Data Processing and Transformation Analysis

Objective:

The objective of this analysis was to load, transform, and analyze two datasets: "Video Game Sales" and "Video Games Traffic on Steam". The goal was to perform correlation analysis between the global sales of video games and their viewership on Steam.

1. Dataset Information:

- Video Game Sales Dataset:

- Number of rows: 16,598

- Number of columns: 11

- Data types: Integer, string, and double

- Issues identified: Year column is assigned as a string data type instead of an integer.

- Video Games Traffic Dataset:

- Number of rows: 199,999

- Number of columns: 5

- Data types: Integer, string, and double

- Issues identified: First row contains header values, "0" column is irrelevant.

2. Data Cleaning and Preprocessing:

- Video Game Sales Dataset:

- Converted the "Year" column from a string to an integer data type.

- Identified and removed duplicate rows. Number of duplicate rows: 16,598.

- No null or missing values found.

- Video Games Traffic Dataset:

- Removed the first row as it contained header values.

- Identified and removed duplicate rows. Number of duplicate rows: 199,292.

- Removed the "0" column as it was irrelevant.

- No null or missing values found.

3. Correlation Analysis:

- Calculated the correlation between the global sales of video games and their viewership.

- Result: Weak Positive Correlation

- The correlation coefficient suggests a weak positive correlation (0.0247) between global sales and viewership.

- As viewership increases, there is a slight tendency for global sales to increase.

- However, the correlation is relatively low, indicating that other factors influence video game sales apart from viewership.

4. Dataset Merge:

- Merged the datasets based on the common columns "Name" from the Video Game Sales dataset and "game-title" from the Video Games Traffic dataset.

- Result: Merged Dataset

- Number of rows: 85,114

- Number of columns: 15

5. Exploratory Data Analysis:

1. Genre-wise Sales:

- Action video games are the most sold games globally, followed by shooter video games.

- This indicates a strong demand for action-oriented and shooter genres in the gaming market.

2. Distribution of Global Sales:

- The majority of games have global sales values between 0 and 2 million units.

- There is a long tail on the right side, suggesting a few games with significantly higher global sales.

3. Global Sales vs. Value:

- There is a positive relationship between global sales and value, indicating that higher global sales correspond to higher value.

- However, there is variation in value for games with similar global sales, suggesting other influencing factors.

4. Variation in Global Sales:

- Different gaming platforms exhibit varying levels of global sales, with some platforms having a wider range and higher median sales.

- Platform choice can impact the overall sales performance of games.

5. Outliers:

- Outliers in the box plot indicate exceptional cases where certain platforms have achieved significantly higher global sales.

- These outliers may represent platforms with a strong user base or successful game titles.

6. Sales Trends Over the Years:

- The line graph shows a general increasing trend in video game sales across all regions (NA, EU, JP, and Global) over the years.

- This indicates a growing market and opportunities for game sales.

7. Top 5 Publishers by Video Game Sales:

- Take-Two Interactive, Electronic Arts, Ubisoft, Bethesda Softworks, and Vivendi Games are among the top publishers for video game sales.

- These publishers have achieved high sales figures globally and in specific regions like NA, EU, and JP.

8. Most Played Games, Genres, and Publishers:

- The top 5 most played games are Terraria, Counter-Strike, Grand Theft Auto V, Sid Meier's Civilization V and Borderlands 2.

- The top 5 genres with the most played games are Action, Shooter, Role Playing, Strategy, and Racing.

- The top 5 publishers for which most games are played are Take-Two Interactive, 505 Games, Electronic Arts, Ubisoft, and THQ.

9. Correlations:

- Higher-ranked games tend to have higher sales in different regions.

- There is a weak negative correlation between Year and Rank, indicating that game ranking tends to decrease over time.

10. Global Sales and Viewership:

- There is a weak positive correlation between global sales of video games and the number of viewers.

- This suggests that an increase in viewership is associated with a slight tendency for global sales to increase.

These findings provide valuable insights into the video game sales patterns, genre preferences, and the influence of variables such as platform, publisher, and viewership on sales.

**Conclusion:**

Based on the analysis, we found a weak positive correlation between the global sales of video games and their viewership on Steam. However, the correlation coefficient suggests that viewership alone has limited impact on video game sales. Other factors such as marketing strategies, game quality, pricing, and platform availability likely play a more significant role in driving sales.