Technical Report: Customer Loyalty Analysis for Turtle Games

By: Annora Ng

Background and Context

Turtle Games, a global game retailer and manufacturer, seeks to improve customer engagement by identifying the factors that drive loyalty point accumulation. This analysis uses customer demographic, behavioural, and sentiment data to explore who earns the most loyalty points and why. The goal is to provide insights that support personalised loyalty strategies, segment targeting, and marketing optimisation. It assumes the dataset is representative of recent customer behaviour, with loyalty points reflecting both purchase activity and engagement. The report applies statistical and text analysis techniques to generate recommendations for refining loyalty programmes and targeting high-value customer segments.

Analytical Approach

This analysis was conducted using both Python and R to address Turtle Games' business questions around customer loyalty, segmentation, and sentiment. Python was used for data cleaning, regression modelling, clustering, and natural language processing. R was used for visual analysis, statistical checks, and final regression modelling.

1. Data Preparation and Cleaning (Python)

Initial exploration was performed in Python to assess null values, duplicates, and basic descriptive statistics. Two non-informative columns, language and platform, were removed. Columns were then renamed for clarity. The cleaned dataset was saved and imported into R for further analysis.

2. Regression and Tree-Based Modelling (Python)

Linear regression models were first applied using statsmodels.api to evaluate whether spending_score, remuneration, and age were statistically significant predictors of loyalty points. This was followed by a decision tree regressor using sklearn to visualise data splits and identify key customer groups. Spending score and remuneration consistently emerged as the strongest factors influencing loyalty points.

3. Clustering for Segmentation (Python)

K-means clustering was used to identify customer segments based on remuneration and spending_score. The optimal number of clusters (k = 5) was determined through elbow and silhouette methods. Each cluster was then profiled to identify strategic targeting opportunities.

4. Natural Language Processing (Python)

TextBlob was used to compute sentiment polarity scores (-1 to 1) for both the review and summary fields. Word clouds and histograms were generated to compare sentiment patterns.

Reviews showed a more balanced and useful spread of sentiment than summaries, which appeared repetitive and overly positive.

5. Statistical Analysis and Modelling (R)

The cleaned dataset was loaded into R for detailed exploration. Boxplots and histograms were used to assess distributions and detect outliers. The moments, corrplot, and car libraries supported checks on skewness, correlation, and multicollinearity. A multiple linear regression model was built using the log-transformed loyalty points. Model diagnostics confirmed reasonable fit and assumption validity.

The combined use of Python and R enabled robust exploratory work, predictive modelling, and interpretable insights that support business decision-making.

Visualisations and Insights

Distributions and Trends

Distribution of Loyalty Points (Histogram)

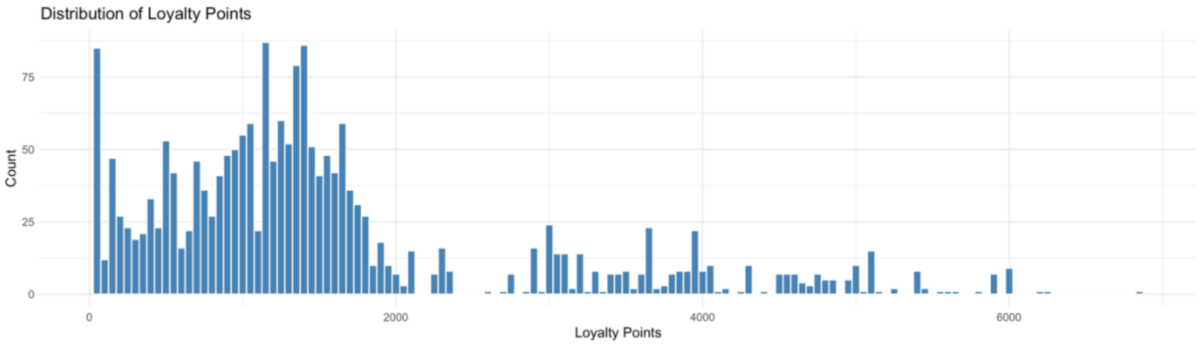


Figure 1: Histogram Distribution of Loyalty Points

Loyalty points distribution is right-skewed. Most customers have fewer than 2,000 points.

Distribution of Boxplot of Loyalty Points (Boxplot)

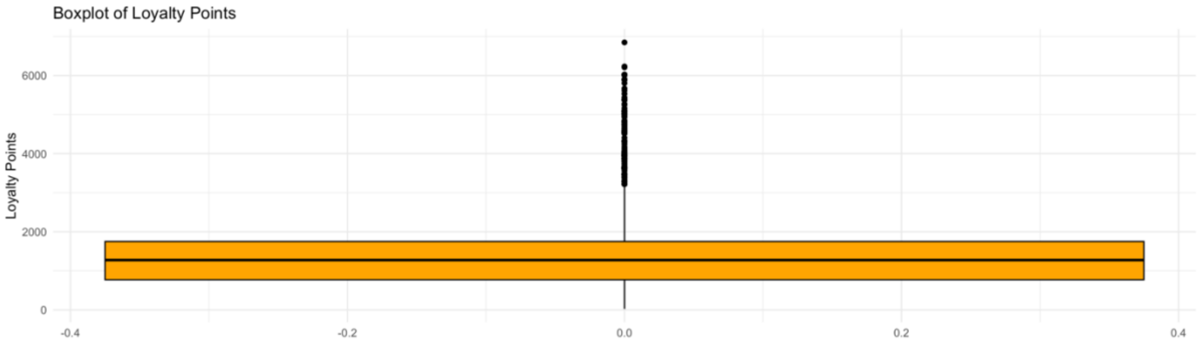


Figure 2: Boxplot Distribution of Loyalty Points

Boxplots show presence of outliers up to 6,800 points. These form the top 1% of customers that could be targeted separately.

Demographics and Loyalty Points

Loyalty Points vs Remuneration

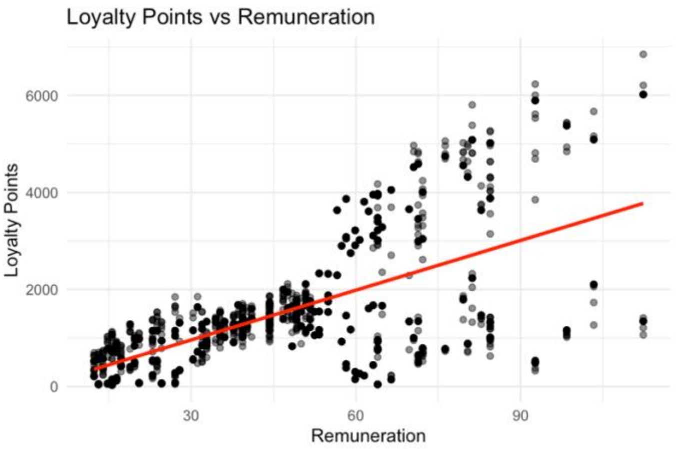


Figure 3: Scatterplot of Loyalty Points vs Remuneration

Remuneration: Positive correlation with loyalty points. Higher income, higher points.

Loyalty Points vs Spending Score

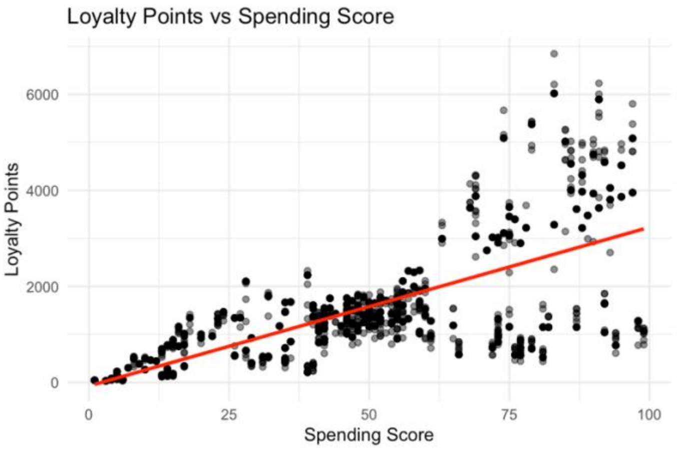


Figure 4: Scatterplot of Loyalty Points vs Spending Score

Spending Score: Strong upward trend with loyalty points. Scatterplot and regression line support correlation.

Loyalty Points vs Age

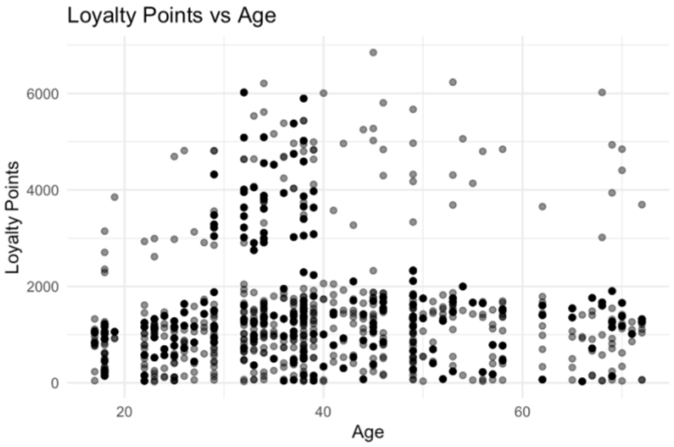


Figure 5: Scatterplot of Loyalty Points vs Age

Age: Loyalty points are spread across age groups. No meaningful trend.

Loyalty Points by Gender

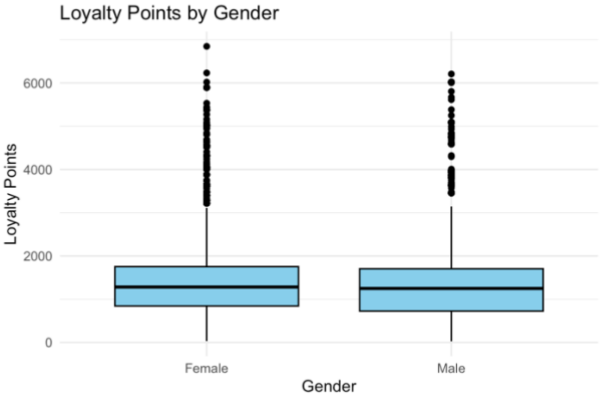


Figure 6: Boxplot of Loyalty Points by Gender

Gender: Similar distributions, not a meaningful predictor.

Loyalty Points by Education Level

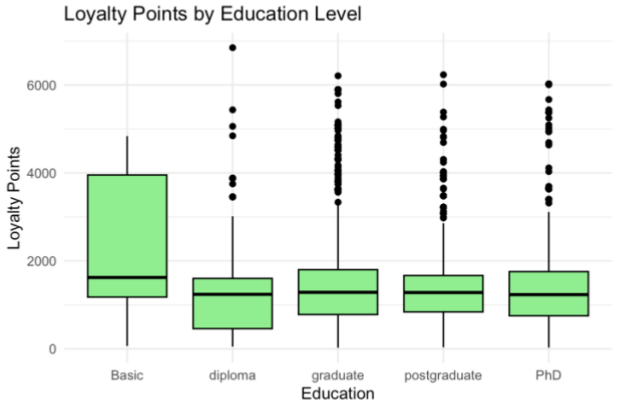


Figure 7: Boxplot of Loyalty Points by Education Level

Education: Median values across education levels are comparable. Not a meaningful predictor.

Correlation Matrix

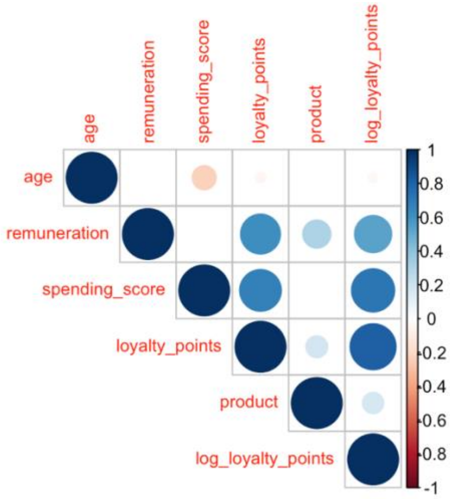


Figure 8: Correlation Matrix

- Highest correlations observed between loyalty points and remuneration/spending score.
- No multicollinearity between predictors.
- Age and education show weak correlation with loyalty scores.

Outliers and Top 1% Loyalty Earners

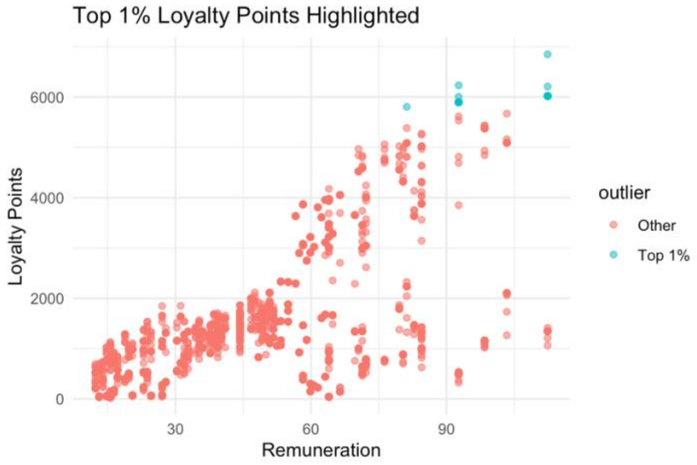


Figure 9: Scatterplot of Loyalty Points vs Remuneration with Outliers highlighted

- Customers in the top 1% (highlighted in light blue) had high remuneration and spending scores.
- This group can be targeted for premium campaigns or loyalty upgrades.

Text Sentiment and Word Analysis

<u>Histogram of Review Polarity</u>

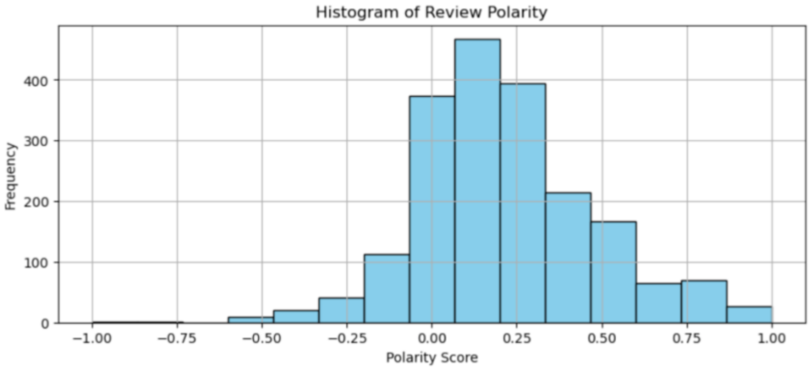


Figure 10: Histogram of Review Polarity

Review polarity scores are more positive, ranging from 0 to 1.

Histogram of Summary Polarity

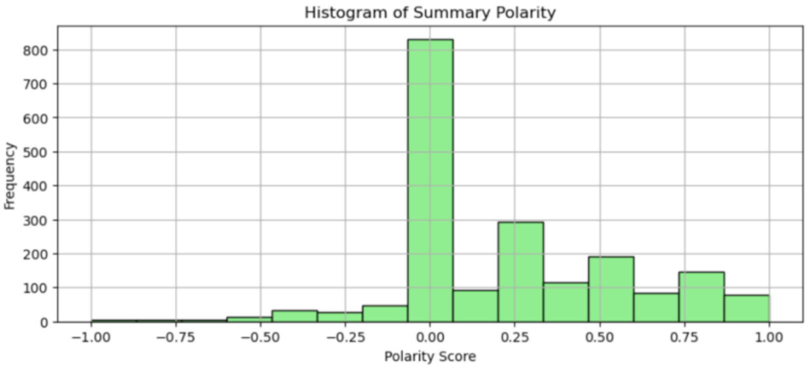


Figure 11: Histogram of Summary Polarity

Summary polarity was mainly clustered around 0, suggesting more neutral sentiment.

Word Cloud for Reviews

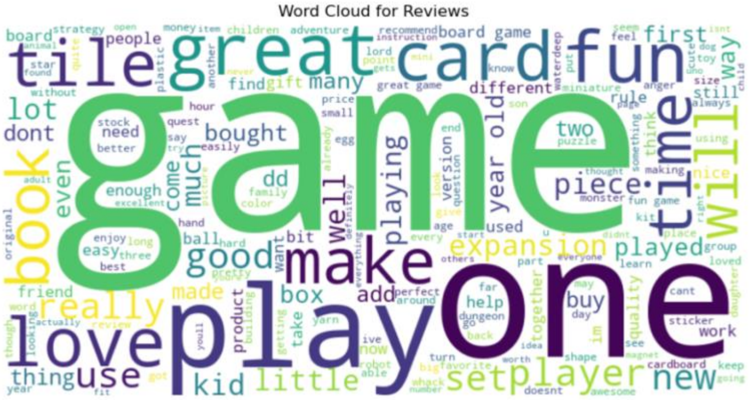


Figure 12: Word Cloud for Reviews

Word Cloud for Summary

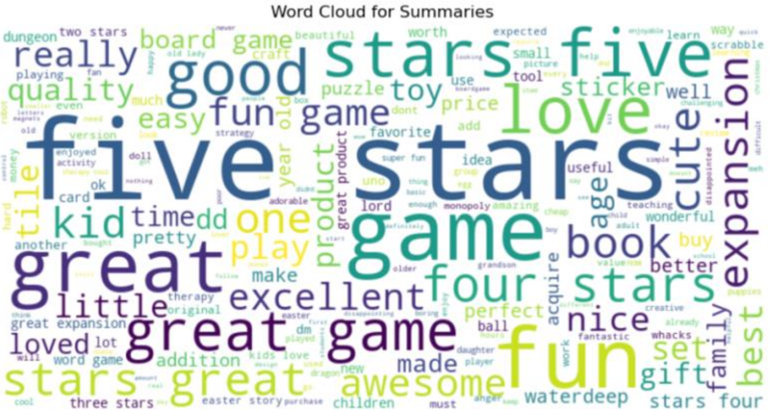
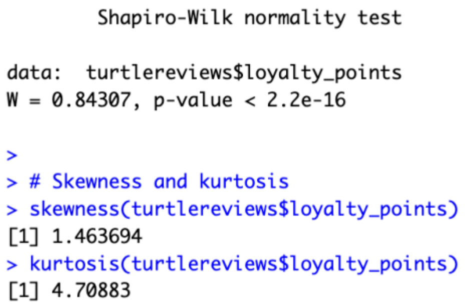


Figure 13: Word Cloud for Summaries

Word clouds confirmed this: summaries contained repeated high-level terms like "great," "fun," and "five stars." Reviews offered more expressive content: "tile," "player," "family," "set", which could be used to infer product preference. Moving forward, review is recommended over summary due to its natural variation.

Patterns and Predictions

A multiple linear regression model was built using R to predict log-transformed loyalty points based on remuneration and spending score. The log transformation was necessary due to non-normality of the original distribution (Shapiro-Wilk p < 2.2e-16; skewness = 1.46; kurtosis = 4.71).



```
                Shapiro-Wilk normality test

data:  turtlereviews$loyalty_points
W = 0.84307, p-value < 2.2e-16

>
> # Skewness and kurtosis
> skewness(turtlereviews$loyalty_points)
[1] 1.463694
> kurtosis(turtlereviews$loyalty_points)
[1] 4.70883
```

Figure 14: Shapiro-Wilk Normality Test

Model Summary:

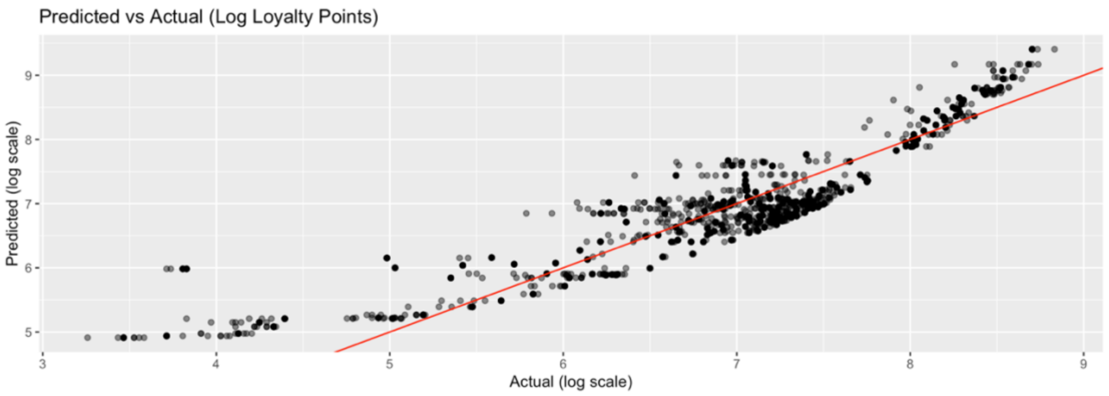
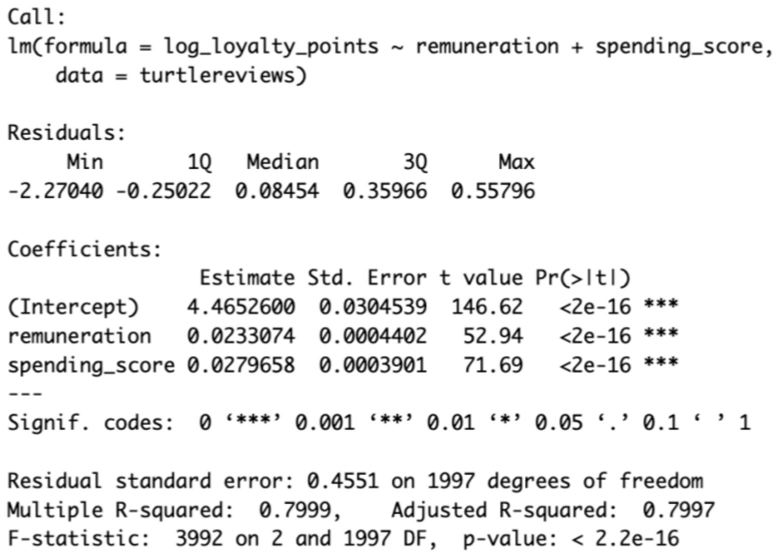


Figure 15: Predicted vs Actual Values

- **Formula**: log_loyalty_points ~ remuneration + spending_score
- **Adjusted R²**: 0.7997 — model explains ~80% of the variance.
- **Coefficients**:
  - Remuneration: β = 0.0233, p < 2e-16
  - Spending Score: β = 0.0280, p < 2e-16
- **Residual Std. Error**: 0.4551
- **F-statistic**: 3992 on 2 and 1997 DF, p < 2.2e-16

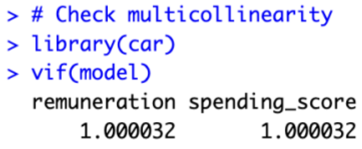


```
Call:
lm(formula = log_loyalty_points ~ remuneration + spending_score,
    data = turtlereviews)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.27040 -0.25022  0.08454  0.35966  0.55796 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    4.4652600  0.0304539  146.62   <2e-16 ***
remuneration   0.0233074  0.0004402   52.94   <2e-16 ***
spending_score 0.0279658  0.0003901   71.69   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4551 on 1997 degrees of freedom
Multiple R-squared:  0.7999,	Adjusted R-squared:  0.7997 
F-statistic:  3992 on 2 and 1997 DF,  p-value: < 2.2e-16
```

Figure 16: Regression Model

Multicollinearity Check:

- VIF scores for both predictors were ~1.000, indicating no multicollinearity.



```
> # Check multicollinearity
> library(car)
> vif(model)
  remuneration spending_score 
      1.000032       1.000032 
```

Figure 17: VIF

Assumption Diagnostics:

- **Linearity**: Supported by strong R² and visual clustering around line of best fit.
- **Normality**: QQ plot and residuals suggest mild deviations, acceptable post-log transformation.
- **Homoscedasticity**: Residuals vs fitted values and scale-location plots show roughly constant variance.
- **Influential Observations**: No high-leverage points detected; Cook’s distance values remain low.

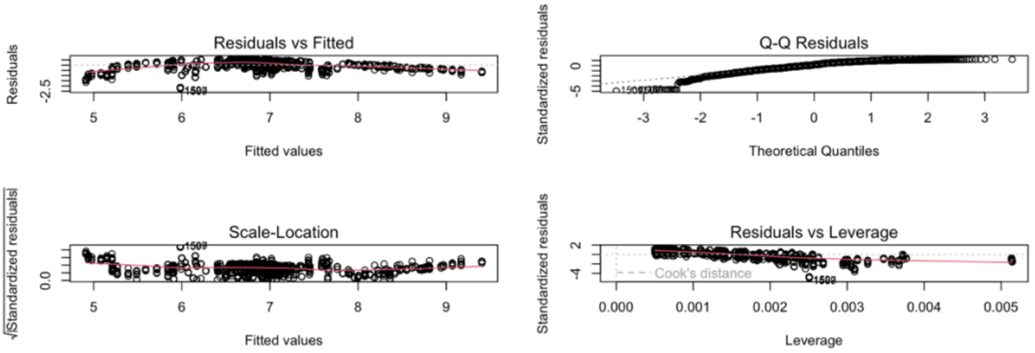


Figure 18: Residual Diagnostics

Implications:

- Spending score and remuneration are strong, independent predictors of loyalty points.
- Behavioural features outperform demographics like age or gender in predictive strength.
- This model offers a stable foundation for loyalty targeting and customer scoring, especially for use in campaign allocation or CRM updates.

Recommendations

Loyalty Programme Targeting

- Focus 70% of loyalty campaign resources on customers with remuneration and spending scores ≥ 60. This group showed the highest loyalty point accumulation and includes most of the top 1% earners identified in the regression and clustering analysis.
- Introduce premium-tier perks (e.g. early access, exclusive bundles) for customers in the top 1% of loyalty points (> 4000 points), who have both high income and high spend levels.
- Set up churn-prevention prompts for customers with remuneration ≥ 70 but loyalty points < 1000. These high-potential customers are under-engaged despite their spending capacity.

Sentiment Insights Application

- Discontinue reliance on the summary column for sentiment monitoring. Sentiment scores were overly clustered around zero or mildly positive, with repetitive vocabulary.
- Use polarity scores from the review column to build a monthly sentiment tracker. Prioritise negative reviews (polarity < 0) for product feedback analysis and service intervention.
- Include aggregate sentiment trends in quarterly campaign reviews to support brand and product perception insights.

Segment-Based Campaigns

- Deploy marketing campaigns based on the five clusters identified from K-means:
  - High-income, low-spend customers → individualised promotions
  - Low-income, high-spend customers → volume-based incentives
- Redirect 40% of digital ad budget toward segments with spending scores ≥ 60 but loyalty points < 2000 to encourage more program interaction.

Future Data Enhancements

- Add behavioural features like visit frequency, customer ID, product category, and purchase channel in the next data cycle to improve segmentation granularity.
- Implement lightweight NPS-style surveys post-purchase to gather structured feedback (target 5% response rate/month).
- Allocate one analyst resource to begin compiling customer-level transaction histories for model refinement in future phases.

Appendix

Appendix A: Limitations

Data Limitations

- **Manual Summary Bias**: The summary column appears to be manually entered, often using generic descriptors (e.g. “five stars”, “great”), which skews sentiment scores and limits its reliability for NLP.
- **Lack of Mechanism for Loyalty Accumulation**: The data does not specify *how* loyalty points are calculated or earned (e.g. spending thresholds, engagement activities). This limits interpretation of causality in modelling.
- **Missing Product Metadata**: The product column consists of codes without an accompanying product table. This prevents analysis based on product category, price, or type.
- **Potential Sampling Bias**: It’s unclear whether the dataset is a full or partial representation of all customers. If it’s a subset, insights may not generalise to the entire customer base.

Modelling and Statistical Limitations

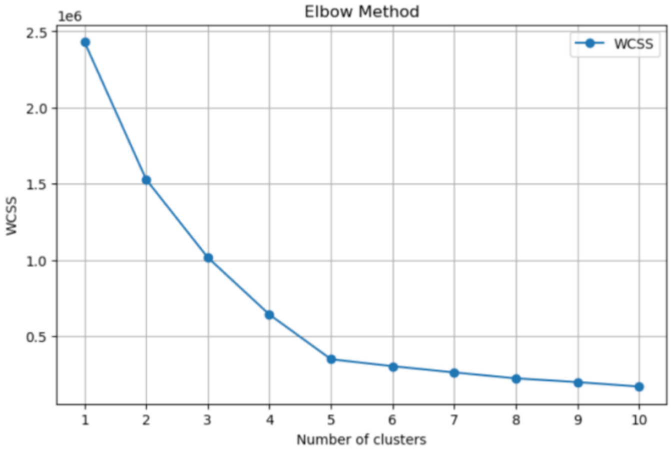
- **Non-normality of Raw Loyalty Points**: The original loyalty points variable was highly skewed, requiring log transformation for regression modelling.
- **Residual Deviations**: Although the final regression model assumptions were mostly met, there were mild deviations from normality and signs of heteroscedasticity in the residuals.
- **Limited Feature Set**: Core predictors like remuneration and spending_score were useful, but other valuable behavioural features (e.g. transaction history, frequency, recency) were not available.
- **Outlier Sensitivity**: High-value loyalty customers (top 1%) exert influence on model outputs. These outliers were not excluded but could distort general patterns if not handled cautiously.

Analytical Scope Limitations

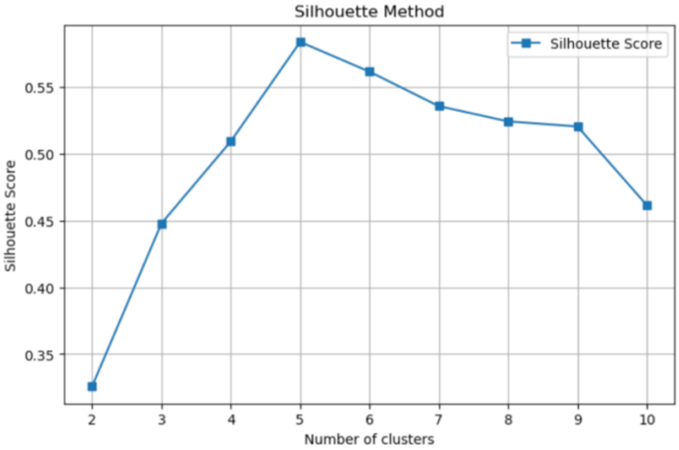
- **Review Sentiment Not Linked to Specific Products**: Sentiment scores were calculated globally across all reviews, not tied to specific product experiences or actions.

Appendix B: Additional Visualisations

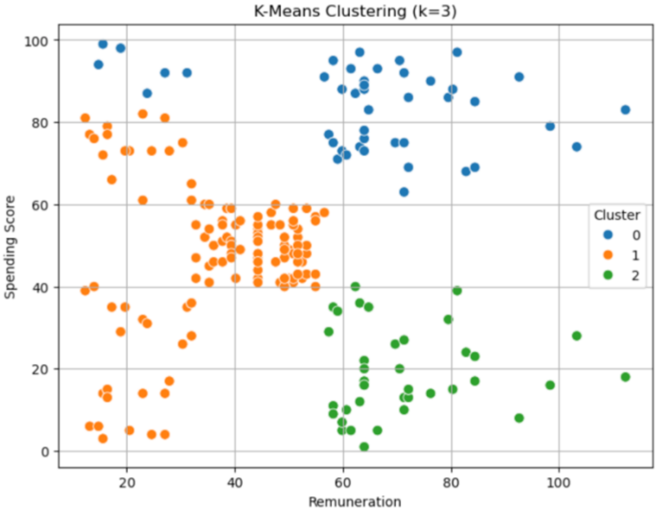
Elbow plot for determining optimal k in clustering

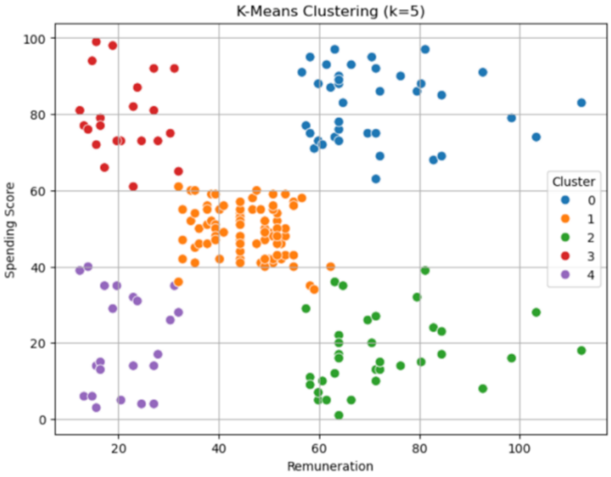


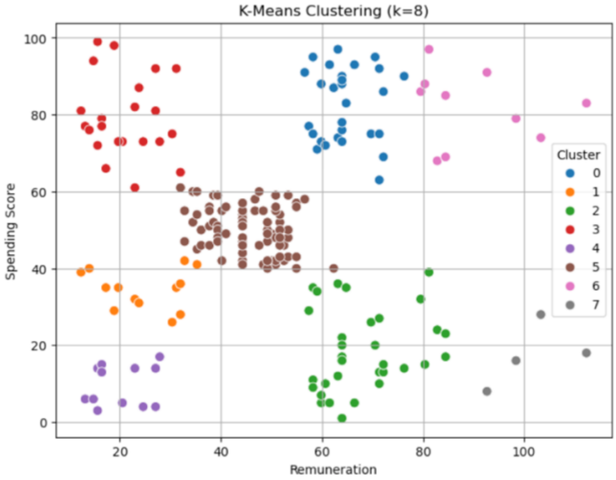
Silhouette plot for clustering evaluation



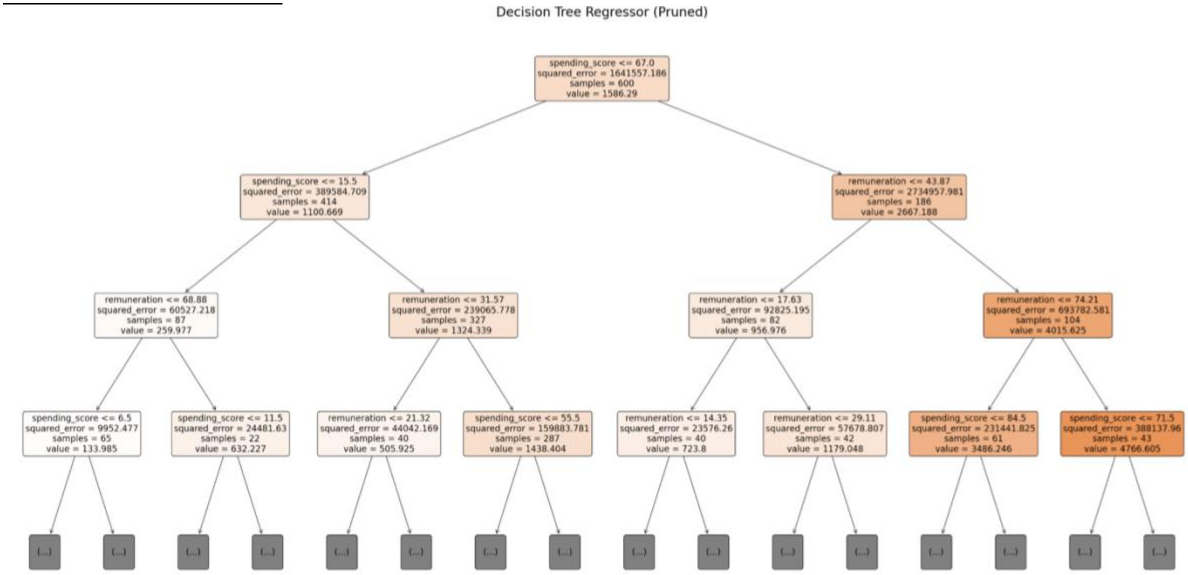
Different k-means outputs (k=3, k=5, k=8)



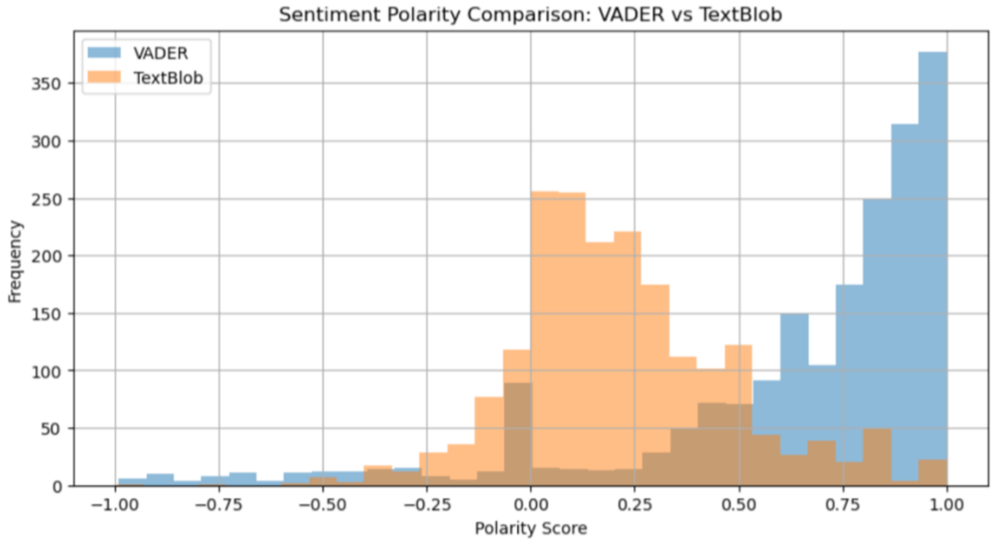




Pruned Decision Tree



VADER vs TextBlob Comparison



The histogram comparison reveals a divergence in sentiment classification between TextBlob and VADER. While TextBlob's polarity scores cluster around 0.1–0.3, VADER shows a strong skew toward scores above 0.7, suggesting higher sensitivity to emotional cues or informal language. VADER’s compound score may better reflect customer enthusiasm in casual reviews, whereas TextBlob provides more neutral and balanced estimates. This discrepancy can influence interpretation of customer satisfaction and should be considered when designing text-driven loyalty or feedback models.

Customer Group Segmentation from k=5

Cluster	**Key Traits**	**Recommended Action**
0	High Income, High Spending	Prioritise with premium loyalty tiers, exclusive launches, early access
1	Mid Income, Mid Spending (Average Customer Base)	Maintain general loyalty programme communications, test cross selling offers
2	High Income, Low Spending	Upsell with personalised offers, highlight value-driven bundles
3	Low Income, High Spending	Use retention strategies (points boosters, timed discounts)
4	Low Income, Low Spending	Deprioritise for resource intensive campaigns, use low cost digital nudges

Appendix C: Assumptions

This analysis was conducted based on the following assumptions:

- **Loyalty points reflect customer engagement and purchase activity.**
    - Without access to the exact formula used to compute loyalty points, we assume they are meaningfully tied to customer behaviour (e.g. purchases, frequency, or other engagement metrics).
- **The dataset represents recent and typical customer behaviour.**
    - We assume the turtle_reviews.csv dataset is current, complete, and representative of Turtle Games' broader customer base during the analysis period.
- **Text sentiment analysis reflects general satisfaction.**
    - Sentiment polarity scores derived from TextBlob are used as a proxy for overall customer satisfaction, despite the absence of product-level or time-stamped sentiment granularity.
- **Clusters identified using K-means are stable and business-relevant.**
    - The 5-cluster solution is assumed to represent meaningful segmentation for marketing and loyalty programme strategy, even without external validation.