BANGABANDHU SHEIKH MUJIBUR RAHMAN SCIENCE AND TECHNOLOGY UNIVERSITY, GOPALGANJ, BANGLADESH

Introduction to Programming with Python

Project Report On

Video Game Sales Analysis: Predicting Trends with a Regression Model

Report Submitted By:

Name: Nisat Sultana Nira

Batch Id: B05020

EDGE-BSMRSTUCSE Digital Skills Training

ICT DIVISION
FUTURE IS HERE

BANGLADESH COMPUTER COUNCIL
COMPUTER FOR EVERYTHING

EDGE
ENHANCING DIGITAL GOVERNMENT AND ECONOMY

Table of Contents

1. Introduction

Since early years of 1980s, video game industry emerged as one of the major entertainment industry. In the last few decades, they took the world by storm with the enthalpies of our grand little globes, and the video games, playing a diversified role in running the waves, have reached far into the people's face. And this has resulted in a lot of player engagement, it is not just in terms of how many they play but also in terms of how much revenue they generate every year within the industry, billions of dollars.

In this project, we analyse historical sales data for video games from 1980 till 2020 to understand meaningful trends and patterns. By examining this data, we aim to answer key questions:

- What percent of global sales do different regions and genres generate?
- What are the top performing games and publishers?
- What charts are thrown at the wall to answer the question of how do video game sales trends change?

To get these results, we use data visualization techniques, like line charts and bar graphs, to show trends by region, genre as well as single games. Second, we run a regression on regional sales data to predict future sales. By utilizing this predictive approach, we're able to understand how each market relates to the overall sales of games, which helps industry stakeholders make data driven decisions about when to release each game, how to market these games, or what to invest as a whole.

The project is built using data analysis and predictive modeling to give a complete picture of the video game industry to help stakeholders make informed decisions amidst such a competitive and volatile market.

2. Features

I. **Data Analysis**

- **Breakdown by Region:** We divide sales data on major markets: North America, Europe, Japan and other markets. This breakdown enables the examination of regional differences in what consumers want, and what they can afford to spend on games, to glean meaningful knowledge about the difference regarding distribution across different demographics and market sizes.
- **Breakdown by Genre, Platform, and Publisher:** The sales looking across all genres (Action, Role-Playing, Sports) gives us an idea of what popular game types are and how trends are changing. Analyzing by platform (PlayStation, Xbox, etc.) allows to identify which gaming console was most conquering and analyzing by publisher helps to understand who is the big player of the industry and which are the ones standing out with main keys trends.

II. **Top-Selling Games**

- **Identification and Visualization:** This feature identifies the top five best selling games in the world — ensuring that a snapshot of the games that have the greatest sales impact are shown. Bar charts help visualize these games, highlighting how much of the market they capture, and patterns in consumers spending preferences (for example, between family friendly or iconic franchise games).

III. **Sales Trends Over Time**

- **Analysis of Trends by Year:** This feature shows the sales over the years (1980–2020) and how the sale changes from decades. So it will contain interesting ones like the rise of video gaming in the 2000s, and periods of decline. We also show how regional preferences evolve over time, as well as which genres are gaining or losing popularity.

IV. Correlation Analysis

- **Examination of Regional Correlations:** This feature explores the relationship between sales in each region and global sales based on a correlation matrix. For example, a high correlation between North American and international sales implies that to a large degree North America drives the global performance. It's about understanding these correlations because companies that based their actions on regional trends have had to guess at what success will look like globally.

V. Predictive Modeling

- **Regression Model for Global Sales Prediction:** Secondly, a linear regression model was created to predict the global sales based on sales from North America, Europe and Japan. This model uses regional sales as independent variables and predicts the global sales very well. Given their usefulness, they let the industry stakeholders anticipate whether their games will sell and make such decisions as when to launch a game, how to market it and where to invest in resources.

3. Implementation

This project uses Python and essential data science libraries for analyzing, visualizing, and modeling video game sales data:

- **Data Loading and Processing:** The dataset is imported, cleaned, and manipulated with `Pandas` to subject it to analysis by structure, base regions, base genres, and years.
- **Visualization:** In `Matplotlib` and `Seaborn`, we use line plots, bar charts and heatmaps to show trends over time, genre popularity and correlations between regions allowing us to see trends clearly visually.
- **Regression Model:** Global sales are predicted by a linear regression model based on North American, European and Japanese sales, at relatively high accuracy. This predictive model is used to help with strategic decision making by predicting future sales possible.

Key code snippets:

Data Load:

```
df = pd.read_csv("vgsales.csv")
```

Top-Selling Games:

```
top_5_games = df[['Name', 'Global_Sales']] .sort_values (by =
'Global_Sales', ascending=False).head(5)
top_5_games.plot(x='Name', y='Global_Sales', kind='bar', color=
'gold')
```

Game Sales by Year (1980-2020)

```
sales_columns = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales']
yearly_sales = df.groupby('Year')[sales_columns].sum()
yearly_sales.plot(kind='line', figsize=(10,4))
```

Game Sales by Genre

```
genre_sales = df.groupby('Genre')[sales_columns].sum()
genre_sales_total = genre_sales.sum(axis=1)
print(genre_sales_total)
genre_sales_total.plot(kind='pie', figsize=(10, 6),
autopct='%1.2f%%', startangle =90 ,cmap='Set3')
```

Correlation Matrix:

```
sales_columns = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales']
yearly_sales = df.groupby('Year')[sales_columns].sum()
yearly_sales.plot(kind='line', figsize=(10,4))
```

Regression Model:

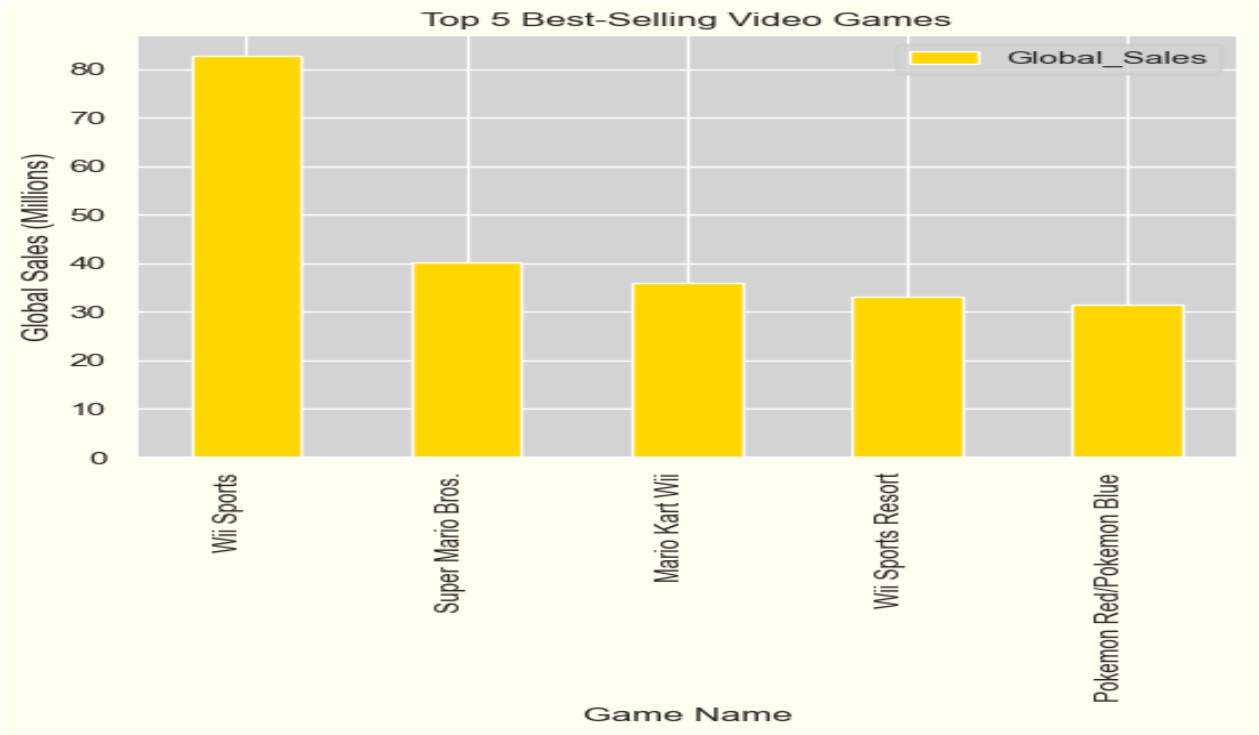
```
corr = df[sales_columns + ['Global_Sales']].corr()
plt.figure(figsize=(8,4))
sns.heatmap(corr, annot=True, cmap='YlGnBu', fmt='.2f')
```

4. Reports

- **An explanation of each variable:**

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

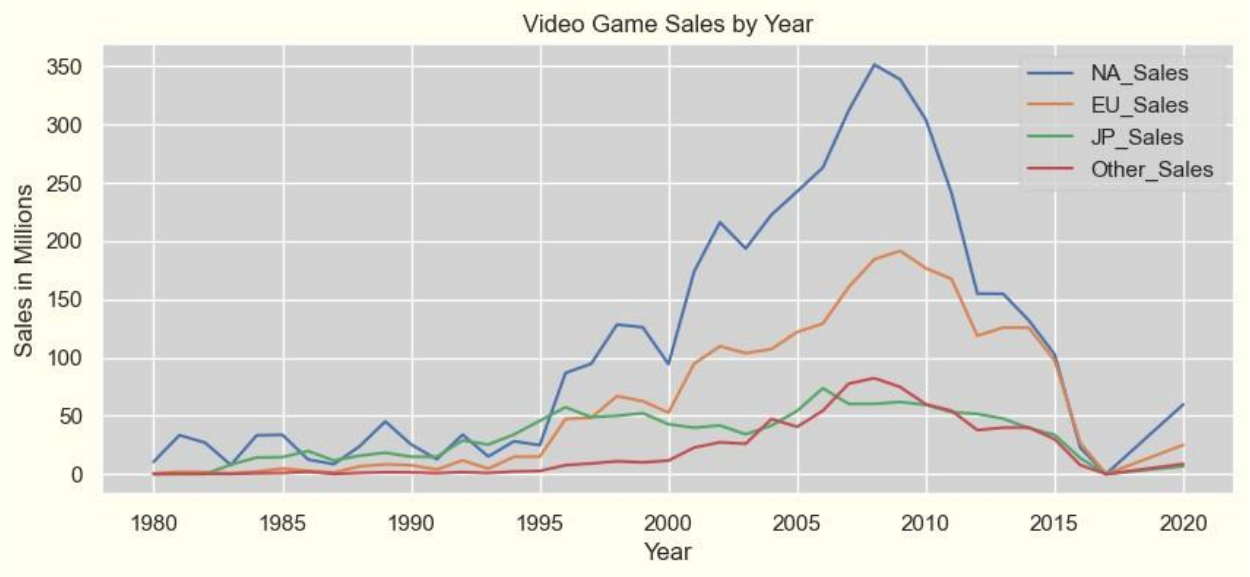
- **Top-Selling Games**



I. **Overview:** Wii Sports, Super Mario Bros., Mario Kart Wii, Wii Sports Resort and Pokemon Red/Blue are the top five best sellers of all time.

II. **Insight:** These were each titles to reach out into the millions of players on a worldwide scale, reaffirming Nintendo as a games powerhouse throughout gaming. Dominated by family-friendly themes, accessibility, and broad appeal, the games are fairly easy to play and accomplish great global sales.

- **Sales by Region**



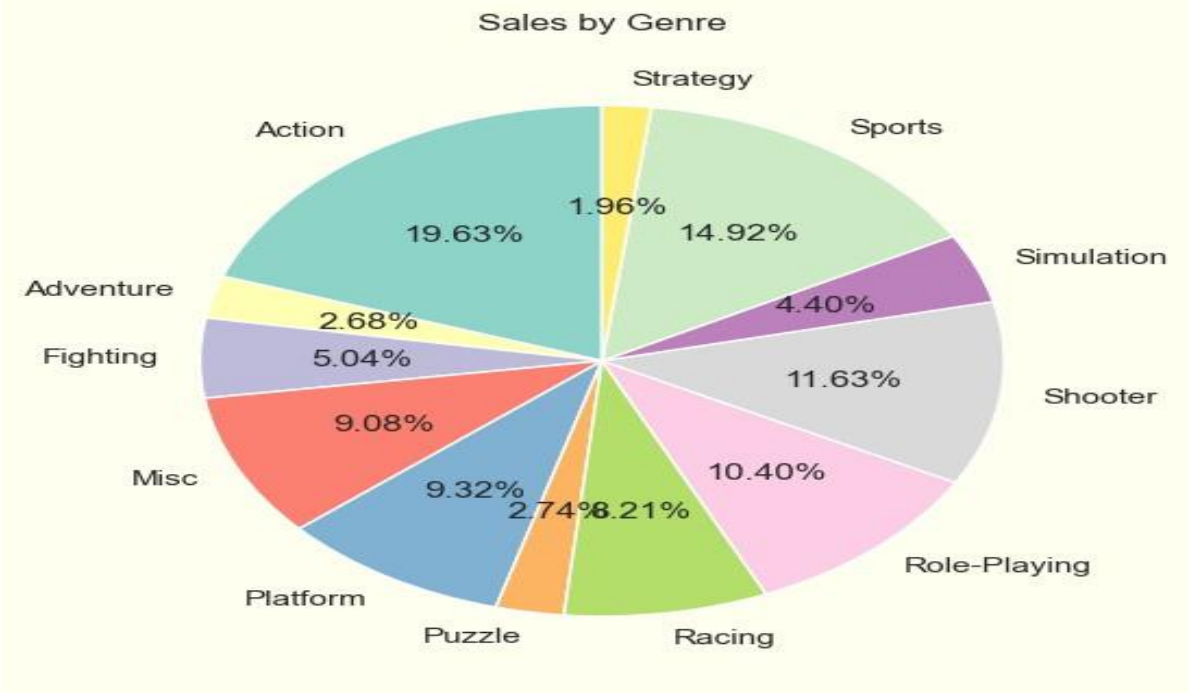
I. North America: With 350 million units in peak sales, North America has always been leading video game sales and they peaked in the mid 2000s. Strong consumer demand and an already robust gaming industry comprise the major reasons for the region's sales dominance.

II. Europe: The sales trend in Europe matches North American but with a slightly late peak in the 2000s at around 200 million units. Although to a lesser extent, global sales are driven by Europe.

III. Japan: The smallest of the major markets, Japan has a known demand for video games whose peak was in very early 2000s. But sales have been dropping off year by year, which suggests a more crowded or more exclusive market.

IV. Other Regions: While sales in other regions always tend to be lower, they increasingly grow in the 2010s reflecting growth potential in the emerging markets.

- **Sales by Genre**

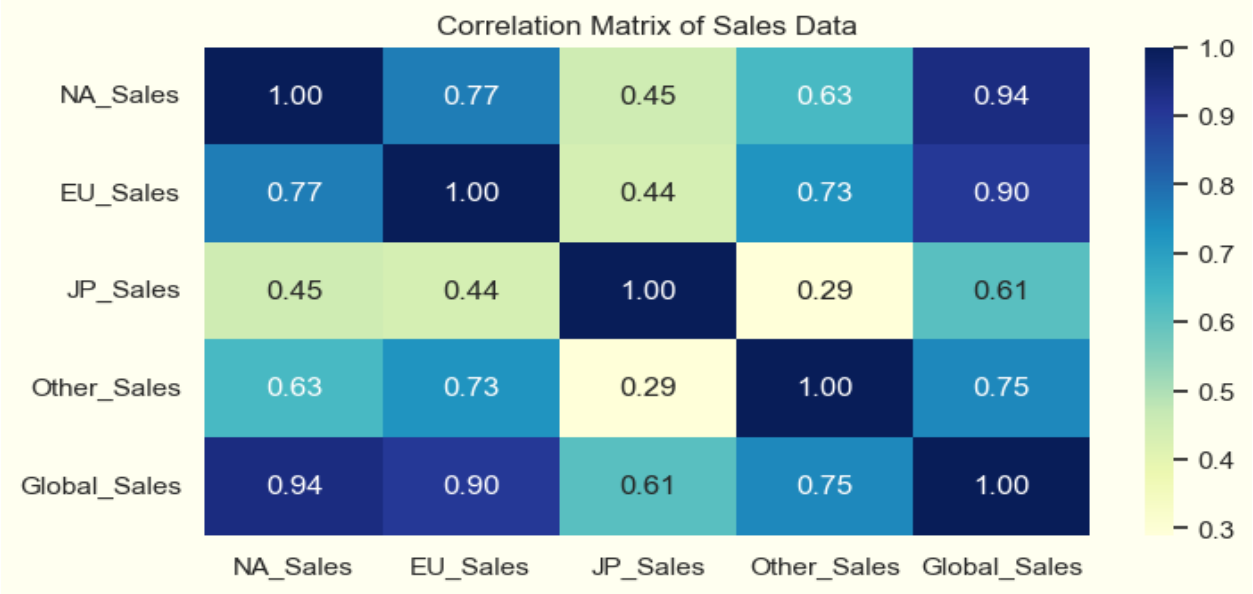


I. **Top Genres:** Global sales are led by Action and Role-Playing games, Action games with 19.63 per cent of total sales and Role-Playing games at 18.21 per cent. Clearly this indicates that consumers are very attracted to play, games that are entertaining, interactive and imbued with a storyline.

II. **Strong Performers:** Sports, Shooter and Racing are also prominent genres, attracting strong share from audience interested in various categories.

III. **Lower Performing Genres:** Yet there are far smaller sales derived from Puzzle, Fighting, and Adventure genres, though they retain a faithful growing subspecies. Strategy games make up a tiny sliver of the market, with only a small share of the market.

- **Correlation Analysis**



I. **Findings:** The correlation analysis shows that the highest correlation to global sales is exhibited from North American sales (correlation coefficient of 0.94), indicating that to be a successful global brand, you must be successful in North America.

II. **Other Regions:** While sales in Europe also highly correlate with global sales (0.90), Japan's correlation is only moderate (0.61), so that sales trends in Japan are not very predictive of global success.

- **Predictive Model Performance**

I. **Model Overview:** High predictive accuracy can be achieved by the regression model that uses North American, European and Japanese sales as predictors for global sales.

II. **Performance Metrics:** The model explained 99.33% of the variance in global sales, using an R squared value of 0.9933. The low Mean Squared Error (MSE) also leads us to believe that it's an accurate tool to forecast global sales upon regional data.

Let,

Y = Global sales

X_1 = NA Sales , X_2 = EU Sales , X_3 = JP Sales

The model will be :-

$$Y = 0.0063 + 1.0469X_1 + 1.2226X_2 + 0.9623X_3 + \varepsilon$$


Actual vs. Predicted Global Sales
Predicted Global Sales
Actual Global Sales
0
20
40
60
80
0
20
40
60
80

5. Conclusion

This work analyzes video game sales statistics from 1980 till 2020 and reveals key trends in terms of market, regions, and genres. Global sales were predominantly underpinned by the leading market of North America, and Action, Role-Playing, and an increasing number of Games dominated genres. Multiple top selling games worldwide came from Nintendo. The predictive model performs well in prediction and demonstrates that stakeholders can predict global sales using regional data, and that such a model is a powerful strategic planning tool for the gaming industry.