Data Analytics Capstone

Analysis of Video Game Sales

BHN1 Task 2: Project Proposal

Jared C. Plaisance

Western Governors University

04/23/2024

Table of Contents

A. Proposal Overview

A.1 Research Question

How do various factors such as platform, genre, and publisher influence video game sales across different regions of the world? This information can assist video game developers, publishers, and marketers in making informed decisions about game development, marketing strategies, and audience segmentation.

A.2 Context and Background

The video game industry is dynamic and continuously evolving in a multi-billion dollar annual market. With an array of gaming platforms spanning consoles, personal computers (PC's), and mobile devices, the industry witness's fierce competition among game developers and publishers clawing for consumer attention. In this landscape, understanding the web of factors that influence video game sales is necessary for success. Market dynamics, consumer behavior, and technological advancements all play pivotal roles in shaping the industry.

Within this context, data analytics emerges as a powerful tool, offering insights for strategic decision-making processes. By diving into the data of video game sales across different regions of the world accompanied by attributes such as platform, genre, and publisher, this project will seek actionable insights. These insights will have the potential to drive innovation, optimize marketing strategies, and pinpoint growth opportunities for stakeholders. As the video game market continues to grow the need for data-driven decision making becomes increasingly pronounced.

A.3 Summary of Published Works

The 3 published works to be summarized include:

- "The Impact of Platform on Global Video Game Sales" by Jeffery Babb, Neil Terry, and Karrem Dana
- "Video Game Market Segmentation Based on User Behavior" by Kaveh Khalehgi and Artur Lugmayr
- "Consumer Perceptions & Video Game Sales: A Meeting of the Minds" by John Sacranie

A.3.1 Summary of Published Work 1

"The Impact of Platform on Global Video Game Sales"

This paper dives into the dynamics of video game sales worldwide from 2006 to 2011, focusing on various gaming platforms. It reveals that Nintendo's Wii emerged as the dominant platform during this period, followed closely by Nintendo DS, Xbox 360, PlayStation 3, and PC. The study explores the significant evolution of the video game industry over the past four decades, transforming from a niche market to a multi-billion dollar global industry rivaling the film sector. It also explores the factors driving video game sales, including network effects, innovation, platform strategies, vertical integration, and gamification.

A.3.2 Summary of Published Work 2

"Video Game Market Segmentation Based on User Behavior"

This paper explores market segmentation in the video game industry based on user behavior. It discusses the historical evolution of video games and the transition from simple entertainment to a profitable industry. Market segmentation is defined as dividing the market into smaller segments to respond to different customer needs. The authors categorize games into genres and analyze how personality traits influence game preferences. They use the Five Factor Model to predict user preferences for different genres based on traits like extraversion and conscientiousness.

A.3.3 Summary of Published Work 3

“Consumer Perceptions & Video Game Sales: A Meeting of the Minds”
This article explores the determinants of video game sales in the competitive and evolving video game industry. It discusses the impact of factors such as consumer preferences, hardware platforms, genre, exclusivity, sequels, licenses, and advertising on video game sales. The article employs a demand model to analyze weekly sales data for 100 randomly selected US game releases over their first ten weeks of availability. The study aims to provide insights into the market's inner workings and identify factors contributing to the success of video game titles.

A.3.4 Relation of Published Works to Project

Each published work contributes uniquely to the development of the project by providing a comprehensive understanding of the determinants of video game sales. John Sacranie's article serves as the cornerstone, delineating the key factors such as consumer preferences, hardware platforms, genre dynamics, exclusivity, sequels, licenses, and advertising strategies. Sacranie's empirical model offers a structured approach for analyzing sales data and unraveling the intricacies of the video game market. Supplementary works by Clements & Ohashi, Corts & Lederman, Vany A., and Burrato & Viscolani further enrich the project by diving into specific facets like network effects, platform exclusivity, sequel trends, and advertising impacts. By forming insights from these diverse studies, the project gains depth and clarity, enabling exploration of consumer behavior and market dynamics within the video game industry.

A.4 Deliverables of Data Analytics Solution

The deliverables for the data analytics solution aimed at determining the determinants of video game sales include a comprehensive report, visualizations, and a t-test. The report will provide detailed insights into the factors influencing video game sales, their impacts, and recommendations for stakeholders. Visualizations, such as charts and graphs, will facilitate understanding of the analysis results. A t-test will explain differences between sales and independent variables like platform or genre. Together, these deliverables will empower stakeholders to make informed decisions, optimize strategies, and enhance their competitive edge in the video game market.

A.5 Benefit and Support of Decision-Making Process

The data analytics solution offers valuable insights into the determinants of video game sales, enabling strategic decision-making. By understanding factors like platform availability and genre preferences, the organization can optimize resource allocation and marketing strategies, minimizing waste and maximizing ROI. This evidence-based approach fosters agility and adaptability, allowing the organization to capitalize on emerging opportunities and navigate market dynamics effectively. Overall, the solution serves as a powerful tool for driving growth, mitigating risks, and maintaining competitiveness in the ever-changing video game industry.

B. Data Analytics Project Plan

B.1 Goals, Objectives, and Deliverables

The project aims to understand the factors influencing video game sales to inform strategic decision-making. By analyzing consumer behavior, preferences, and market trends, we seek to provide actionable insights for product development, marketing, and overall business strategy. Key deliverables include a comprehensive analysis report, visualizations, and statistical tests to forecast sales and guide decision-making. Ultimately, the project aims to empower the organization with data-driven insights to drive growth and competitiveness in the video game market.

B.2 Scope of Project

The project scope encloses the analysis of factors influencing video game sales, focusing on consumer behavior, platform dynamics, genre preferences, and marketing strategies. It involves collecting and analyzing data on video game sales, platform popularity, and genre trends. The analysis will cover a diverse range of video game titles across multiple platforms. The scope also includes developing t-tests to identify factors and providing actionable insights for strategic decision-making. However, the scope does not extend to the analysis of digitally distributed games or the impact of piracy on sales.

B.3 Planning Methodology

The CRISP-DM methodology will structure the project, ensuring clarity and efficiency. Beginning with Business Understanding, I'll define objectives and stakeholder needs. Data Understanding follows, acquiring and assessing relevant data on video game sales and related factors. Then, Data Preparation involves cleaning and exploring the data. Moving to Modeling, t-tests and graphs will be developed and evaluated in the Evaluation phase. Finally, in Deployment, the concluded analysis will be explained, with thorough documentation provided. This iterative process will allow for flexibility and adaptability throughout the project

B.4 Timeline and Milestones

Milestone	Duration	Anticipated Start Date	Anticipated End Date
Initiation	1 Day	4/22/2024	4/23/2024
Data Collection	1 Day	4/23/2024	4/24/2024
Data Analysis	3 Days	4/24/2024	4/27/2024
Evaluation	2 Days	4/27/2024	4/29/2024
Statistic Modeling	2 Days	4/29/2024	5/1/2024
Visualization	2 Days	5/1/2024	5/3/2024
Documentation	2 Days	5/3/2024	5/5/2024
Feedback	1 Day	5/5/2024	5/6/2024
Presentation	1 Day	5/6/2024	5/7/2024

B.5 Resources and Costs

Personnel:

- Me: $0 (Student)

Technology:

- Microsoft Word: $0 (Included Through Microsoft 365 from WGU)
- Microsoft Excel: $0 (Included Through Microsoft 365 from WGU)
- Jupyter Notebook: $0 (Open Source)
- Python: $0 (Open Source)

Infrastructure:

- Personal Computer: $0 (Already Owned)
- Internet: $0 (Already Paid For)

The estimated work hours are 75 for tasks like data collection, analysis, and reporting. These resources ensure the project's success within the specified timeline and budget. Adjustments can be made as needed based on project requirements and resource availability.

B.6 Success Criteria

Project success will be evaluated based on several key criteria. Data accuracy and completeness will be ensured through validation against trusted sources. Timeliness will be monitored by comparing actual progress to the planned timeline for each milestone. Budget compliance should be no issue since we intend to not incur any costs.

The performance of analytical models will be measured using metrics like accuracy, precision, recall, and F1 score. We'll also evaluate the relevance and usefulness of insights generated from data analysis. Finally, the impact of implementation will be assessed by examining how effectively insights are integrated into decision-making processes within the organization. These criteria will guide the project's execution and ensure that it delivers value according to its objectives.

C. Design of Data Analytics Solution

C.1 Hypothesis

Sales patterns in the video game industry are influenced by a combination of factors including platform popularity, game genre, regional preferences, and release year. Top-selling video games tend to share common characteristics such as popular publishers, most used platform per region, and alignment with consumer preferences in specific regions.

C.2 Analytical Methods

The data analytics solution will primarily employ descriptive analytics methods. Descriptive analytics will be utilized to summarize and interpret the relationships between the independent variables and video game sales. This will involve generating summary statistics, visualizations, and reports to provide insights into the patterns and trends observed in the data.

C.2.1 Justification

The selected analytical method, descriptive analytics, is appropriate for this project considering its focus on interpreting video game sales and understanding the factors influencing them. Descriptive analytics will provide valuable insights into the relationships between variables and sales patterns, enhancing understanding of market dynamics. This method will offer a comprehensive approach to analyzing the dataset and deriving actionable insights for the video game industry.

C.3 Tools and Environments

For this data analytics solution, I'll primarily use Python for data preprocessing, analysis, and modeling, leveraging libraries like Pandas and Scikit-learn. Jupyter Notebooks will handle statistical analysis, hypothesis testing, and database querying. Visualization will be done using Matplotlib and Seaborn.

C.4 Methods and Metrics for Analytical Solution

The evaluation of the data analytics solution will involve using cross-validation techniques like k-fold cross-validation to assess the model's generalization performance. For classification tasks, metrics such as accuracy, precision, recall, and F1-score will be used.

C.4.1 Justification

The selected methods and metrics are suitable for evaluating the data analytics solution's output as they offer a well-rounded assessment of its performance. Metrics like accuracy, precision, recall, and F1-score are crucial for classification tasks, providing insights into the model's ability to classify instances accurately across different classes.

C.5 Practical Significance

Assessing the practical significance of the data analytics solution involves evaluating its real-world impact and effectiveness in meeting the predefined objectives. Specific criteria to determine its success include measuring improvements in decision-making processes and benchmarking performance against industry standards. Analyzing these criteria will help determine whether the solution has provided the expected benefits and effectively supported decision-making processes.

C.6 Tools and Graphical Representations

To visually communicate the findings of the data analytics solution, I will utilize various tools and graphical representations tailored to the nature of the data and the audience. This may include statistical software such as Python with libraries like Matplotlib, Seaborn, and Plotly for data visualization. I'll employ bar charts, line graphs, scatter plots, histograms, heatmaps, and pie charts to represent trends, distributions, correlations, and comparisons within the data. Interactive visualizations will allow stakeholders to explore the data dynamically, enabling deeper insights and facilitating better decision-making processes.

D. Description of Dataset

D.1 Source of Data

The data for this project will primarily be sourced from VGChartz, a website that provides video game sales data, including information on the number of units sold for various gaming platforms. Other sources may include industry reports, academic studies, and publicly available datasets related to the video game industry.

D.2 Appropriateness of Dataset

This dataset is ideal for the project's goals because it contains detailed information on video game sales, essential for understanding consumer preferences and market trends within the gaming industry. With data on units sold from VGChartz, the project can analyze sales patterns and game performance across different platforms. This comprehensive dataset enables the exploration of factors influencing video game sales, aiding strategic decision-making and market understanding for industry stakeholders.

D.3 Data Collection Methods

The data collection methods involved obtaining information from VGChartz for video game sales. VGChartz provided weekly sales figures for 100 randomly selected US game releases over their first ten weeks of availability. These sales numbers were crucial for understanding the performance of each game in terms of units sold over time. By obtaining the sales data from VGChartz, a comprehensive dataset was created for analysis, allowing for a deeper understanding of the factors influencing video game sales.

D.4 Observations on Quality and Completeness of Data

The dataset obtained from VGChartz provides valuable insights into video game sales, focusing on releases over the initial ten weeks of availability. However, it's crucial to note that the data has some quality and completeness issues. There are missing values in the "year" and "publisher" columns, which can affect the accuracy of the analysis and interpretation of results. Additionally, the "year" column is formatted as a float instead of an integer, which may lead to inconsistencies in calculations and visualizations. These data quality issues need to be addressed through data cleaning and preprocessing steps to ensure the reliability and accuracy of the analysis.

D.5 Data Governance

Ensuring data governance, privacy, security, and ethical compliance is paramount when dealing with datasets, such as the one obtained from VGChartz. Establishing clear policies and procedures for data management, access, and usage is essential to maintain data integrity and quality. Ethical considerations involve ensuring transparency in data handling and avoiding biases or discrimination in analysis. Legal, and regulatory compliance with laws such as GDPR and CCPA is crucial, requiring organizations to adhere to data protection regulations and industry standards. Addressing these considerations is vital to ensure responsible data usage, maintain stakeholder trust, and derive valuable insights from the dataset for the project.

D.5.1 Precautions

When working with the data, precautions are essential to ensure governance, privacy, security, and ethical standards. This involves implementing clear governance policies, robust security measures, and ethical guidelines. Measures include data encryption, access controls, and transparency. Legal compliance with regulations like GDPR or CCPA is crucial, and communication should prioritize clear explanations and anonymization of sensitive data. These precautions collectively uphold integrity, privacy, and legal standards in data handling and communication.

F. References

Babb, Jeffry & Terry, Neil & Dana, Kareem. (2013). The Impact Of Platform On Global Video Game Sales. International Business & Economics Research Journal (IBER). 12. 1273. 10.19030/iber.v12i10.8136. https://www.researchgate.net/publication/297754899_The_Impact_Of_Platform_On_Global_Video_Game_Sales

Khaleghi, Kaveh & Lugmayr, Artur. (2012). Video game market segmentation based on user behavior. 283-286. 10.1145/2393132.2393194. https://www.researchgate.net/publication/259287561_Video_game_market_segmentation_based_on_user_behavior

Sacranie, John (2010) "Consumer Perceptions & Video Game Sales: A Meeting of the Minds," The Park Place Economist: Vol. 18 https://digitalcommons.iwu.edu/parkplace/vol18/iss1/12/