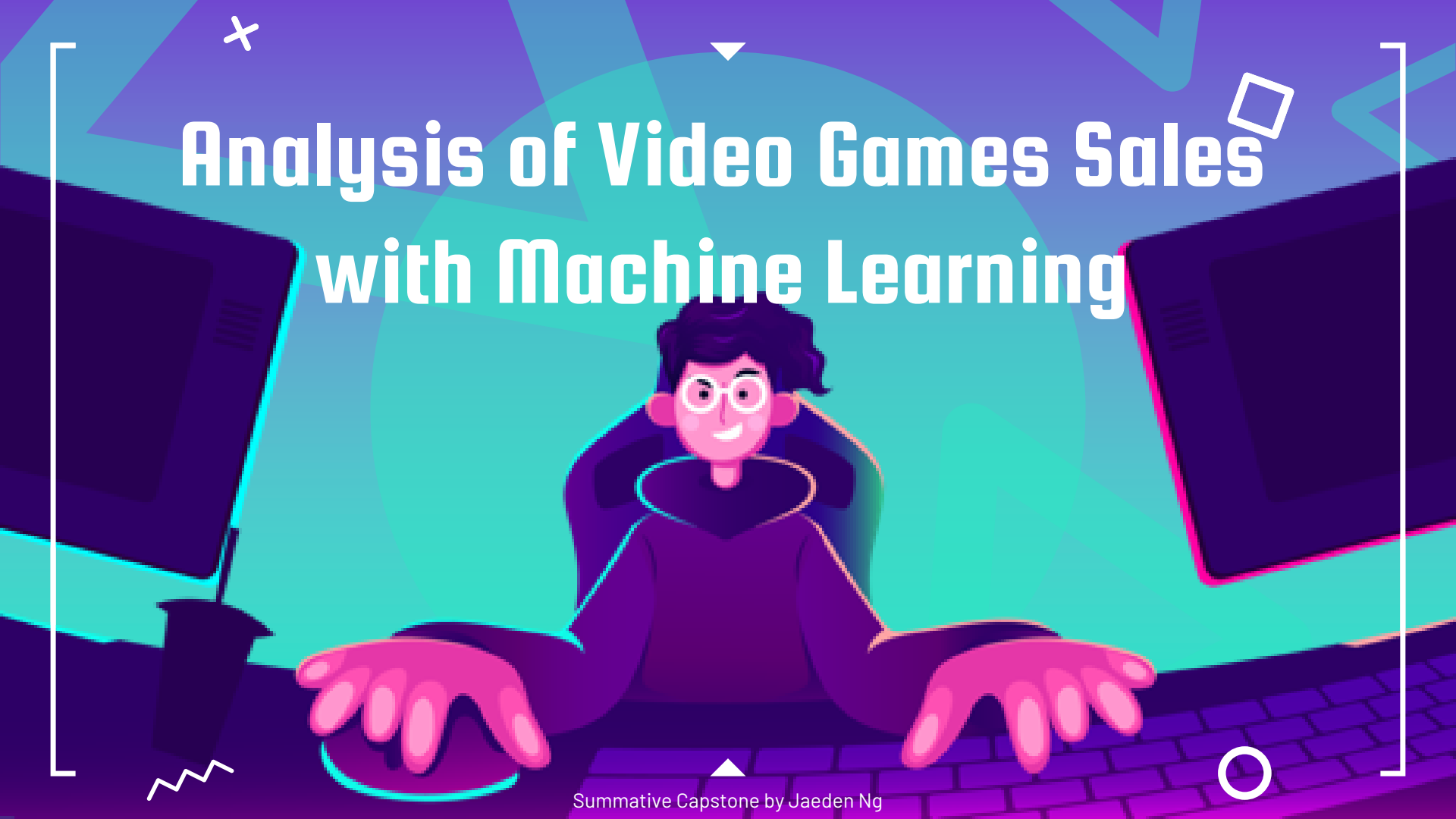
Analysis of Video Games Sales with Machine Learning



Summative Capstone by Jaeden Ng

INTRODUCTION

Problem Statement & Definition

Methodology

Datasets, Models, Metrics, Tools

Process Flow

Data Preparation, EDA, Data Analysis, ML model training/evaluation

Results

Prediction & Test Results

CONCLUSION

Recommendations

Introduction – Problem Statement

As the product sales manager of the video **game distribution company**, I want to know what are the viable products that I can propose to introduce to the market. This analysis will help **predict which product category** and how much **sales revenue** this product can bring for the company.

Introduction – Problem Definition

Some of the areas to analyze:

- What genre of games product to introduce.
- What are the top games publishing company.
- Which market region will have the most impact for global sales.

“We are all our own worst enemy.
But also, our best teacher.”

—Super Street Fighter IV

Methodology

Datasets, Models, Metrics, Tools

Datasets

11563 unique values

https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings

(Data updated as of 22 Dec 2016)

Machine Learning for Video Games Sales

ML Models

Linear Regression

Metrics

Coefficients
Mean Square error
Coefficient of determination

Tools

Google Colab
Numpy
Pandas
Matplotlib
Seaborn
Sklearn

Process Flow

Data Preparation & Transformation

```
check_na(df_vgames)
```

Name	2
Platform	0
Year_of_Release	269
Genre	2
Publisher	54
NA_Sales	0
EU_Sales	0
JP_Sales	0
Other_Sales	0
Global_Sales	0
Critic_Score	8582
Critic_Count	8582
User_Score	6704
User_Count	9129
Developer	6623
Rating	6769

Data Cleaning

- Checking for NULL values and dropping rows
- Dropping of unused columns

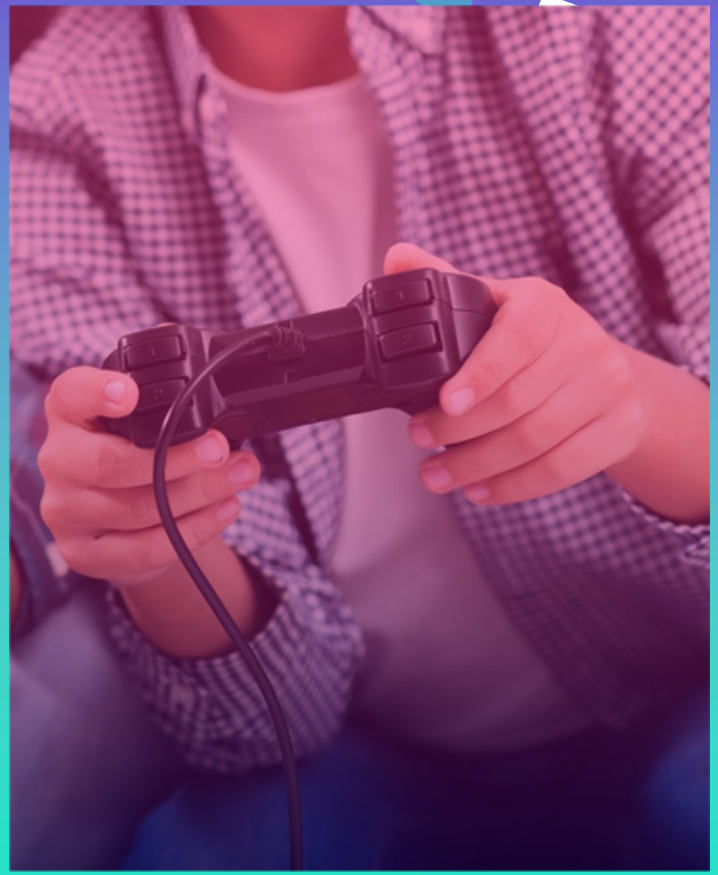
Balancing Target Variable

- Balancing the records ensure more accurate prediction

Exploratory Data Analysis

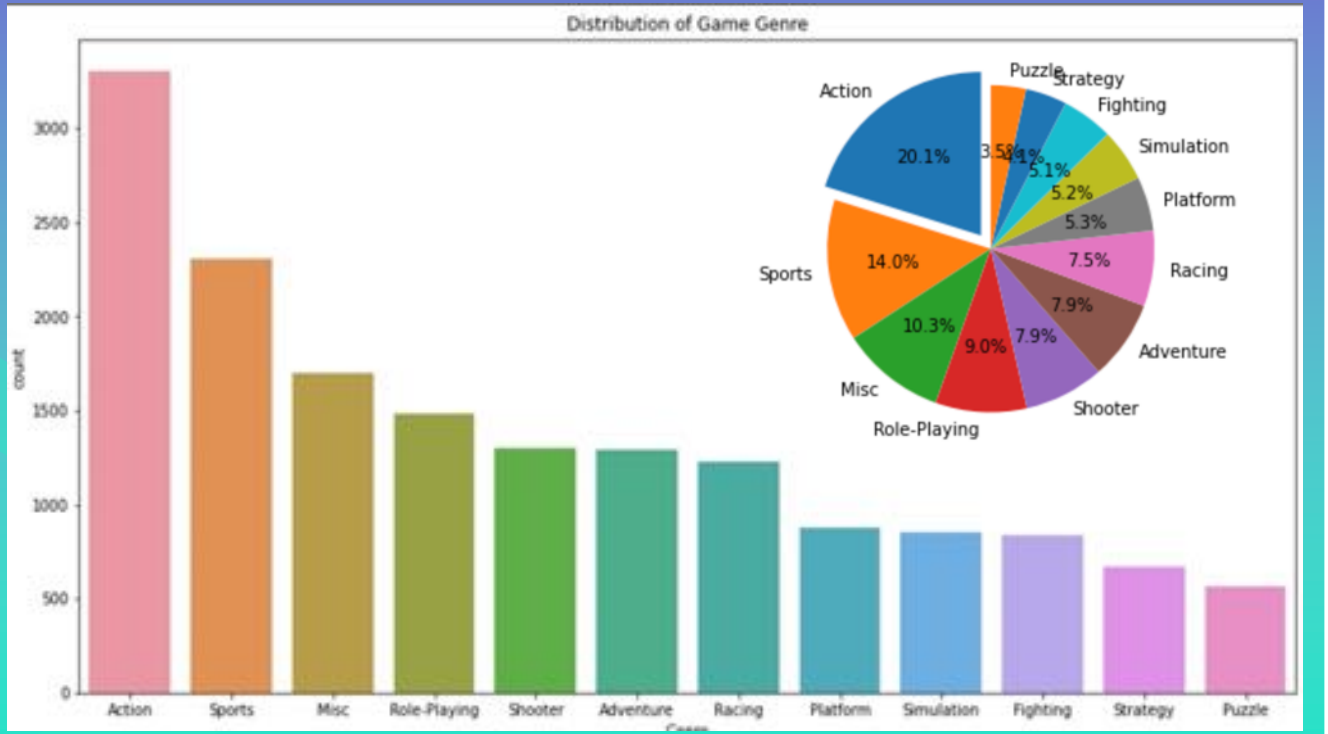
Exploratory Data Analysis

Performed initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

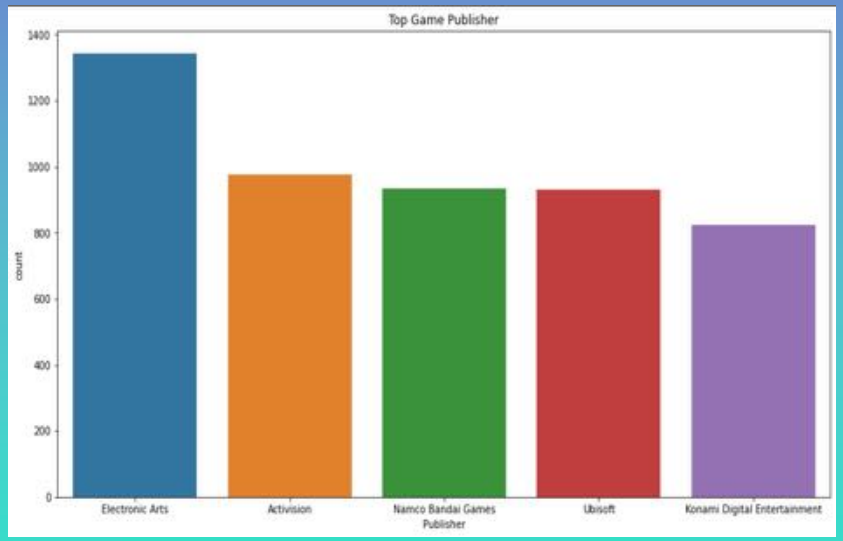


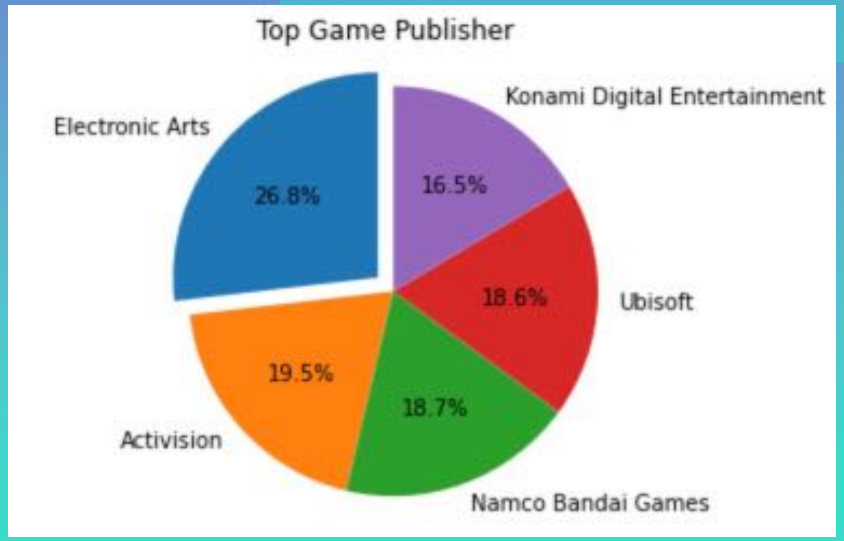
Distribution of Game Genre

Action	3307
Sports	2306
Misc	1697
Role-Playing	1483
Shooter	1296
Adventure	1291
Racing	1225
Platform	878
Simulation	855
Fighting	837
Strategy	672
Puzzle	569

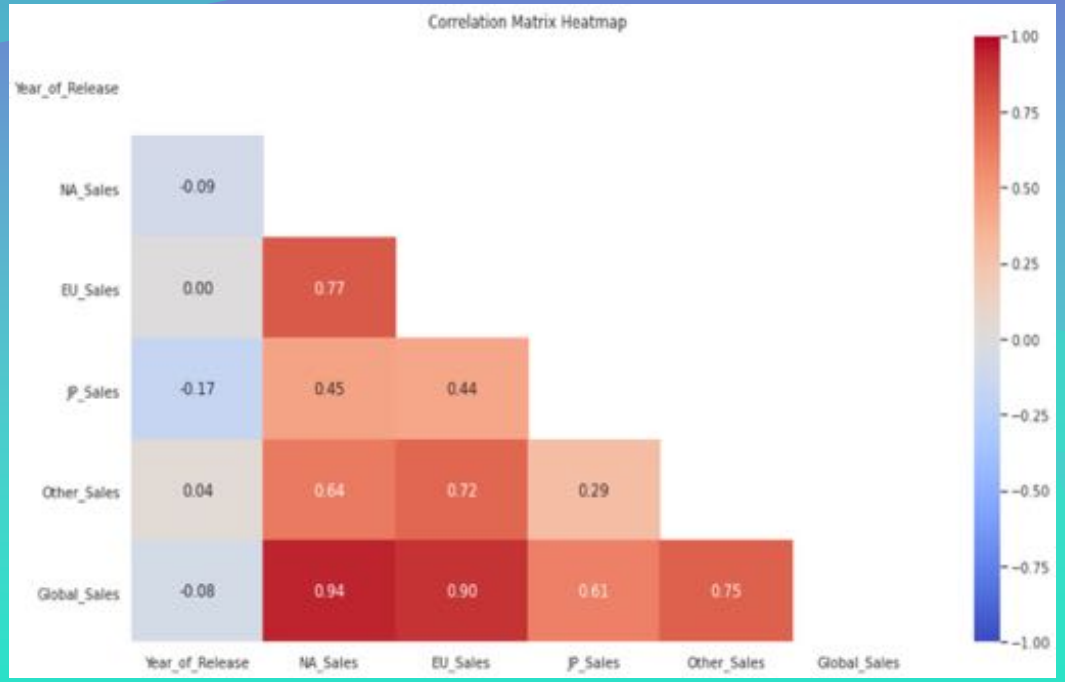


Top Game Publishers





Correlation check for Machine Learning

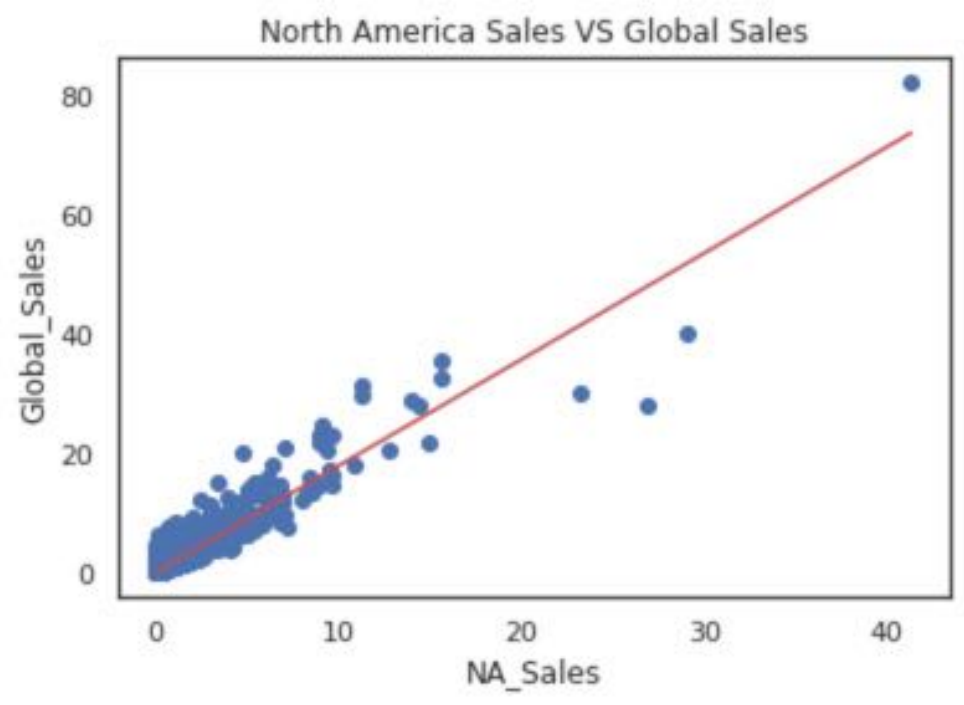


Base on the correlation check, we found that the column "**NA_Sales**" has the highest correlation with "**Global_Sales**" with a score of **0.94**

Machine Learning

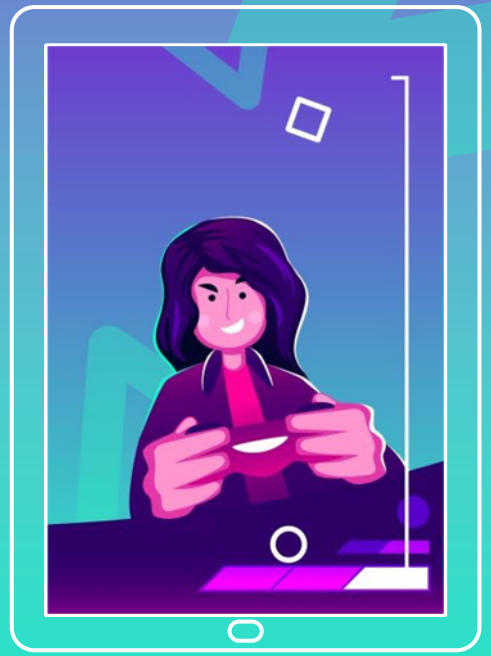
Using Linear Regression

North America Sales versus Global Sales



Training Results/Prediction

Desired Output (Actuals)	Predicted Output
0.6	0.747262
0.6	1.025940
0.6	0.816932
0.6	1.165279
0.6	0.503419
0.6	0.660175
0.6	0.869184
0.6	0.520836
0.6	0.816932
0.6	0.834349



Validating Prediction

```
Coefficients:
 [1.74173742]
Mean squared error: 0.09
Coefficient of determination: -2.85
```

```
Regression model's training score = 0.86
Regression model's test score     = -2.85
```

Exporting/Saving

To CSV/SQL Server

The training and test data are exported/save into csv files for future use. It may also be save into an SQL database.

```
compare_df.to_csv("vgames_pred.csv")
compare_df.head(10)
```

	Desired Output (Actuals)	Predicted Output
3333	0.6	0.747262
3334	0.6	1.025940
3335	0.6	0.816932
3336	0.6	1.165279
3337	0.6	0.503419

Conclusion/Recommendation

- Top popular video games genres are **action** and **sports**
- Consider importing games published by **Electronics Arts**
- Recommend to push more sales to **North America** where the main market is.

THANKS!

Do you have any questions?