Introduction to Machine Learning: Project

VIDEO GAMES SALES ANALYSIS

By: Kushal Therokar, Nimesh Wankhede, Snehal Kamble, Vamshi Krishna Perabathula

TABLE OF CONTENTS

1. [**INTRODUCTION**](#INTRODUCTION)
2. [**DATA EXPLORATION**](#DATA_EXPLORATION)
3. [**VISUALIZATIONS**](#VISUALIZATIONS)
4. [**MACHINE LEARNING MODELS**](#ML_MODELS)
5. [**SUMMARY**](#SUMMARY)
6. [**FUTURE SCOPE**](#FUTURE_SCOPE)
7. [**REFERENCES**](#REFERENCES)

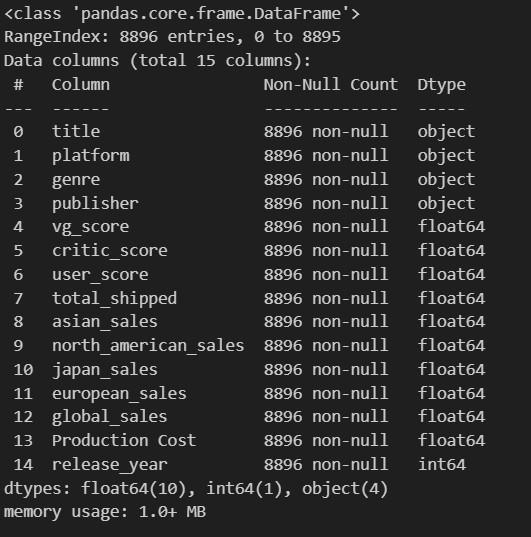
**INTRODUCTION**

* This research aims to study a dataset about video game sales to find out valuable information for people in the gaming industry, like game makers and sellers. The dataset includes details about games, like their names, platforms, release years, type and how many copies they sold in distinct parts of the world. We want to figure out which types of games are most popular, which game systems make the most money, and how these things have changed over time.
* We will also look at the sales data from different regions to understand what kinds of games people in various places like to play. This will help game companies make games, advertise them, and sell in a way that matches what people in different areas like. Additionally, we want to see how things like the type of game and the game system affect how well it sells. The information is important for game developers and sellers to make their games and marketing better and to focus on the most profitable parts of the market.
* The main people interested in this research are game companies that want to make money and have the biggest share of the market. The information that we find will help them decide what kinds of games to make, which game systems to concentrate on, and which areas to target with their marketing and distribution plans. This could result in more sales, a stronger presence in the market, and better advertising for these companies.
* This research will give a detailed look at trends in gaming and what people like about it. It will help the game companies be competitive, adapt to changes in the market, and create games that people want to play, leading to continued success.

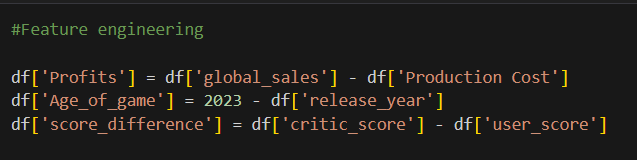
**DATA EXPLORATION**

* At the start, we carefully organized the data. We made sure there were no missing details and adjusted some features to make the dataset more useful for our analysis.
* We checked out how sales are spread out globally and looked at some stats to understand the sales figures better. A correlation heatmap helped us see connections between various aspects of the data.

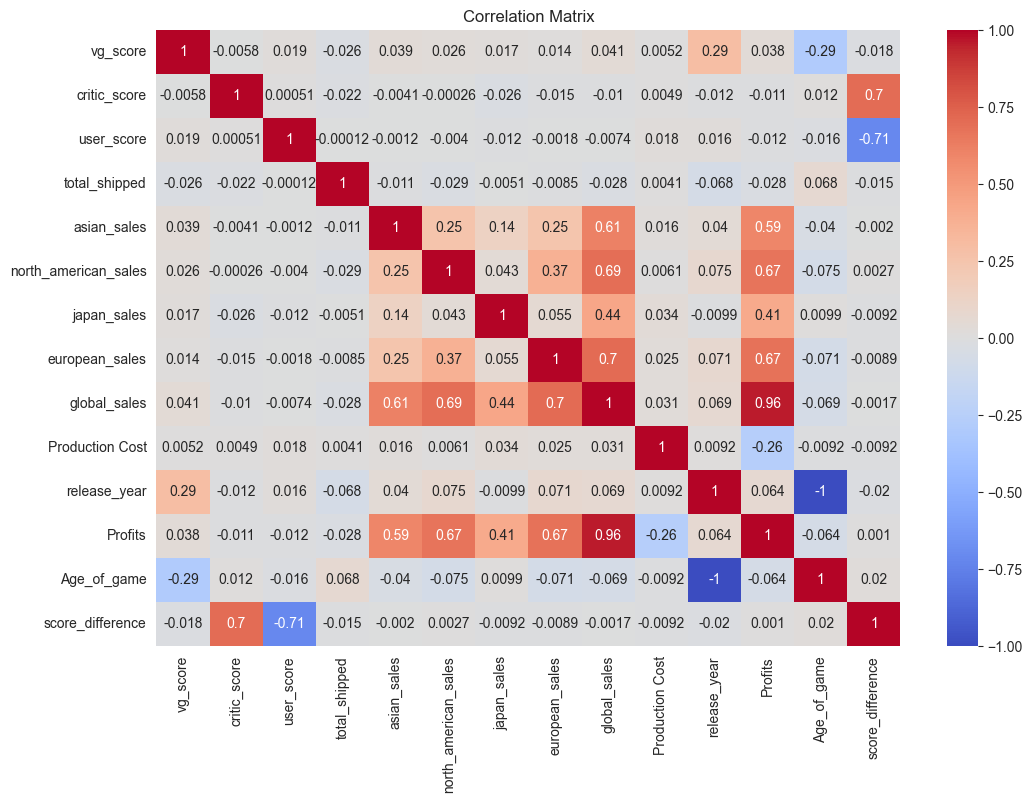
**INFORMATION OF DATA**

****

**FEATURE ENGINEERING**

****

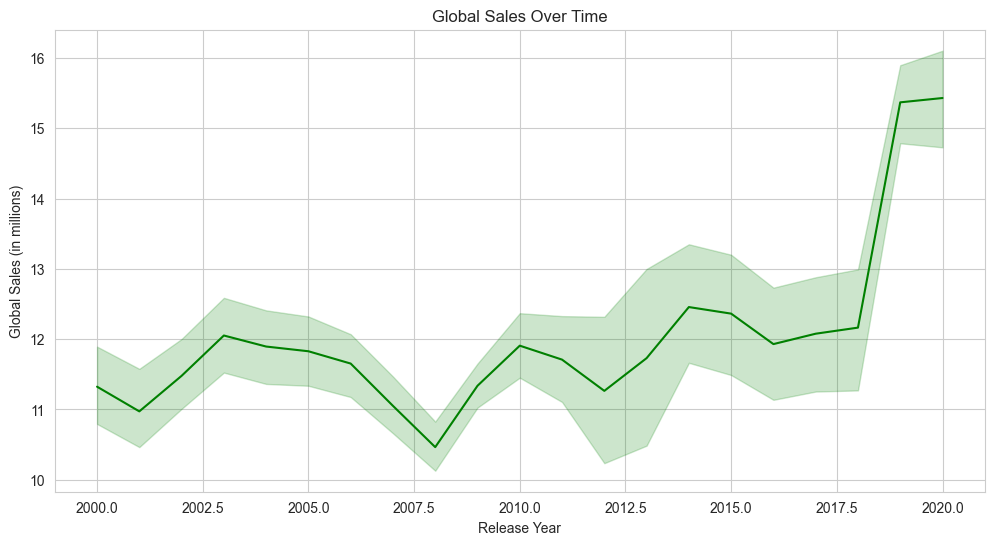
**CORRELATION HEATMAP**

****

**VISUALIZATION**

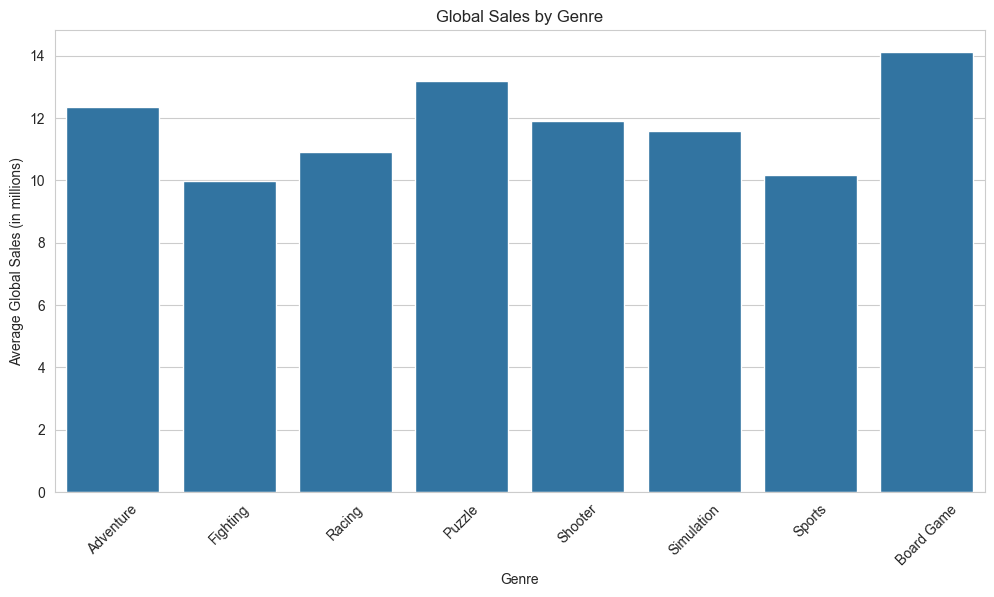
We created a line graph to show how global sales change over time. Bar graphs helped us see sales by game genre, average sales in various places, and the top game publishers. These visuals help make complex data easier to understand.

**LINE PLOT: GLOBAL SALES OVER TIME**



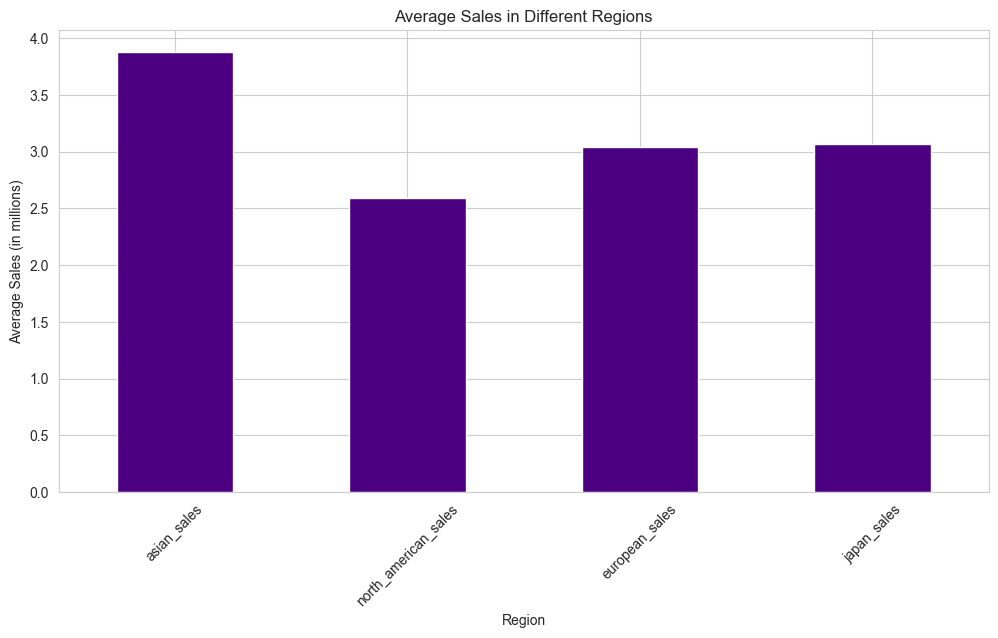
* This line plot here shows the net sales of video games over the period. In the 2000s the net sales were close to 11.5 million, the sales kept increasing till 2003 to 12 million and then started dropping gradually till the middle of 2008 to 10.5 million.
* The net video game sales have increased since then and in 2020 we have net video games sales of 15.5 million

**BAR PLOT: GLOBAL SALES BY GENRE**

****

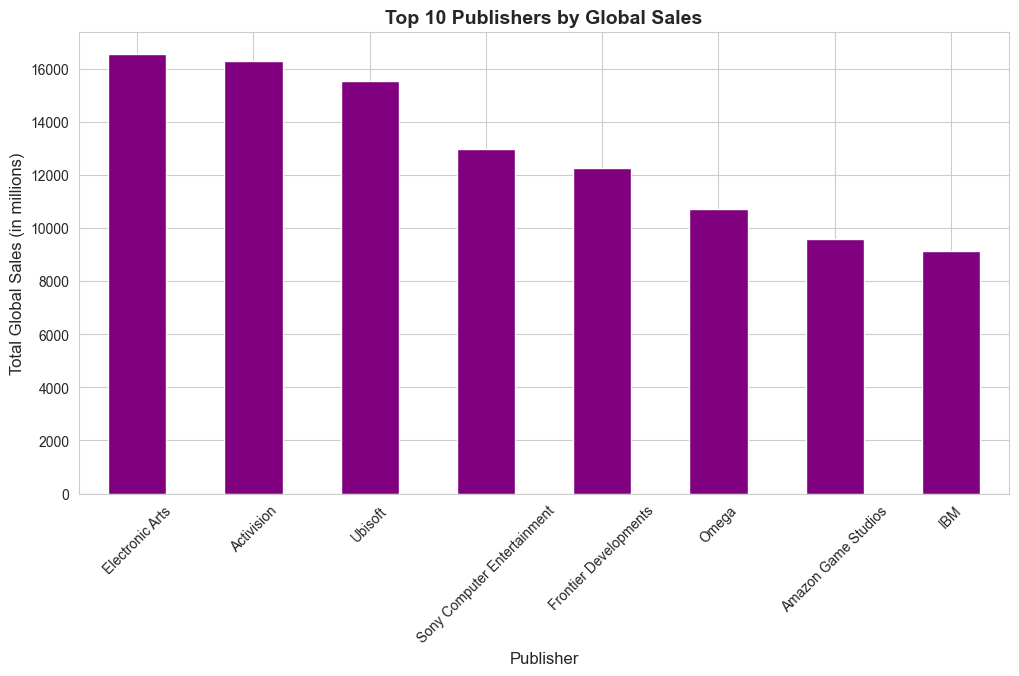
* Here in this histogram, we aim to see the distribution of sales according to the genre to see which genre is popular among the audience.
* We can clearly see that board games have the highest sales of 14 million followed by puzzle games and adventure games
* This is an extremely useful insight for game makers to continue making games in such genres for higher sales.

**BAR PLOT: AVERAGE SALES IN DIFFERENT REGIONS**

****

* This histogram shows us the distribution of video game sales according to the region.
* The sales in Asia are the highest, followed by Japan and Europe
* Game makers should keep this in mind and give more priority to the customers in these regions and make sure that the requirements of people from these regions are met.

**BAR PLOT: TOP 10 GAME PUBLISHERS**

****

* The top 10 game publishers can be seen here in this bar plot.
* These are the most reliable and most consistent game publishers in the industry.
* Game players can rely on these publishers to have the best gaming experience as these publishers have the most sales, and looking at the sales they have the experience of delivering flawless content.

**MACHINE LEARNING MODELS**

In our project, we have used two supervised learning algorithms, *linear regression,* and *random forest regression.* We performed these two models on our data against our target variable of global sales. Both algorithms performed well on our data with close to 99% accuracy. This high accuracy, however, indicates that our model may be overfitting. To fix this problem, we used lesser features for linear regression, and selected the most relevant ones based on domain knowledge. For random forest, we performed hyperparameter tuning using k-fold validation.

**SUMMARY**

Our dataset is solid—no missing info, making our analysis trustworthy. We found interesting stuff, like how sales vary a lot globally and the different scores games get. Looking at profits, not just sales, gives us a broader view of financial success in the gaming world. The dataset has games from different years, letting us see how sales trends have changed.

**FUTURE SCOPE**

As we wrap up, there is more we can explore in the future. Predicting game sales based on scores, genre, and publisher is something to investigate. Tailoring games to specific groups of players and finding ways to make games more cost-effective are exciting possibilities. Also, digging into how release timing and market conditions affect sales over time could reveal interesting insights.

**REFERENCES**

* [www.vgchartz.com](http://www.vgchartz.com) : It is a web scraping website which gathers sales data across the globe for video games.
* Used chatgpt and stackoverflow for debugging our code.
* [www.geeksforgeeks.org](http://www.geeksforgeeks.org) : for getting general help with python