**Assignment 12.2**

**DSC 530: DATA EXPLORATION AND ANALYSIS**

**Term Project - VIDEO GAMES SALES Analysis**

**Author - Amita Shah**

**Instructor - Cary Jim**

**Date - 02/22/2024**

As part of the project on video game sales data, extensive Exploratory Data Analysis (EDA) and statistical testing were conducted to understand the dynamics of video game sales globally and across different regions. The analysis included descriptive statistics, probability mass functions (PMFs), cumulative distribution functions (CDFs), hypothesis testing, and regression analysis. This comprehensive approach aimed to uncover patterns and relationships within the data, guiding insights into the video game industry's sales trends.

**Outcome of EDA**

The EDA revealed significant insights, such as the distribution of game sales across different regions and the impact of factors like platform, genre, and release year on sales performance. The regression analysis highlighted the predictive power of regional sales on global sales figures, indicating strong correlations, particularly between North American sales and global success. The hypothesis further underscored changes in market dynamics over time, suggesting an evolution in consumer preferences and market reach.

**Missed Opportunities in Analysis**

While the analysis covered various aspects, there was a missed opportunity in exploring the role of critical factors such as game reviews, player ratings, and social media influence on sales. Incorporating user engagement metrics could have provided a deeper understanding of consumer preferences and their impact on sales success.

**Additional Variables for Enhanced Analysis**

Incorporating variables related to digital vs. physical sales, pricing strategies, and promotional efforts could have enriched the analysis. Given the industry's shift towards digital distribution, examining this aspect could reveal trends in consumer purchasing behavior and the impact of digital platforms on global sales.

**Assumptions and Their Validity**

The analysis assumed linear relationships in regression models, which might not fully capture the complexities of the video game market. For example, the assumption that past sales performance predicts future success does not account for shifts in consumer technology or sudden changes in market conditions. Acknowledging non-linear relationships and market volatility could provide a more nuanced understanding.

**Challenges and Areas of Improvement**

One of the primary challenges was dealing with incomplete data, especially missing values in the year of release and sales figures for some games. This limitation might have skewed the results or omitted valuable insights. Additionally, the assumption of equal variances in sales across regions could be reconsidered, given the diverse market conditions and consumer behavior across different geographical areas.

Understanding the impact of external factors, such as economic conditions or global events, on video game sales was another area that could have been explored further. These factors likely influence consumer spending patterns and, consequently, video game sales, suggesting a more complex interplay of variables than initially considered.

**Conclusion**

The project provided valuable insights into the video game industry's sales dynamics, emphasizing the importance of regional sales in predicting global success. However, incorporating additional variables and considering non-linear relationships could enhance the analysis. Addressing the challenges of incomplete data and re-evaluating assumptions about market behavior would further refine the understanding of factors influencing video game sales. This exploration underscores the multifaceted nature of consumer markets and the need for a holistic approach to data analysis.