Turtle Games

By: **Mirasha Perera**

Table of Contents

1.0 Introduction

Turtle Games, a global manufacturer and retailer of games, aims to enhance its sales performance through data-driven strategies. This report outlines our analysis of customer review data, focusing on the following objectives:

- Understanding customer engagement and loyalty point accumulation.
- Segmenting customers into targetable groups.
- Using text data (customer reviews) to inform marketing campaigns and business improvements.
- Evaluating the suitability of loyalty points data for predictive modelling using descriptive statistics.

2.0 Analytical Approach

2.1 Data Import and Cleaning

I started by immersing myself in Turtle Games' business objectives and delving into the provided dataset. Using R, I performed initial data exploration, cleaning, and analysis using essential tools like tidyverse, skimr, and DataExplorer package

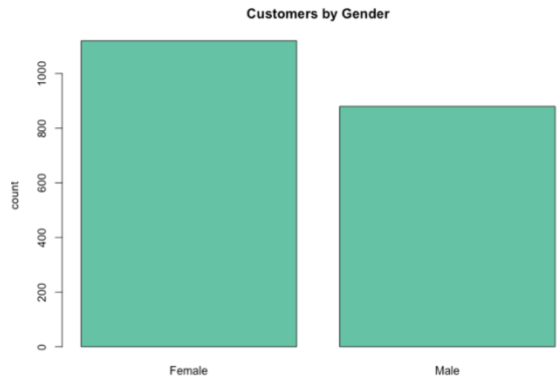
- **Data Loading**: The turtle_reviews.csv file was loaded into R and viewed using head() and as_tibble().
- **Descriptive Statistics**: We used summary(), skim(), and create_report() for an initial review (details in the appendix).
- **Data Cleaning**: We checked for null values and duplicates (none found), and removed the language and platform columns as they contained only one type of data.
- **Renaming Columns**: The remuneration..k.. column was renamed to Salary and spending_score..1.100. to Spending.

2.2 Exploratory data analysis

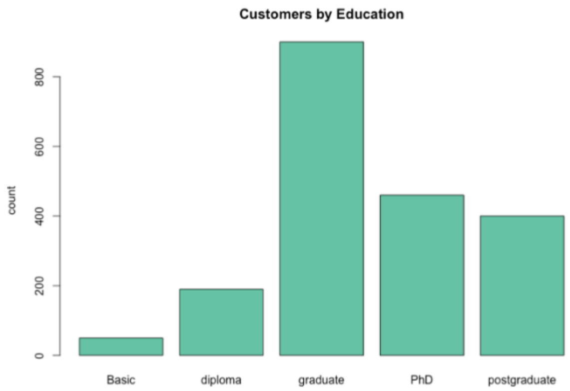
Categorical and quantitative data were explored through univariate and bivariate analyses to uncover distributions and patterns. Distributions and patterns uncovered:

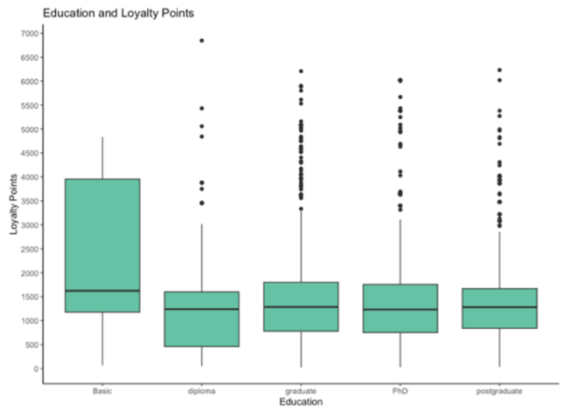
Gender Distribution

A bar plot illustrated that there are 1120 female customers and 880 male customers. This indicates that female customers make up a larger portion of the customer base, which could be significant for understanding spending behaviours and loyalty point accumulation patterns.



Education Level

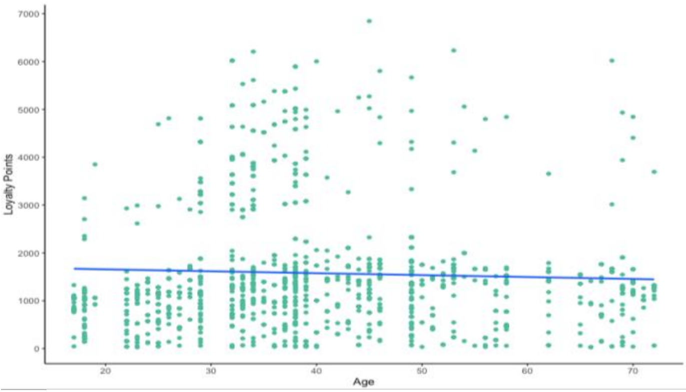




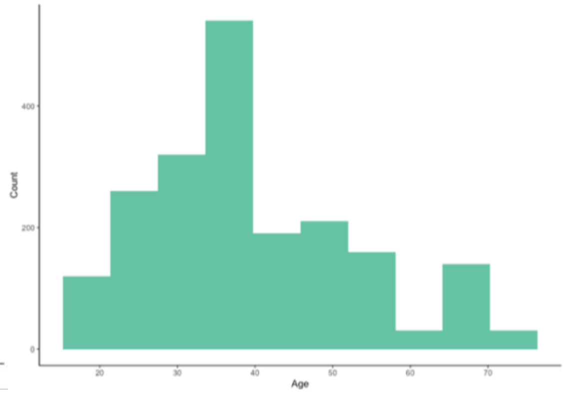
Utilizing bar plots, it was revealed that most customers hold Graduate degrees, followed by PhD and Postgraduate degrees. Fewer customers have Diploma and Basic education levels. This distribution suggests that higher education may influence spending behaviour and loyalty points accumulation.
A box plot was used to show that customers with basic education levels tend to have a higher median of loyalty points compared to those with Graduate, Postgraduate, and PhD degrees. There's not much variability in loyalty points among different education levels.

Age Distribution

Histogram analysis showed that most reviews were provided by customers in the 35-40 age group.



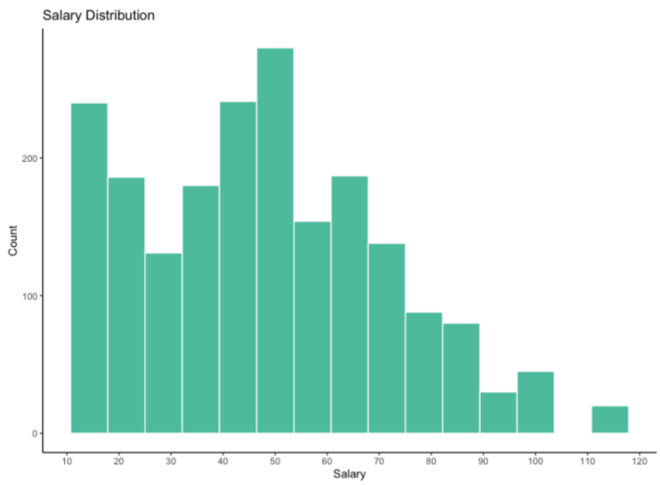
2.2 Distribution

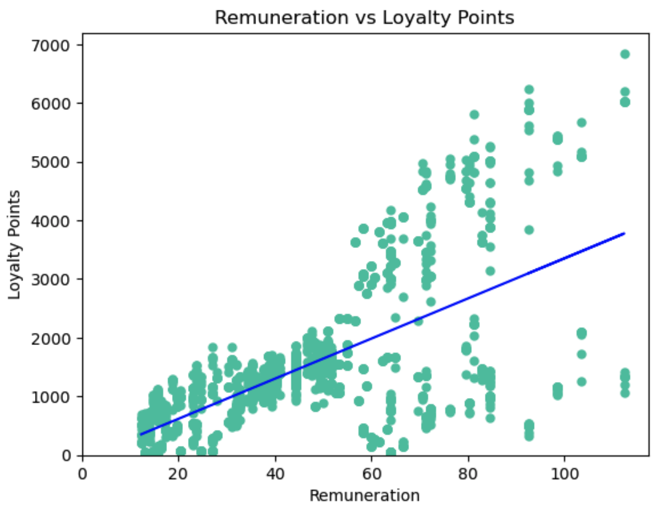


2.1 Distribution by Age

Scatterplots displayed a downward slope in Age vs. Spending Score and Age vs. Loyalty Points plots, suggesting that older customers tend to have lower spending scores and accumulate fewer loyalty points.

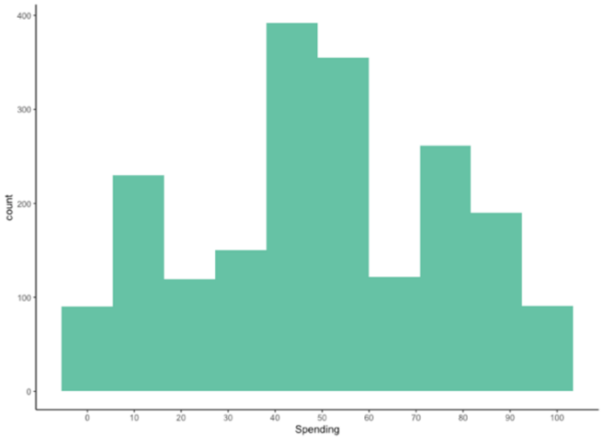
Salary Distribution

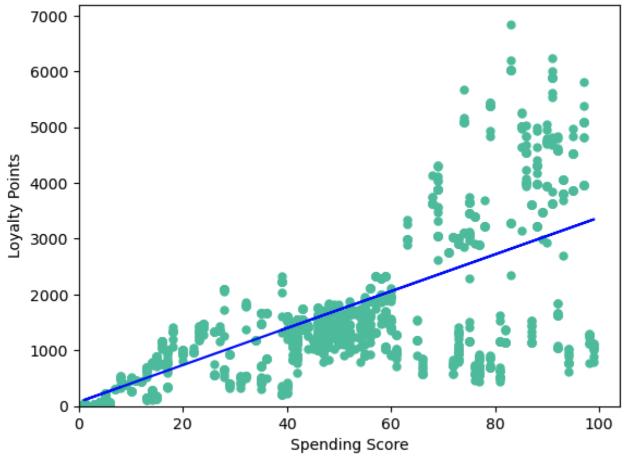




The salary distribution is skewed, with most customers earning between 15k and 35k, and fewer high-income customers. A box plot confirmed there were no income outliers in the 110-120k range. Analysis indicated that customers with higher incomes tend to accumulate more loyalty points, potentially due to higher disposable incomes and greater awareness and responsiveness to loyalty rewards.

Spending vs. Loyalty Points:



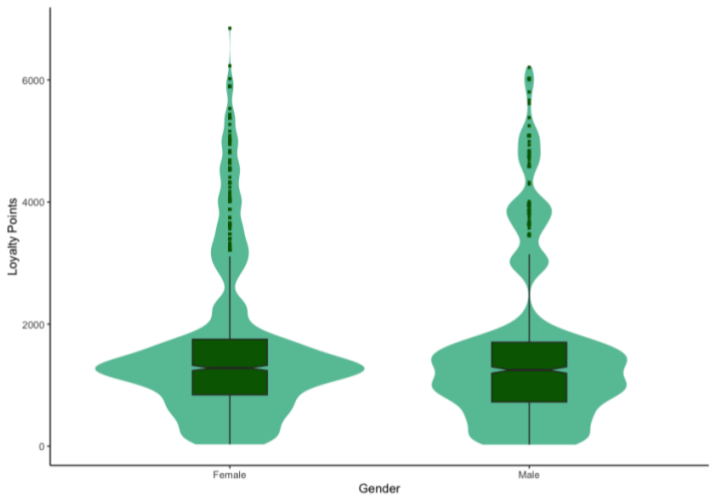


Visualizations highlighted that the highest spending category among reviewers was in the 40-50 range. Most reviewers had accumulated between 1200-1800 loyalty points, with Spending Score vs. Loyalty Points plots showing a positive correlation between spending score and loyalty points.

Age vs. Loyalty Points

Older customers tended to have fewer loyalty points compared to younger customers, potentially indicating differences in awareness and engagement with loyalty programs. Insights suggested that younger customers may be more active in earning points through various activities compared to older customers.

Loyalty Points by Gender



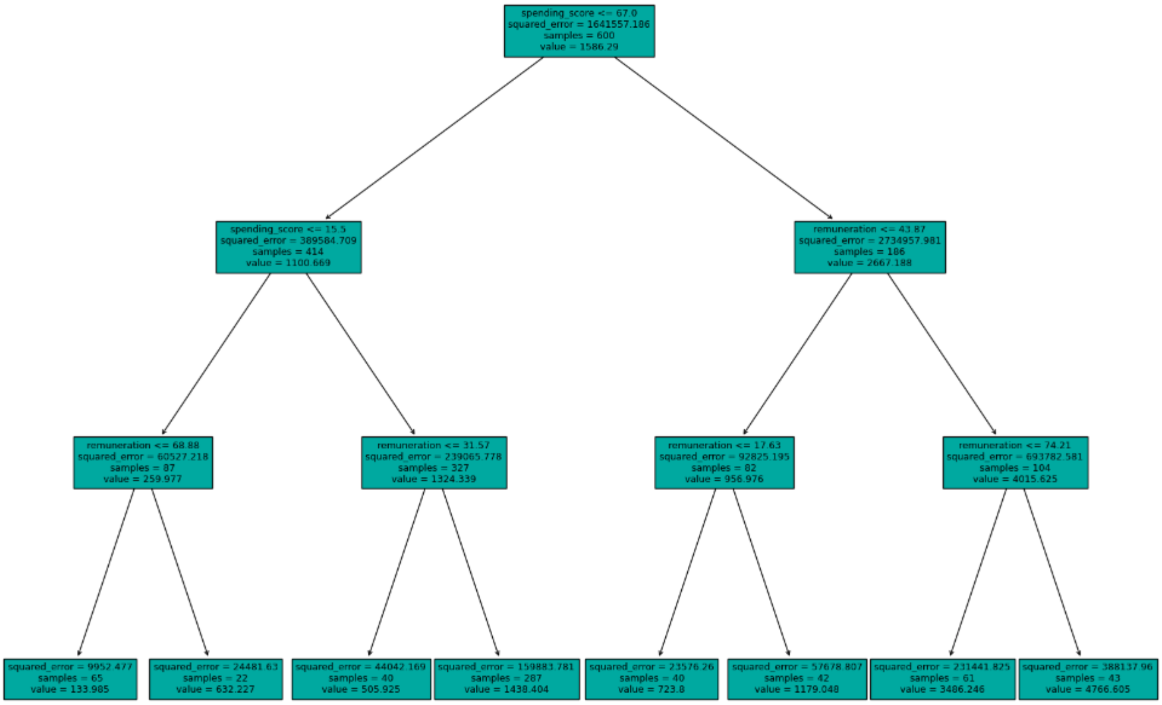
2.3 Loyalty points by gender

A violin plot overlaid with a box plot visualized the relationship between loyalty points and gender, highlighting the distribution and outliers in loyalty points across genders.

Outliers in loyalty points were not removed. The video game industry typically values loyalty points highly as they indicate player engagement and can drive retention. Many companies use tiered loyalty systems or reward frequent players with exclusive content or early access to new features. (Loyaltylion, 2024)

2.3 Feature Importance Analysis

Next a decision tree regressor was employed to evaluate predictors of loyalty points. After training on the dataset, the model's performance was assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Feature importance analysis identified **Remuneration** as pivotal, indicating higher income leads to more loyalty points. **Spending Score** also showed strong correlation with loyalty points. A pruned decision tree (max depth 3) simplified the model, revealing key decision points and reinforcing the influence of income and spending score on loyalty points, whereas **Age** had minimal impact.



3.0 Addressing Business Objectives

3.1 Segmenting Customer Behaviour: Remuneration and Spending Scores Analysis

Objective: Investigate customer segmentation to inform targeted marketing strategies.

Methodology: Utilized k-means clustering to segment customers based on remuneration and spending scores. Scatterplot, elbow, and silhouette methods determined k=5 as the optimal number of clusters.

Customers can be segmented into five groups:

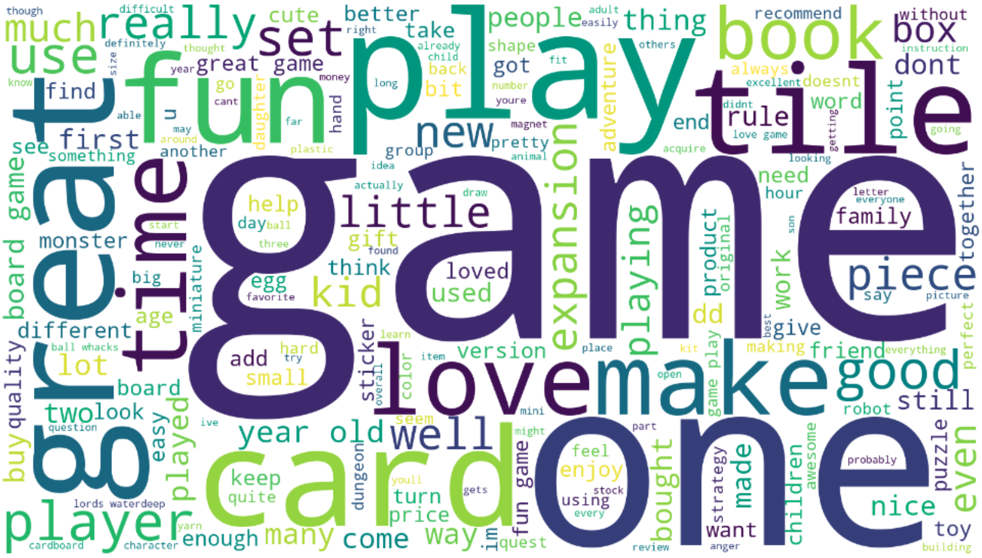
1. High Income, High Spending (Segment 0)
2. Average Income, Medium Spending (Segment 1)
3. High Income, Low Spending (Segment 2)
4. Low Income, High Spending (Segment 3)
5. Low Income, Low Spending (Segment 4)

More details in appendix.

3.2 Leveraging Customer Reviews to Inform Marketing Campaigns and Business Improvements

Customer reviews offer valuable insights into Turtle Games' brand perception and product preferences. Using NLP, reviews were pre-processed and analysed, revealing a generally positive sentiment. Frequent terms like "five" and "stars" indicate satisfaction. Sentiment analysis with TextBlob and VADER confirmed the positive trend, despite neutral or negative scores for star ratings.

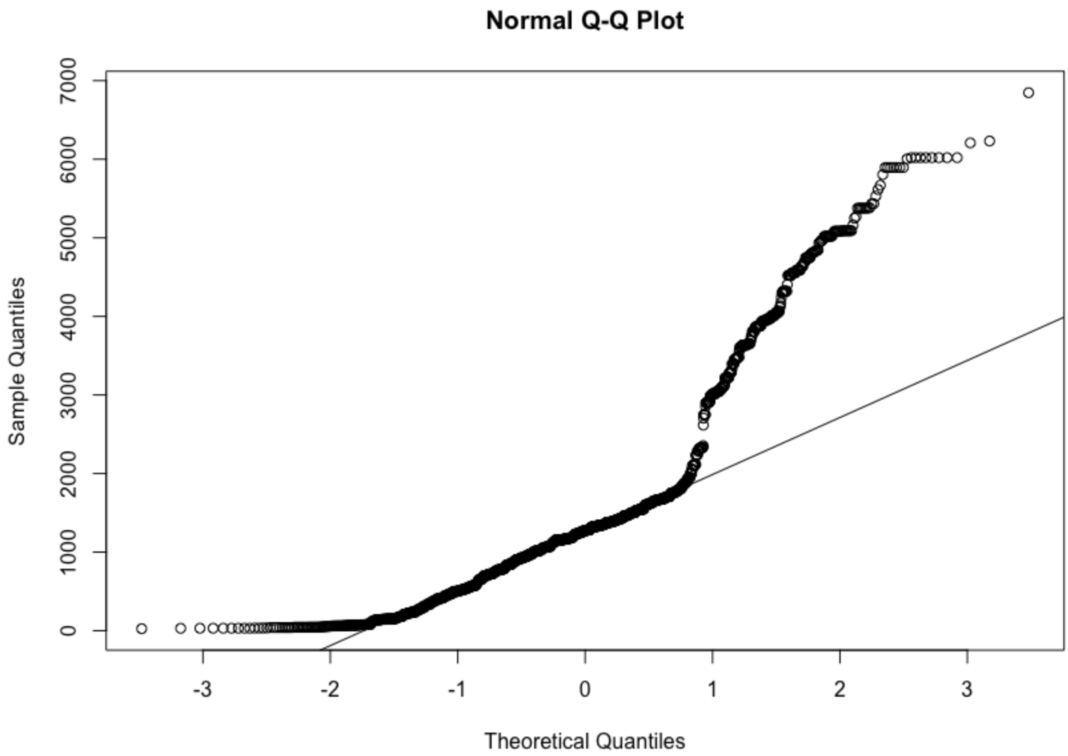
Insights from frequent positive and negative terms help identify popular and underperforming products, informing targeted marketing strategies and product enhancements. Additionally, feedback on loyalty points provides guidance for improving customer loyalty programs, aligning marketing efforts with customer preferences and expectations.



Word clouds will visually represent the most common terms from reviews, highlighting key areas for business focus and potential improvement.

3.3 Descriptive statistics for loyalty points

The Q-Q plot shows that the data points deviate from the straight line, indicating that the loyalty points data does not follow a normal distribution.



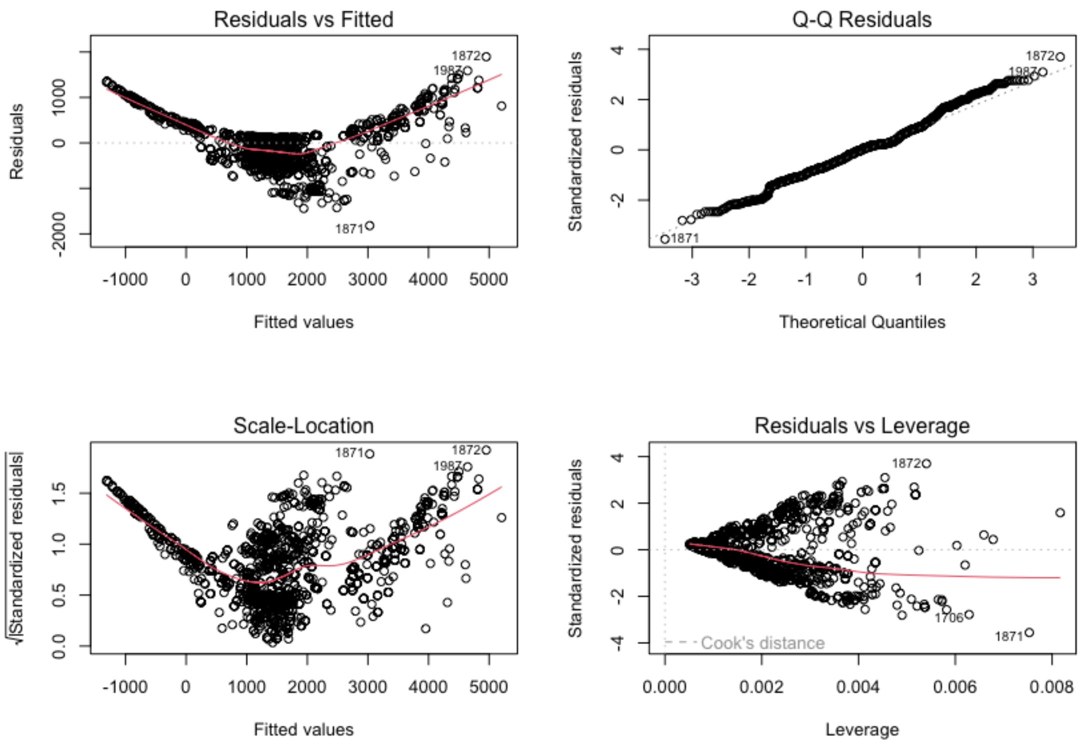
It further confirmed by the Shapiro-Wilk test ($p < 0.05$). It exhibits positive skewness (1.46) and higher kurtosis (4.71), suggesting a right-skewed distribution with heavier tails compared to a normal distribution.

4.0 Predictive Analysis and Model Evaluation

4.1 Multiple Linear Regression Model

A multiple linear regression model was developed using age, salary, and spending scores to predict loyalty points. The model demonstrated strong predictive capability with an R-squared of 0.8399, indicating that approximately 84% of the variance in loyalty points is explained by the model.

Diagnostic plots were used to validate model assumptions, including the residuals vs. fitted values plot to assess homoscedasticity and independence of residuals.



Evaluation metrics such as MAE (394.98), MSE (263486.6), and RMSE (513.31) were computed to assess model performance. These metrics underscore the model's accuracy in predicting loyalty points and highlight areas for potential refinement.

Predictions were made for a sample scenario and visualized against actual loyalty points:

Scenario 1: A 20-year-old customer with a salary of $20,000 and a spending score of 20.

Predicted Loyalty Points: -618.01

Scenario 2: A 45-year-old customer with a salary of $50,000 and a spending score of 60.

Predicted Loyalty Points: 2,046.08

Scenario 3: A 60-year-old customer with a salary of $80,000 and a spending score of 90.

Predicted Loyalty Points: 4,257.74

These predictions illustrate how the model can be used to estimate loyalty points based on specific customer attributes, aiding in targeted marketing strategies and customer segmentation. The negative predicted value in Scenario 1 suggests the model's limitations with certain input ranges, indicating a potential area for model refinement.

5.0 Recommendations

Based on the analysis and predictive modelling, we offer the following recommendations for Turtle Games:

1. **Enhance Loyalty Programs**:
    - **Target High Spending Customers**: Develop exclusive rewards for high spenders, such as premium memberships and early access to new products.
    - **Engage Low Income, High Spending Customers**: Offer value-driven rewards and promotions tailored to their spending habits.
2. **Personalized Marketing Campaigns**:
    - **Age-based Campaigns**: Younger customers are more engaged with the loyalty program. Create targeted marketing campaigns that highlight the benefits of loyalty points for younger demographics.
    - **Income-based Promotions**: High-income customers show a higher accumulation of loyalty points. Design promotions that cater to high-income customers' preferences and spending patterns.
3. **Leverage Text Data for Customer Insights**:
    - **Sentiment Analysis**: Utilize customer reviews to gauge customer sentiment and identify popular products and areas for improvement.
    - **Feedback-driven Improvements**: Implement changes based on customer feedback to enhance product offerings and customer satisfaction.
4. **Further Model Improvements**:
    - **Feature Engineering**: Create new features or transform existing ones to improve model accuracy.
    - **Non-linear Models**: Explore alternative models like polynomial regression or decision trees to capture non-linear relationships.
    - **Outlier Handling**: Address outliers to enhance model robustness.

6.0 Appendix

Customer Segmentation

Customers can be segmented into five groups:

1. **High Income, High Spending (Segment 0):**

   This consists of 356 customers who have a high disposable income and spend significantly on Turtle Games products. These customers are prime candidates for exclusive and premium rewards, early access to new products, and personalized services. Implementing a VIP loyalty program with benefits such as concierge services, special events, and limited-edition products can further engage this lucrative group, maximizing their loyalty and lifetime value.

2. **Average Income, Medium Spending (Segment 1):**
   The largest segment, comprising 774 customers, represents the majority with average income and moderate spending behaviour. For this group, the focus should be on value-driven offers like discounts, bundle deals, and reward points for regular purchases. Regular promotions and loyalty point multipliers can encourage more frequent purchases and deeper participation in the loyalty program, fostering a steady revenue stream.

3. **High Income, Low Spending (Segment 2):**
   With 330 customers, this segment features individuals with high income but relatively low spending. The marketing strategy for these customers should aim to increase engagement by promoting premium product lines and emphasizing the benefits of loyalty programs that could motivate higher spending. Personalizing marketing efforts to showcase high-value items and introducing loyalty point bonuses for higher spending tiers can help unlock their spending potential.

4. **Low Income, High Spending (Segment 3):**
   This segment includes 269 customers who, despite their lower income, spend considerably on Turtle Games products. Marketing efforts should highlight value-for-money propositions and loyalty rewards that offer tangible benefits for their spending. Special loyalty rewards and discounts tailored to their spending habits can reinforce their loyalty and ensure they feel valued, potentially increasing their overall expenditure.

5. **Low Income, Low Spending (Segment 4):**
   Comprising 271 customers, this segment has low income and minimal spending. To engage these customers, focus on entry-level products and cost-effective promotions that can incentivize them to spend more. Gradually introducing them to the benefits of the loyalty program and offering promotions that align with their financial capacity can help increase their spending gradually and integrate them more fully into the customer base.

7.0 Future Work

To further enhance the analysis and recommendations:

- **Expand Data Collection**: Collect more data points to refine the model and capture additional customer behaviours.
- **Implement Advanced Analytics**: Use machine learning techniques to identify complex patterns and provide more accurate predictions.
- **Monitor and Adjust Strategies**: Continuously monitor the effectiveness of implemented strategies and adjust based on real-time data and customer feedback.

8.0 Bibliography

Loyaltylion, 2024. *Loyaltylion.* [Online]
Available at: https://loyaltylion.com/blog/calculating-loyalty-point-value