Dataset Description

Dataset Name: Video Game Sales

Source: Kaggle

Description: This dataset contains sales data for over 16,500 video games. It includes various fields such as the game's name, platform, release year, genre, publisher, and sales figures across different regions (North America, Europe, Japan, and other regions) as well as global sales.

Reasons for Selection

Relevance: This dataset is highly relevant for analyzing trends in the video game industry, understanding the popularity of different genres and platforms, and identifying key factors that contribute to high sales.

Availability: The dataset is readily available on Kaggle and is well-documented, making it easy to use for analysis.

Interest: The video game industry is a dynamic and rapidly evolving field, making it an interesting subject for analysis. Additionally, the dataset's comprehensive nature allows for a wide range of analyses, from sales trends to market segmentation.

Processing Problems and Solutions

Problem 1: Missing data for certain games or regions.

Solution: Our team used imputation techniques to estimate missing values based on available data. For example, if sales data for a particular region was missing, we estimated it using the average sales for that region across similar games.

Problem 2: Inconsistent data formats, such as different date formats for release years.

Solution: Our team standardized the data formats by converting all dates to a consistent format (e.g., YYYY) and ensuring all numerical values were in millions for consistency.

Problem 3: Outliers in sales data that could skew the analysis.

Solution: Our team identified and removed outliers using statistical methods, such as calculating the interquartile range (IQR) and removing data points that fell outside 1.5 times the IQR.