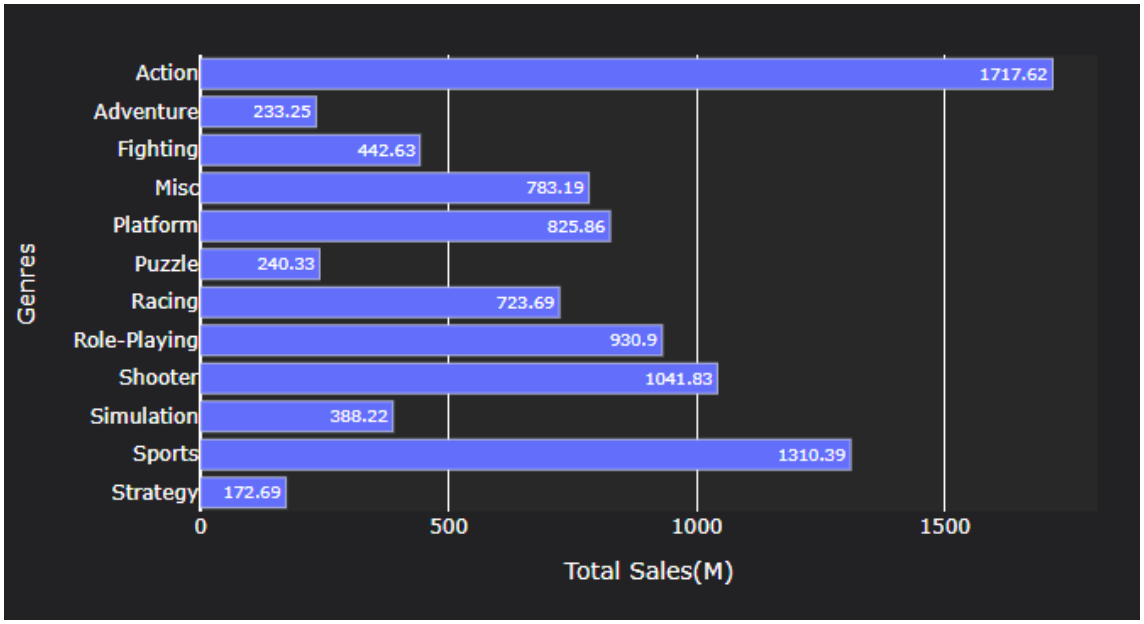

Figure 5.5: Bar Chart for genre

Another bar chart have been built to explain that objective. Only the top 10 publishers are shown, with "Nintendo" outselling every publisher with close to 1500M sales above electronic arts in second with 2199M sales. Last in the top 10 list are "Namco Bandai Games" with 508M sales, not too far away from place nine.

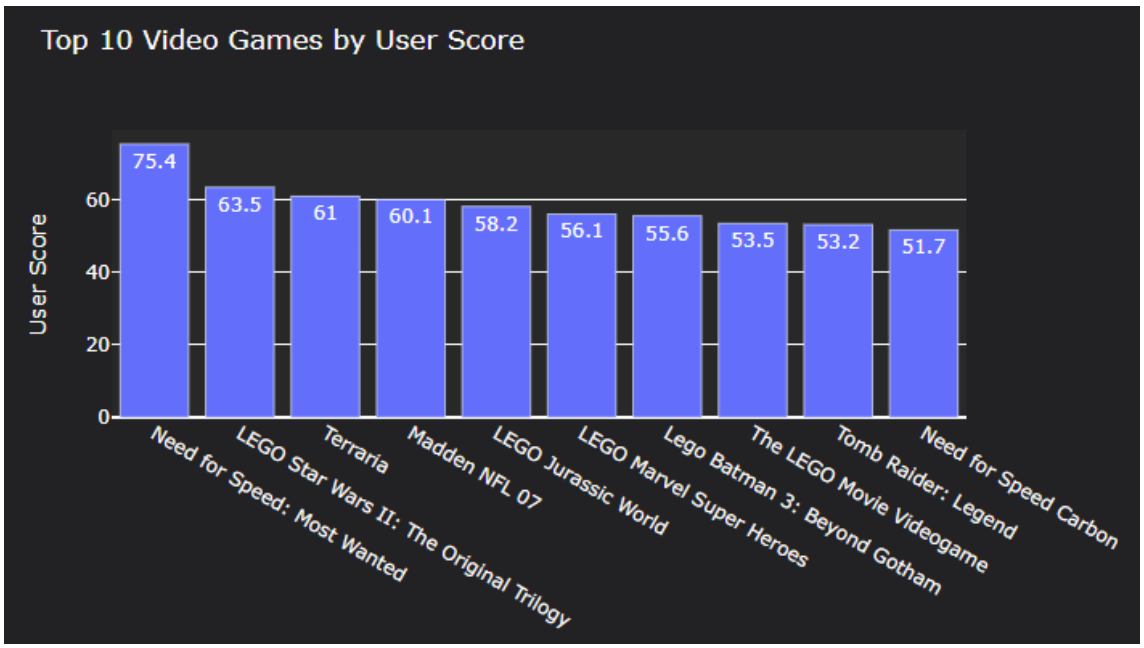

Figure 5.6: User score Bar Chart for popularity

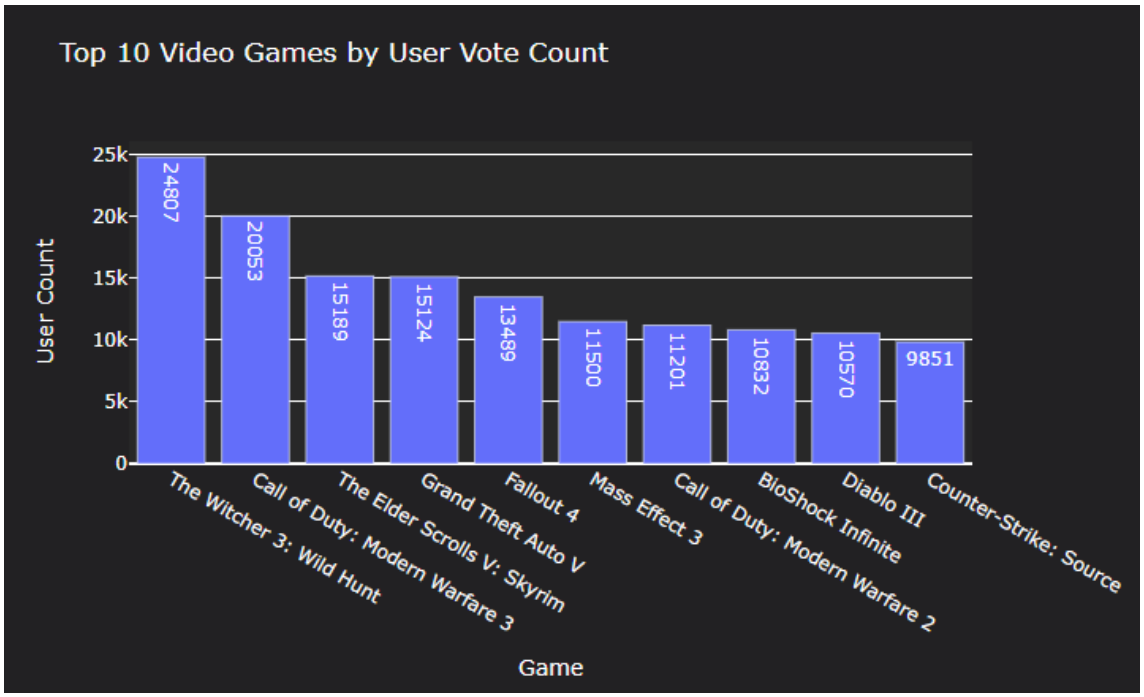

Figure 5.7: User count Bar Chart for popularity

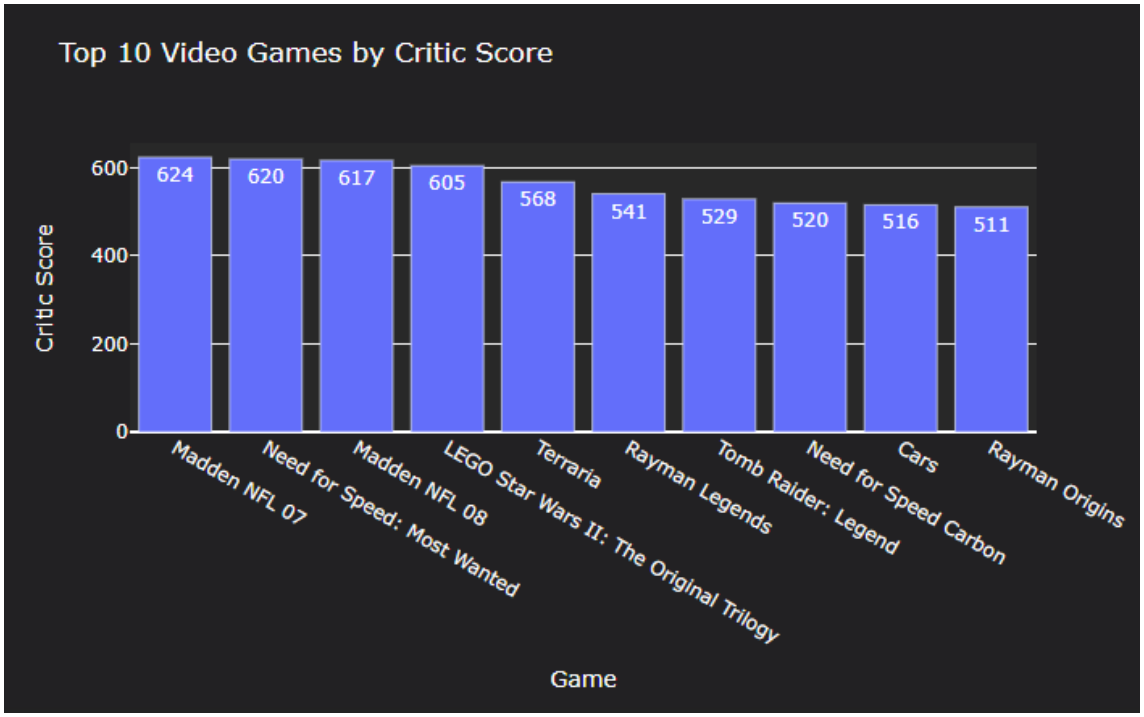

Figure 5.8: Critic score Bar Chart for popularity

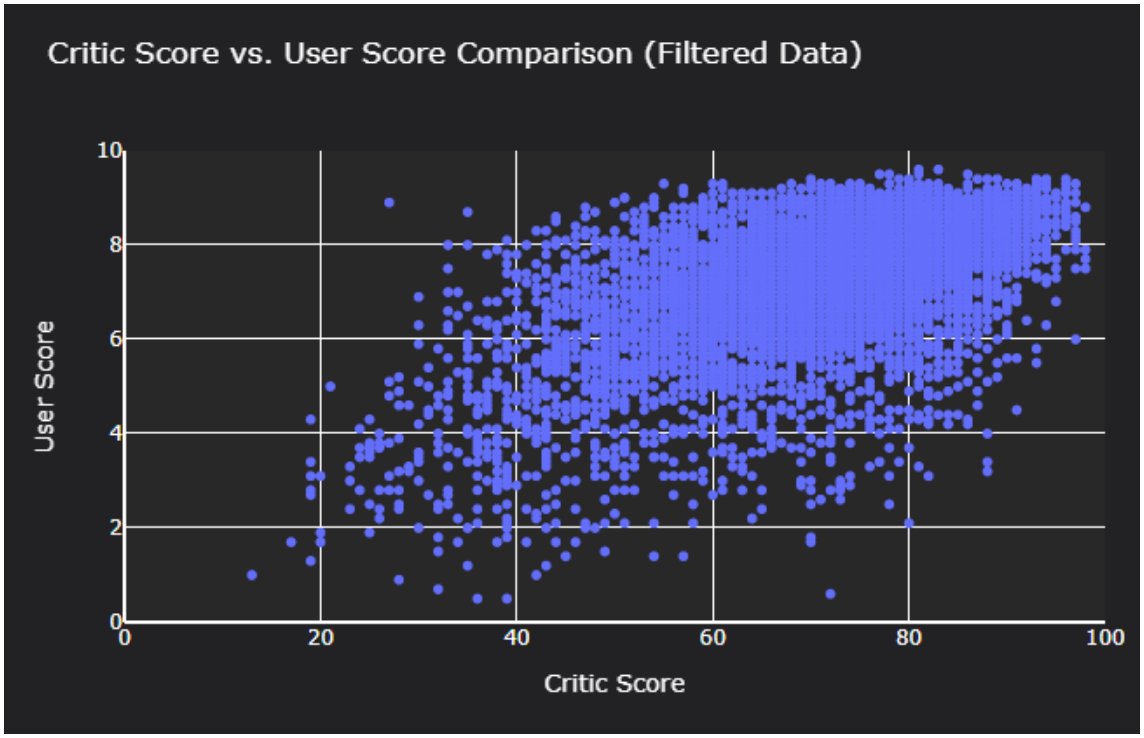

Figure 5.9: Scatter graph for popularity

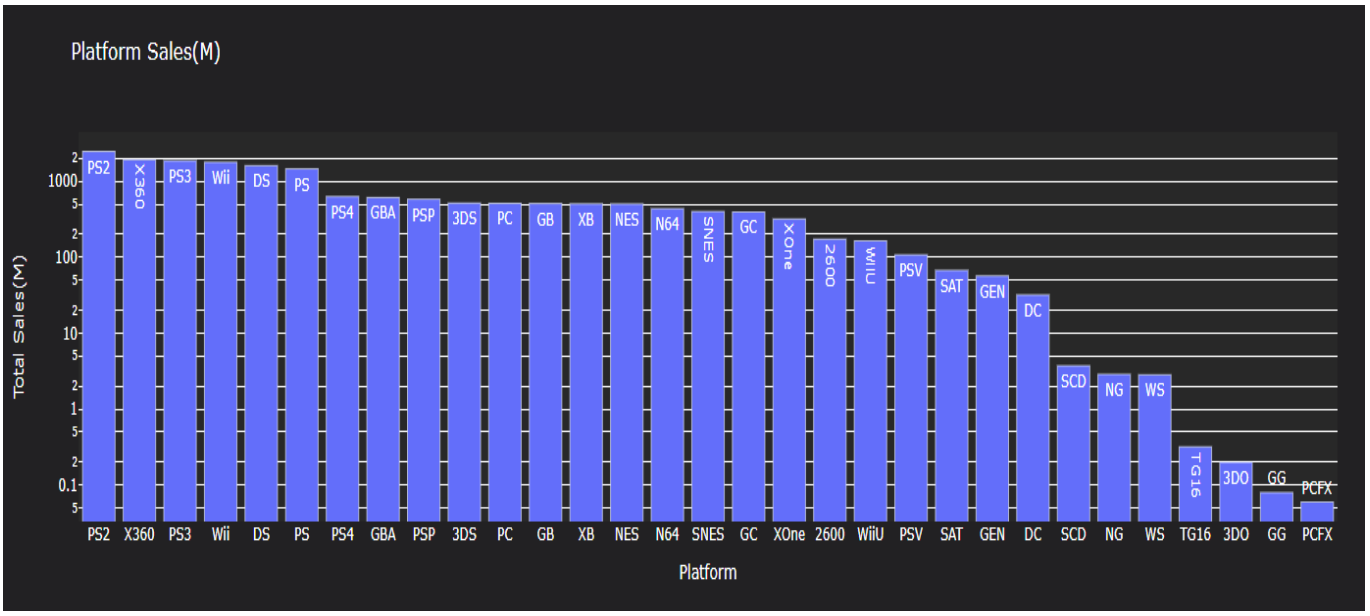

Figure 5.10: Bar Chart for platform

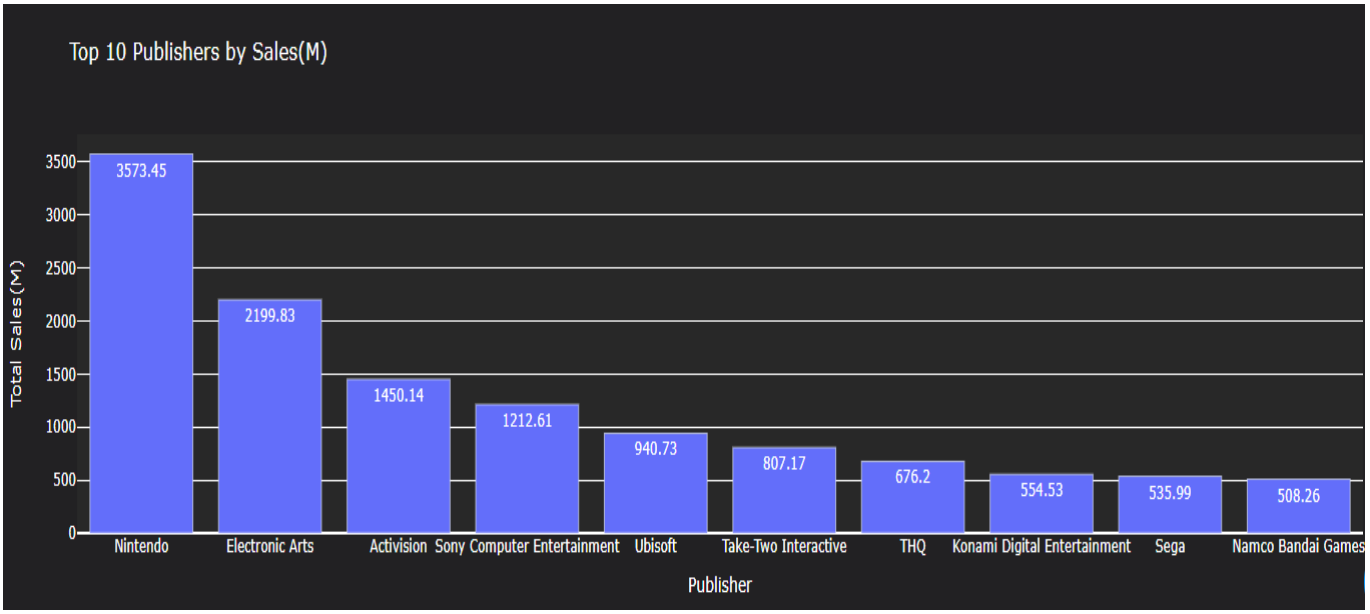

Figure 5.11: Bar Chart for publisher

Chapter 6

Conclusion and Discussion

6.1 Conclusion

An interactive dashboard with several different graphs, including an animated one have been created, and it accurately visualize the data from the chosen data set with the help of Python Dash. All the questions and curiosities have been answered by these graphs, and all the requirements have been met. The group thus conclude on the observations made, and unanimously agree that the task has been wholeheartedly fulfilled.

6.2 Discussion

One of the challenges the group encountered was our initial attempt to create a dashboard. Our first attempt was made with the programming language R, instead of our current solution which is in Python. Non of the group members have an in depth and extensive knowledge base in R, which resulted in us having a harder time initially than what the group had planned for. This eventually lead to the group swapping to Python, which everyone believed was the clear and correct choice for us. After that change, the creation of the dashboard felt a lot easier.

Appendix A

Appendix

Here is the link to the github which contains the code for the dashboard, and a readme file that explains how to set it up and run it.
https://github.com/martinloevborg/Data-Visualization-project

Grading contribution is 100% contribution form each member, no differences between anybody.