

**IS 525 Final project: Video Games Sale Analysis**

**Pete Chen**

**Bohan Shan**

**JingChen Liu**

**Shchool of Information Science , UIUC**

**IS525 Data Warehousing and Business Intelligence BC**

**Professor *Wonderlich***

**May 11,2025**

## **Introduction & Scenario**

With the gaming industry becoming increasingly saturated and diversified, developers and publishers need tools to track long-term genre trends and identify shifts in market focus. Our intended audience includes game industry analysts, publishing strategists, and product managers who seek to understand evolving trends in game release patterns, genre activity, and consumer engagement. This project is designed to support decision-making by offering insights into which genres are rising or declining in popularity, how publisher output aligns with market shifts, and how player demographics influence game preferences across time and regions.

The first dashboard focuses on the business and market performance of the Top 10 game publishers during the pivotal decade between 2000 and 2010. By examining best-selling titles, genre specialization, and regional market dependence, highlights the strategic dominance of major companies and their shifting roles in a rapidly evolving industry. The second dashboard explores long-term trends in PC gaming, integrating sales data, genre release patterns, and Steam platform engagement. It compares sales figures with estimated player ownership, revealing discrepancies between commercial success and player reach. This analysis is contextualized within 5-year intervals from 1995 to 2020, enabling a clear view of how different genres have grown or declined across time. The third dashboard brings in the demographic perspective, analyzing how variables such as age, gender, and region shape player behavior and genre preferences. By aggregating user behavior data by genre, it provides insights into who plays what, and how different audiences interact with various types of games, offering valuable context for market segmentation and targeted game development.

By integrating these three dashboards, our project provides a view of the gaming industry that moves beyond isolated metrics. Instead of analyzing sales, genre trends, or user behavior in isolation, we connect these dimensions to offer context-rich insights that are relevant to both strategic decision-makers and academic researchers. Whether the goal is to identify which publishers led the market during key periods, understand how genre preferences shift over time, or explore how demographic patterns influence gameplay behavior, this unified analytical dashboard is designed to support deeper exploration of the evolving video game landscape.

## Data Sources

The data used for this project was sourced primarily from Kaggle, consisting of multiple complementary datasets that together offer a broad and multifaceted view of the video game industry. The datasets cover a wide range of variables, including game metadata (titles, genres, release years), commercial performance (global and regional sales, recommendation counts), platform engagement (Steam player ownership estimates), publisher information, and user demographic behavior. These datasets were originally compiled from industry databases and platform APIs such as Steam, allowing for both traditional sales based analysis and user engagement evaluation.

One key dataset, titled *Gaming Companies*, served as the foundation for our publisher-focused analysis. It provided detailed information on game titles, publishers, genres, release years, and sales figures across regions such as North America, Japan, and globally. This dataset enabled us to assess the market dominance of top publishers and their reliance on specific geographic regions. Another dataset focused on PC games and Steam-based features, enabling the comparison of digital ownership data with traditional sales metrics. This was crucial for identifying discrepancies between commercial success and platform popularity, particularly in cases where free-to-play games had high user engagement but little to no revenue data. Lastly, a population dataset was used to support demographic analysis. This dataset included fields such as player age, country, game genre preference, playtime duration, and session frequency. It allowed us to explore how different population groups interact with various types of games, offering behavioral insights.

Collectively, these datasets span a wide range from the early 1970s to the 2020s. However, due to data sparsity before 1995 and after 2020, we limited our analysis to the 1995–2020 period to ensure consistency and reliability. Together, the datasets are diverse, well-structured, and suitable for comparative and interactive visual analysis across genre trends, publisher behavior, and player demographics.

### **Informal Client**

To ensure the dashboard addressed the needs, we consulted several informal clients throughout the development process, drawing feedback from both strategic minded individuals and general gaming audiences. One of our reviewers was a Business Analytics student who offered insights from a market strategy perspective. She was particularly interested in identifying which genres showed long-term growth, how platform engagement, such as Steam ownership, compared to actual sales, and which underperforming genres had the potential for market entry. Her feedback encouraged us to include performance indicators, correlation charts, and genre-level analysis that support business decision-making.

In addition, we invited two avid gamers to provide usability and content feedback. Their reactions highlighted the importance of visual clarity and intuitive chart design. Charts showing top titles and publisher output were noted as helpful for identifying standout games and trends at a glance. They also noted that while demographic breakdowns were rarely visible in typical platforms, the data revealed clear differences in genre preferences by age and region that matched their personal experiences. For example, they weren't surprised that players aged 31–45 preferred simulation and strategy games, but they appreciated how the dashboard quantified them. These user-driven feedbacks and data-backed analysis helped refine both the content and structure of the dashboards, ensuring they provide both business users and everyday players with relatable insights.

## Steps Taken & Analysis

### *Bohan Shan's Business Dashboard:*

We began by identifying our core objective: to analyze how the video game industry evolved over time, particularly through the lens of sales data, publisher performance, and market trends. Given the importance of video games in entertainment, culture, and global business, we wanted to explore this subject in a way that would be insightful for both casual players and industry professionals.

The “Gaming Companies” dataset offered a wide range of information, including game titles, release years, genres, platforms, regional and global sales, publishers, developers, and ratings. We quickly realized that organizing our analysis around game publishers would be the most effective strategy. Publishers are typically the central players in the industry, as they often oversee development, marketing, and distribution. Unlike developers, who may release only a few titles or work under a publisher, major publishers consistently release numerous games across genres and platforms. This made them a solid anchor for the dashboard.

Given the North American market's significant share of global video game sales, we decided to compare each publisher's performance in this region against their global performance. This allowed us to better understand regional dependence and market influence.

### *Pete Chen's PC Gaming Enterprise Dashboard:*

The dashboard was built in several key steps, beginning with data cleaning and transformation. Multiple datasets from Kaggle were standardized and preprocessed, including renaming inconsistent fields, handling null values, and creating a [Years by 5] field to group release dates into 5-year intervals. Next, calculated fields were developed, including [PC Sales], Genre by group, and a genre activity rate comparing each genre's output in a given period against the average. We also constructed publisher-related fields and a Steam ownership-to-sales gap to explore the platform dynamics. Visualizations were then created using bar charts, bubble plots, diverging bar charts, and performance indicators. Interactive filters for genre and time period allowed for flexible exploration.

Throughout the process, the dashboard was refined based on business-focused feedback, with an emphasis on trend detection and clarity in genre-level comparisons. One core analysis involved measuring genre activity rate: a calculated percentage that compares

the number of games released in a given genre and time period to the overall genre average for that period. This helped identify which genres were over- or under-performing relative to the market baseline. Another key analysis looked at Steam ownership versus PC sales, highlighting the gap between player engagement and commercial performance. Games with high player counts but low sales, often pointed to free-to-play models or bundled promotions. This comparison revealed how traditional revenue-based success metrics can differ from platform engagement. The dashboard also examined publisher output by genre and time, showing which companies dominated particular niches and whether their activity aligned with broader trends. Additional visualizations analyzed year-over-year changes in sales and release volume using diverging bar charts and performance indicators.

#### *JingChen Liu's Demographic Dashboard:*

The demographic portion of our dashboard was developed to explore how different population groups interact with games across dimensions such as age, region, gender, and genre preference. The first step involved identifying which variables from the dataset were most relevant for capturing behavioral differences, specifically playtime, in-game purchases, and genre engagement. We also added regional and gender-based comparisons to round out the behavioral portrait. The intent was to detect market-specific tendencies, such as the U.S. market's high engagement with Easy and Medium difficulty games, which may inform localization strategies. In addition, a gender-genre matrix was developed to visualize gender distribution across top game genres, using stacked bars. From there, the dashboard was structured to present these variables through a combination of five plots to offer layered insights.

## Discoveries

### *Bohan Shan's Business Dashboard:*

Upon exploring the dataset, we found a variety of useful fields that enabled rich analysis: game developer, genre, platform, publisher, rating, global sales, regional sales (North America, Japan), and year of release. One major realization was that focusing on publishers gave us a coherent and consistent way to compare success metrics across companies.

Initially, we considered analyzing data from 1985 to 2024, as the dataset covered this full range. However, the data prior to 2000 and after 2010 had significant gaps, particularly in fields like genre, sales data, and publisher consistency. Many older entries were incomplete, and newer entries (post-2020) were sparse or missing key information. To maintain analytical integrity and focus on a period with high-quality data, we narrowed our scope to the 2000–2010 decade. This 10-year span contained the most complete and reliable data, making it ideal for our purposes.

### *Pete Chen's PC Gaming Enterprise Dashboard:*

The dashboard revealed several key insights about the video game industry's evolution over time. One major discovery was that genre activity is highly cyclical, with certain genres like Action and Sports peaking in the early 2000s and then gradually declining, while others like Casual and Strategy gained sales in more recent years. This suggests shifting player interests and potential oversaturation in formerly dominant categories. Another important finding was the disconnect between Steam ownership and actual sales. Several games (free-to-play) had millions of owners on Steam but low or zero recorded sales, emphasizing that platform popularity doesn't always translate to revenue. This insight gives understanding of the need to look beyond traditional sales metrics when evaluating a game's success.

We also found that a small group of publishers consistently dominated output across time, especially in popular genres like Action-Based and Sports. However, newer genres showed more variety in publishers, hinting at potential entry points for smaller studios. Additionally, by comparing each genre's release volume to time period average, we discovered underperforming genres that were slowly gaining activity, potential indicators of market opportunities.

*JingChen Liu's Demographic Dashboard:*

The demographic dashboard uncovered several non-obvious trends in player behavior that have strategic implications for game developers and publishers. One major insight was that the 31–45 age group consistently exhibited the highest engagement across multiple behavioral metrics, including total playtime hours and in-game purchases. This challenges the industry's frequent overemphasis on the 18–30 segment and suggests that midlife players should be considered a high-value target audience, especially for games requiring long-term retention.

Another discovery came from genre-specific playtime analysis across age groups. The Simulation and Strategy genres showed high engagement from older age groups (31–45 and 46+), suggesting that these genres may benefit from deeper gameplay mechanics or narrative complexity tailored to more experienced gamers. In contrast, Action and RPG titles performed more evenly across age brackets, with slight skews toward younger audiences. While most genres were male-dominated in absolute player counts, Simulation and Strategy games had the most balanced gender distribution. These discoveries support informed content development and marketing strategy.



## Challenges & Adjustments

### *The Business Dashboard:*

Throughout the project, we encountered several challenges that required creative solutions and adjustments:

#### **Calculating Market Reliance:**

To create the “Reliance on the North American Market” metric, we had to manually calculate the ratio of North American sales to global sales for each publisher. This required generating a calculated field that allowed us to visualize the relative importance of the NA market to each publisher's business.

#### **Platform Duplication Issue:**

Initially, we planned to include a breakdown by platform ( Xbox, PlayStation, PC) along with top games and publishers. However, we discovered that many top games were released across multiple platforms. For instance, *Grand Theft Auto: San Andreas* appeared multiple times on different platforms, which led to redundancy in our top lists. As a result, we decided to shift the focus from platforms

### *The PC Gaming Enterprise Dashboard:*

There are several obstacles that we encountered while constructing the PC Gaming Enterprise Dashboard:

#### **Year Range Limitations**

Although the dataset spanned a wide timeline, from the 1970s to the 2020s, data before 1995 and after 2020 was extremely sparse and inconsistent. To ensure clean, meaningful time-based analysis, we restricted the [Years by 5]field to only include periods from 1995 to 2020, which balanced completeness with reliability.

#### **Extreme Outliers Skewed Visualizations**

Games with unusually high Steam ownership distorted scatter plots and bar charts, making most other data points unreadable. We resolved this by excluding outliers conditionally or using window functions like WINDOW\_MAX to normalize values and express them as a proportion of the highest value, rather than absolute counts.

### **Filter Sensitivity Broke Some Views**

Some charts, especially those using percent difference or comparison against average, broke when only one genre or one time period was selected. This was because calculations like WINDOW\_AVG require multiple data points for context. To solve this, we replaced fragile window calculations with more stable FIXED expressions and added logic to return placeholders like “—” when meaningful comparison wasn’t possible.

#### *The Demographic Dashboard:*

While building the Demographic Dashboard, we encountered several challenges related to segmentation logic and metric interpretation:

### **Overlapping Age Groups with Uneven Distribution**

The dataset included multiple overlapping or narrowly defined age brackets, some of which had very low player counts. This made it difficult to compare engagement meaningfully across all age segments. To address this, we regrouped the data into broader, more balanced age ranges (e.g., 0–18, 19–30, 31–45, 46+), which allowed for clearer trend visualization and reduced noise in charts like playtime and genre preference by age.

### **Genre Bias in Small Subgroups**

In some regions or gender groups with fewer entries, genre preferences appeared skewed or unreliable due to sample size limitations. This was especially noticeable when comparing female player data across niche genres. To ensure integrity, we added tooltip disclaimers and avoided over-interpreting thin slices of data, focusing instead on broader trends across higher-volume groups.

## Conclusion

This project brought together three complementary dashboards to analyze the video game industry through the lenses of market strategy, platform dynamics, and player demographics. By focusing on publisher performance, genre evolution, and population behavior, we constructed a unified framework that highlights how different forces shape the gaming industry.

From a business standpoint, we identified the critical role of top publishers during the 2000–2010 decade and revealed how regional markets, particularly North America, influenced their global success. On the PC gaming front, our analysis showed that genre activity is very cyclical, and that Steam ownership doesn't always align with commercial revenue, which underscores the growing importance of alternative monetization models. The demographic dashboard further explored the understanding by illustrating how engagement patterns vary significantly by age, region, and gender. Such as that female players showed above-average engagement in specific genres often overlooked by mainstream development.

By integrating insights from these three perspectives, our project goes beyond static dashboards, it offers a multidimensional view of the industry that can inform strategic decisions across publishing, game design, marketing, and audience targeting. This approach reflects the thinking that is required in real-world analytics and offers practical value to stakeholders across publishing, product management, and game development. Moving forward, further exploration could involve incorporating dynamic behavioral data such as multiplayer participation, achievement tracking, and retention over time to build a richer profile for the gaming industry clients.

## References

Ali, H. M. (2025, February 17). *Video game sales 2024*.

Kaggle. [Video Game Sales 2024](#)

Devastator, T. (2023, December 20). *Steam games user statistics and features*.

Kaggle. <https://www.kaggle.com/datasets/thedevastator/steam-games-user-statistics-and-features>

Chaves, M. F. (2023b, July 11). *Popular video games* .

Kaggle.

<https://www.kaggle.com/datasets/matheusfonsecachaves/popular-video-games/data>

Kharoua, R. E. (2024, June 21). *predict online gaming behavior dataset*.

Kaggle.

<https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset>

Kirubi, R. (2016, December 30). *Video game sales with ratings*.

Kaggle.

<https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings%20>