

Discover Report

Video Game Sales Analysis

1. Topic Discussion

Question : How do game features (playtime, genre, platform, ratings) influence global sales performance?

The video game industry generates billions in revenue annually, yet the relationship between game characteristics and commercial success remains complex. In this topic we try to investigate whether longer games sell better, which genres dominate specific platforms, and how critical reception correlates with sales figures. Understanding these patterns can help game development decisions and marketing strategies.

Key Analysis Goals:

- Examine relationship between multiple features and sales performance
- Identify platform-specific trends in genre popularity and sales
- Analyze the impact of review scores and ratings on commercial success
- Compare regional sales patterns

2. Dataset Descriptions

Dataset 1 : IGDB Game Metadata

Source : Platform-specific game databases (PlayMyData - <https://zenodo.org/records/10797761>)

Size : 129,380 Total Games (final shape – 97785 x 14)

Time Period : Games with release dates spanning multiple decades

Key Features:

- **Game Information :** Title, release date, genres (multi-label), platforms
- **Quality Metrics:** User ratings, review scores, review count, people polled
- **Gameplay Data:** Playtime statistics (main story, extra content, completion hours)
- **Metadata:** Game summaries, storylines, unique game IDs

Dataset 2 : Video Game Sales Data (VGChartz)

Source: VGChartz sales tracking database (<https://zenodo.org/records/5898311>)

Size : 16,598 x 11

Time Period: 1980-2020

Key Features:

- **Sales Figures :** Regional breakdown and global totals
- **Commercial Metadata :** Publisher, release year, platform, primary genre
- **Market Coverage:** Focuses on commercially successful titles with tracked sales

Conclusion:

Deriving relationships and getting insightful results of any domain requires proper combination of data and I hope to achieve sufficient results with this combination of datasets. I feel there is a lot of insightful information lying underneath the structure of this data and this domain, which I will extract in the upcoming tasks.