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

**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.