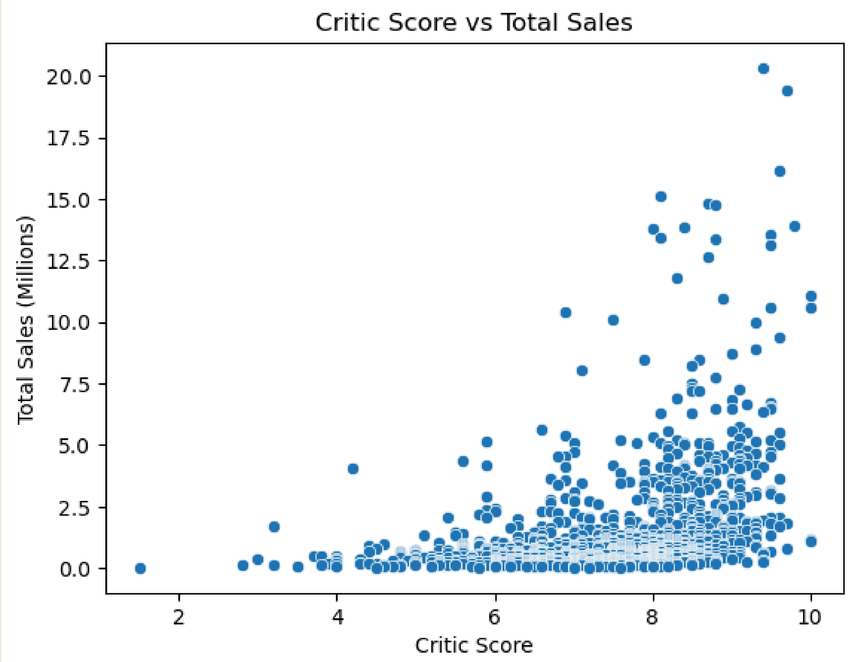
Video Game Sales 2024: Do Critic Scores Predict Success?

Exploring How Reviews and Region Impact Game Success

AUTHORS

Group 8
Members: Xingyu Luo, Geonho Lee
Course: CS131 - Spring 2025



01. Introduction

Video game sales are influenced by many factors, including genre, platform, and marketing—but how much do professional critic reviews actually matter? In this project, we analyze data from the Video Game Sales 2024 dataset to investigate whether higher critic scores lead to higher global and regional sales.

02. Project Objective

This project investigates whether a video game's critic score can be used to predict its global sales performance. We also explore differences in regional sales across North America and Japan.

03. Dataset Metadata

The dataset contains sales and rating information for thousands of video games released across different platforms and regions.

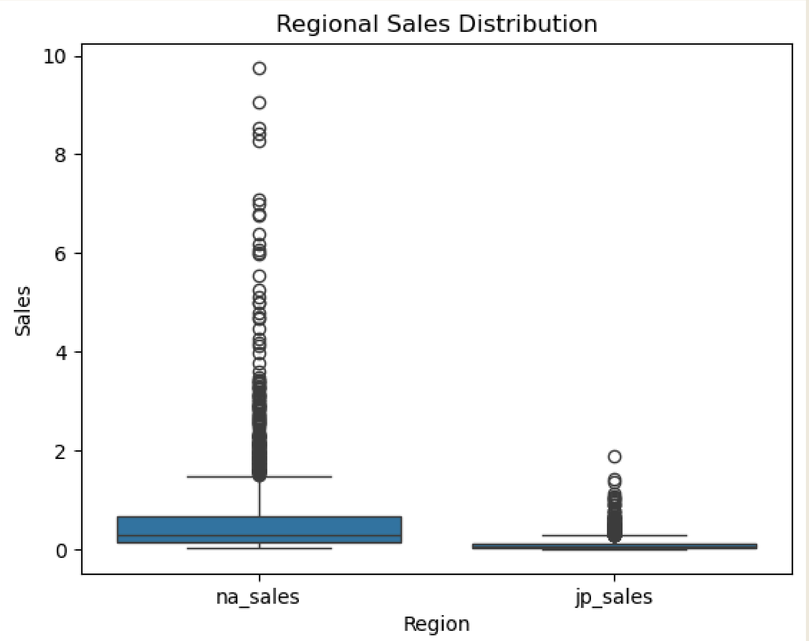
- **Dataset: Video Game Sales 2024 (Kaggle - asaniczka)**
- **Entries after cleaning: ~5,000**
- **Selected Features: title, console, genre, critic_score, total_sales, na_sales, jp_sales**
- **Value Ranges:**
  - **critic_score: 4.0 – 10.0**
  - **total_sales: 0.01 – 20.3 million**
  - **na_sales average: 0.42 million**
  - **jp_sales average: 0.11 million**

04. Results/Findings

Using Spark MLlib, we applied a linear regression model to determine the relationship between critic scores and total sales. The model achieved:

- **R^2 score**: 0.13
- **Coefficient**: 0.67
- **Intercept**: -3.58

This result indicates that critic scores are somewhat positively associated with sales, but they only explain ~13% of the variance. Additionally, box plot analysis shows that North American sales are generally higher and more variable than Japanese sales, suggesting market-specific preferences.



05. Visualizations

Scatter Plot - Critic Score vs Total Sales
Shows weak positive trend between game ratings and global sales

Box Plot - NA vs JP Sales
Shows North American market has higher and more variable sales than Japan

06. Tools and Process Summary

- Unix Shell Tools (awk, cut): We used shell commands to clean the raw CSV file by removing rows with missing critic scores and selecting relevant columns. This preprocessing helped reduce noise and focus the analysis.
- Google Cloud Storage: Stored cleaned CSV file for distributed processing
- PySpark + MLlib: Used to cast data types, remove nulls, and train linear regression models
- Pandas + Seaborn: Used to create the visualizations exported as PNG

07. Conclusion

The scatter plot shows a weak but positive relationship between critic scores and total game sales—higher-rated games tend to sell more, but the trend is not strong. This supports our regression findings.

The box plot comparing North American and Japanese sales reveals that games in North America generally sell more and show greater variability, while Japanese sales are lower and more consistent. This highlights regional differences in market size and consumer behavior.