UNIVERSITY OF BIRMINGHAM | BIRMINGHAM BUSINESS SCHOOL

Assessment and Feedback: Student Template

Student ID Number(s): 2659883

Programme: MSc in Business Analytics

Module: Data Analytics and Predictive Modelling

Name of Tutor: Dr Hannan Amoozad Mahdiraji

Assignment Title: Individual Report (60%)

Date and Time of Submission: 15/01/2024 – 07:00 AM

Actual Word Count: 3025

Extension: N

I do not wish my assignment to be considered for including as an exemplar in the **School Bank of Assessed Work.**

The purpose of this template is to ensure you receive targeted feedback that will support your learning. It is a requirement to complete to complete all 3 sections, and to include this completed template as the first page of every assignment that is submitted for marking (your School will advise on exceptions).
Section One: Reflecting on the feedback that I have received on previous assessments, the following issues/topics have been identified as areas for improvement: • Writing Pattern • Problem Solving • Critical thinking
Section Two: In this assignment, I have attempted to act on previous feedback in the following ways: • Focused on Critically analysing problems. • Effective writing skills • Time management

Section Three: Feedback on the following aspects of this assignment (i.e., content/style/approach) would be particularly helpful to me: (3 bullet points)

- Literature Review
- Data Analysis
- Content

By submitting your work online you are confirming that your work is your own and that you understand and have read the University's rules regarding authorship and plagiarism and the consequences that will arise should you submit work not complying with University's Code of Practice on Academic Integrity.

I confirm that I have / have not used a proof-reader(s)

Programme Title	MSc in Business Analytics
Module Title	Data Analytics and Predictive Modelling
Module Code	38157
Assignment Title	Individual Report (60%)
Level	PG - Semester 1 (2023-2024)

Report Title

Hidden trends In Videogames and its Future

Student ID

2659883

Abstract/Executive Summary

The Purpose/Objective of this research is to analyse the dynamics of the video game industry by studying the patterns across different regions and platforms on which the games are played. This analysis aims to classify the games based on sales trends and to classify them by using different relevant features so as to predict the future sales which will be helpful for the industry stakeholders as it will offer actionable insights. The variables are Game Title, Platform, Year, Genre, Publisher, Sales (By region & Global), Review. The Data Types are Categorical (Genre, Platform), Numerical (Sales, Year, Review). The Data Sets are Video Games Sales dataset spanning Years 2000 – 2022. The Framework is Exploratory Data Analysis, followed by different analyses. The methods are clustering for the game segmentation, Classification for the genre prediction, Prediction for the sales forecasting.

Table of contents

Introduction

This business problem about the video game industry is very interesting. It's a little confusing because, at one point, it was on top of the entertainment industry. In recent years, though, they've been getting hit with challenges. While some titles still seem to be doing fine, there's this perception that this industry is slowly starting to lose its charm. As Ian Bogost said in "How to Work with Video Games" (2011), combine quantitative analysis with qualitative insights. That's why I think this dataset on the video game sales could be a unique opportunity to investigate this problem and give us a deeper understanding of things like sales patterns, genres, platforms, etc. If we're able to find out what happened, then maybe stakeholders can adapt and continue despite these challenges. The worst-case scenario for the industry is job losses and reduced technological advancements in the field, so it's best we get ahead of this now rather than later. I became interested in this field due to my love for video games. When I first got into them, it was such a booming market in the 2000s. However, as we go into these next few years, I'm interested in seeing if there will be a twist in how people react to video games, especially since there have been articles talking about how developers and publishers are struggling all over real life "There was just this unbelievable cascade effect," (Bailey, 2020). The objectives of this research is to find out different business aspects of videogames, like are there certain games more declining than others. Another one we have is Can we predict the success of the games based on its platform, genre, and other features? Lastly we have can the reviews help in influencing the games sales. At the core of this analysis are several characteristics and aspects that may perhaps affect video game sales. These include game title, which offers brand recognition and franchise power; release platform here reflecting technology choice and user base; year of production-a good indicator for market evolution over time, genre conventionally used to estimate consumer preferences points. Publisher as an agent measure the effectiveness in marketing a company's product (Baltezarević et al., 2018). The retail data is both regional ,international Moreover, review scores are taken into consideration which provide the quality of game and consumer responses. These variables are important in providing a generalized map of the video game industry landscape. Here a suitable data analysis framework is used to get insights from the dataset. At first the EDA that helps to understand basic trends followed by the visualisations, with clustering, classification and prediction of the games sales based on their features. This research is done with factors

that influence the sales of video games. In its attempt to make the prediction and insights valuable, this report categorizes games by sales trends as it makes predictions about market movements. Such outcomes are expected to be critical for game developers, publishers and marketers in the formulation of strategies that can either be defensive or offensive. In addition, this study aims to increase our knowledge about digital gaming changes by focusing on the changing consumer demand patterns and technological advancements. This study seeks to provide an in-depth analysis of the video game series or genres, platform and market trends. For this reason, it acts as an educational and guiding influence in the interpretation of a rapidly changing form of digital entertainment

Data Preprocessing

Descriptive Analysis

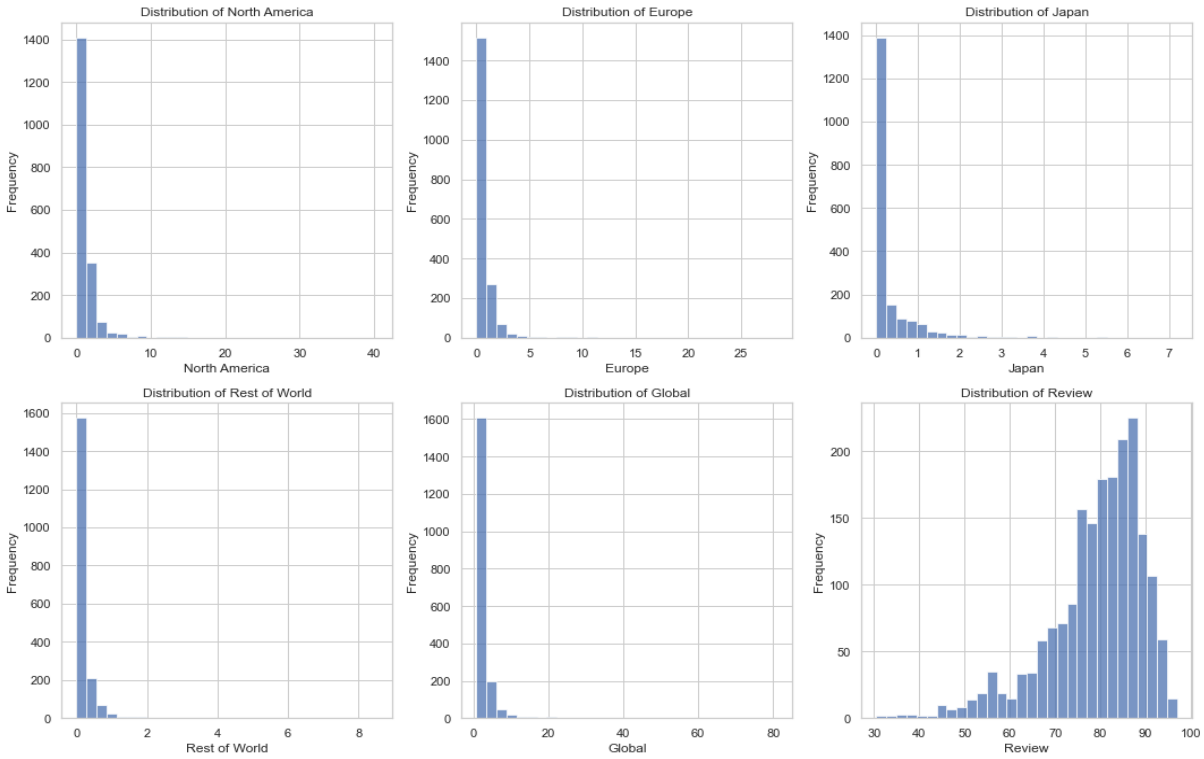
Numerical Features:

- Rank: Ranges from 1 to 1907.
- Year: Games are from 1983 to 2012, with an average release year around 2004.
- North America Sales: Average sales are 1.26 million units, with a maximum of 40.43 million.
- Europe Sales: Average sales are 0.71 million units, with a maximum of 28.39 million.
- Japan Sales: Average sales are 0.32 million units, with a maximum of 7.2 million.
- Rest of World Sales: Average sales are 0.21 million units, with a maximum of 8.54 million.
- Global Sales: Average global sales are 2.49 million units, with a maximum of 81.12 million.
- Review: Average review score is around 79 out of 100.

Categorical Features:

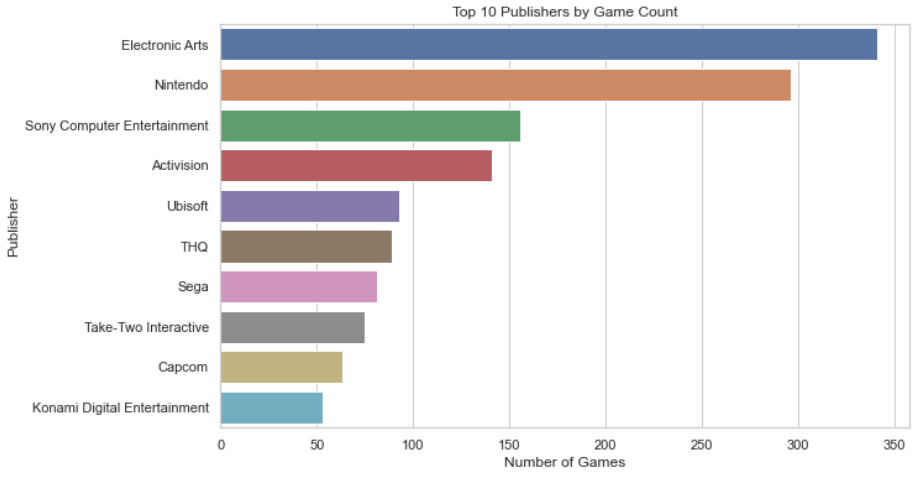
The game title has 1519 unique titles in the data set. The games are distributed in 22 platforms out of which Play station being the highest of all. There are all together 12 generes out of them sports is the most common ones.There are 94 unique publishers with EA sports which is the most frequent.

Visualization:

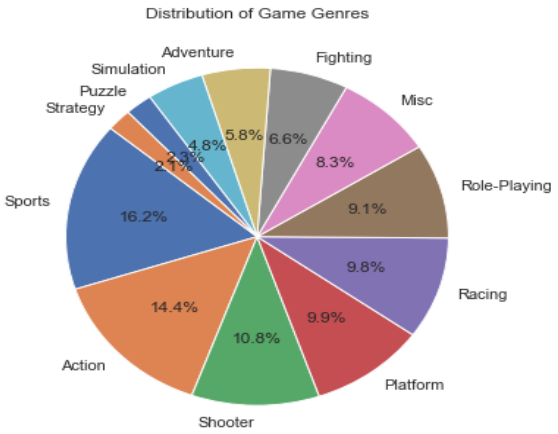


Sales Distributions: The histograms for sales in different regions (North America, Europe, Japan, Rest of the World, Global) show that most games have relatively low sales, with a few outliers having exceptionally high sales.

Review Scores: The distribution of review scores shows a left-skewed pattern, indicating most games have high review scores.



The bar chart displays the top 10 publishers by the number of games. 'Electronic Arts' leads, followed by 'Activision' and 'Nintendo'.

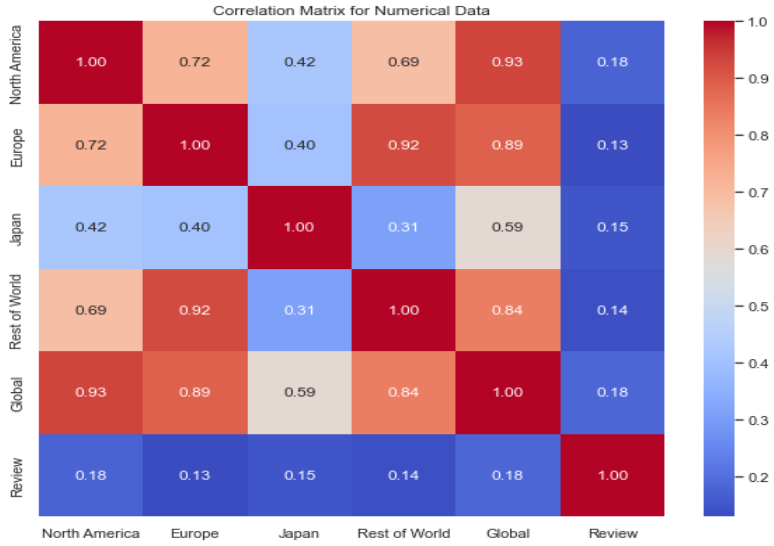


The pie chart shows the different proportion of different game genres. 'Sports' and 'Action' genres which has significant portions of the dataset.

Proximity Analysis:

We'll use a correlation matrix to identify relationships between numerical features. For categorical data, given the complexity and variety of methods available, we will focus on a high-level approach to understand the relationships, such as looking at the most common combinations of platform, genre, and publisher.

Correlation Matrix: To identify relationships between numerical sales data and review scores.



The correlation matrix provides insights into how different numerical features are related to each other: In **Sales Correlation t**here are strong positive correlations between sales in

different regions and global sales. This suggests that games which perform well in one region tend to perform well globally. In **Review Scores and Sales** there is a generally low correlation between review scores and sales, indicating that high scores do not necessarily translate to high sales.

Categorical Data Analysis: Identify common patterns in platform, genre, and publisher.

PlayStation 2 (PS2) Sports Games by Electronic Arts is the most common combination, with 57 games. PlayStation (PS) Sports Games by Electronic Arts with 21 games, which indicates the popularity of sports games by Electronic Arts on PlayStation platforms. Then Xbox 360 (X360) Sports and Shooter Games by Electronic Arts these both genres are popular on this platform with 18 and 15 games, respectively. Nintendo DS Role-Playing Games by Nintendo ties with Xbox 360 shooters with 15 games. PlayStation 2 (PS2) Racing Games by Electronic Arts and PlayStation 3 (PS3) Sports Games by Electronic Arts each with 14 games. Game Boy (GB) Platform Games by Nintendo and PC Simulation Games by Electronic Arts are each with 13 games. PlayStation 2 (PS2) Miscellaneous Games by Sony Computer Entertainment is also with 13 games.

Discussion of Findings:

- **Sales Trends**: The sales data shows a few highly successful games, with most games achieving moderate sales. This skew in sales distribution is typical in the gaming industry.
- **Genre Popularity**: Sports and action games are quite popular, with sports games especially favored by Electronic Arts.
- **Platform Preferences**: Certain publishers have clear preferences for platforms, like Electronic Arts with PlayStation and Nintendo with its own platforms.
- **Review Scores**: The low correlation between review scores and sales suggests that factors other than game quality (such as marketing, brand recognition, or nostalgia) might play a significant role in a game's commercial success.

Pre-Processing

To implement the data preprocessing stage, we'll follow these steps:

1. **Data Cleaning**: Address missing or incorrect data and outliers.
2. **Data Reduction**: Simplify the data without losing informative features.
3. **Data Transformation**: Modify data formats or create new variables as needed.
4. **Data Discretization**: Convert continuous data into categorical bins if beneficial.

- **Handle Missing Values**: For 'Year', we will impute missing values with the median year. For 'Publisher', we'll label missing entries as 'Unknown'.
- **Data Reduction**: Missing Values: Missing 'Year' values have been imputed with the median year, and missing 'Publisher' entries are labeled as 'Unknown'. The 'index' column has been removed.
- **Data Transformation**: Sales and review score columns have been standardized. One-Hot Encoding: Categorical data (Platform, Genre, Publisher) have been encoded, resulting in a wide format with additional columns for each category.
- **Data Discretization**: The 'Global' sales column has been discretized into three categories ('Low', 'Medium', 'High') based on quantiles. This categorization can be useful for analyses that require categorical interpretation of sales data.

Discussion of Pre-processing results:

The preprocessing steps have transformed the original dataset into a format more suitable for advanced analytical tasks. The normalization of numerical data ensures that variables with larger scales do not dominate the analysis. The one-hot encoding of categorical variables allows machine learning algorithms to process these data points effectively. Lastly, the discretization of sales data provides a categorical perspective that can be useful in certain types of analysis, such as classification tasks.

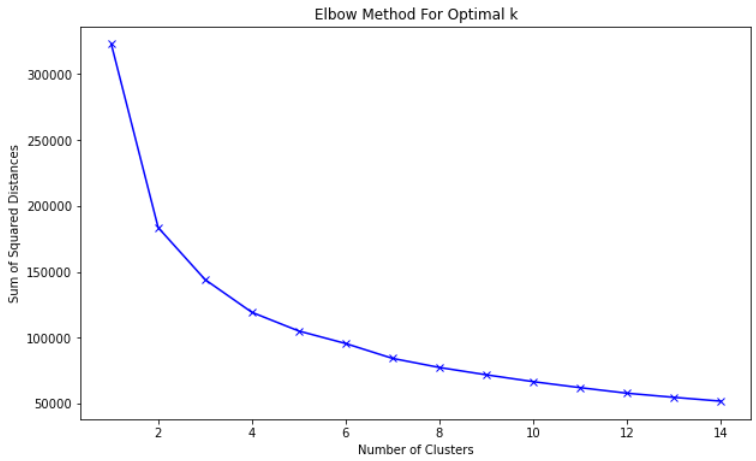
Data Processing

First the cleaned dataset after pre-processing was loaded, including the details likegame titles, platforms, release years, genres, publishers, regional sales (North America, Europe, Japan, Rest of World), global sales, and review scores. The aim was to create the cluster of videogames based on their sales performance across various regions, which will aim for

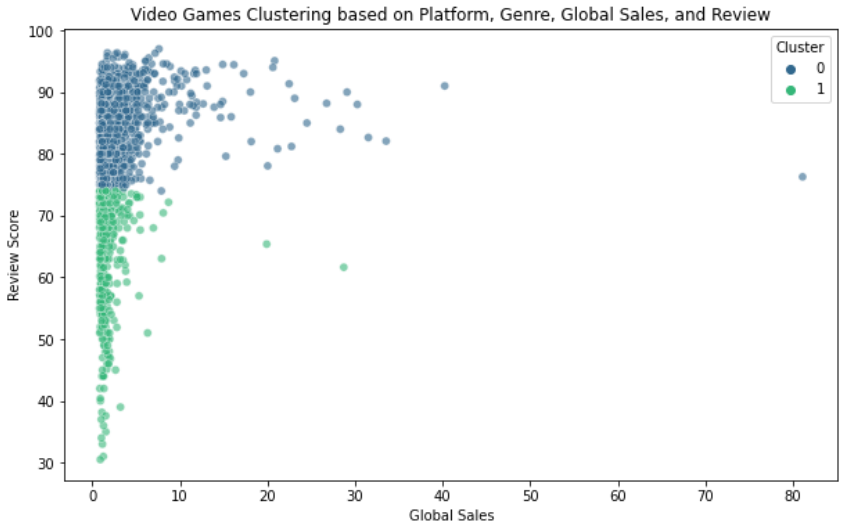
uncovering different patterns in regional preferences and the performance of different games on different platforms over the period of time.

Clustering

K-means



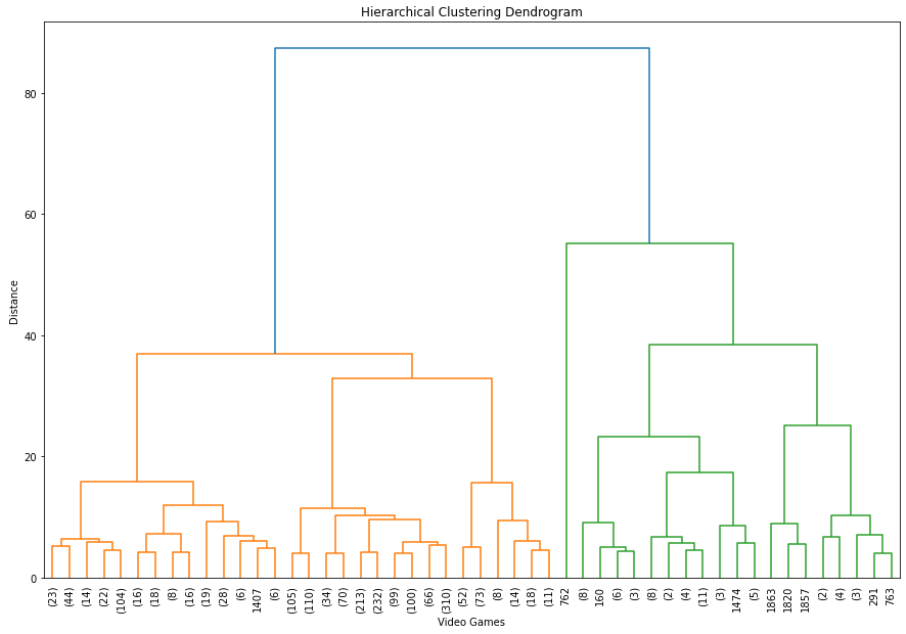
Above is a visualisation of elbow method for finding the optimal key. For this, the optimal k to, so only two clusters will be formed from this data set.



This is a visualisation of the video game data, which is clustered into two groups based on the sample of the data set. The main focus of the sample is on the review score, Global sales, platform, genre. Both of the clusters are shown in different colours. In this, the average review score of 79 is taken and then it classified into two categories which is 0 and 1. Cluster 0 here defines the high performing video games in the market with a good number of global sales and positive reviews, where as cluster 1 shows the low performing video games in the market, which has comparatively less number of global sales and they lack in good reviews.

Hierarchy method

This method is used so as to decide the number of clusters which shows the best inherit structure of the data. This is a diagram which is a visual representation of the hierarchial clustering process.



In the above figure the X axis is the videogames and the Y is the distance. This suggests that here are 2 clusters as Orange and Green. The node represents the merging of two clusters and the height shows the distance at which the two clusters are merged. The longer the branches the more dissimilarity and the shorter the branches the more the similarity.

Classification

K-Nearest Neighbors (KNN)

The KNN algorithm was used to predict the videogame's success with the use of the pre-processed data. Then a binary target variable was created which defined this success based on exceeding the mean review score. This also included essential steps like categorical variable encoding and feature scaling. This model was done with the initial setting of 5 neighbours, this parameter is very crucial as it influences the model's accuracy. The performance of this model was evaluated by using a classification report and confusion matrix, this provided different insights on the precision, recall and F1 scores. This approach also helped us to enhance the dataset's predictive capability.

Classification Report for KNN

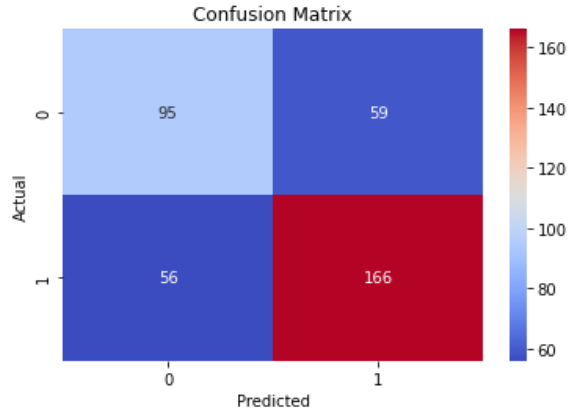
Precision: This score shows the accuracy of positive predictions. For the class '0'(Not Successful), it's 0.62, and for class '1' (successful), it's 0.68

Recall: This specially shows the ability of the model so as to find all the positive samples. For class '0' its 0.52, and for class '1' its 0.77

F1 Score: The score for class '0' is 0.57, and for class '1' is 0.72

Accuracy: Model's accuracy is 0.66 (66%)

Confusion Matrix (Heatmap)



This model gives a basic classification with a moreover moderate accuracy.

Decision Tree

The decision tree model was applied to the data set to find out that which games are successful or unsuccessful based on the review scores. In this a new column in 'game success' was created in on the basis of total average score of the reviews. Games below 79 score were labelled as unsuccessful (0) where is the games above 79 score were labelled as successful (1), The overall accuracy which was achieved by the model is 63%.

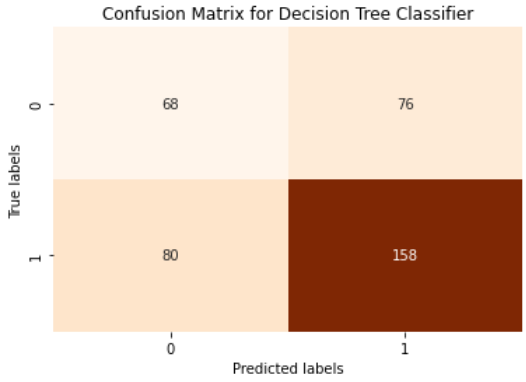
Precision: For unsuccessful games (0): 46%, For successful games (1): 68%

Recall: For unsuccessful games (0): 48%, For successful games (1): 66%

F1-Score: For unsuccessful games (0): 47%, For successful games (1): 67%

Overall Accuracy: 63%

Confusion Matrix (Heatmap)



- Darker Shades: Shows the higher numbers of observations.
- True Labels (Y-axis): Shows the actual categories of the games.
- Predicted Labels (X-axis): Shows the categories predicted by the model.

The numbers in the squares tells the count of observations. This visualization should help us in understanding that how well the model is performing and where it might be making the errors.

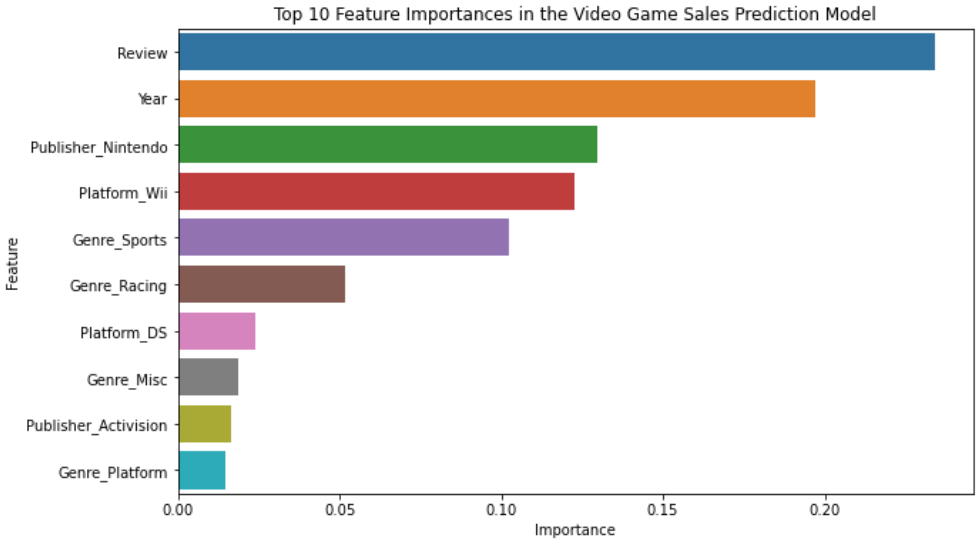
Predictive Modelling

For creating the predictive model, the data was taken and thenthe model selection was done. Next the linear regression model was explored to predict the sales of the games, the model was then evaluated with the help of cross validation, so as to estimate the performance of the model. Then the performance were measured by means Square error R squared metric.

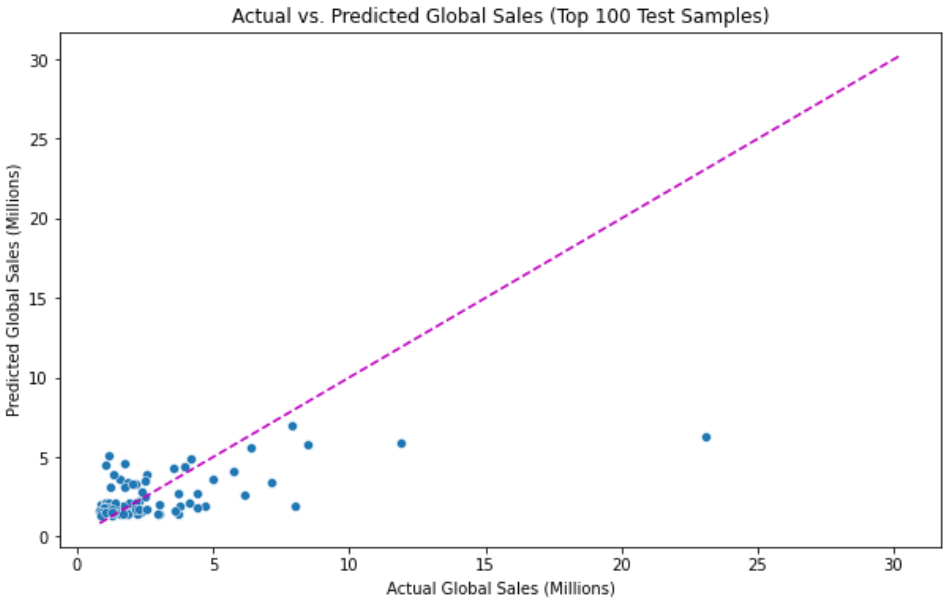
The results we got from the five fold Cross validation using the linear regression model was: Mean of MSE (12.22): The mean of MSE was 12.22 in the prediction. The linear regression model predicted a good average prediction, this indicates that the linear model is the best fit for this dataset than others. Standard deviation of MSE (6.84): The result was comparatively lower, which shows that the performance was consistent across different subsets of the data. Consistency tells us that this model's prediction is more stable.

The R^2 value that the model gave is 0.384 which suggests that the model explains about 38% of the variance globally in the video game sales.Then the prediction was continued with the

help of visualisation in which it was found out that features like review, year, publisher please a major rule in influencing the sales prediction model it is because they have majorly a huge impact on the sales.



This is the visualisation which shows the top 10 most important features for prediction of the video game sales. These features majorly contribute the most which impacts the models decision. This is done so as to analyse the importance in measuring how much each of the feature contributes so as to reduce the uncertainty in the model's prediction.



The scatterplot shows that the linear regression model tends to showcase the data clustering at the lower end which tells that only few games achieve a very high sales. The Pink line

represents the line of prediction, Some of the dots lie above the line which shows the games have a very high selling beyond the prediction. This visual is helpful for understanding how good the prediction model lines-up with the actual sales.

Conclusion

This research, aimed at discovering the unknown patterns of video game sales has given valuable information about gaming industry dynamics. Identification of patterns is important for analysis (Schudey et al., 2023). Pre-processing and processing of data has shown important patterns and trends that are integral to understand the market dynamics across regions as well as Smartphone platforms.The performance of sales trends in different genres and platforms has been clearly depicted, with a significant difference in consumers' taste across regions. Significant findings are of the form preferred game genres in specific markets and shifting popularity of platforms across time. These outcomes, however, show not only market status but also indications of future trends. Some of our research findings are aligned with known trends like those we see in studies done by well-known gaming market analysts, while others reveal new insights. This study builds on the current literature by providing a more insightful picture of regional variation and trending patterns in game popularity. These insights for the industry stakeholders are priceless as strategic decisions. The data-driven approach helps synchronize marketing strategies with customers' needs (Liang, 2018). Design efficient channels of delivery and set the right direction for development games that follow current trends These tendencies need to be understood for the purpose of retaining competitiveness as a rapidly changing industry.This research, however thorough it may be has its limits. Use of secondary data sources could have created biases; further, since the gaming industry changes often and rapidly trends can change fast. Furthermore, the predictive models applied while being strong cannot cover unanticipated market cycles and technological breakages.

Bibliography

Anderson, J. &. (2019). The Economics of the Video Game Industry

Baltezarević, R., Baltezarević, B., & Baltezarević, V. (2018). The video gaming industry: From play to revenue. *International Review, 3-4*, 71–76. https://doi.org/10.5937/intrev1804071b

Grand View Research. (2020, May). *Video Game Market Size, Share | Industry Report, 2020-2027*. Www.grandviewresearch.com. https://www.grandviewresearch.com/industry-analysis/video-game-market

Liang, Y. (2018). *Analysis of the video gaming industry*.

Mahoney, L. M., & Tang, T. (2020). The Rowman & Littlefield Handbook of Media Management and Business. In *Google Books*. Rowman & Littlefield. https://www.google.co.uk/books/edition/_/KeUGEAAAQBAJ?hl=en&gbpv=1&pg=PA285&dq=gaming+industry

Schudey, A., Kasperovich, P., Ikram, A., & Panhans, D. (2023, June 7). *Game Changer: Accelerating the Media Industry's Most Dynamic Sector*. BCG Global. https://www.bcg.com/publications/2023/drivers-of-global-gaming-industry-growth

Appendices

1. Clustering:

```python
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns

# Selecting relevant columns for clustering
data_for_clustering = video_games_sales[['Platform', 'Genre', 'Global', 'Review']]

# Converting categorical data into numeric using dummy variables
data_for_clustering = pd.get_dummies(data_for_clustering, columns=['Platform', 'Genre'])

# Performing clustering
kmeans = KMeans(n_clusters=2, random_state=0)
clusters = kmeans.fit_predict(data_for_clustering)

# Adding cluster information to the original dataframe
video_games_sales['Cluster'] = clusters

# Visualizing the clusters
plt.figure(figsize=(10, 6))

# We'll use Global Sales and Review for visualization purposes
sns.scatterplot(data=video_games_sales, x='Global', y='Review', hue='Cluster', palette='viridis', alpha=0.6)
plt.title('Video Games Clustering based on Platform, Genre, Global Sales, and Review')
plt.xlabel('Global Sales')
plt.ylabel('Review Score')
plt.legend(title='Cluster')
plt.show()

```

```python
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Calculating the sum of squared distances for different numbers of clusters
sum_of_squared_distances = []
K_range = range(1, 15)  # Considering 1 to 14 clusters

for k in K_range:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans = kmeans.fit(df)
    sum_of_squared_distances.append(kmeans.inertia_)

# Plotting the Elbow Graph
plt.figure(figsize=(10, 6))
plt.plot(K_range, sum_of_squared_distances, 'bx-')
plt.xlabel('Number of Clusters')
plt.ylabel('Sum of Squared Distances')
plt.title('Elbow Method For Optimal k')
plt.show()
```

```python
import pandas as pd
import matplotlib.pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.preprocessing import StandardScaler

# Load the data (replace 'your_file_path.csv' with the path to your dataset)
video_games_sales = pd.read_excel('data.xlsx')

# Data Preparation
# Selecting only the sales data columns for clustering
clustering_features = video_games_sales[['North America', 'Europe', 'Japan', 'Rest of World', 'Global']]

# Standardizing the data
scaler = StandardScaler()
scaled_features = scaler.fit_transform(clustering_features)

# Hierarchical Clustering using Ward's method
linkage_matrix = linkage(scaled_features, method='ward')

# Dendrogram
plt.figure(figsize=(15, 10))
# Limiting the number of labels for readability
dendrogram(linkage_matrix, truncate_mode='lastp', p=50, leaf_rotation=90, leaf_font_size=10)
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Video Games')
plt.ylabel('Distance')
plt.show()
lines = [plt.Line2D([0], [0], marker='o', color='w', markerfacecolor=color, markersize=10, label=label)
         for color, label in zip(colors, labels)]

# Display legend
plt.legend(handles=lines, loc='upper right')
```

2. Classification:

```python
import matplotlib.pyplot as plt
import seaborn as sns

# Plotting the distribution of review scores
plt.figure(figsize=(10, 6))
sns.histplot(video_games_data['Review'], kde=True, bins=30)
plt.title('Distribution of Review Scores')
plt.xlabel('Review Score')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
video_games_data['Review'].describe()
```

```python
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 1. Create the KNN Model
knn = KNeighborsClassifier(n_neighbors=5)

# 2. Train the Model
knn.fit(X_train, y_train)

# 3. Make Predictions
y_pred_knn = knn.predict(X_test)

# 4. Evaluate the Model
report_knn = classification_report(y_test, y_pred_knn)
conf_matrix_knn = confusion_matrix(y_test, y_pred_knn)

# Print the evaluation results
print("KNN Classification Report:\n", report_knn)
print("KNN Confusion Matrix:\n", conf_matrix_knn)

# Evaluation
accuracy_knn = accuracy_score(y_test, y_pred_knn)
print("KNN Accuracy:", accuracy_knn)

sns.heatmap(conf_matrix, annot = True, fmt = 'd', cmap = 'coolwarm')
plt.ylabel('Actual')
plt.xlabel('Predicted')
plt.title('Confusion Matrix')
plt.show()
```

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix
import numpy as np

# Create the 'Game Success' column
data['Game Success'] = np.where(data['Review'] >= 79, 1, 0)

# Dropping the non-numeric and target columns for simplicity
X = data.drop(['index', 'Rank', 'Game Title', 'Platform', 'Year', 'Genre', 'Publisher', 'Review', 'Game Success'], axis=1)
y = data['Game Success']

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Creating and training the Decision Tree Classifier
dt_classifier = DecisionTreeClassifier()
dt_classifier.fit(X_train, y_train)

# Predicting the Test set results
y_pred = dt_classifier.predict(X_test)

# Evaluating the model
classification_report_result = classification_report(y_test, y_pred)
confusion_matrix_result = confusion_matrix(y_test, y_pred)

classification_report_result, confusion_matrix_result

```

```
import seaborn as sns
import matplotlib.pyplot as plt

# Creating a heatmap for the confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)

sns.heatmap(conf_matrix, annot=True, fmt='g', cmap='Oranges', cbar=False)

# Adding labels and titles
plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.title('Confusion Matrix for Decision Tree Classifier')
colors = ["#4CAF50", "#FFC107", "#F44336"]
plt.show()

```

3. Prediction:

```
from sklearn.linear_model import LinearRegression


linear_regression_pipeline = Pipeline(steps=[('preprocessor', preprocessor),
                                             ('model', LinearRegression())])

# Using 5-fold cross-validation for Linear Regression
cv_scores_lr = cross_val_score(linear_regression_pipeline, X_multi, y_multi, cv=5, scoring='neg_mean_squared_error')

cv_mse_scores_lr = -cv_scores_lr
cv_mse_mean_lr = cv_mse_scores_lr.mean()
cv_mse_std_lr = cv_mse_scores_lr.std()

cv_mse_mean_lr, cv_mse_std_lr

```

```python
import matplotlib.pyplot as plt
import seaborn as sns

linear_regression_pipeline.fit(X_train_multi, y_train_multi)

# Make predictions
y_pred_lr = linear_regression_pipeline.predict(X_test_multi)

plt.figure(figsize=(10, 6))
sns.scatterplot(x=y_test_multi[:100], y=y_pred_lr[:100])
plt.plot([y_test_multi.min(), y_test_multi.max()], [y_test_multi.min(), y_test_multi.max()], 'm--')
plt.title('Actual vs Predicted Global Sales (Top 100 Test Samples) - Linear Regression')
plt.xlabel('Actual Global Sales (Millions)')
plt.ylabel('Predicted Global Sales (Millions)')
plt.show()

```

4. Data Pre-Processing:

```python
import matplotlib.pyplot as plt
import seaborn as sns

# Setting up the visualization style
sns.set(style="whitegrid")

# Creating histograms for numerical data
numerical_columns = ['North America', 'Europe', 'Japan', 'Rest of World', 'Global', 'Review']
plt.figure(figsize=(15, 10))

for i, column in enumerate(numerical_columns, 1):
    plt.subplot(2, 3, i)
    sns.histplot(video_games_data[column], kde=False, bins=30)
    plt.title(f'Distribution of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')

plt.tight_layout()
plt.show()

# Creating a bar chart for the top 10 Publishers
top_publishers = video_games_data['Publisher'].value_counts().head(10)
plt.figure(figsize=(10, 6))
sns.barplot(x=top_publishers.values, y=top_publishers.index)
plt.title('Top 10 Publishers by Game Count')
plt.xlabel('Number of Games')
plt.ylabel('Publisher')
plt.show()

# Creating a pie chart for Game Genres
genre_counts = video_games_data['Genre'].value_counts()
plt.figure(figsize=(10, 6))
plt.pie(genre_counts, labels=genre_counts.index, autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Game Genres')
plt.show()
```

```python
# Creating a correlation matrix for numerical data
numerical_data_for_correlation = video_games_data[numerical_columns]
correlation_matrix = numerical_data_for_correlation.corr()

# Plotting the correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix for Numerical Data')
plt.show()

```