**Abdullahi Mohamed (M00810926)**

**"Analysing Regional Sales Patterns and Genre Influence on Xbox One Games"**

Table of contents

Section 1: Research Goal ......................................................................Page 3

Section 2: Data Processing....................................................................Page 4-6

Section 3: Data Mining ...........................................................................Page 7-11

Section 4 Data Visualizations .................................................................Page 12-13

Section 5: Workflow Impact ....................................................................Page 14

Section 6: Interesting Findings................................................................Page 15-16

Section 7: Conclusions ..........................................................................Page 18

Section 8: Future Work ..........................................................................Page 19

**Section 1: Research Goal**

**1.1 Does regional sales in North America, Europe, Japan, and the Rest of the World correlate with global sales for Xbox One games?**

The analysis discovers whether regional markets play a role in global sale and identify regions with the most impact on a game’s success. Then an examination of sales patterns within these regions, the project aims to identify key markets that publishers should prioritise when launching new XBOX games.

**1.2 Does the genre of the game affect its sales performance in different regions?**

The project explores if genres such as action, shooter or sports perform better in specific regions. Understanding regional-genre relationships will help publishers tailor their marketing strategies to target the most appropriate regions for each gaming genre. It also explores whether regional preferences are more influential than the genre itself when determining a game’s success.

**1.3 How can these regional-genre sales patterns be used to identify distinct market segments?**

By utilising data mining techniques such as linear regression, decision trees and k-means clustering, the analysis aims to uncover market segments based on regional sales and genre preferences. Identifying these segments enables publishers to create a more focused and effective marketing strategies that cater to specific needs and preferences of different markets.

**1.4 What role does the year of release play in the sales performance of Xbox One games?**

The project will investigate whether the year release affects sales performance. The analysis will include w of marketing or do older games still perform well.

**Section 2: Data Processing**

A diagram of a machine

AI-generated content may be incorrect.

Figure - (Data Processing) - Workflow

This is the workflow needed to process data of the XBOX One sales dataset. This step-by-step procedure that helps get rid of missing values, delete unnecessary rows, create new column using an IF rule, filter columns and convert string columns to numbers.

**2.1 Missing Value**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Processing) - Missing Value

After the CSV reader file is imported and read, the next step is to find missing values, to which all columns were fine barring publisher as it contained missing rows, I decided to remove rows that were empty.

**2.2 Row Filter**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Processing) - Row Filter

To answer my research question, I filtered out the global sales to just values above 0 meaning they must have value.

**2.3 Rule Engine**

A screenshot of a computer

AI-generated content may be incorrect.

Figure (Data Processing) - Rule Engine

If global column has over 5 sales, it means it is a high-profile success, but if it greater than or equal to 0.1 it is limited success, the aim is to categorise sales into group and give it a name that suits it.

**2.4 Column Filter**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Processing) - Column Filter

This step filters out columns, the game doesn't have much use, the released year or the position also don’t, therefore they have been excluded from this stage.

**2.5 Category to Number**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Processing) - Category to Number

Converting the genre and publishing columns is necessary for the project.

**Section 3 Data Mining**

**3.1 DECISION TREE - Workflow**

A diagram of a computer

AI-generated content may be incorrect.

Figure (Data Mining) - Decision Tree Workflow

This is the workflow of decision tree, carrying on from the data being processed, just 2 steps needed to complete this.

**3.2 Partitioning – Decision Tree**

A screenshot of a computer

AI-generated content may be incorrect.

Figure (Data Mining) - Partitioning Decision Tree

The partitioning tools splits dataset into two parts, to which 80% of the data into the first partition and using it to train model. Whilst 20% is for testing, the column I decided to use for sampling is Genre meaning that it will maintain the same proportions of each category in both partitions. This ensures the training and testing have similar distribution, the random seed also makes the split reproducible.

**3.3 Decision Tree Learner**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Mining) - Decision Tree Learner

This step is the configuration of the decision tree learner to predict the genre, the quality measure being used is “Gain ratio”, with a reduced error pruning enabled, a minim of 5 record per node. This interface allows to store 10,000 records for visualization, using 8 processing threads and skipping nominal columns.

**3.4 k-Means Workflow**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Mining) - k-Means Workflow

This is the workflow of K-Means, the purpose is to find and segregate groups with similar traits and assign them into clusters.

**3.5 Column Filter – k-Means**

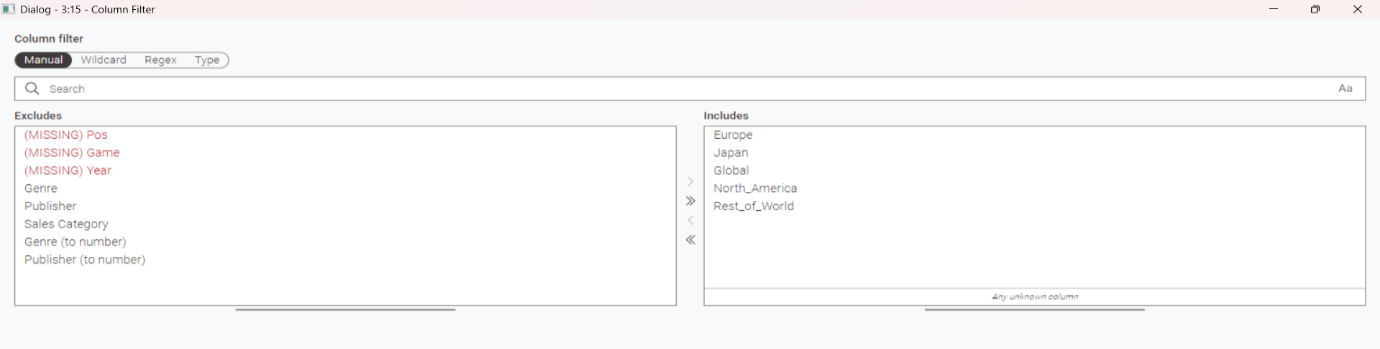


Figure (Data Mining) - Column Filter k-Means

The exclusion of columns that don’t serve a purpose in the findings, with the inclusion of numeric columns that aligns with research questions.

**3.6 Normalizer – k-Means**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Mining) - Normalizer k-Means

Normalizing the filtered column with a minimum of 0 and a maximum of 1.

**3.7 k-Means – Clustering**

A screenshot of a computer

AI-generated content may be incorrect.

Figure (Data Mining) - Clustering k-Means

4 clusters will be tested and 100 iterations with the filtered column.

**3.8 Manual Aggregation - k-Means**

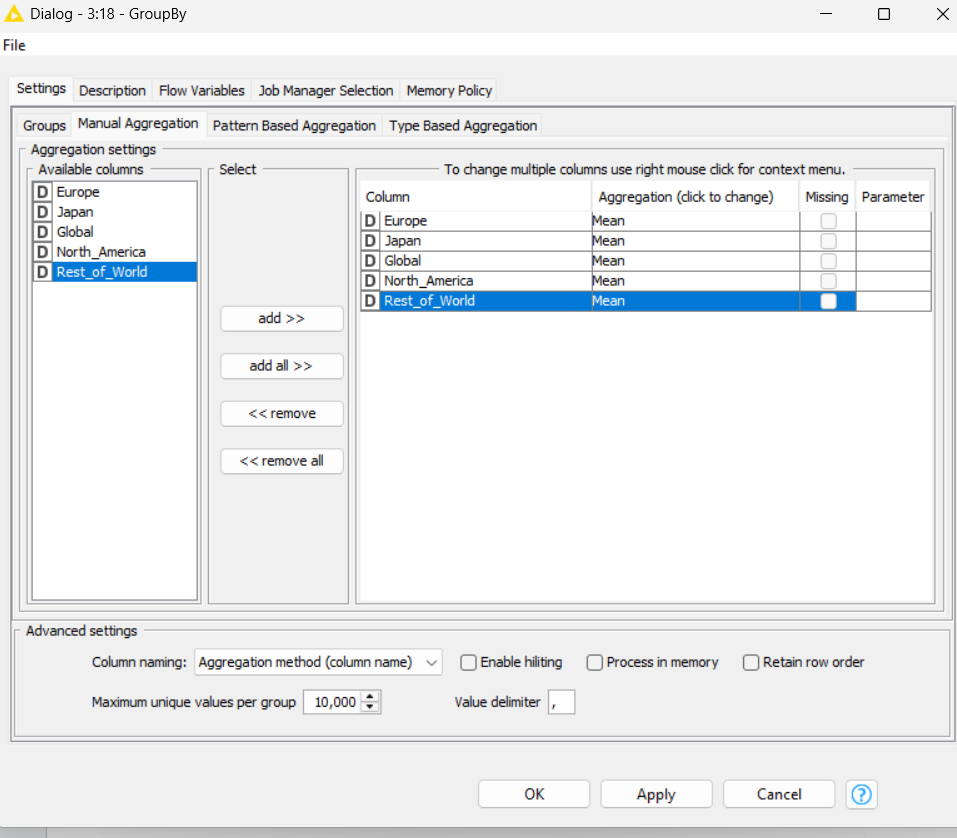


Figure (Data Mining) - Manual Aggregation k-Means

Aggregate numeric columns to mean, this helps with the k-Means results.

**3.9 Group by – k-Means**

A screenshot of a computer

AI-generated content may be incorrect.

Figure (Data Mining) - Group By k-Means

Group by all five columns into one.

**3.10 Linear Regression - Workflow**

A diagram of a number

AI-generated content may be incorrect.

Figure - (Data Mining) - Linear Regression Workflow

This is the flow of how to use linear regression to study the linear relationships

**3.11 Linear Regression Learner**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Mining) - Linear Regression Learner

Exclude publisher as it does not have value in this mining of data.

**Section 4: Data Visualisations**

**4.1 k-Means Scatter Plot**

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure - (Data Visualisations) - Scatter Plot k-Means

This is the data visualisations of k-Means, with the horizontal dimension being North America and the vertical dimension being Europe.

**4.2 Linear Regression**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Visualisations) - Linear Regression

This is the visualised table of results with Europe and America having a strong impact on global sales.

**4.3 Decision Tree**

A screenshot of a computer

AI-generated content may be incorrect.

Figure - (Data Visualisations) - Decision Tree

This is the visualised decision tree results identifying relationships between game genres.

**Section 5: Workflow Impact**

**5.1 How does your workflow help achieve the goal?**

My workflow addresses the research question through key components:

**5.2 Data Processing Pipeline**:

* CSV Reader imports Xbox games sales data to start foundation for analysis.
* Missing Values to ensure data integrity, addressing inconsistencies and removing rows.
* Row Filter removes irrelevant columns and has no relevance to reseach question.
* Rule Engine to apply logic and understanding of data representations.
* Column Filter selects only relevant attributes to focus on the analysis
* Category to Number transforms string categories and converts into numeric values
* Normalizer scales regional sales data, preventing a set region from dominating the analysis due to market size differences.

**5.3 Data Mining Components**:

* Decision Tree Learner identifies relationships between game genres and their market performance, revealing which genres use similar distribution patterns.
* K-Means clustering identifies grouping of games based on their regional sales distributions, discovering market segments that transcend to genre classifications'.
* Linear regression to predict the relationship between regional and global sales.
* Group By node aggregates results, allowing for the examination of genre performance by region at a higher level.

**5.4 Data Visualization**

* Scatter plot to represents the relationship between North American and European sales with a color-coding (via Colour Manager) to identify cluster membership.
* Table view to statistically to understand the numeric results of regional and global sales.
* This visualization makes regional sales pattern apparent showing how games perform differently across markets.

This workflow transforms raw sales data into actionable insights identifying both genre-based patterns and market segments based on regional performance.

**Section 6: Interesting Findings**

**6.1 Interesting findings achieved by applying the workflow on the data**

The workflow reveals several interesting findings on the data:

**6.2 Regional Sales and Global Sales Correlation (Linear Regression)**

* A strong correlation between global and regional sales.
* **North America**: Coeff: 1.092, p-value = 0 (strongest impact).
* **Europe**: Coeff: 1.044, p-value = 0 (second strongest impact).
* **Japan**: Coeff: 1.022, p-value = 0 (third strongest impact).
* **Rest of the World**: Coeff: 0.239, p-value = 0.002 (poor, but still relevant).

**6.3 Implications for Publishers (Linear Regression)**

* A priority should be given to North America and Europe when creating new XBOX games, because these regions contribute the most to global sales to highlight their current usefulness to XBOX.

**6.4 Impact of Game Genre (Linear Regression)**

* The set genre does not affect sales (Coeff: 0, p-value = 0.361).
* The publishers should focus on regional preferences rather than specific genres for marketing efforts because it will be more useful to XBOX.

**6.5 Impact of Year of Release (Linear Regression)**

* The release years does not have an impact on sales performance (Coeff: 0, p-value = 0.21).
* Mature games can still compete well depending on region.

**6.6 Important Points (Linear Regression)**

* Sales patterns are decided by the regions and have the most impact on global sales.
* The release year and genre of games does not have an impact on sales.
* Publishers should focus on segmentation of markets based on regional trends to maximise sales.

**6.7 Regional-Genre Correlations (k-Means)**

* k-Means clustering to identify market segments, (clusters 0-3) is based on regional sales patterns between North America and Europe.
* Cluster 0 (green) represents games with a good performance in both regions but particularly high European sales.
* Cluster 3 (grey points) is games with minimal sales in both regions, containing most points in scatter graph.
* Cluster 1 (red) and Cluster 2 (brown) portray a moderate success with different regional balances.

**6.8 Genre-Based Market Segmentation (Decision Tree)**

* Decision tree reveals a split in the market using genre with a value of (5.5)
* Left branch contains 63.8% of games) with a focus on action-based games.
* Right branch (36.2% games) contains sports and simulation games.

**6.9 Market Insights**

* Action and shooter games perform better in North America then Europe.
* Sporting and simulation title a more balanced within regions.
* Clustering analysis highlights the regional performance is not determined by genre alone, as games with the same genre can fall into different clusters.
* Highest performance outliers (1.0 on either axis points) tend to be specific genres, suggest that success follows genre-specific patterns.

These insights suggest that publishers should adopt different marketing strategies based on the genre of the game. Action-oriented titles need greater marketing investment in North America, while sports and simulation games benefit from balanced regional marketing approaches. Cross region high performers provide a valuable case for optimizing publishing strategies.

The combination of decision tree, clustering provides different perspectives, one identifying market segments based on sales performance and the other connecting patterns to specific game genres and helps design a framework for strategic decision-making in game publishing.

**Section 7 Conclusion**

The analysis of Xbox One games sales data has unveiled several great insights that can inform publishers about strategic decision-making:

**Regional Impact**: Europe and North America dominate as the most influential markets for Xbox One games’ global success, with correlation coefficients of 1.092 and 1.044. These markets should be the main preference in marketing and distribution strategies.

**Genre**: The genre does not significantly impact the sales performance (p-value = 0.361). This highlights regional preferences and market-specific approaches are more important than genre-based segmentation.

**Gaming Market**: The release year has no impact on sales performance (p-value = 0.21), indicating that well-developed games can maintain commercial viability regardless of their tenure.

**Distinct Market Segments:** k-Means clustering revealed four distinct market segments based on regional sales patterns, with interesting findings in Cluster 0 (strong performance in Europe).

**Strategic Marketing**: The decision tree analysis demonstrates that publishers should adopt different marketing approaches based on game genres, with action titles warranting greater investment in North America whilst sports and simulation games benefit from a balanced regional market.

The findings highlight that Xbox One games’ success is determined by regional market dynamics rather than genre or release timing. Publishers should focus on understanding and leveraging regional patterns to optimise their global sales strategies and understand and leverage those regional patterns to optimise their global sales strategies.

**Section 8 Future Work**

The analysis contains valuable insights, with several avenues that can lead to future research to enhance Xbox gaming sales.

Pricing Strategy: Examine the relationship between price points and sales performance across regions to provide insights and enable an efficient pricing strategy.

Competitiveness: Expand analysis to include competitor platforms such as PlayStation and provide context about platform-specific preferences and identify exclusive opportunities for Xbox One publishers.

Digital vs Physical Sales: As the industry transitions towards digital distribution analysing the differential impact of digital versus physical sales across regions can reveal important trends for distribution.

Demographics: Including data of age, genders and income level of each buyer enables market segment to target specific users.

Content and Feature Analysis: Examining specific gaming features (multiplayer, limited edition) can identify and drive sales in different regions, providing guidance for game development priorities.

Machine Learning Applications: Using predictive algorithms can improve accuracy and identify patterns that traditional statistical methods can miss.

By pursing these additional research directions, publishers can develop a refined strategies for maximising Xbox One game sales.