CIND820: Project Design

Student: Eddie Li

Student #: 500820042

Supervisor: TBA

Date of Submission: January 25, 2026

Contents

1 Problem Statement and Context

The video game industry has become an integral part of the entertainment sector and is expected to continue growing [1][2]. With such drastic growth in recent times, accurately forecasting the titles that will achieve high sales remains a challenge for developers and publishers since sales are influenced by complex factors. These factors can include things like: gaming platform, genre, and review scores [3]. Historical data demonstrates that total sales vary widely depending on these factors [1][4]. This project aims to enhance our understanding on the specific factors that video game sales and explore the extent to which sales outcomes can be reliably forecasted [5].

2 Research Questions and Justification

This project will investigate the factors that influence global video game sales and evaluate the extent to which sales outcomes can be predicted using game characteristics and review metrics. To achieve this goal, the project will focus on the following research questions and why they are important:

- Among platform, genre, review score, which features are most influential in predicting global sales?
  - Understanding which features are important helps to quanity each variable's impact which allows accurate sales forecasting [1][3].
- Can global video game sales be predicted using characteristics such as platform, genre, review score, and release year?
  - Examining predictive accuracy addresses the challenge posed by complex interactions among game features and consumer behavior [5].
- Are certain platforms or genres inherently more predictable in terms of their sales outcomes than others?

– Investigating predictability across genres and platforms will determine whether predictive models perform equally well on each [6].

3 Dataset Selection and Rationale

The dataset selected for this project is "Global Video Game Sales and Reviews" dataset obtained from Kaggle. This dataset combines historical sales data, qualitative game data, and review-related information, making it well-suited for predictive analytics applications. The variables in the dataset include things such as: global sales, platform, genre, release year, and review scores. All these variables are ones which align with the research questions posed in this project. As such, the dataset enables analysis of feature importance and predictability of total video game sales.

This dataset is appropriate because prior research has identified platform, genre, and review metrics as significant determining factors of video game sales [1][6]. With all these variables in a single dataset, the project can evaluate their relative influence on global sales while also examining potential interactions among them. Additionally, the inclusion of review scores allows for analysis of the relationship between public reception and commercial success, which has been shown to have a complex relationship [1][5]. Lastly, the dataset is feasible for analysis due to its manageable size. It should be noted however, that the data set does not include some other important factors that are hard to quantify and less public such as: marketing budget, monetization method, and publisher/developer reputation [2]

4 Proposed Methodology and Tools

This project will use predictive analytics to examine patterns in video game data and to predict global video game sales based on game characteristics. Predictive analytics is appropriate for this project because the dataset contains historical sales data that can be used to model and forecast outcomes.

The methodology will begin with data pre-processing which will handle missing values

and normalizing data types and values. Afterwards, data analysis will be used to observe overall trends and relationships between these features and global sales. Regression analysis will then be applied to predict global sales and assess feature importance. The model performance will be evaluated using metrics such as mean squared error (MSE) and the coefficient of determination (R^2), alongside a train-test split to ensure predictive reliability.[5].

The tools expected to be used are: pandas and NumPy for data pre-processing, matplotlib and seaborn for data visualization and trend analysis, and sklearn for linear regression [5]. Each methodological step is directly aligned with one or more research questions, ensuring that the analysis addresses feature importance, predictive accuracy, and predictability across platforms and genres.

References

[1] T. Kraemer, W. H. Weiger, and S. Heidenreich, "Do all stars shine the same? investigating the nonlinear effects of user and critic reviews on video game sales," en, *J. Bus. Res.*, vol. 188, no. 115034, p. 115 034, Feb. 2025.

[2] H. Cheng, "Video game sales trends and stats," *BCP Business & Management*, vol. 25, pp. 222–225, Aug. 2022.

[3] J. Cox, "What makes a blockbuster video game? an empirical analysis of US sales data," en, *MDE Manage. Decis. Econ.*, vol. 35, no. 3, pp. 189–198, Apr. 2014.

[4] J. Babb and N. Terry, "Comparing video game sales by gaming platform," *Journal of the Southwestern Society of Economists*, vol. 40, pp. 25–46, 2013, Accessed: 2026-01-20.

[5] A. Singh, Z. Hasan, Z. Vishwakarma, A. Banshal, and S. Kushwaha, "Evaluating linear and non-linear regression techniques for video game sales prediction," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 11, no. 03, pp. 1507–1511, Nov. 2023.

[6] J. Babb, N. Terry, and K. Dana, "The impact of platform on global video game sales," *Int. Bus. Econ. Res. J. (IBER)*, vol. 12, no. 10, p. 1273, Sep. 2013.