LSE_D301_Assignment
Dimas Rizqy Firdiansyah

Turtle Game Report

Introduction

This report concerns Turtle Games, a globally recognised game manufacturer and retailer. As a unique business, Turtle Games not only produces its line of products but also sources and retails goods from various manufacturers. The main objective is to explore avenues for enhancing overall sales performance by leveraging customer trends and behaviours. The analysis will cover loyalty points, customer segmentation, social data for marketing, product impact on sales, data reliability, and sales relationships. Utilising sales data, customer reviews, and related datasets, the report aims to provide actionable insights to refine strategy and bolster sales for Turtle Games.

Analytical Approach

The analytical approach entailed the use of two datasets: 'turtle_reviews.csv' and 'turtle_sales.csv'. The former offers vital information about customer reviews and demographic details to inform customer segmentation and assess the loyalty points system. The latter provides insights into sales performances, thus facilitating a detailed examination of product impact on sales, data reliability, and global sales patterns.

The data analysis was conducted using Python and R for data preparation and analysing, including data wrangling, cleaning, and exploratory data analysis (EDA). Python libraries such as pandas, numpy, seaborn, matplotlib, and statsmodel, and R libraries like tidyverse, dplyr, ggplot2, and psych were instrumental in this process.

Analytical Findings & Visualisation

The analytical findings are categorized into different sections based on the analysis' scope: statistical analysis, sentiment analysis, exploratory data analysis, data reliability, and regression analysis.

This report employs key statistical tools to uncover intricate relationships within the data. Regression analysis gauges the influence of the loyalty on spending score and income fig.1 Based on these findings, it's evident that spending score and remuneration strongly influence loyalty points accumulation, whereas age plays a lesser role. Further research is needed to explore how additional factors, like purchasing habits or types of transactions, affect loyalty point accrual.

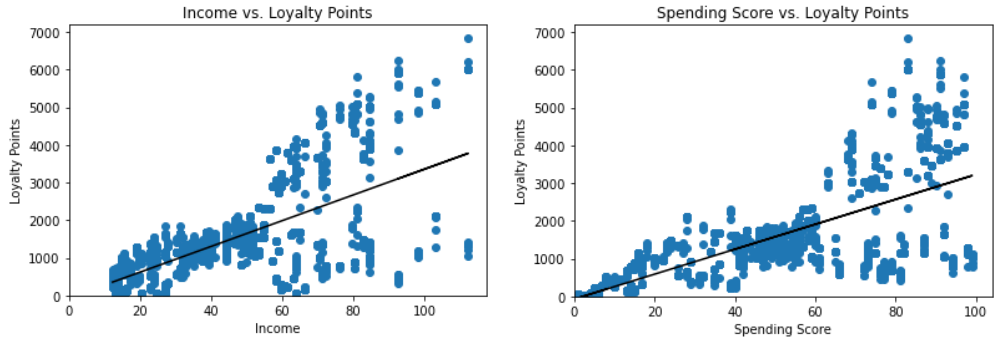


Figure 1 (a). Income relationsship with loyalty points, (b) spending relationship with loyalty points

While in clustering methods segregate the customer base, informing targeted marketing strategies fig.2. Our analysis utilized the k-means clustering algorithm to divide our customer base based on their remuneration levels and spending scores, the k-means value resulting in 5 distinct clusters which most effective cluster number determined from elbow and silhouette method, each representing different customer behaviour patterns:

1. High income, High spending

2. High income, Low spending

3. Low income, High spending

4. Low income, Low spending

5. Moderate income, Moderate spending

This segmentation enables us to tailor marketing strategies effectively, catering to the distinct needs and preferences of each cluster. More in-depth analysis can be conducted to explore factors like age, loyalty points, and others, providing a deeper understanding of each segment's characteristics.

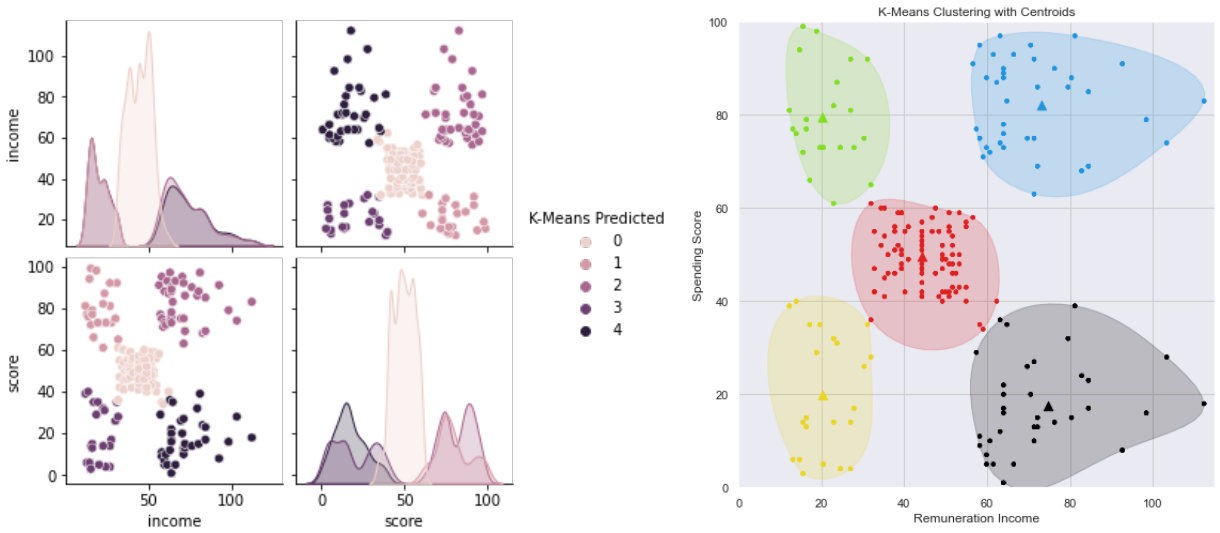


Figure 2. Clustered result

Natural Language Processing (NLP) and frequency distribution helped interpret customer sentiments and identify frequently used words in the reviews and summary in form of wordcloud element in fig.3.

Figure 3. wordcloud of frequent words in Reviews and Summary

An exhaustive analysis of customer reviews, performed using NLP, categorized sentiments as positive, neutral, or negative as shown on fig.4. This process revealed prevalent themes in the dataset's reviews and summaries, highlighting aspects like product quality, unmet expectations, product design, value for money, and functionality. Moreover, it indicated that the reviews tended to maintain a neutral position, with promising potential to increase positive sentiment in the future.

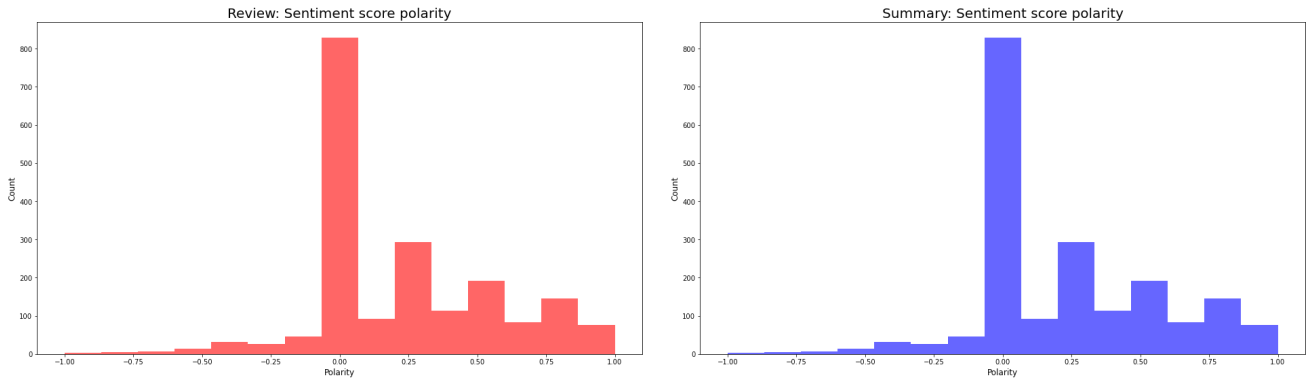
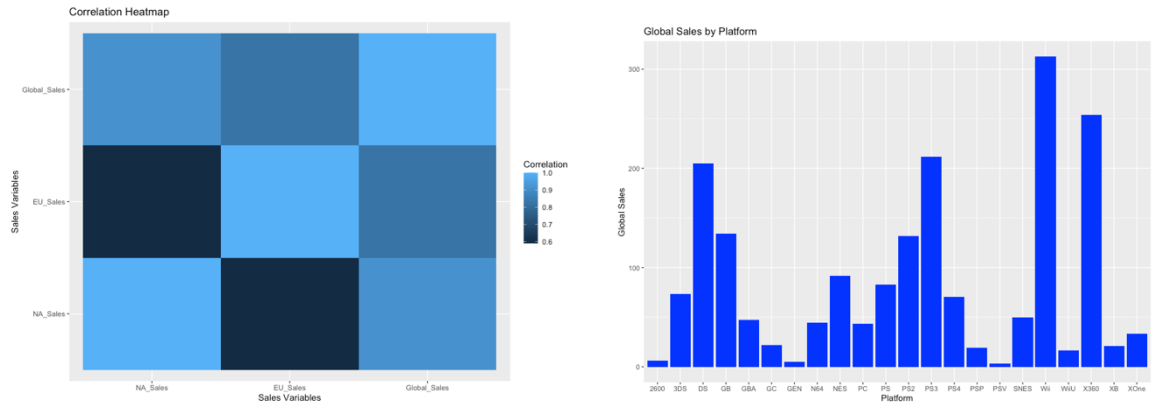


Figure 4. Sentiment score on Review and Summary

EDA was performed on the sales data to unearth key patterns and relationships. This step included studying correlations between regional and global sales shown using correlation heatmap fig.5(a) that resulted North America have a high correlation with the global sales By comparing impact on global sales. an insight such as market demographic or most used platform on each region can be determine by identifying best-selling game platforms as shown in fig 5(b). The insights gained from EDA can guide data-driven decision-making and future sales strategies.



(a) (b)

Figure 5. (a) Correlation Heatmap, (b) Global sales by latform

Assessing data reliability utilising various statistical tests, such as the Shapiro-Wilk test, skewness, kurtosis, and correlation matrices. Fig.6 presents visual representations. Our analysis involved scatter plots and histograms to explore relationships between sales regions and the distribution of sales across global, North America, and Europe markets. Positive skewness values indicate right-skewed distributions with longer tails on the right side due to high outliers or extreme values. Higher kurtosis

values suggest distributions with heavier tails and more peaking than a normal distribution, indicating more extreme values or outliers. In summary, all three datasets (Global Sales, NA Sales, and EU Sales) exhibit positive skewness, indicating the presence of high outliers on the right side. Additionally, the higher kurtosis values imply a higher number of extreme values or outliers in the data.

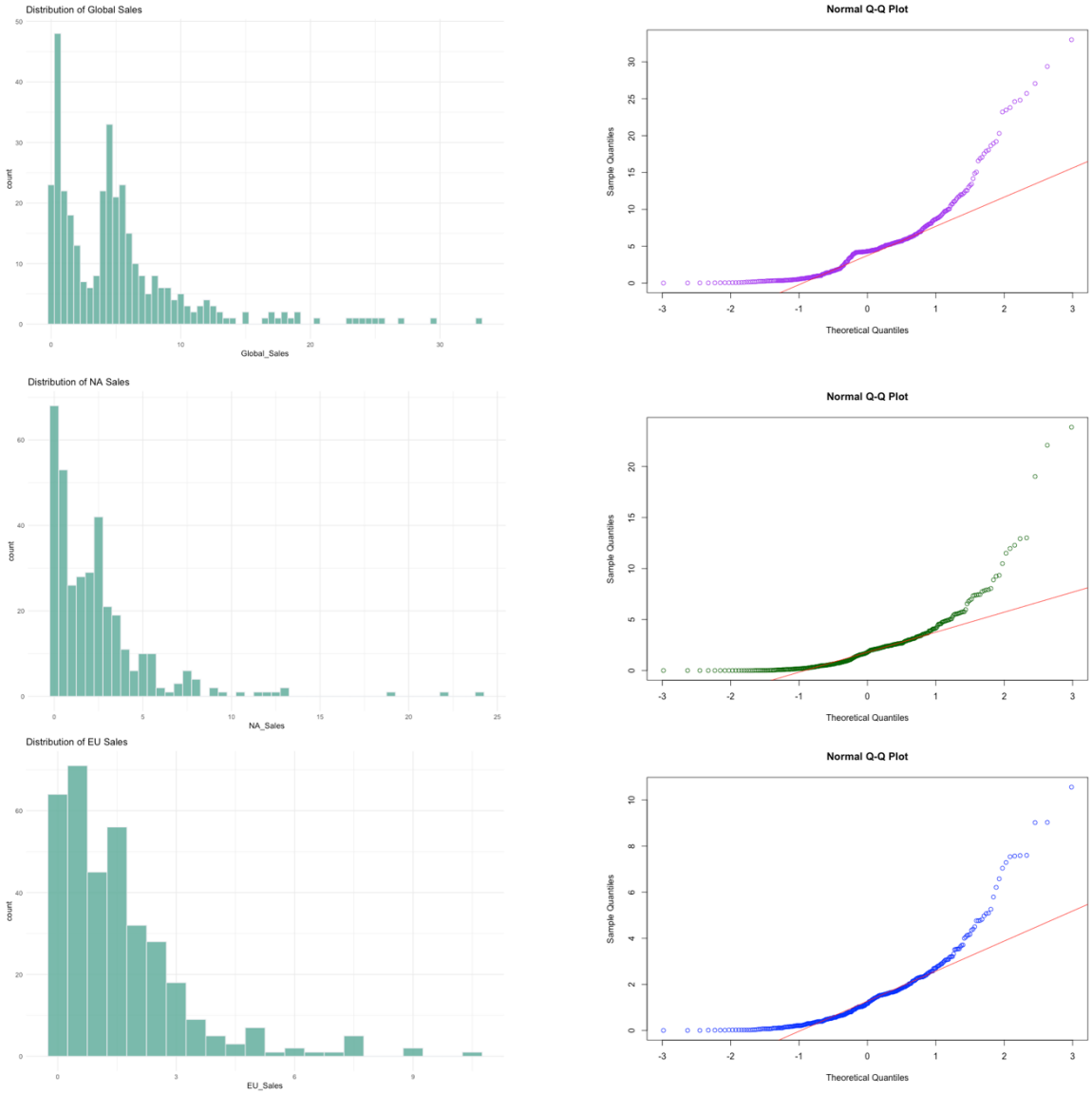


Figure 6. Data reliability

This report includes a multiple linear regression model that predicts global sales based on sales in Europe (EU_Sales) and North America (NA_Sales). The model is highly accurate, accounting for 97% of the variation in global sales. The outcomes derived from the multiple linear regression model, designed to forecast global sales, NA sales, and EU sales.

For NA Sales, the model reveals a positive correlation, indicating that an increase of approximately 34.02 units in the logIndex variable is associated with a corresponding rise of approximately 23.80 units in NA Sales. Moreover, the Index variable exhibits a significant influence, with each unit increment resulting in a sizeable increase in NA Sales.

In the case of EU Sales, the model demonstrates a stronger positive correlation, suggesting that a unit increase in the logIndex variable corresponds to a remarkable surge of approximately 73.48 units in EU Sales. However, it is vital to exercise caution when interpreting the effect of the Index variable on EU Sales, as the magnitude of approximately 8.13e+31 units raises concerns about potential outliers or atypical data points.

The logIndex variable emerges as a pivotal factor in both NA Sales and EU Sales predictions, with higher logIndex values strongly contributing to substantial sales growth. Nonetheless, the exceptional value associated with the Index variable for EU Sales requires further scrutiny to ascertain its validity and assess potential impact on the model.

Conclusion

In conclusion, this comprehensive analysis of Turtle Games has provided valuable insights to enhance sales performance and customer satisfaction. Leveraging customer trends and behaviors, we explored various factors, including loyalty points, customer segmentation, social data for marketing, and product impact on sales. Our analytical approach, utilizing Python and R, delivered actionable insights through statistical analysis, sentiment analysis, and exploratory data analysis. Clustering methods enabled targeted marketing strategies, catering to distinct customer behaviour patterns. In studying sales data, we identified correlations between regional and global sales, guiding data-driven decision-making and sales strategies. Additionally, the multiple linear regression model accurately predicted global sales based on Europe and North America sales. Key recommendations include focusing on customer satisfaction through high-quality products and prioritizing sales growth in North America and Europe. Continuous refinement of predictive models will be essential to adapt to market dynamics. This report's emphasis on data reliability and accurate predictive modeling underscores the significance of informed decision-making. By embracing data-driven strategies, Turtle Games can thrive and remain competitive in the ever-evolving gaming industry.