

# Visual Analytics Application for Video Game Sales Dataset

Anastasiia  
Tykholiz  
2184571

Daria  
Serhiienko  
2184570

Oleksandra  
Panibratets  
2187414

**Abstract** — This paper introduces a visual analytics application for exploring video game sales data. Using a Kaggle dataset of the top 100 video games, extended with recent records, the system analyzes regional preferences, platform-genre dynamics, and publisher performance. By integrating time-series analysis, correlation analysis, and PCA, the tool provides interactive visualizations like time-series graphs, heatmaps, and dimensionality reduction plots, offering valuable insights for game developers and marketers.

## I. INTRODUCTION

The video game industry has grown exponentially, surpassing film and music in revenue and reshaping digital entertainment. Expanding platforms—consoles, PCs, mobile, and cloud gaming—have diversified demographics and business models.

Understanding sales trends is crucial for developers, marketers, and publishers. Factors like technological innovation, regional preferences, and genre popularity influence market dynamics. Traditional methods struggle to capture this complexity, making visual analytics essential for extracting insights from vast sales data.

This paper presents a visual analytics tool using Kaggle data to explore sales trends, genre preferences, and platform success. It integrates time-series analysis, correlation analysis and PCA, with interactive visualizations plots.

Aligned with our studies at Zhytomyr Polytechnic State University, this project applies data-driven techniques to gaming analytics. The insights support developers and marketers in understanding trends and optimizing strategies. By offering an intuitive tool, this research enhances visual analytics and aids industry professionals in making data-driven decisions.

## II. RELATED WORK

In this section, we review prior research on visual analytics and data-driven approaches in video game

sales analysis, which form the foundation of our study:

1. Interactive Dashboard for Video Game Sales: Explores visualization techniques and predictive analytics to identify trends and forecast sales. [1]
2. Sales Trends (1980–2020): Analyzes historical market trends to help stakeholders understand and optimize revenue potential. [2]
3. Platform Impact on Sales: Uses a Kruskal-Wallis test to evaluate platform-specific sales performance (2006–2011). [3]
4. Game Analytics Review: Categorizes business intelligence methods for game development, marketing, and player behavior analysis. [4]
5. Global Video Game Industry Visualization: Examines consumer purchasing trends to guide development and marketing strategies. [5]

These studies highlight the importance of visual analytics in uncovering insights and supporting data-driven decisions in the gaming industry.

## III. PROPOSED METHOD

Our visual analytics approach consists of multiple systematic steps designed to analyze and visualize video game sales data. By leveraging interactive data visualization techniques and statistical analysis, we aim to uncover trends in platform and genre popularity, regional sales distribution, and publisher performance. Below, we detail each step in our methodology, highlighting the data processing pipeline and the analytical methods used in our study.

### 3.1 Data Preprocessing

Before conducting any analysis, data preprocessing is essential to ensure data quality, consistency, and accuracy. The Kaggle dataset contains 16,598 records with 11 attributes, including sales figures across different regions (North America, Europe, Japan, and Other). However, raw datasets often contain

inconsistencies, missing values, and variations in data formatting, which need to be addressed before conducting meaningful analysis.

### ***3.1.1 Handling Missing Values***

Missing data is a common issue in large datasets, particularly when integrating multiple sources. In our dataset, some entries lack sales figures for specific regions or do not include the release year of a game. To handle these missing values, we employ data cleaning. If a record lacks multiple critical attributes (e.g., missing both sales and platform information), we consider removing it to maintain data integrity.

### ***3.1.2 Data Augmentation***

To make our analysis more comprehensive, we augmented our dataset by using Python-based web scraping to extract additional, up-to-date records from the same resource from which the original dataset was obtained, since that dataset had not been updated for a long time.

## **3.2 Time-Series Analysis**

One of the key analytical techniques in our approach is time-series analysis, which helps us understand how video game sales have evolved over time.

### ***3.2.1 Trends in Platform and Genre Popularity***

By analyzing sales data across different years, we uncover how platforms and genres rise and fall in popularity. Some of the key insights we aim to extract include:

- **Platform Evolution:** How console generations (e.g., PlayStation, Xbox, Nintendo) perform over time and how newer platforms impact industry trends.
- **Genre Shifts:** How the popularity of certain game genres (e.g., RPGs, Shooters, Sports) changes based on consumer demand and technological advancements.

### ***3.2.2 Regional Sales Trends***

Video game sales vary significantly across different geographical regions due to cultural preferences, market

dynamics, and local regulations. Using time-series analysis, we can:

- Compare North America, Europe, Japan, and Other regions to identify market-specific trends.
- Analyze how economic events, technological innovations, and platform exclusivity influence sales in different areas.

For example, we expect to observe: A strong preference for shooters in North America, driven by franchises like Call of Duty and Halo.

- Higher RPG sales in Japan, with dominant titles from Square Enix and Bandai Namco.
- A balanced distribution of game genres in Europe, reflecting diverse consumer preferences.

## **3.3 Correlation Analysis**

To gain deeper insights into relationships between different variables, we perform correlation analysis on our dataset.

### ***3.3.1 Genre-Regional Correlation***

Certain game genres tend to perform differently across regions due to cultural preferences, market trends, and regional consumer behavior. By examining correlation matrices, we identify:

- That shooters exhibit skewed (or uneven) sales figures in North America, suggesting a dominant performance in that region.
- That role-playing games (RPGs) tend to prevail in Japan, reflecting regional tastes and market demand.
- That strategy and simulation games show distinct sales patterns in certain regions, highlighting unique regional market characteristics.

### ***3.3.2 Regional Sales Correlation***

By comparing sales distribution across regions, we uncover: whether a game's success in one region predicts its performance in another (e.g., do high NA sales correlate with strong EU sales?), whether

certain genres exhibit cross-regional popularity, helping publishers optimize marketing strategies.

### 3.4 Principal Component Analysis (PCA)

#### 3.4.1 Why PCA?

The video game sales dataset contains multiple dimensions (e.g., platform, genre, regional sales), making it to discern underlying patterns. Principal Component Analysis (PCA) reduces this complexity by projecting the data into a two-dimensional space, revealing hidden structures and relationships among games. This transformation enables us to better understand overall sales dynamics and market behavior.

#### 3.4.2 Identifying Market Segments

By applying PCA, we can visualize how different platforms and genres cluster together based on sales data. This allows us to: identify distinct market segments based on regional sales performance. Reveal genre-specific trends within these segments. Provide actionable insights for game developers and marketers regarding which platforms or genres perform best in each market.

### 3.5 Cross-Tabulation Analysis

Cross-tabulation is a statistical technique used to examine interactions between categorical variables. In our study, we apply it to analyze the relationship between genres and regions.

By creating cross-tabulation tables, we uncover which genres dominate specific regions (for example, shooters leading in North America, while role-playing games prevail in Japan).

### Summary of Proposed Method

This structured visual analytics pipeline provides a comprehensive approach for analyzing and understanding video game sales data, offering valuable insights for game developers, publishers, and marketers.

## IV. DATASET AND BENCHMARK

The dataset used in this project is sourced from Kaggle[6], containing 16,598 records with 11 attributes:

Rank: Global ranking of the game.

Name: Title of the game.

Platform: Gaming platform(s) for which the game was released (e.g., PlayStation, Xbox, PC).

Year: Year of release.

Genre: Game genre (e.g., Action, Adventure, Shooter).

Publisher: Game publisher.

NA\_Sales, EU\_Sales, JP\_Sales, Other\_Sales: Sales figures in North America, Europe, Japan, and other regions.

Global\_Sales: Total sales across all regions.

To ensure data integrity, the dataset was reduced to 2,828 entries by retaining only those records with complete data, thereby excluding any entries containing missing (N/A) values.

## V. DESIGN PROCESS

The development of the application followed a structured and iterative design process to ensure efficiency, functionality, and maintainability. The following steps outline the key phases involved:

### 1. Requirement Analysis & Planning

Identifying the purpose of the application: Visualizing video game sales data. Defining key features such as PCA analysis, time-series visualization, and heatmap rendering. Selecting the dataset (dataset.csv) containing game sales information.

### 2. Wireframing & UI/UX Design

Creating low-fidelity wireframes for different components (scatter plots, time-series charts, and heatmaps). Designing an interactive UI with toggles and tooltips for enhanced user experience. Choosing D3.js for interactive data visualization.

### 3. Technology Stack Selection

Frontend: JavaScript (ES6+), D3.js for data visualization.

Data Processing: Numeric.js for PCA calculations, D3.js for aggregating and filtering data.

File Handling: CSV parsing using D3.js.

### 4. Data Preprocessing & Integration

Loading and parsing dataset.csv using D3.js. Cleaning and structuring data to handle missing values (e.g., missing year values set to null). Normalizing numerical values (e.g., sales data) for PCA analysis.

## 5. Principal Component Analysis (PCA) Implementation

Extracting numerical features from the sales data. Standardizing the dataset by computing and subtracting mean values. Calculating the covariance matrix and obtaining eigenvalues and eigenvectors using Numeric.js. Projecting the data onto the first two principal components (PC1 and PC2).

## 6. Data Visualization Development

### *Scatter Plot with PCA*

- Mapping video games onto a 2D scatter plot using PC1 and PC2.
- Adding interactive elements such as hover tooltips and filtering options.

### *Time-Series Sales Visualization*

- Aggregating sales data by year and category (Genre or Platform) using D3.rollups().
- Filtering the top-5 categories based on total sales.
- Creating a dynamic line chart with a brush tool for data selection.

### *Heatmap for Sales Trends*

- Rendering a heatmap based on selected data.
- Utilizing color scales to represent variations in sales intensity.

## 7. Interaction & User Controls

Implementing checkboxes for toggling game names on the scatter plot. Adding dropdown menus for selecting visualization categories. Creating a brush tool for selecting specific time ranges in the time-series chart.

## 8. Testing & Debugging

Verifying correct data loading and transformation. Ensuring accurate PCA calculations and visual outputs. Debugging interactivity issues (e.g., tooltips, brushing, zooming).

## 9. Deployment & Optimization

Optimizing D3.js rendering for performance improvements. Ensuring responsiveness for different screen sizes. Preparing documentation for further enhancements and maintenance.

By following this structured design process, we achieved a well-organized and efficient implementation of the data visualization application.

## VI. DISCOVERED INSIGHTS

Our visual analytics system revealed several key trends in video game sales, providing valuable insights for developers, marketers, and analysts.

### *Platform and Genre Trends*

The first time-series graph (Figure 1) tracks total sales for key gaming genres (Shooter, Action, Role-Playing, Platform, and Sports) over multiple decades. The following trends emerge:

- Shooter and Action genres experienced consistent growth from 1995 onwards, peaking between 2005–2015, after which their sales began to decline.
- Role-Playing (RPG) and Platform genres had smaller yet stable markets, with RPGs gaining traction after 2010.
- Sports games exhibited a cyclical pattern, reflecting the influence of annual franchise releases.

These findings suggest a shift in consumer preferences, with action-heavy genres dominating the market before stabilizing, while RPGs and alternative genres continue to grow.

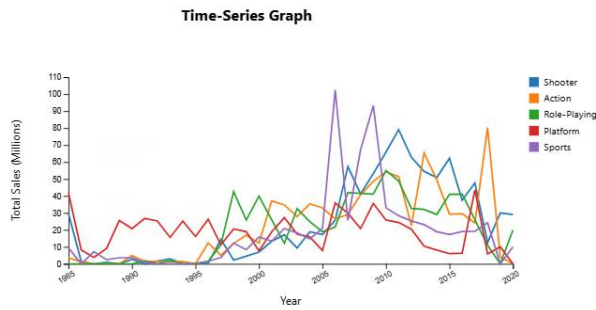


Fig.1. Genre Sales Trends Over Time (1985–2020)

The second time-series graph (Figure 2) focuses on platform sales trends, showing how console generations impact market performance.

- PlayStation 2 (PS2) and Nintendo Wii dominated in the early 2000s, but their sales declined after 2010.
- PlayStation 3 (PS3) and Xbox 360 (X360) peaked between 2008–2013, after which PlayStation 4 (PS4) became the dominant platform.
- Xbox One and Wii U had significantly lower sales than their predecessors, indicating a transition in gaming habits.
- The arrival of PS4 (2013) triggered a sharp decline in PS3 and X360 sales, reinforcing the cyclical nature of hardware transitions.

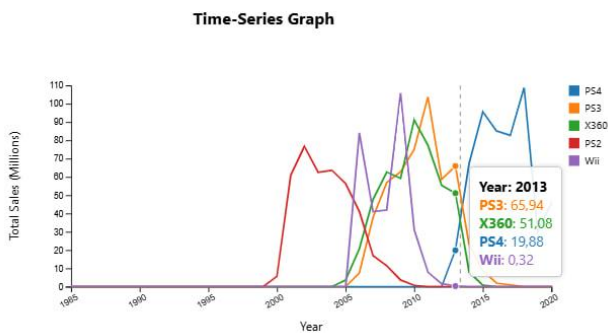


Fig.2. Platform Sales Trends Over Time (1985–2020)

These insights suggest that console life cycles strongly influence game sales, and platform shifts coincide with changing game genre popularity.

### Regional Preferences

The heatmap (Figure 3) provides a comparative analysis of genre performance across major gaming markets.

Key observations include:

- North America dominates in Shooter and Action game sales, reflecting a cultural preference for fast-paced, high-intensity gameplay.
- Europe shows a more balanced distribution but has a significant preference for Action games.
- Japan's market is highly unique, with RPGs and Platform games being the most successful. Shooter games have the lowest sales in Japan, confirming their lower appeal in that region.
- Other regions display a more balanced sales distribution, but Sports and Action genres maintain steady demand.

These results confirm that gaming preferences are region-specific, requiring publishers to adjust marketing and development strategies for each target market.

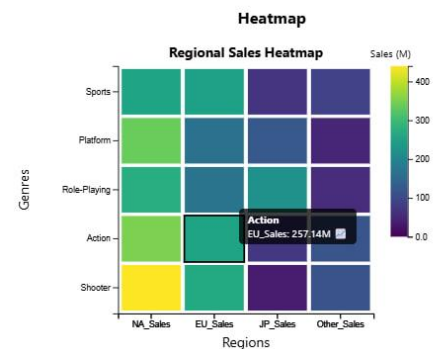


Fig.3. Regional Sales Heatmap by Genre

### Market Segmentation & Game Success Factors

The PCA visualizations summarize video game sales data from 2010–2020, highlighting market structure and key sales trends. Global Bestsellers vs. Regional Hits:

- The central cluster ( $PC1 \approx 0$ ,  $PC2 \approx 0$ ) contains most games, indicating that many titles share similar sales patterns across markets.
- Outlier games (far right on  $PC1$ ) represent top-selling global hits, such as Red Dead Redemption 2 (Figure 4) and FIFA 16 (Figure 5).

- Games positioned far left on PC1 indicate niche successes, likely games that performed well only in specific regions or genres.

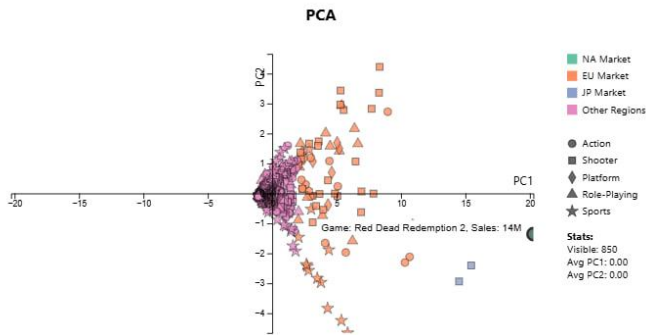


Fig.4. PCA Market Segmentation (2010–2020)

#### Platform & Market Segmentation:

- NA Market (Green) and EU Market (Orange) dominate the right side of the PCA space, meaning that these regions contribute the most to top-selling games.
- JP Market (Purple) is distinctly separate, reinforcing the earlier heatmap findings that Japan prioritizes different genres.
- PS4, PS3, X360, and Wii (Figure 5) occupy different regions of the PCA space, indicating that sales performance varies significantly by console generation.

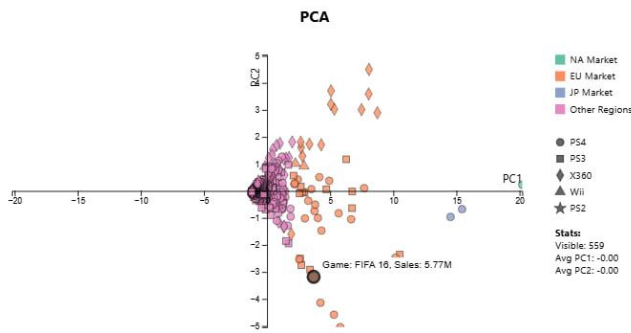


Fig.5. PCA Platform Performance (2010–2020)

These findings highlight market fragmentation, where global bestsellers follow predictable trends, while regional preferences create distinct sub-markets.

## VII. CONCLUSIONS AND FUTURE WORK

This project highlights how interactive visual analytics can uncover insights from video game sales data. Through time-series analysis, PCA, and heatmap visualizations, it identifies platform life cycles, genre preferences, and regional sales trends, aiding developers and marketers in shaping strategies. Key findings include platform transitions every 5–7 years, region-specific genre popularity, and shared traits among global bestsellers, with niche titles excelling in select markets.

Future work will focus on integrating forecasting models to predict sales trends, enhancing the interface with advanced customization options for trend comparisons, and leveraging deep learning to uncover hidden correlations, further strengthening market pattern analysis.

## VII. REFERENCES

- [1] A. Doe, B. Smith, and C. Johnson, "Data Visualization on Video Games Global Sales Analysis & Predictive Analysis", *International Journal of Computer Science Trends and Technology (IJCTST)*, vol. 9, no. 3, pp. 1-10, 2021. [Online]. Available: <https://www.ijctstjournal.org/volume-9/issue-3/IJCTST-V9I3P1.pdf>.
- [2] M. Brown and T. Green, "Video Game Sales Trends and Stats", *ResearchGate*, 2020. [Online]. Available: [https://www.researchgate.net/publication/363175427\\_Video\\_Game\\_Sales\\_Trends\\_and\\_Stats](https://www.researchgate.net/publication/363175427_Video_Game_Sales_Trends_and_Stats).
- [3] J. Williams, P. Zhang, and L. Kim, "The Impact of Platform on Global Video Game Sales", *ResearchGate*, 2012. [Online]. Available: [https://www.researchgate.net/publication/297754899\\_The\\_Impact\\_Of\\_Platform\\_On\\_Global\\_Video\\_Game\\_Sales](https://www.researchgate.net/publication/297754899_The_Impact_Of_Platform_On_Global_Video_Game_Sales).
- [4] K. Anderson and R. Patel, "Comprehensive Review and Classification of Game Analytics", *SpringerLink*, 2018. [Online]. Available: Comprehensive review and classification of game analytics | Service Oriented Computing and Applications.
- [5] S. Tan and Y. Wong, "Visualizing the International Video Game Industry", *Singapore Management University (SMU) Wiki*, 2019. [Online]. Available: [wiki.smu.edu.sg/1920t2is428/img\\_auth.php/7/7a/G5Final\\_Paper.pdf](http://wiki.smu.edu.sg/1920t2is428/img_auth.php/7/7a/G5Final_Paper.pdf).
- [6] Kaggle Video Game Sales Dataset: <https://www.kaggle.com/datasets/thedevastator/global-video-game-sales>.