**Data wrangling – project report**

**PC game sales throughout the years**

**Author 1:** **Author 2:** **Author 3:** **Author 4:**

Dimitar Bachvarov Gustavo Amatsahip Kamil Glabinski Boyi Zhang

*[dbv900]* *[ggo380]* *[kgl202]* *[bzh201]*

1. **INTRODUCTION**

The gaming industry has undergone significant evolution since its creation in the mid-20th century. What began as a niche hobby has transformed into a multi-billion-dollar global phenomenon, reaching nearly every individual with access to technology, as in 2020, the global gaming industry’s revenue was projected to reach $165 billion, with 2.7 billion gamers worldwide [1]. Starting with simple pixelated games on arcade machines, the industry has expanded to encompass a variety of platforms, including consoles, virtual reality, mobile devices, and PCs.

PC gaming in particular has played a significant role in the development of video games, evolving from early text-based and strategy games into sophisticated, visually stunning experiences. The rise of esports, online multiplayer gaming, and streaming platforms, coupled with the creation of digital distribution services like Steam, has solidified PC gaming's dominance in the industry [2]. Today, PC gaming is not merely a form of entertainment; it has become a social and cultural phenomenon, connecting millions of players of all ages worldwide [3].

The importance of PC gaming was further highlighted during the COVID-19 pandemic, when everyone had to stay home, without being able to socialize with others in person, leading many to turn to gaming as a way to connect with others and alleviate isolation. This period also marked shifts in gaming trends, including sales patterns and genre popularity. Reports suggest that during 2020, global video game revenue grew by 20%, driven by an increase in digital sales and in-game spending, while platforms like Steam reported a 50% rise in user activity compared to 2019 [4][5].

This research aims to analyze PC game sales over the years, addressing the following questions:

1. **MAIN RESEARCH QUESTION**

How have PC game sales evolved over the years?

Sub questions:

What was the impact of COVID-19 on PC game sales?

How has the popularity of different game genres changed over time?

By addressing these questions, this project seeks to provide insights into the factors influencing PC game sales and uncover broader trends within the gaming industry.

1. **DATA WRANGLING METHODS**

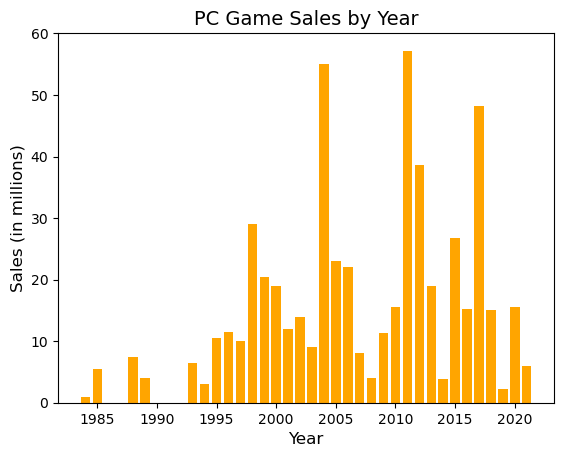
To clean and prepare the data, we started by filling in missing values in the "Series" column with "Unknown" so that no important rows were left out because of incomplete data. We then worked on the "Release" column, converting the dates into a standard "YYYY-MM-DD" format and splitting the year and month into separate columns to make it easier to analyse trends over time. If the dates couldn’t be processed, we marked them as "Invalid Date" and left the year and month blank. In order to avoid duplicates, we checked each row against a set of already-seen rows and then added the clean non-duplicate rows to the cleaned dataset. This left us with a cleaned and prepared dataset for further use and analysis.

1. **PC GAME SALES OVER THE YEARS**

From the early 1980s to the early 2000s, PC game sales showed steady growth, driven by the increasing accessibility of personal computers and advancements in gaming technology. A notable spike in 1998 can be attributed to the release of Starcraft, a genre-defining title that set new gaming standards. Although sales dipped slightly after this peak, overall growth persisted, reflecting the expanding interest in PC gaming. 

Three additional major spikes in sales highlight the impact of industry-changing games: World of Warcraft in 2004, Minecraft in 2011, and PlayerUnknown's Battlegrounds (PUBG) in 2017. These games not only revolutionised their respective genres but also boosted PC gaming's global prominence. However, such massive successes are outliers rather than the norm.

In the mid-2000s, the rise of the free-to-play model fundamentally shifted the gaming industry. Games relying on in-game purchases rather than upfront payments are not directly reflected in the sales data, potentially giving the impression of a decline. This transformation in business models highlights a redefinition of how success is measured in the gaming sphere.



In recent years (2015–2020), the data shows several smaller peaks but also sharper declines. This could reflect shifts in consumer preferences, increased competition from platforms like mobile and console gaming, or the growing dominance of digital and subscription-based gaming models—many of which focus on sustained revenue rather than upfront sales. Together, these trends underscore the evolving nature of PC gaming sales and the importance of considering alternative revenue streams to fully understand the industry's trajectory.

The sales chart underscores the dynamic nature of PC gaming, with periods of explosive growth driven by standout titles and underlying trends such as the adoption of new business models.

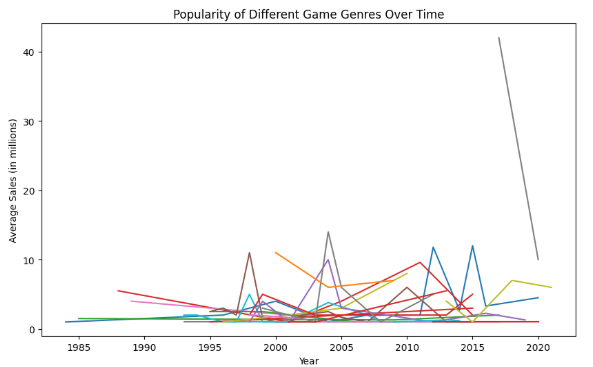
1. **COVID-19 IMPACT ON THE PC SALES**

To examine the impact of COVID-19 on PC game sales, the data was divided into two distinct periods: Pre-COVID (2018–2019) and COVID (2020–2021). The total sales for each period were calculated by summing up the Sales column for all games released within the respective periods. These totals were then visualized using a bar chart to compare the results. The analysis revealed a 95.93% decrease in sales during the COVID period, with total Pre-COVID sales amounting to 528.2 and COVID sales dropping to 21.5. A graph of sales

Description automatically generatedThe percentage decrease was calculated using the formula: (PD = percentage decrease)

This resulted in a substantial decline of 95.93%, suggesting a sharp drop in salesThis result defies expectations, as stay-at-home measures during the pandemic likely boosted gaming activity. The decline may reflect dataset limitations, discussed in the next paragraph. While global video game revenue grew 20% in 2020 due to digital sales and in-game spending [6], and Steam user engagement surged 50% [7], physical sales likely dropped due to restricted retail access, skewing the dataset. A broader dataset, including digital sales and post-2021 data, is needed to fully assess the pandemic's impact on PC game sales.

1. THE POPULARITY OF DIFFERENT GAME GENRES CHANGED



Correlation between Genre and Sales:Genre\_encoded: 1.000000 -0.051367

The popularity of different game genres has fluctuated significantly over time, with no single genre maintaining dominance throughout the years. The data suggests that different genres rise and fall in popularity, often experiencing short-lived peaks before declining, which could be influenced by market trends, technological advancements, or the release of blockbuster titles. The correlation analysis further supports this observation, as the correlation coefficient between genre and sales is -0.051367, indicating a very weak negative relationship. This suggests that no particular genre consistently outperforms others in terms of revenue, and game sales are not strongly tied to a specific genre. Instead, player preferences appear to shift over time, possibly driven by evolving gaming mechanics, changing storytelling trends, and the emergence of new gaming platforms.

The overall pattern suggests that the gaming industry is highly dynamic, with different genres gaining traction at different times rather than following a linear trajectory of popularity.

1. LIMITATIONS

One significant limitation of this project is that our dataset only extends to 2020. As a result, no subsequent developments or trends beyond 2020 are captured in our results. Additionally, in 2020, the COVID-19 pandemic was still actively influencing daily life and the gaming industry. Extending the analysis to include the post-pandemic period could provide a more comprehensive understanding of its long-term impact on PC game sales. This aspect will be addressed further in the "Future Works" section.

Another limitation is the exclusion of free-to-play (F2P) games, as the dataset focuses only on paid PC games. This creates a gap between our results and industry trends, where F2P titles drive significant growth. By omitting these games, the analysis may not fully capture the gaming landscape, particularly the impact of in-game purchases on engagement and revenue. This limitation skews the data and should be considered when interpreting the findings.

1. **DATA SOURCES**

Link: <https://www.kaggle.com/datasets/khaiid/most-selling-pc-games>

1. **CONCLUSION AND FUTURE WORK**

Our research explores the evolution of PC game sales, focusing on key industry trends and shifts in consumer behavior. By analyzing refined data from Kaggle through systematic data wrangling and metric analysis, we identified notable transformations driven by technological advancements, and external economic forces. One significant finding is the ongoing transition from physical to digital distribution channels, which has become a dominant trend in the industry. Fluctuations in sales patterns were observed in alignment with pivotal industry milestones, including the widespread adoption of online multiplayer gaming, the evolution of monetization models, and macroeconomic events such as the COVID-19 pandemic.

Future research should extend beyond 2021 to capture the influence of emerging technologies, such as AI-driven game development and cloud gaming, as well as the evolving patterns of consumer engagement in the post-pandemic era

1. **REFERENCES**

*[1] Statista. Evolution of the gaming industry.* [*https://www.statista.com/topics/868/video-game-industry/*](https://www.statista.com/topics/868/video-game-industry/)*.*

*[2] PC Gamer. Steam’s growth in digital PC gaming.* [*https://www.pcgamer.com/steam-growth-2020/*](https://www.pcgamer.com/steam-growth-2020/)*.*

*[3] Newzoo. Gaming: The social and cultural phenomenon. https://newzoo.com/gaming-trends.*

*[4] Statista. Global video game revenue during the COVID-19 pandemic. https://www.statista.com/statistics/1225093/global-video-game-market-value-covid/.*

*[5] PC Gamer. Steam saw record-breaking growth in 2020. https://www.pcgamer.com/steam-record-growth-2020/.*

*[6] Statista. Global video game market value during COVID-19 pandemic. Available at: https://www.statista.com/statistics/1225093/global-video-game-market-value-covid/.*

*[7] PC Gamer. Steam saw record-breaking growth in 2020. Available at: https://www.pcgamer.com/steam-record-growth-2020/.*