

Exploratory Data Analysis Report: Video Game Dataset

1. Dataset Overview

The dataset contains data about video games, including attributes like their name, release date, price, and user reviews. The primary goal of this analysis is to uncover patterns related to game sales, pricing, and the impact of user reviews.

- **Total records:** 4,931
 - **Total attributes (columns):** 20
 - **Key attributes:**
 - **Name:** The name of the video game.
 - **Release Date:** The release date of the video game.
 - **Price:** The retail price of the game.
 - **DLC Count:** The number of downloadable content (DLC) packs available for the game.
 - **Positive Reviews / Negative Reviews:** User feedback in the form of reviews.
 - **Average Sales:** Estimated average sales of each game.
-

2. Descriptive Statistics

The descriptive statistics give an initial sense of the dataset's distribution and scale:

- **Average Price:** \$10.98
- **Maximum Price:** \$124.99
- **Average Sales:** 835,063 units
- **Highest Sales:** 150 million units
- **Average Positive Reviews:** 9,741
- **Highest Positive Reviews:** 1,477,153

The price of most games lies in the lower range (below \$20), but some high-end games reach up to \$124.99. The sales figures also show a skewed distribution, with a few games reaching extremely high sales numbers, while the majority have moderate to low sales.

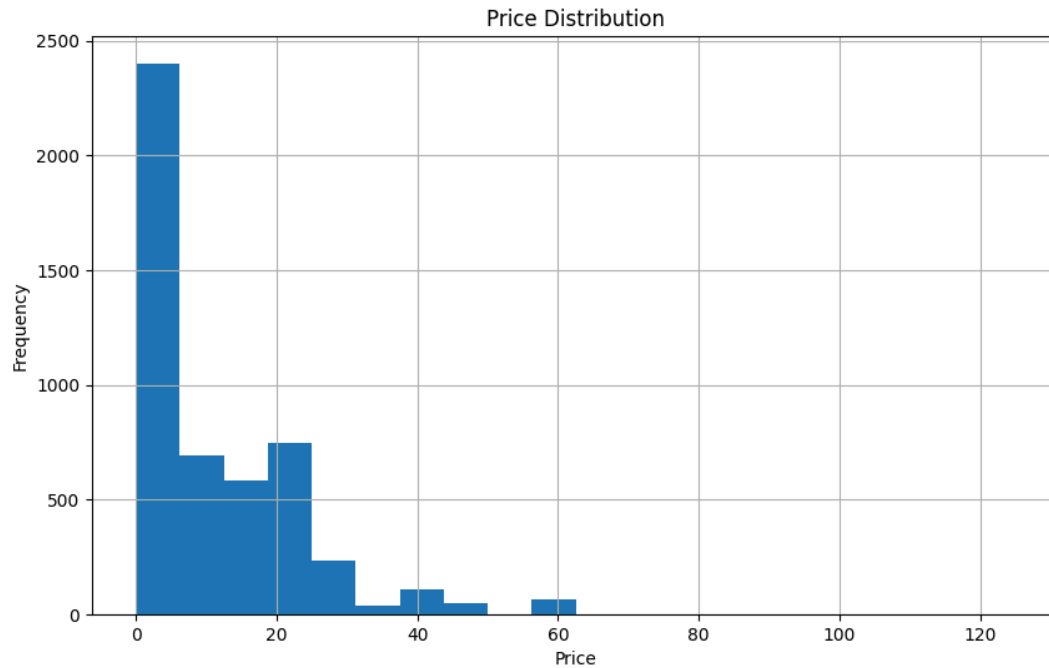
3. Missing Values

No missing values were detected in the dataset, which ensures a complete analysis without any need for data imputation or handling missing data.

4. Data Visualizations

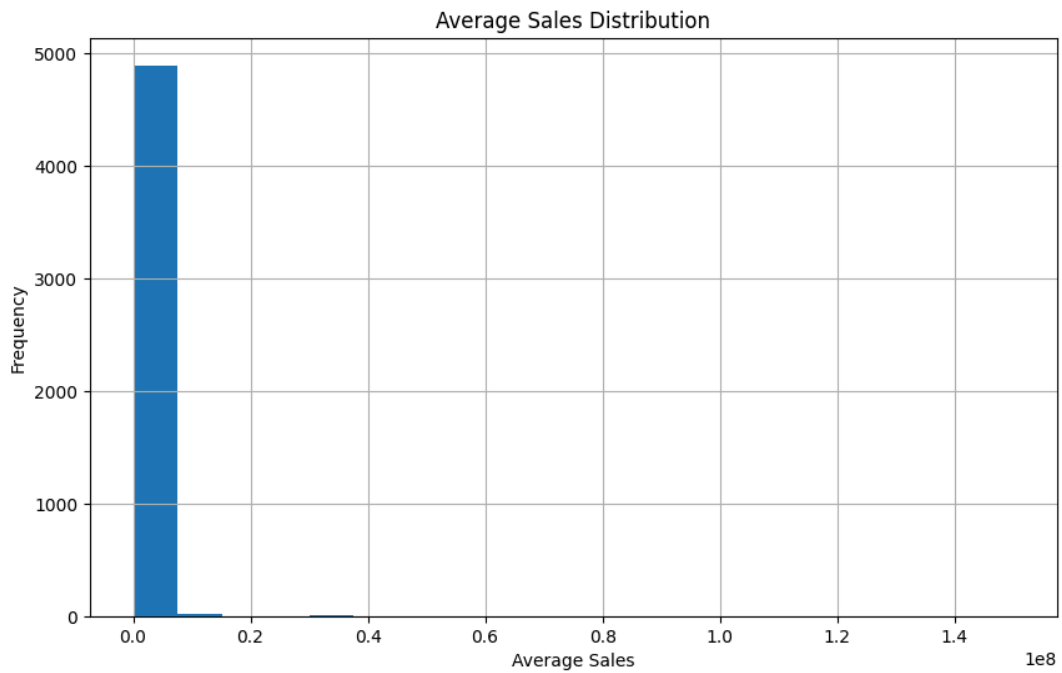
To further explore the data, several visualizations were created:

Price Distribution:



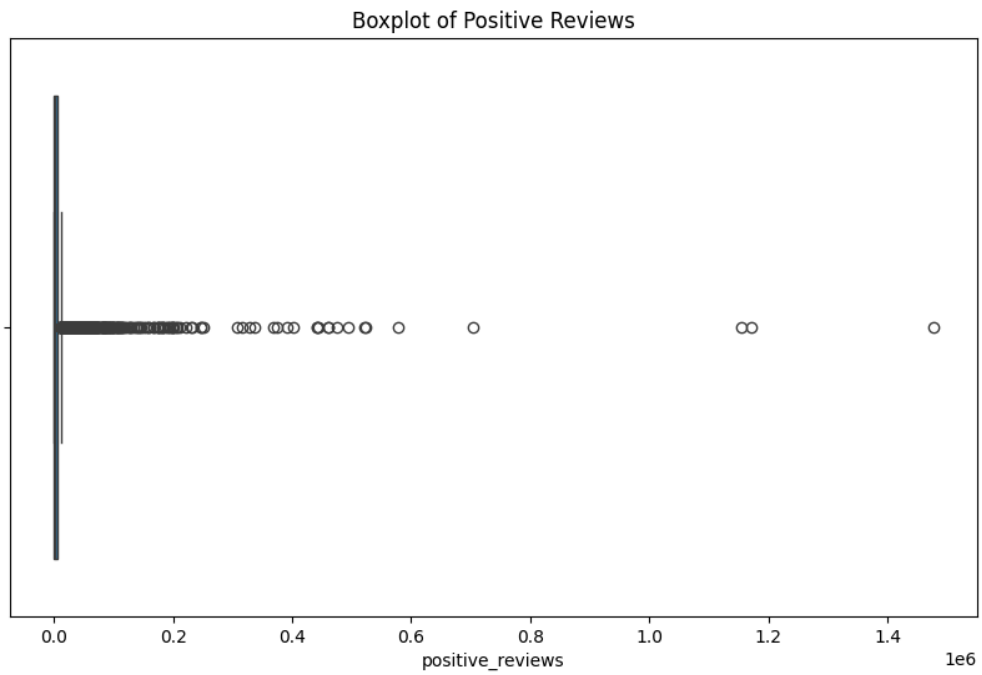
The histogram shows that the majority of games are priced below \$20. A few outliers exist with prices up to \$124.99, indicating premium-priced games.

Average Sales Distribution:



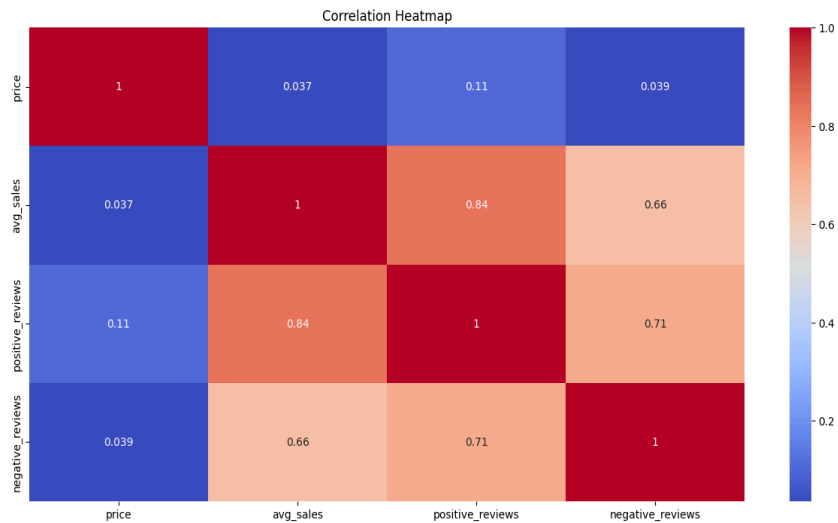
The sales distribution is highly skewed. A large number of games have relatively low sales figures, while a few outliers have extremely high sales, with some games reaching up to 150 million units sold.

Boxplot of Positive Reviews:



The boxplot reveals that the distribution of positive reviews is wide, with many games receiving extremely high numbers of positive reviews. This highlights that a few top-tier games dominate in terms of user feedback.

Correlation Heatmap:

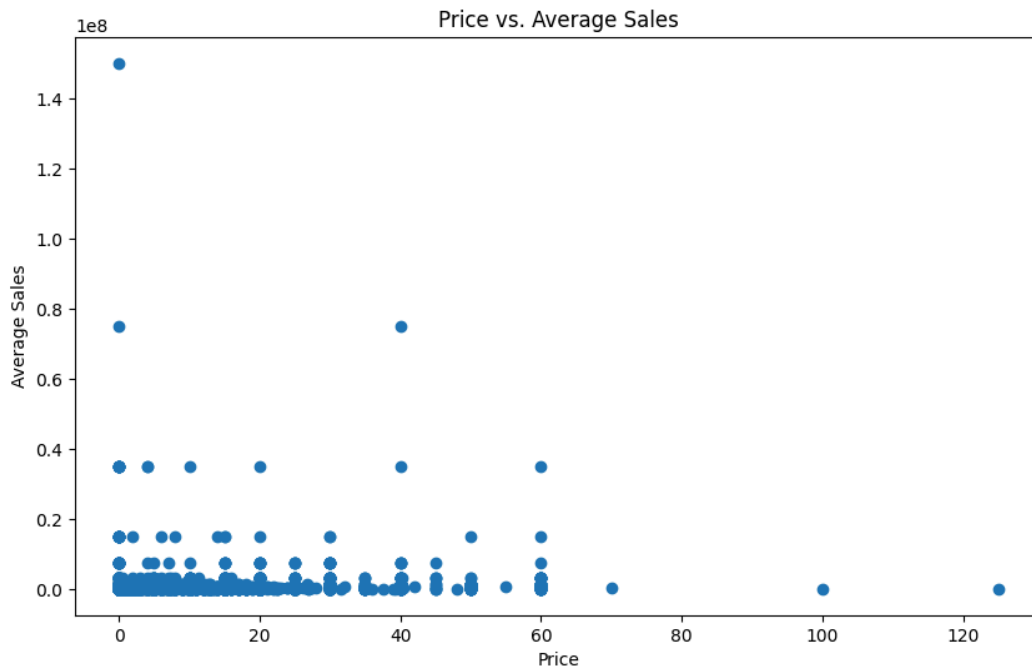


A correlation heatmap was created to explore the relationships between numerical variables (e.g., price, sales, positive and negative reviews). The key findings include:

- A positive correlation between average sales and positive reviews, indicating that games with higher sales generally receive more positive feedback.
 - Weak correlations between price and other variables, suggesting that pricing doesn't significantly influence sales or reviews.
-

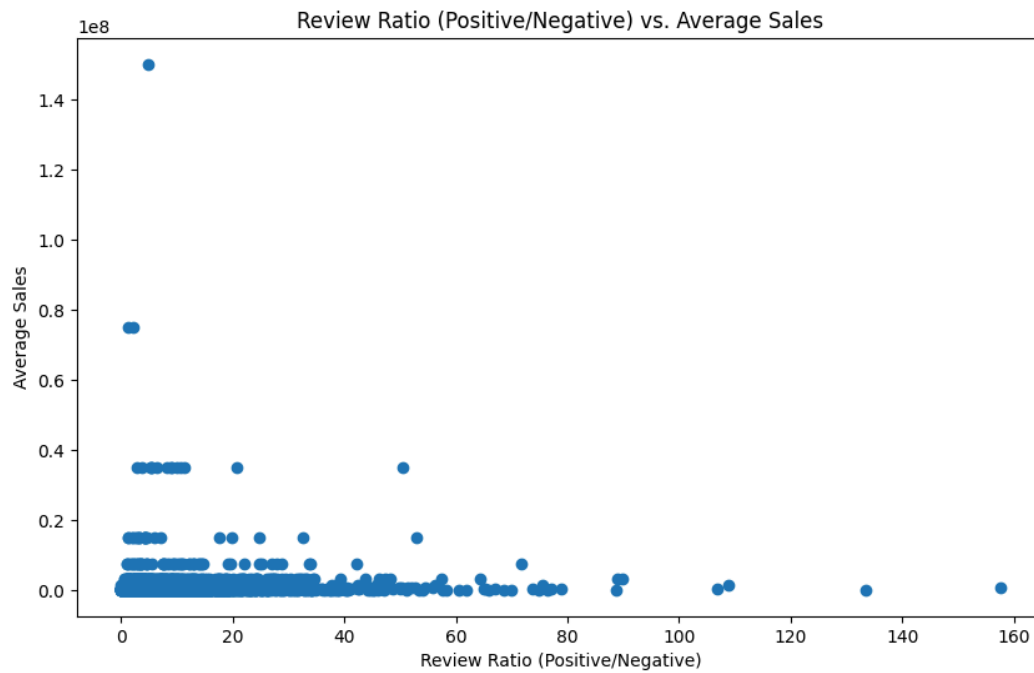
5. In-Depth Analysis

Price vs. Average Sales:



There is no clear relationship between the price of a game and its sales. Many of the top-selling games are **free-to-play**, which likely skews the sales data towards higher volumes for lower-priced games.

Review Ratio (Positive to Negative Reviews) vs. Sales:



The ratio of positive to negative reviews has a positive relationship with sales. Games with higher ratios of positive feedback tend to achieve better sales figures. This suggests that user sentiment plays a significant role in driving game sales.

6. Genre Popularity Among Top-Selling Games

An analysis of genre popularity among the top-selling games shows that certain genres dominate the market:

Genre	Count
Action	9
Free-to-Play	6
Adventure	5
RPG	4
Massively Multiplayer	3

- **Action** games are the most popular, with 9 out of the top 10 best-selling games belonging to this genre.
 - **Free-to-play** games are prevalent in 6 of the top 10 games, showing that games with no upfront cost tend to drive large sales volumes.
 - **Adventure** and **RPG** genres also have strong representation among the top sellers.
-

7. Publisher Impact on Sales

The analysis of publishers shows that certain game publishers dominate the market.

Publisher	Total Sales
Valve	150 Million
Amazon Games	75 Million
KRAFTON, Inc.	75 Million
Bungie	35 Million
Gaijin Distribution KFT	35 Million
Smartly Dressed Games	35 Million
Digital Extremes	35 Million

- **Valve** dominates the leaderboard with **Dota 2**, which has 150 million sales.
- **Amazon Games** and **KRAFTON, Inc.** also feature prominently, with major releases like **PUBG** and **New World**.

These results suggest that well-established publishers with strong game franchises significantly impact sales.

8. Top 10 Best-Selling Games

The top 10 best-selling games represent a mix of free-to-play and premium titles. Most of these games receive a large number of positive reviews, and their sales are closely aligned with their overall user reception.

Game	Price	Average Sales	Positive Reviews	Negative Reviews
Dota 2	\$0.00	150 million	1,477,153	300,437
New World	\$39.99	75 million	154,914	73,900
PUBG: BATTLEGROUNDS	\$0.00	75 million	1,154,655	895,978
Destiny 2	\$0.00	35 million	403,109	77,006
War Thunder	\$0.00	35 million	246,926	67,506
Unturned	\$0.00	35 million	443,320	41,385

9. Conclusion

The analysis of the video game dataset reveals several interesting trends:

1. **Free-to-play dominance:** Free-to-play games dominate the sales charts, with several of the top-selling games having no upfront cost.
2. **User reviews:** There is a strong relationship between positive user reviews and game sales. Games with a high ratio of positive to negative reviews tend to sell better.
3. **Action genre:** The action genre leads in popularity among the top sellers, followed closely by free-to-play and adventure games.

These insights can be used to guide decisions in game development, marketing, and pricing strategies.

This report provides a comprehensive analysis of the video game dataset, using descriptive statistics, visualizations, and in-depth analysis to uncover patterns in game sales, pricing, and user reviews. The findings can serve as a guide for further research or business strategy in the video game industry.