Turtle Games
Technical Report

Eloise Farmer
LSE Career Accelerator
Advanced Analytics for Organisational Impact

Table of Contents

Background: **2**

Analytical Approach: **2**

Visualisation and Insights: **3**

Patterns and Predictions: **5**

Appendix List: **6**

Appendix 1: 5 Whys Analysis - Turtle Games 6
Appendix 2: Notebook Structure 6
Appendix 3: Data Preparation Details (Python) 7
Appendix 4: Loyalty Points Distribution 7
Appendix 5: Correlation Matrix of Key Variables 9
Appendix 6: Simple Linear Regression (Python, OLS) 10
Appendix 7: Multiple Linear Regression 12
Appendix 8: Decision Tree Regression 15
Appendix 9: Comparative Results of Regression and Decision Tree Models 18
Appendix 10: Model Comparison - Results and Reasoning 19
Appendix 11: K-means Clustering - Detailed Results and Visualisations 20
Appendix 12: NLP & Sentiment Analysis (Python) 25
Appendix 13: Product Sales & Sentiment Proxy Analysis 33
Appendix 14: Further Exploratory Analysis in R 36
Appendix 15: Further Analysis Opportunities 41

Reference List: **41**

Background:

Turtle Games is a manufacturer and retailer of books, board games, video games, and toys. It collects customer purchase and review data but has yet to fully leverage this information (see Five Whys in Appendix 1). The business objective is to improve sales by analysing customer behaviour, loyalty patterns, and feedback, providing insights to inform marketing, enhance engagement, and drive growth.

To guide the analysis, we focused on four key questions:

Business Questions:

1. How customers earn and use loyalty points.
2. How customers can be segmented into clear groups for targeted marketing.
3. How reviews can inform campaigns and product improvements.
4. Whether loyalty data is suitable for predictive modelling.

These framed the modelling and exploratory work carried out in Python and R, and link directly to the recommendations presented later in this report.

Analytical Approach:

Analysis combined Python (data cleaning, regressions, trees, clustering, NLP) and R (exploratory analysis and multiple regression), with a fixed random state for reproducibility.

Data preparation (Python)
• Imported with pandas/NumPy; no missing values found.
• Dropped irrelevant columns (language, platform), renamed fields, and saved cleaned dataset.

Exploratory Analysis (R)
• Used ggplot2 for scatterplots, distributions, and heatmaps.
• Confirmed remuneration and spending as main loyalty drivers.

Regression
• Python OLS tested spending, remuneration, and age.
• Final MLR (Python & R) combined spending + remuneration, validated with RMSE, MAE, MSE.
• Checked assumptions (linearity, residuals, Durbin-Watson).

Decision Trees (Python)
• Built across depths 2-6; depth 3 selected as best balance of accuracy and interpretability.

Clustering (Python)
• K-means (k=3-6) with elbow + silhouette (best at k=5).
• Hierarchical clustering confirmed k=5.

NLP (Python)
• Processed reviews/summaries: lowercasing, punctuation removal, tokenisation, stopword removal.

• Retained 39 duplicates as genuine repeated feedback.
• Ran sentiment analysis (VADER, polarity, compound distributions).

Product Analysis (Python)
• Grouped products by spending_score (proxy for sales).
• Combined with VADER sentiment, producing scatter plots to map products into four performance quadrants.

Notebook structure - Appendix 2. Further detail of analytical approach - Appendix 3.

Visualisation and Insights:

Loyalty Points Distribution
Most customers earn 0-2,000 loyalty points, showing typical engagement levels. The graph highlights the outlier threshold (~3,220 points), with about 13% of customers above it. These are the most loyal, high-value buyers, while the long tail of low-point customers signals opportunities to drive wider engagement (Appendix 4).

Correlation Analysis
Before running regression models, I examined correlations between key variables. Loyalty points are most strongly associated with spending score and income, confirming these as the main financial drivers of loyalty outcomes. In contrast, age and product have little explanatory power (Appendix 5).

This analysis guided the regression modelling: only spending score, remuneration, and age were tested directly, with age quickly excluded after showing negligible predictive value. The correlations also confirm that financial behaviour, rather than demographic attributes, is central to loyalty performance - a theme reinforced throughout the regression and clustering results.

Linear and Multiple Regression Analysis
To test how customer attributes predict loyalty points, I applied Ordinary Least Squares (OLS) regression, followed by a train/test split to assess predictive accuracy. Based on the feature importance analysis (which showed only remuneration and spending score matter, while age and demographics contributed negligibly), I limited regression to three candidate predictors: spending score, remuneration, and age (Appendix 6).

The simple regressions confirmed that spending score is the strongest single predictor (R^2 = 0.45), remuneration explains less (R^2 = 0.38), and age has almost no predictive value ($R^2 \approx 0$, $p = 0.058$). Each linear model showed mild assumption violations (table below).

Combining remuneration and spending score in a Multiple Linear Regression (MLR) model produced a step change in performance, explaining 83% of the variation in loyalty with substantially lower errors (RMSE ≈ 534, MAE ≈ 415). Both predictors were highly significant ($p < 0.001$) and retained independent explanatory power. Assumption checks confirmed this model is robust, with residuals approximately normal and no multicollinearity (Appendix 7).

Assumption Testing Summary (Linear & Multiple Regression)

Model	Linearity	Independence (DW)	Homoscedasticity	Normality	Outliers / Influential Points	Overall Assessment
Linear: Spending Score → Loyalty	Clear positive trend; slight curvature at higher values	DW ≈ 1.19 → mild positive autocorrelation (not severe)	Variance increases at high spending → heteroscedastic	Residuals skewed, heavy tails (tests reject normality)	Some extreme loyalty scores >6000; inflate variance but not dominant	Strong predictor ($p<0.001$); assumptions mostly hold with mild violations
Linear: Remuneration → Loyalty	Broadly positive, but weaker fit at high income	DW ≈ 3.62 → negative autocorrelation (likely noise)	Higher variance in wealthy customers → heteroscedastic	Residuals skewed with heavy tails ($p<0.001$)	High-income customers with low loyalty distort variance	Statistically significant, but weaker and less reliable than spending score
Linear: Age → Loyalty	No meaningful trend; slope ≈ 0	DW ≈ 2.28 → no autocorrelation	Wide variance across all ages, no clear pattern	Residuals non-normal; fit is weak	Extreme loyalty values appear randomly across ages	Age not significant ($p=0.058$), $R^2 \approx 0$; assumptions confirm irrelevance
MLR (Rem + Spend)	Linearity satisfied for both predictors	DW ≈ 3.48 → slight negative autocorrelation (not critical)	Residuals evenly spread, only minor funnelling	Residuals approx. normal (Jarque–Bera $p \approx 0.098$)	Some outliers, but influence limited	Assumptions largely met; model is robust, interpretable, and most reliable

Overall, MLR emerges as the most defensible regression approach: it balances statistical validity, interpretability, and predictive accuracy, outperforming all single-variable models.

Decision Tree Analysis:
The analysis shows that loyalty points are driven almost entirely by spending score and remuneration. Low-spending, low-income customers earn few points, while high-spending, high-income groups earn the most (~4,800). Deeper branches ($n < 50$) risk fewer representative predictions. Depth 3 was optimal, explaining 90% of variation ($R^2 = 0.92$) while avoiding overfitting from deeper trees (Appendix 8).

This confirms the regression results (where spending and income were the only significant predictors) but adds clear, rule-based insights: loyalty outcomes are structured by **financial behaviour, not demographics**. For business decisions, this makes the model both accurate and interpretable.

Appendix 9 - full performance metrics of all models

Appendix 10 - details for model selection

Customer Segmentation with K-means Clustering

To segment customers, I applied K-means clustering on remuneration and spending score. Both the elbow method and silhouette analysis identified k = 5 as optimal (silhouette = 0.583), balancing separation with interpretability (Kaufman & Rousseeuw, 2009). Hierarchical clustering (Ward linkage, Euclidean distance) cross-validated this, also favouring k = 5 (silhouette = 0.581) (Appendix 11). The five clusters provide clear and actionable customer segments:

K-means Clusters and Business Actions

Cluster	% of Customers	Profile	Strategic Action
0	18%	High income, high spending	Premium retention: VIP perks, exclusive offers
1	39%	Moderate income, average spending	Core market: broad marketing, retention focus
2	17%	High income, low spending	Upsell: use targeted campaigns, bundling
3	13%	Low income, high spending	Value seekers: discounts, loyalty multipliers
4	14%	Low income, low spending	Budget-conscious: low-cost engagement, nurture

Sentiment Analysis
Histograms, bar charts, word counts, and word clouds were used to explore customer reviews. Polarity (TextBlob) showed reviews lean slightly positive, while subjectivity scores revealed reviews are moderately opinionated, but summaries are often objective. VADER sentiment confirmed broadly positive feedback overall, though short one-word summaries (e.g. “Five stars”) often confused the algorithm and were misclassified as neutral (Appendix 12). To connect sentiment with sales, a scatter plot (sales vs sentiment) placed products into four quadrants, distinguishing star performers (high sales, positive sentiment), risk areas (high sales, negative sentiment), and hidden opportunities (low sales, positive sentiment). These visualisations provide actionable insight into how customer perceptions align with product performance (Appendix 13).

Patterns and Predictions:

The analysis showed that loyalty at Turtle Games is driven primarily by financial behaviour. Spending score and remuneration were consistently the strongest predictors, while age, gender, and education had negligible impact. The multiple linear regression (MLR) model explained 83% of loyalty variation with low error, making it both accurate and interpretable. This provides predictive capability at customer onboarding, enabling new customers to be placed into the right segment immediately.

Clustering confirmed five clear customer segments. Premium customers (high income, high spending) should be retained with VIP rewards and exclusivity. Affluent but disengaged customers (high income, low spending) represent the largest growth opportunity and require targeted re-engagement campaigns. The mainstream (~40%) forms the sales backbone, best reached via seasonal campaigns and referral offers. Value seekers can be nurtured with bundles, while budget-conscious customers suit low-cost automated engagement.

At the product level, combining sales proxies with sentiment analysis highlighted four quadrants: protect cash cows, promote hidden gems, investigate risk products (high sales, poor sentiment), and deprioritise low-value items. Importantly, review analysis showed that short “summary” fields (e.g., *“five stars”*) confused sentiment algorithms. Turtle Games should phase out summaries in favour of star rating, while retaining full reviews for NLP monitoring to capture emerging issues.

Considerations for further analysis are outlined in Appendix 15.

Appendix List:

Appendix 1: 5 Whys Analysis - Turtle Games

Problem Statement:
Turtle Games needs to improve sales performance by leveraging loyalty data, customer segmentation, and product feedback more effectively.

1. **Why does Turtle Games need to improve sales performance?**
   Because competition in the gaming market is increasing, with more digital platforms and alternative retailers attracting customers.
2. **Why are competitors attracting Turtle Games' customers?**
   Because competitors use stronger data-driven marketing and loyalty schemes that better target customer needs and behaviours.
3. **Why is Turtle Games not matching these customer-focused strategies?**
   Because its existing use of customer data (purchases, loyalty points, reviews) is underdeveloped, focusing on collection rather than analysis and action.
4. **Why is customer data underutilised?**
   Because the company lacks systematic analytical processes - such as segmentation, predictive modelling, and sentiment monitoring - to translate raw data into actionable insights.
5. **Why has Turtle Games not built these analytical processes?**
   Because of limited investment in analytics capability and underestimation of the value of advanced modelling (e.g., regression, clustering, NLP) for driving targeted marketing and retention.

Root Cause:
Turtle Games has not fully embraced data-driven decision-making, leaving valuable insights from loyalty and review data underexploited.

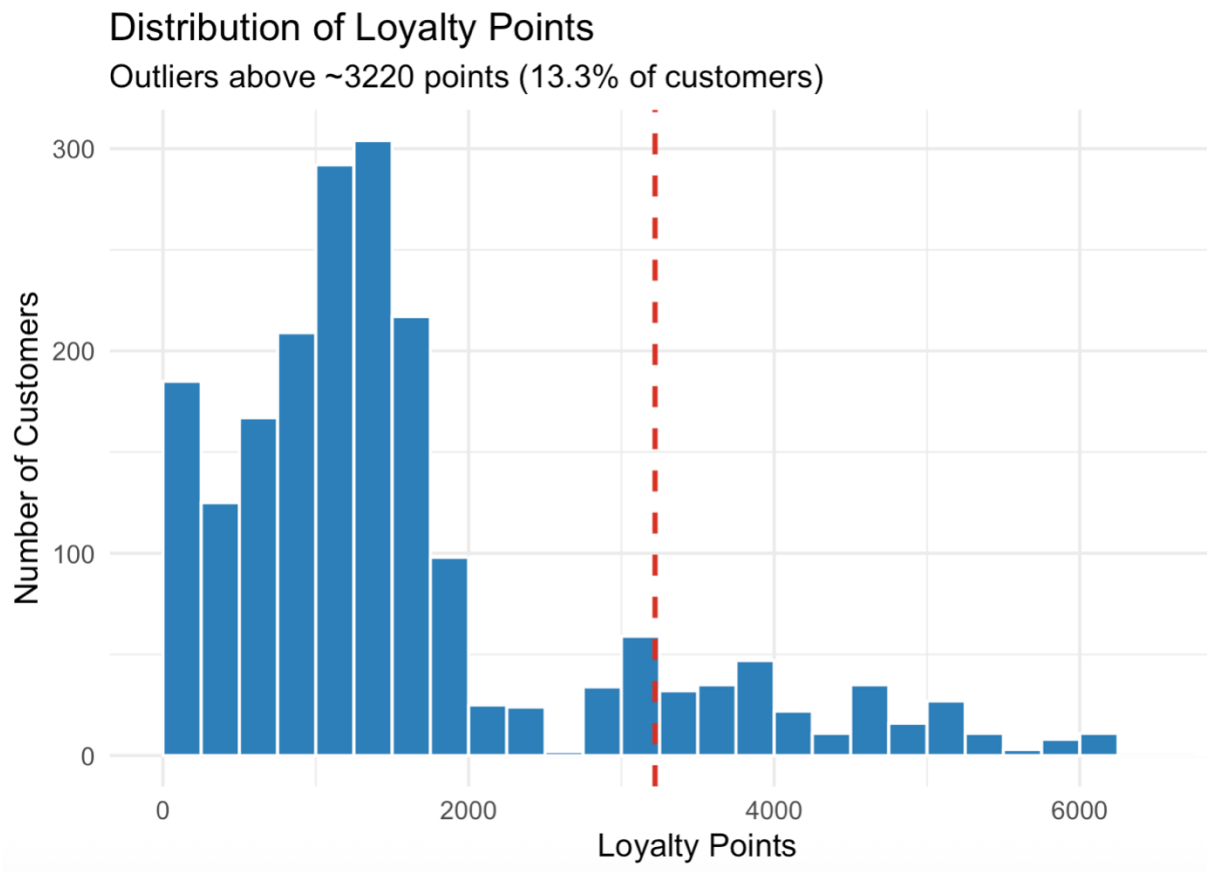
Appendix 2: Notebook Structure

Python Notebook Structure	R File Structure
1. Import, sense check, descriptive stats 2. Linear Regression (Spending, Remuneration, Age) 3. Multiple Linear Regression (Rem + Spend) 4. Decision Tree (Depths 2–6) 5. Feature Importance 6. Clustering (K-means & Hierarchical) 7. NLP Sentiment (reviews & summaries) 8. Product Analysis (sales + sentiment)	1. Import and data check 2. Exploratory graphs (ggplot2) 3. Loyalty points distribution 4. Correlation matrix 5. Multiple Linear Regression

Appendix 3: Data Preparation Details (Python)

- Imported CSV file into Python using pandas and NumPy.
- Conducted a sense check:
    - Checked df.info(), df.describe() and .isnull().sum() to confirm no missing values.
    - Reviewed ranges of continuous fields (e.g., loyalty points, remuneration, spending score).
- Duplicates:
    - Detected 39 full row duplicates (same review + same summary).
    - Retained these deliberately because repeated short reviews such as *"five stars"* likely represent genuine responses from different customers.
    - In NLP, repetition carries information about the frequency and popularity of expressions. Removing duplicates would bias the analysis towards more unique but less representative comments.
- Dropped irrelevant columns (language, platform) that added no analytical value.
- Renamed columns for clarity (e.g., Income → remuneration, Spending → spending_score, Points → loyalty_points).
- Saved the cleaned dataset as turtle_reviews_clean, which became the dataset used throughout analysis.
- Fixed random_state = 42 across models and clustering to ensure reproducibility.

Appendix 4: Loyalty Points Distribution



Methodology:

- The loyalty_points variable was plotted in Python using histograms and boxplots (matplotlib and seaborn).
- The red dashed line marks the statistical cut-off for outliers, calculated as: Q3 +1.5 × IQR, where Q3 is the 75th percentile of loyalty points and IQR = Q3 − Q1.
- This threshold was computed at approximately 3,220 points**,** which separates typical customers from statistical outliers.
- Outlier detection was not used to exclude data, but to highlight the distributional skewness and identify high-value customers.

Key Observations:

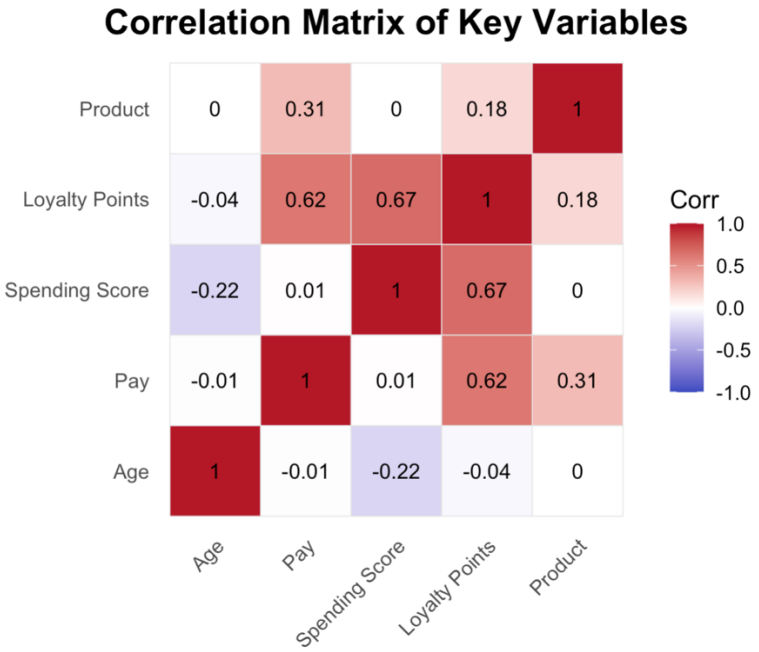
- Main cluster (0-2,000 points): Most customers fall in this range, showing typical engagement with the loyalty scheme.
- Outlier threshold (~3,220 points): Defined using the statistical rule above.
- High-value segment (≥ 3,220 points): Around 13% of customers exceed the outlier cut-off, accumulating unusually high loyalty balances.
- Long tail of low balances**:** A large proportion of customers have loyalty points near zero, indicating low engagement with the program.

Business Implications:

- The top 13% outlier group are the most loyal, high-value customers. They are rare but disproportionately important for retention and targeted VIP rewards.
- The long tail of low loyalty balances represents a growth opportunity - customers could be nudged into higher engagement through campaigns, bonus multipliers, or referral incentives**.**
- The skewed distribution confirms that the current loyalty program is strongly tilted towards rewarding a small subset of financially strong customers. This may be effective for retention but risks excluding broader demographics unless inclusivity measures are added.

Appendix 5: Correlation Matrix of Key Variables

A correlation matrix was generated to explore the linear relationships between loyalty points, spending score, pay (income), product, and age. Correlation values (r) range from -1 (perfect negative) to +1 (perfect positive).



Key Observations:

Loyalty Points

- Spending Score (r = 0.67): Strongest correlation. Customers who spend more accumulate significantly more loyalty points.
- Pay (r = 0.62): Also strongly correlated. Higher earners accumulate more points, indicating the scheme disproportionately rewards income levels.
- Product (r = 0.18): Weak correlation. The type of product purchased has only a minor influence on loyalty accumulation.
- Age (r = -0.04): No meaningful relationship. Older vs younger customers earn similar loyalty points once spending and income are accounted for.

Spending Score

- Loyalty Points (r = 0.67): Strongly positive, confirming spending behaviour is the main driver of loyalty accumulation.
- Age (r = -0.22): Weak negative correlation. Younger customers tend to spend slightly more than older customers.
- Pay (r = 0.01): Essentially no correlation. Income levels do not translate directly into higher spending behaviour.
- Product (r = 0.00): No relationship - product type does not explain differences in spending score.

Pay (Income)

- Loyalty Points (r = 0.62): Higher incomes strongly associated with higher loyalty balances.
- Product (r = 0.31): Moderate correlation. Certain products are more commonly bought by higher earners.
- Spending Score (r = 0.01): No meaningful correlation, indicating disposable income alone does not dictate spending behaviour.
- Age (r = -0.01): No relationship between income and age in this dataset.

Age

- Spending Score (r = -0.22): Slightly younger customers tend to spend more.
- Loyalty Points (r = -0.04): No meaningful effect on loyalty.
- Pay (r = -0.01): No meaningful effect on income.
- Product (r = 0.00): No correlation.

Product

- Pay (r = 0.31): Moderate correlation with income, suggesting some products appeal more to higher earners.
- Loyalty Points (r = 0.18): Weak positive link, implying product type only slightly affects loyalty outcomes.
- Spending Score (r = 0.00): No relationship with spending score.
- Age (r = 0.00): No effect of age on product selection.

Summary:

- Financial behaviour (spending score and income) are the two dominant drivers of loyalty outcomes.
- Demographic factors like age, product, gender and education add little explanatory power
- This justifies the reduced MLR model, which focused only on spending score and remuneration.

Appendix 6: Simple Linear Regression (Python, OLS)

Process

- Linear regression was applied in Python to test whether individual customer attributes could predict loyalty point accumulation.
- Implemented using statsmodels OLS (statsmodels.api.OLS).
- **Three** simple linear regressions were run:
    1. Spending Score → Loyalty Points
    2. Remuneration (Income) → Loyalty Points
    3. Age → Loyalty Points

Each model produced coefficients, R^2 values, standard errors, and p-values. Predictive validity was assessed using a 70/30 train/test split to calculate out-of-sample metrics (R^2, RMSE, MAE) (*Kuhn & Johnson, 2013)*.

OLS Selection

I chose Ordinary Least Squares (OLS) regression via statsmodels because:

- Consistency with teaching**:** Course demonstrations used OLS, aligning my work with the taught approach.
- Interpretability**:** OLS provides inference statistics (coefficients, p-values, confidence intervals), making results transparent for business application.
- Mathematical equivalence**:** OLS and sklearn.LinearRegression solve the same least-squares optimisation, but OLS outputs are richer for analysis.
- Predictive evaluation**:** Although OLS does not have built-in validation, I manually applied a 70/30 train-test split to test generalisability. This produced error metrics (RMSE, MAE) rele*vant to Turtle Games' goal of predicting future outcomes (*Kuhn & Johnson, 2013)*.*

Variable Selection Rationale

- Spending Score**:** Chosen due to strongest correlation with loyalty points in the correlation matrix (Pearson's r = 0.67). This indicated customer spending behaviour was strongly related to loyalty accumulation.
- Remuneration (Income)**:** Second strongest correlation with loyalty points (r = 0.62 from correlation matrix). High earners structurally accumulate more points.
- Age**:** Very weak correlation (r = -0.04). Tested in regression to confirm non-significance.
- Gender/Education**:** Dropped early, as feature importance analysis and correlation heatmaps showed no explanatory contribution.
- Product**:** Not included because only product codes were available (no descriptive categories).

Train/Test Split

- Split ratio: 70% training, 30% testing.
- Purpose: Ensure results generalised beyond the training dataset and identify overfitting.
- Metrics calculated on test set:
    - Spending Score → Loyalty Points: $R^2 \approx 0.45$, RMSE ≈ 934, MAE ≈ 652.
    - Remuneration → Loyalty Points**:** $R^2 \approx 0.32$, RMSE ≈ 1048, MAE ≈ 737.
    - Age → Loyalty Points**:** $R^2 \approx 0.002$, RMSE ≈ 1271, MAE ≈ 901.

Value of Validation

- Demonstrated that spending score was the most reliable standalone predictor (R^2 = 0.45 on test data).
- Showed remuneration was significant but weaker, while age added noise.
- Confirmed generalisation gap: all models performed worse on test data than training data, but spending score remained the strongest out-of-sample predictor.
- This reflects a business-relevant truth: Turtle Games can estimate loyalty behaviour early from spending patterns, but not from demographics.

Summary

- Initial correlations (from the heatmap) guided predictor choice (spending r = 0.67, income r = 0.62, age r = -0.04).
- OLS regression quantified these relationships and confirmed statistical significance.

- Train/test splits validated that spending score is the best single predictor of loyalty points**,** while age, gender, and education are negligible.
- This laid the groundwork for the reduced multiple linear regression (MLR) model, which combined spending score and remuneration for stronger predictive accuracy.

Appendix 7: Multiple Linear Regression

Why MLR Was Performed in Both Python & R

- Cross-Validation of Results: Running the same model in two environments (Python and R) reduces the risk of tool-specific errors. If both implementations return consistent coefficients, significance levels, and fit statistics, it strengthens confidence in the findings.
- Complementary Strengths:
    - Python (statsmodels, scikit-learn) is strong for rapid model building, integration with machine learning, and feature importance checks.
    - R has a long tradition of statistical modelling, offering advanced regression diagnostics (assumption checks, residual plots, VIF, Durbin-Watson, etc.) and high-quality visualisation through ggplot2.
- Reproducibility and Transparency: Using both ensured that the analysis could be reproduced in multiple programming environments. This is particularly valuable in a business setting where different teams may favour different tools.
- Pedagogical Value: Performing MLR in both platforms also clarified the underlying mechanics of regression. Seeing the same relationships play out across two languages reinforced understanding and highlighted the robustness of remuneration and spending as predictors.

Feature Selection

- Started by reviewing all available predictors (remuneration, spending score, age, gender, education, product).
- Used exploratory correlations and feature importance scores (from tree-based models) to identify remuneration and spending score as the two strongest predictors of loyalty points.
- Excluded age, gender, and education due to negligible explanatory value.
- Product codes excluded as categorical identifiers without interpretable meaning.

Model Fitting (Python & R)

- Built regression models using both Python (statsmodels.OLS) and R (lm()) to cross-check consistency.
- Dependent variable: loyalty_points.
- Independent variables: remuneration, spending_score.
- Intercept retained for model fit, though not interpreted.

Validation

- Applied a 70/30 train-test split using train_test_split in Python (random_state = 42 for reproducibility).
- Training data used to fit the model; test data held out for evaluation.
- Performance evaluated with RMSE, MAE, and R^2 on the test set.

Assumption Checking

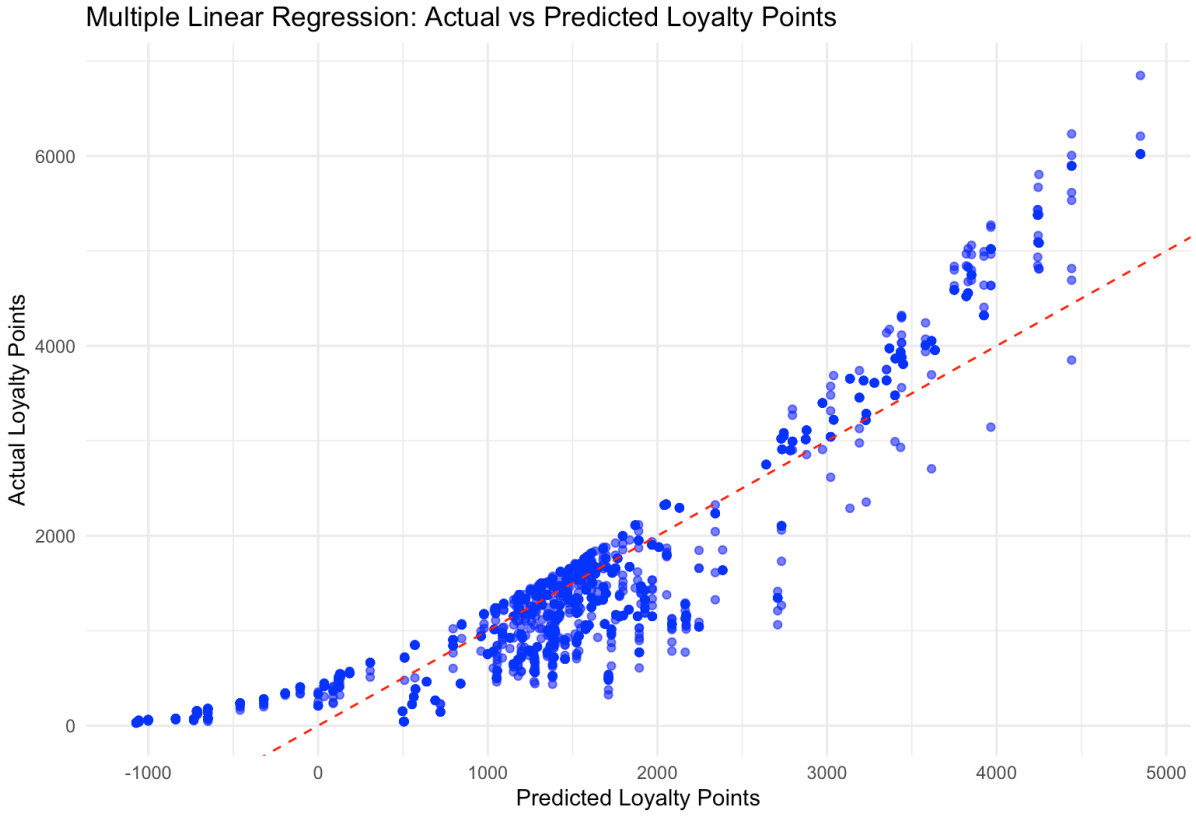
- Linearity: visually checked scatterplots of predictors vs loyalty points to confirm linear trends.
- Normality of residuals**:** generated Q-Q plot of residuals.
- Homoscedasticity: inspected scale-location plot of residuals.

- Multicollinearity: calculated Variance Inflation Factor (VIF) for both predictors.

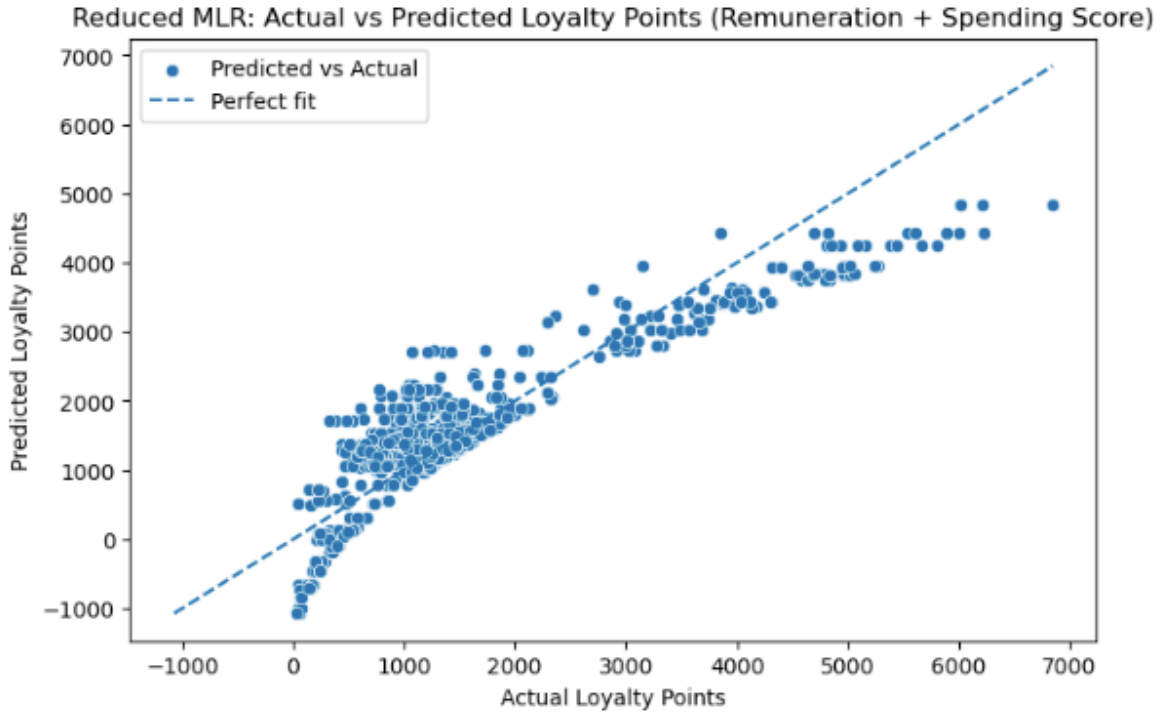
Interpretability

- Plotted actual vs predicted values for test set, overlaying 45° line to visually assess model fit.
- Extracted coefficients for each predictor to show direction and relative contribution.
- Used residual analysis to highlight areas of systematic under/over-prediction.

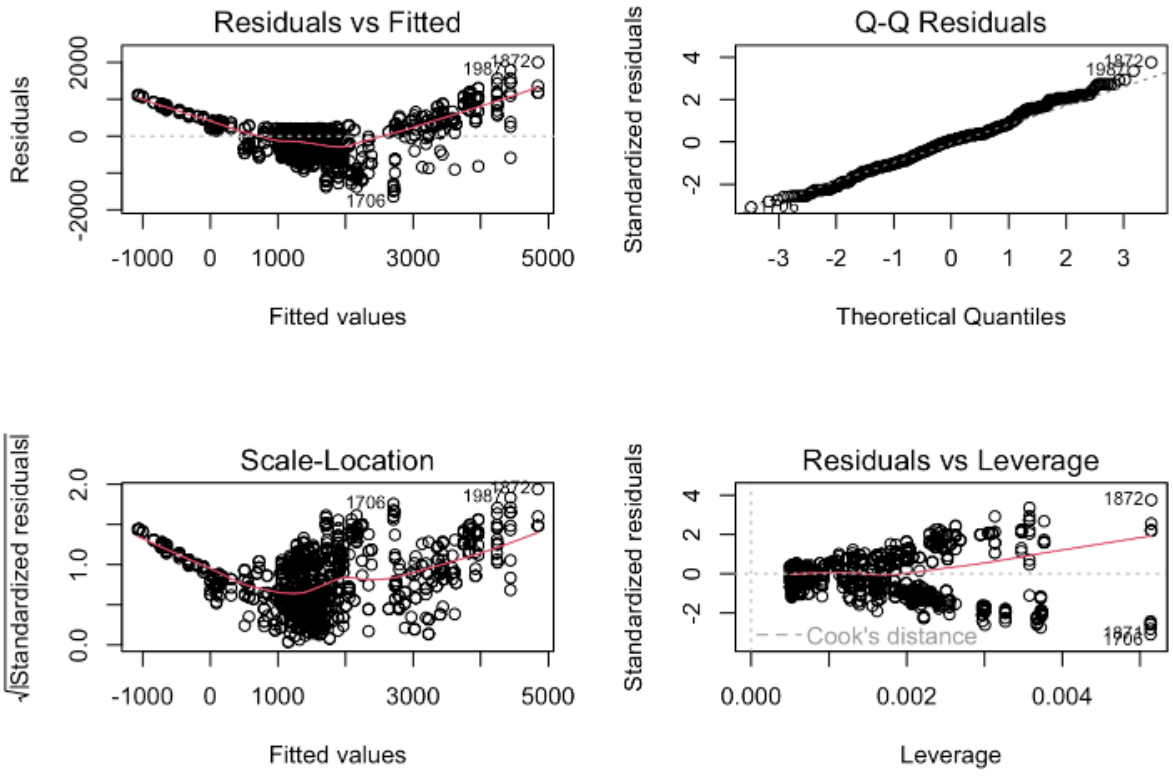
MLR Plot in R:



MLR Plot in Python:



Regression Diagnostics: MLR in R



To ensure the MLR (spending score + remuneration → loyalty points) met its key assumptions, four diagnostic plots were generated:

Residuals vs Fitted

- Purpose: Tests the assumption of linearity. Residuals should be randomly distributed around zero.
- Evaluation: The plot shows residuals curving at the lower and higher fitted values, suggesting some mild non-linearity. However, the majority of values cluster around zero, meaning the linear model is still broadly appropriate.

Q-Q Residuals

- Purpose: Assesses normality of residuals. If residuals are normally distributed, points should follow the 45° line.
- Evaluation: Most residuals lie close to the diagonal, with only minor deviations at the tails. With a large sample size (n = 2000), these small departures are not severe, and the normality assumption can be considered reasonably satisfied.

Scale-Location (Spread vs Fitted)

- Purpose: Checks homoscedasticity (constant variance). Ideally, the red smoothing line should be flat and the spread consistent across fitted values.
- Evaluation: The plot indicates variance increases for higher fitted values, showing mild heteroscedasticity. While not severe enough to invalidate results, it suggests predictions at the high end of loyalty points are less reliable.

Residuals vs Leverage (Cook's Distance)

- Purpose: Identifies influential cases that may disproportionately affect the regression model.
- Evaluation: Most observations fall within acceptable ranges, but a small number of high-leverage points (e.g., ID 18720) exceed the threshold lines. These are outliers with greater influence, but they do not dominate the overall model fit.

Conclusion:
The reduced MLR model broadly satisfies its assumptions. Some mild non-linearity, heteroscedasticity, and influential points are present, but none are severe enough to undermine the model's validity. With $R^2 \approx 0.83$ and low prediction error, the model is both robust and interpretable for business use.

Appendix 8: Decision Tree Regression

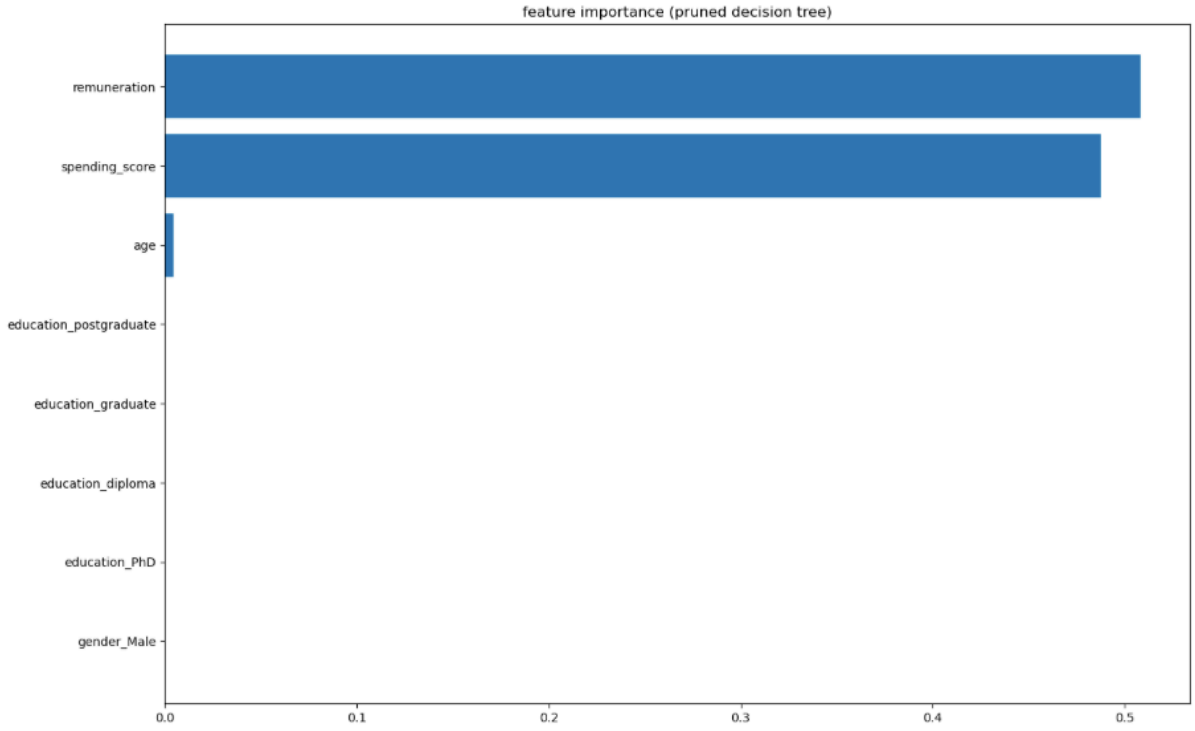
Model Performance (Depths 2-6)

- **Unpruned tree:** $R^2 = 0.995$, RMSE ≈ 90, MAE ≈ 33 → near-perfect fit, but clear overfitting.
- **Depth 2:** $R^2 = 0.83$, RMSE = 522, MAE = 377 → similar to MLR, limited improvement.
- **Depth 3 (chosen):** $R^2 = 0.915$, RMSE = 371, MAE = 267 → large error reduction vs Depth 2, interpretable rules, avoids overfitting.
- **Depth 4:** $R^2 = 0.94$, RMSE = 309, MAE = 217 → higher accuracy, but interpretability declines.
- **Depth 5-6:** R^2 up to 0.97, RMSE ≈ 203, MAE ≈ 138 → marginal gains but based on very small terminal nodes (sometimes $n < 20$), unstable and overfitted.

Feature Importance

- **Remuneration:** 0.51
- **Spending Score:** 0.49
- **Age:** 0.004 (negligible)
- **Gender, Education:** 0.0 (no predictive contribution)

Confirms loyalty points are driven entirely by **financial behaviour**, not demographic attributes.



Interpreting the Depth 3 Tree

- **First split:** Spending score (most important driver).
- **Second split:** Remuneration refines predictions within each spending group.
- **Examples:**
  - Low spend (≤ 15.5) + low remuneration (≤ 68.9) → ~155 loyalty points (n = 155).
  - High spend (> 67) + high remuneration (> 74) → ~4,849 loyalty points (n = 88).
- These rules make financial behaviour the dominant determinant of loyalty outcomes.

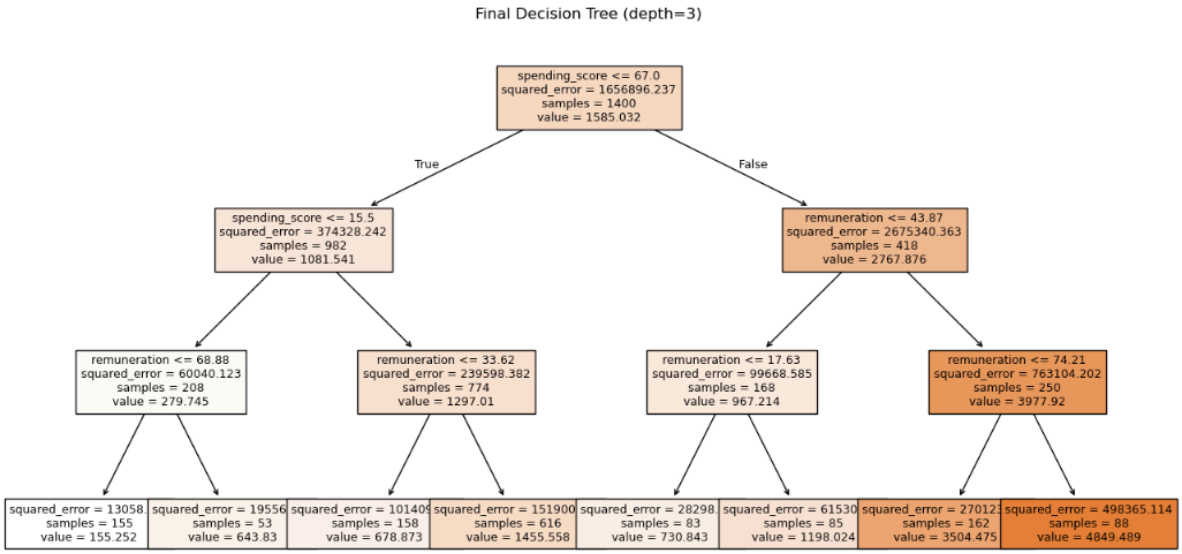
Assumption / Limitation Considerations - Decision Trees (Depths 2-6)

- **Linearity / Normality / Homoscedasticity:** Not required; trees handle non-linear patterns and skewed distributions naturally.
- **Overfitting:** Major risk at higher depths. By Depth 5-6, terminal nodes shrink to very small samples (sometimes fewer than 20 observations), making predictions unstable and non-representative *(*Breiman et al, 1984).
- **Interpretability:** Depth 2 = simple but not useful; Depth 3 = strong balance; beyond Depth 4 = difficult to interpret.

- **Stability:** Deep trees are sensitive to small data changes (splits can shift with one extra observation). Shallow trees (Depth 2-3) are more robust.

Appendix Conclusion:
Decision trees don't rely on regression assumptions (Montgomery et al, (2021), but their validity depends on controlling depth to avoid overfitting. Depth 3 provided the best trade-off: high predictive accuracy ($R^2 = 0.92$), reduced errors, interpretable structure, and node sizes that remain representative (hundreds rather than a handful of cases) *(*Breiman et al, 1984).

Final Tree (pruned to 3):



Appendix 9: Comparative Results of Regression and Decision Tree Models

Model	R^2 (full)	Correlation (r)	Intercept	Slope(s)	Std. Error(s)	p-value(s)	Durbin-Watson	Test R^2	RMSE (test)	MAE (test)	Notes
Simple Linear (Spending Score + Loyalty)	0.45	0.67	-75.05	33.06	(45.93, 0.81)	(0.102, <0.001)	1.19	0.45	934	652	Spending score highly predictive (p<0.001)
Simple Linear (Remuneration + Loyalty)	0.380	0.62	-65.69	34.19	(52.17, 0.98)	(0.208, <0.001)	3.62	0.32	1048	737	Remuneration is significant predictor, but weaker fit
Simple Linear (Age + Loyalty)	0.002	-0.045	1736.52	-4.01	(88.25, 2.11)	(0.000, 0.058)	2.28	0.002	1271	901	Age has no predictive value; slope not significant (p=0.058)
Multiple Linear (Remuneration + Spending Score)	0.827		-1700.31	β_1 = 33.98 (rem), β_2 = 32.89 (spend)	(0.52, 0.46)	(<0.001, <0.001)	3.48	0.83	534	415	Strongest regression; both predictors significant
Tree (Depth 2)								0.83	522	377	Matches MLR
Tree (Depth 3)								0.92	371	267	Higher accuracy
Tree (Depth 4)								0.94	309	217	Overfitting risk starts
Tree (Depth 5)								0.96	252	178	Very high accuracy
Tree (Depth 6)								0.97	203	138	Almost perfect fit, overfitting

Appendix 10: Model Comparison - Results and Reasoning

Model	Notes (reasoning-driven)
Simple Linear (Spending Score → Loyalty)	Explains 45% of loyalty variation ($R^2 = 0.45$) with a strong positive slope (~33 points per unit, $p<0.001$). However, predictive error remains high (RMSE = 934, MAE = 652), and residuals show mild autocorrelation (DW = 1.19). This indicates spending is a key driver but not sufficient alone to capture loyalty outcomes.
Simple Linear (Remuneration → Loyalty)	Explains less variation ($R^2 = 0.38$) than spending, despite a similar slope (~34 points, $p<0.001$). Predictive accuracy is weaker (RMSE = 1048, MAE = 737), and residuals show negative autocorrelation (DW = 3.62), suggesting systematic misfit. Remuneration matters but is less reliable in isolation.
Simple Linear (Age → Loyalty)	Nearly no explanatory power ($R^2 = 0.002$). The slope is small (-4 points per year) and statistically non-significant ($p = 0.058$). Errors are largest here (RMSE = 1271, MAE = 901), confirming age adds noise without predictive value.
Multiple Linear (Remuneration + Spending Score)	Delivers a step change in performance: R^2 jumps to 0.83, meaning over 80% of loyalty variation is explained. Both predictors remain highly significant ($p<0.001$), proving they independently add value. Prediction error drops sharply (RMSE = 534, MAE = 415; MSE ≈ 285k), far lower than any single-variable model. Residuals are better behaved, though DW = 3.48 still suggests some negative autocorrelation. Overall, MLR combines interpretability with accuracy, making it the most robust and defensible choice.
Decision Tree (Depth 2)	Comparable to MLR in performance ($R^2 = 0.83$, RMSE = 522, MAE = 377). While it retains some interpretability, it doesn't provide a meaningful improvement over MLR — accuracy and error reduction plateau at this level. Essentially, Depth 2 adds complexity without significant predictive gains.
Decision Tree (Depth 3)	Delivers a substantial improvement over Depth 2 (R^2 rises from 0.83 → 0.92; RMSE falls from 522 → 371; MAE drops from 377 → 267). This shows the tree is capturing useful non-linear interactions between remuneration and spending. At the same time, it avoids the strong overfitting that occurs at Depths 4–6. Depth 3 therefore represents a "sweet spot": higher accuracy and lower errors than Depth 2, but more generalizable and trustworthy than deeper trees.
Decision Tree (Depth 4)	Further reduces error (RMSE = 309, MAE = 217) and increases fit ($R^2 = 0.94$), but overfitting begins to emerge. The model becomes harder to interpret, with splits tailored too closely to training data rather than generalizable patterns.
Decision Tree (Depth 5–6)	Achieve extremely high accuracy ($R^2 = 0.96$–0.97, RMSE as low as 203), but at the cost of generalization. Terminal nodes are very small, meaning the tree memorizes training patterns. Interpretability is lost, and the gains in error reduction are offset by overfitting risk.

Appendix 11: K-means Clustering - Detailed Results and Visualisations

Model Selection

- Elbow method: clear bend at k = 4-5.
- Silhouette scores:
  - k = 3 → 0.457 (clusters overlap).
  - k = 4 → 0.512 (better separation).
  - **k = 5 → 0.583 (highest, optimal).**
  - k = 6 → 0.563 (slightly worse, unnecessary complexity).

Decision: select **k = 5**.

Cluster Sizes (k = 5)

- Cluster 0 → 356 customers (18%)
- Cluster 1 → 774 customers (39%)
- Cluster 2 → 330 customers (17%)
- Cluster 3 → 269 customers (13%)
- Cluster 4 → 271 customers (14%)

Cluster Centres (remuneration, spending score)

- Cluster 0: (73, 82) → High income, high spending.
- Cluster 1: (44, 50) → Middle income, average spending.
- Cluster 2: (75, 17) → High income, low spending.
- Cluster 3: (20, 79) → Low income, high spending.
- Cluster 4: (20, 20) → Low income, low spending.

Insights

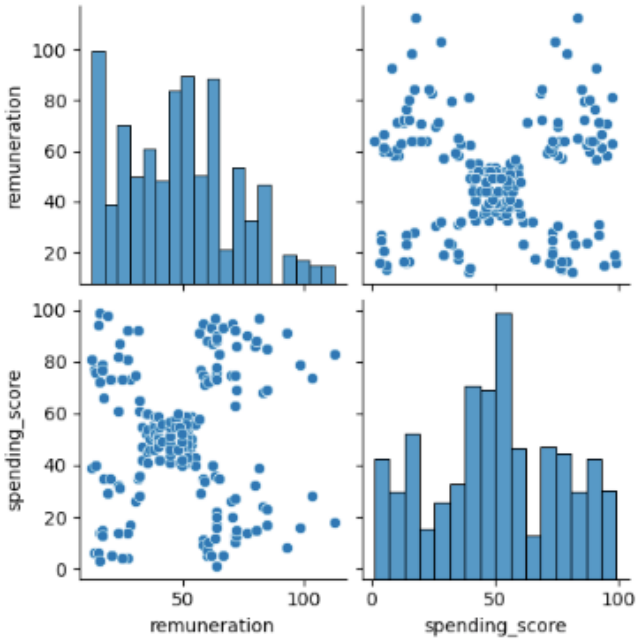
- Mainstream customers (Cluster 1) dominate the dataset.
- High-value premium customers (Cluster 0) require retention focus.
- High-income low spenders (Cluster 2) are untapped potential.
- Clusters 3 and 4 show different behaviours despite both being low income - one spends heavily, the other minimally.

Hierarchical Clustering - Validation

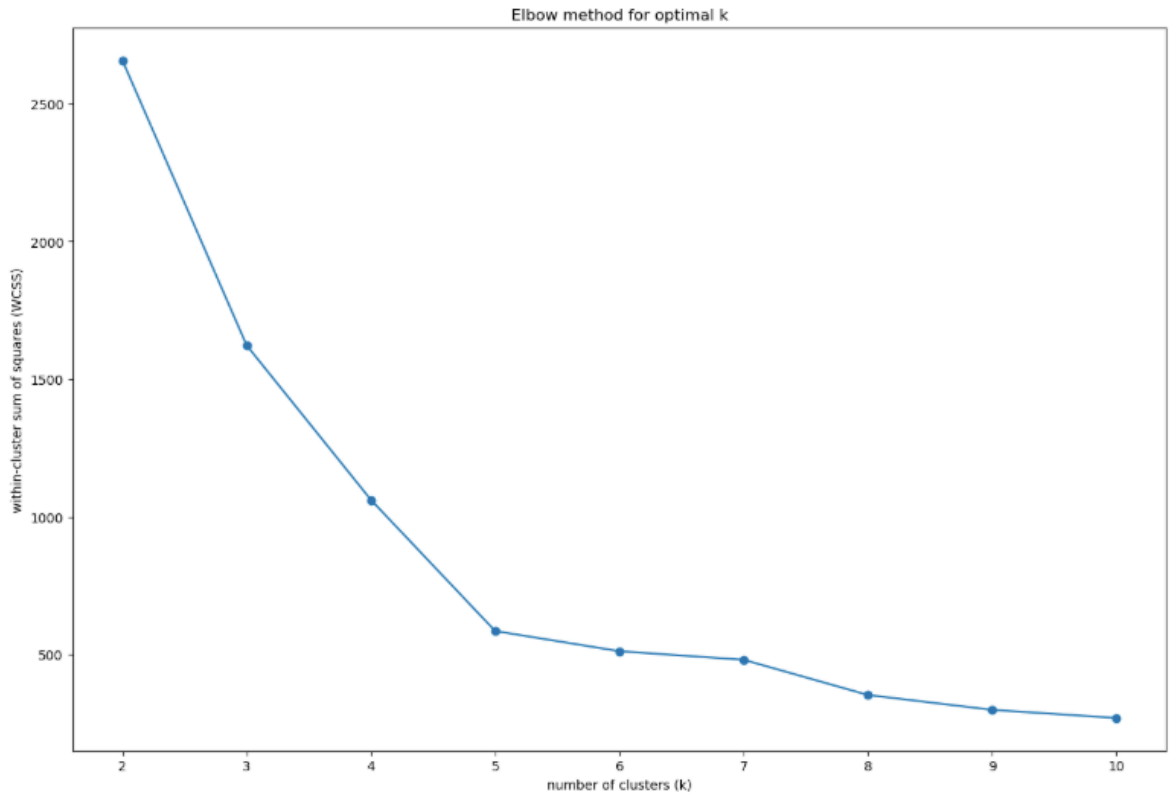
- Tested **Ward, Average, Complete, Single linkage**.
  - Ward + Euclidean performed best (compact, spherical clusters).
- **Silhouette scores (Ward linkage):**
  - k = 2 → 0.390
  - k = 3 → 0.467
  - k = 4 → 0.507
  - **k = 5 → 0.581 (peak)**
  - k = 6 → 0.563
- Cluster sizes at k = 5: one large group (~40%), four smaller groups (13-18% each).

Validation Conclusion:
Both **K-means and hierarchical clustering converge on k = 5** as the best solution. This strengthens confidence in the segmentation and ensures the clusters are not an artefact of one method.

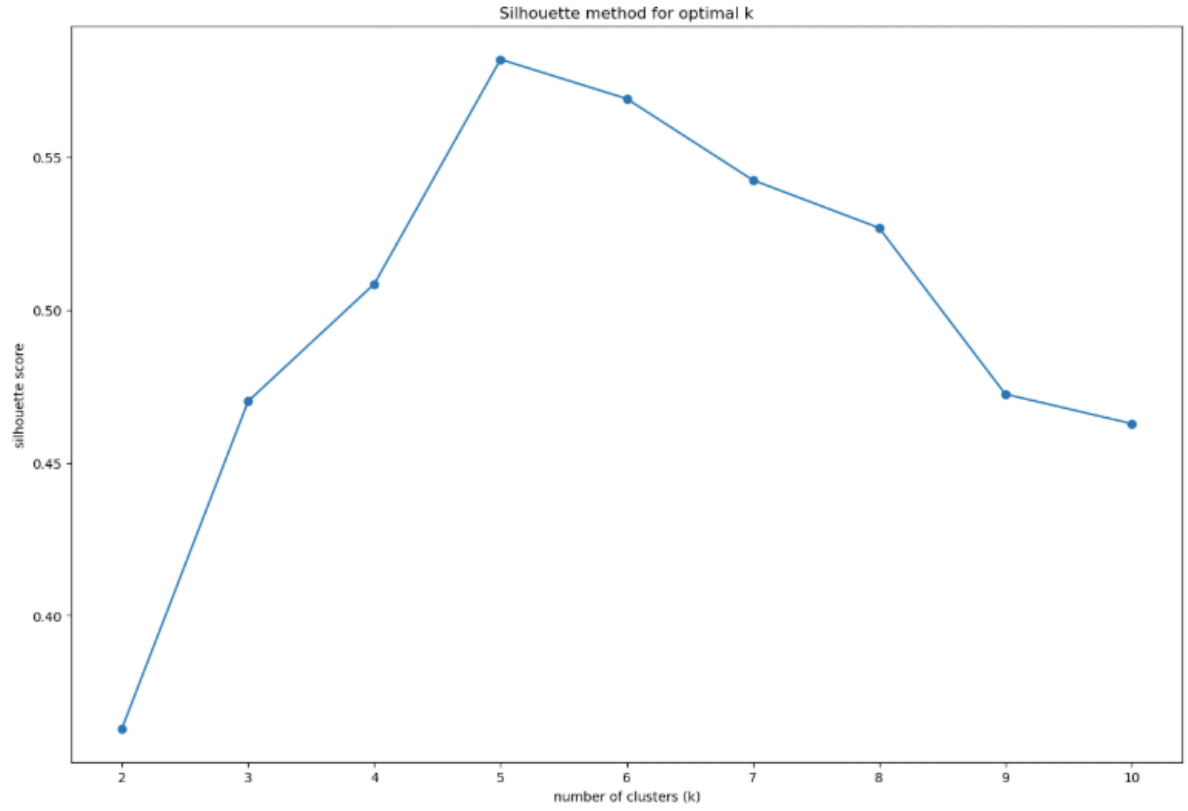
Pair Plot:



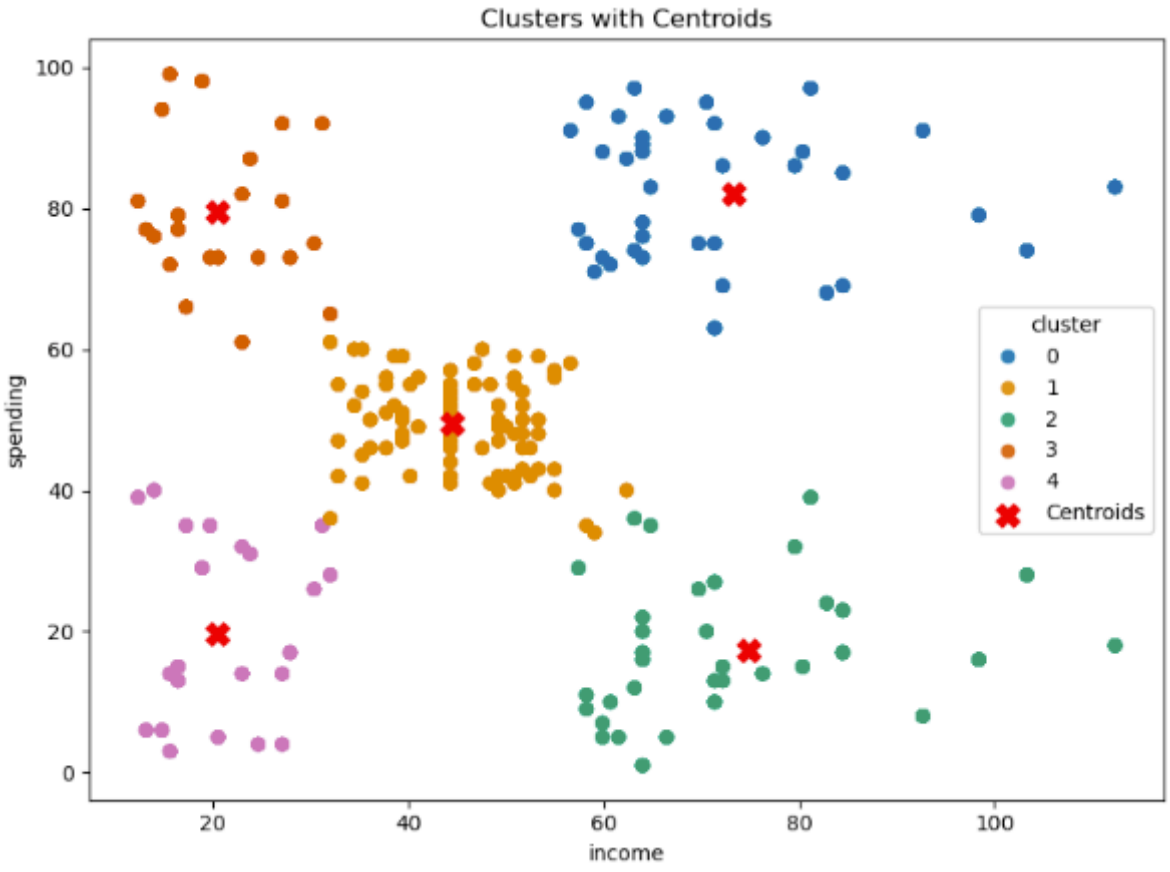
Elbow Method:



Silhouette Method:



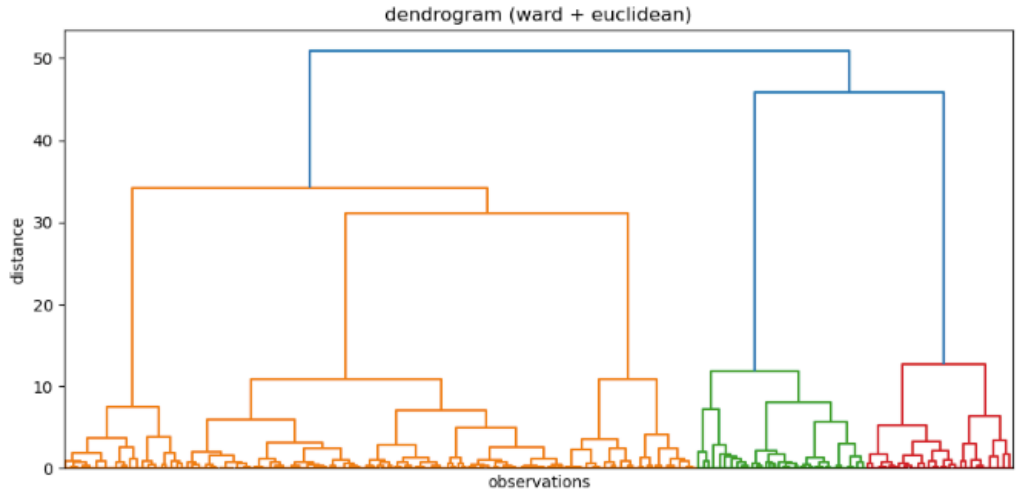
Final Cluster Model with Centroids:

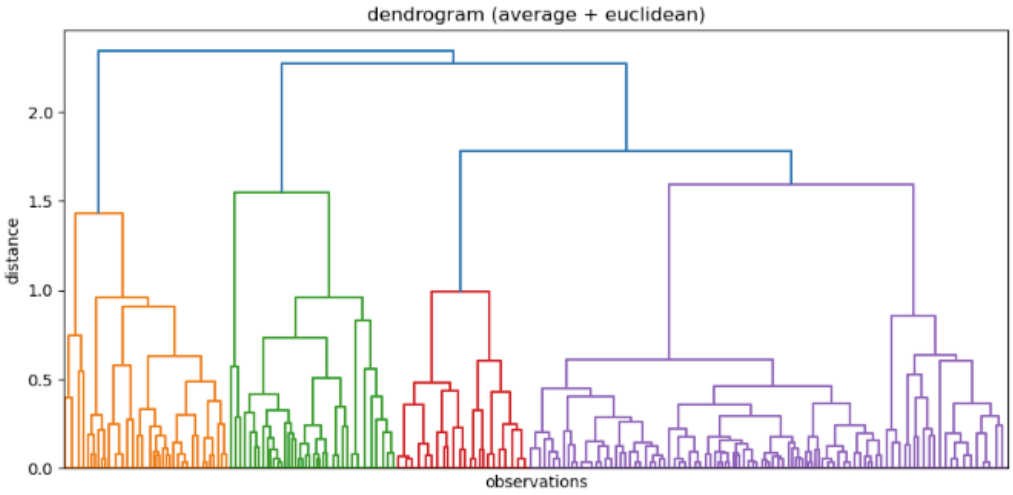


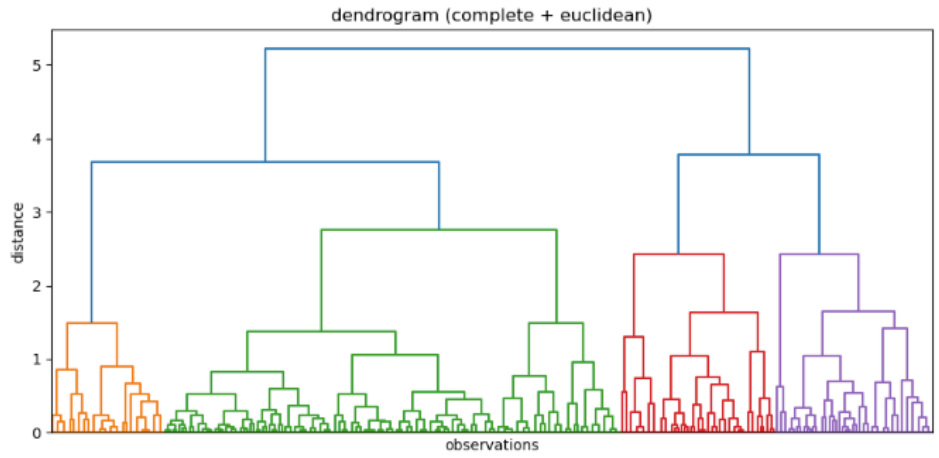
Cluster	% of Customers (n)	Profile	Characteristics	Business Action
0	18% (n=356)	**Premium**	High income (~73), high spending (~82).	Prioritise retention with loyalty perks, VIP programs, exclusivity.
1	39% (n=774)	**Mainstream**	Moderate income (~44), moderate spending (~50). Largest group.	General marketing campaigns, seasonal offers, upsell opportunities.
2	17% (n=330)	**Affluent but disengaged**	High income (~75), low spending (~17).	Targeted re-engagement: personalised recommendations, bundling.
3	13% (n=269)	**Value Seekers**	Low income (~20), high spending (~79). Spend heavily despite limited means.	Discounts, loyalty multipliers, community-driven campaigns.
4	14% (n=271)	**Budget-Conscious**	Low income (~20), low spending (~20).	Low-cost engagement, automated communications.

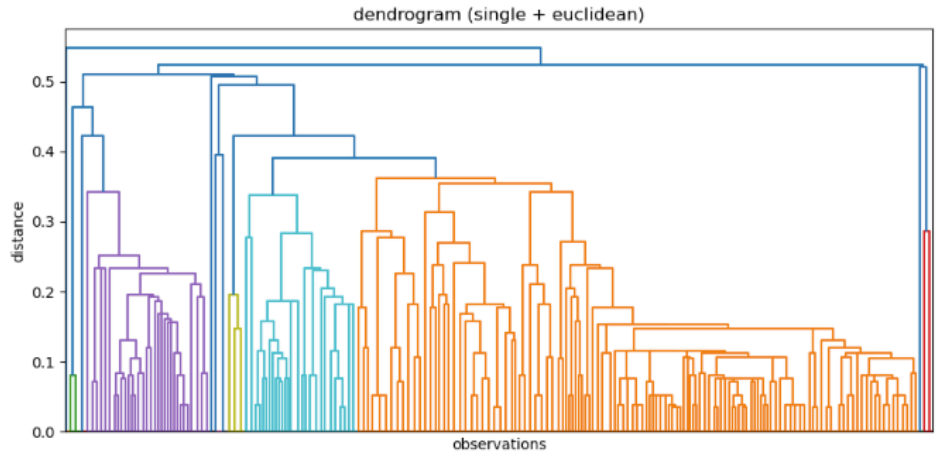
(Kotler & Keller, 2016)

Dendrograms (exploration of linkage methods)

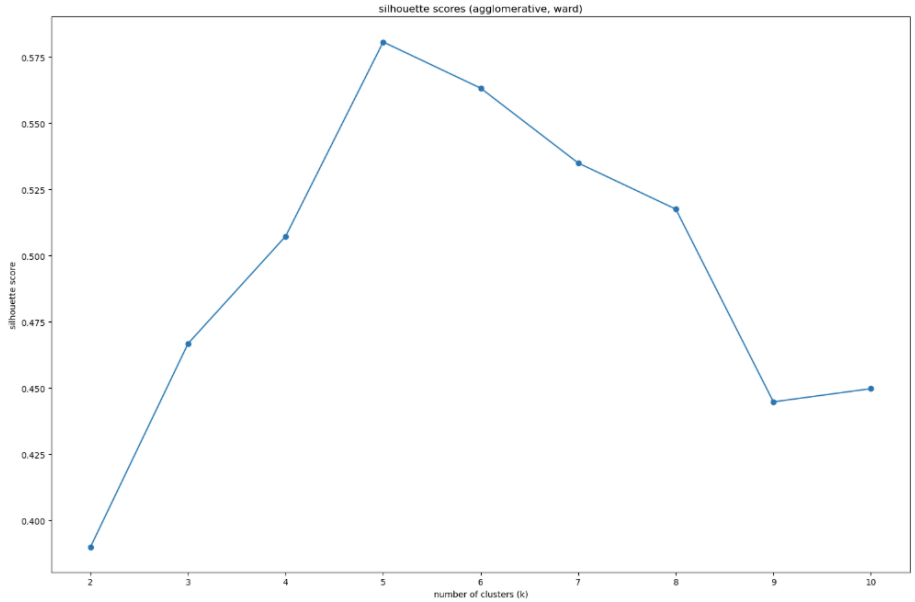



dendrogram (average + euclidean)
distance
2.0
1.5
1.0
0.5
0.0
observations


dendrogram (complete + euclidean)
distance
5
4
3
2
1
0
observations


dendrogram (single + euclidean)
distance
0.5
0.4
0.3
0.2
0.1
0.0
observations

Silhouette Scores (ward and agglomerative)



k=2: silhouette=0.390
k=3: silhouette=0.467
k=4: silhouette=0.507
k=5: silhouette=0.581
k=6: silhouette=0.563
k=7: silhouette=0.535
k=8: silhouette=0.518
k=9: silhouette=0.445
k=10: silhouette=0.450

Appendix 12: NLP & Sentiment Analysis (Python)

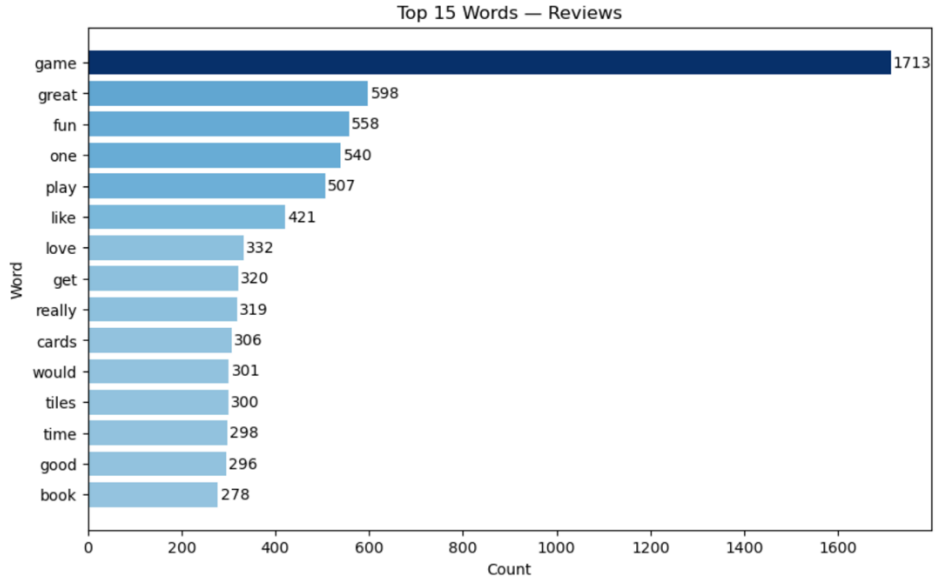
Data Preparation

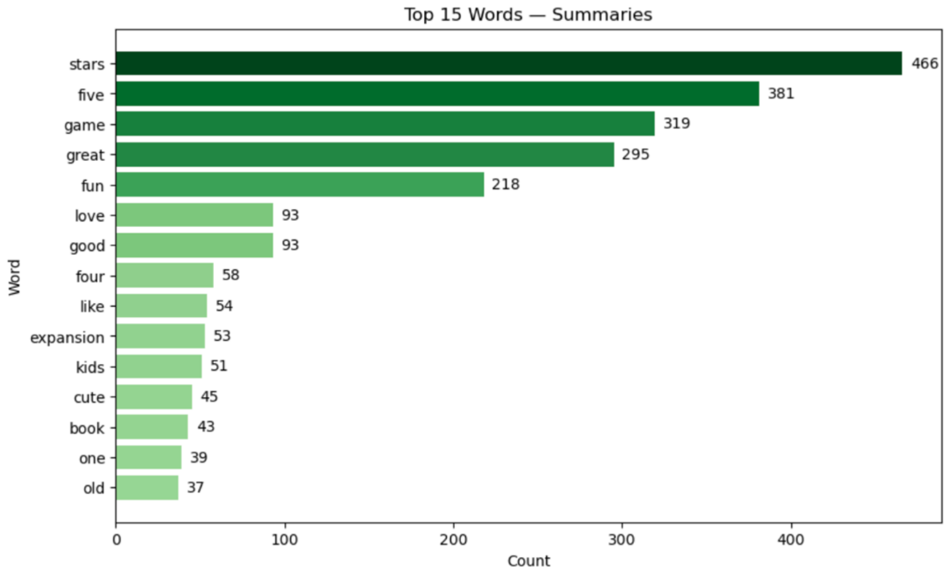
- Both reviews (long-form comments) and summaries (short snippets like *"Five stars"* or *"Great"*) were analysed.
- Preprocessing applied before analysis:
  - Converted all text to lowercase**.**
  - Removed punctuation and special characters**.**
  - Removed stopwords (e.g., "the," "and," "of") as they add noise but no sentiment value.
  - Performed tokenisation (splitting text into individual words).
- Duplicates: 39 duplicate rows were retained. These often consisted of short comments like "five stars," which may have been repeated by multiple customers. Keeping them preserved the frequency of these expressions, which itself is useful for measuring customer sentiment consistency.

Step 1: Word Frequency and Word Clouds

- Generated word clouds separately for reviews and summaries.
- Identified the 15 most frequent words in each and plotted them in bar charts with polarity values.
- Rationale:
  - Word frequency identifies the dominant customer themes.
  - Polarity adds a sentiment dimension, showing whether common words are used in a positive, neutral, or negative context.

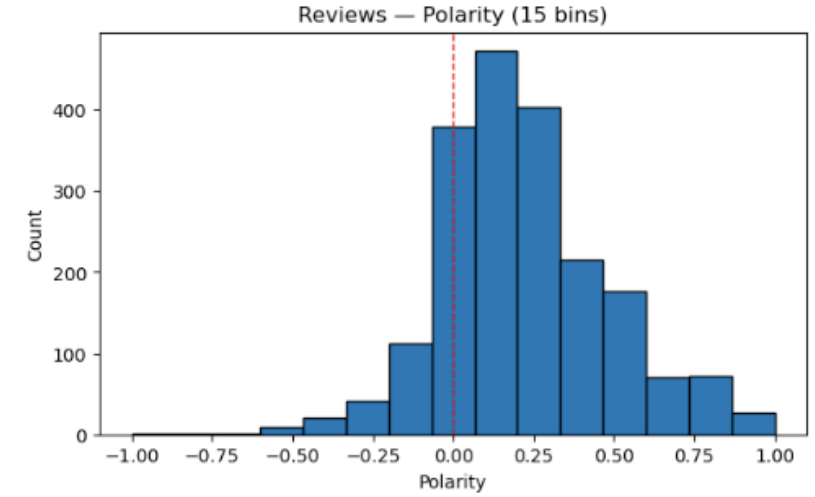
- Results:
  - Reviews: Richer terms like *game, play, fun, family, quality, time*.
  - Summaries: Overwhelmingly *five, stars, great, good*.
  - Confirms that reviews provide useful depth, while summaries are repetitive and shallow.

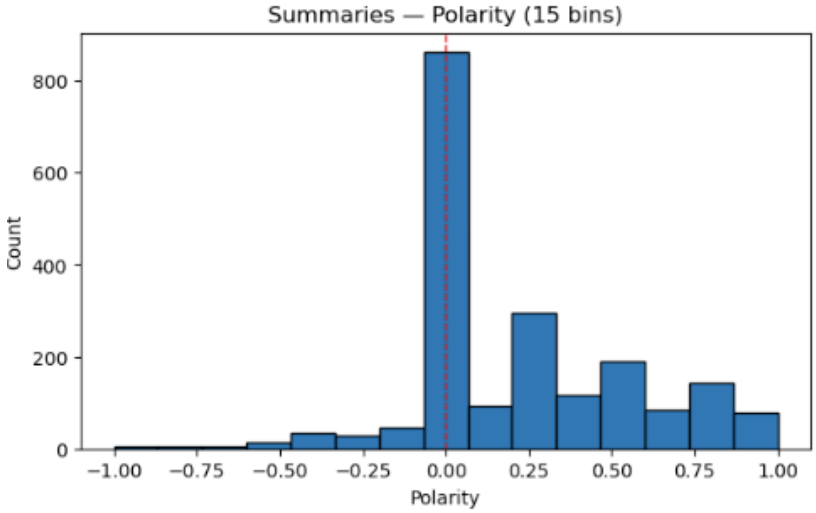




Step 2: Polarity (TextBlob)

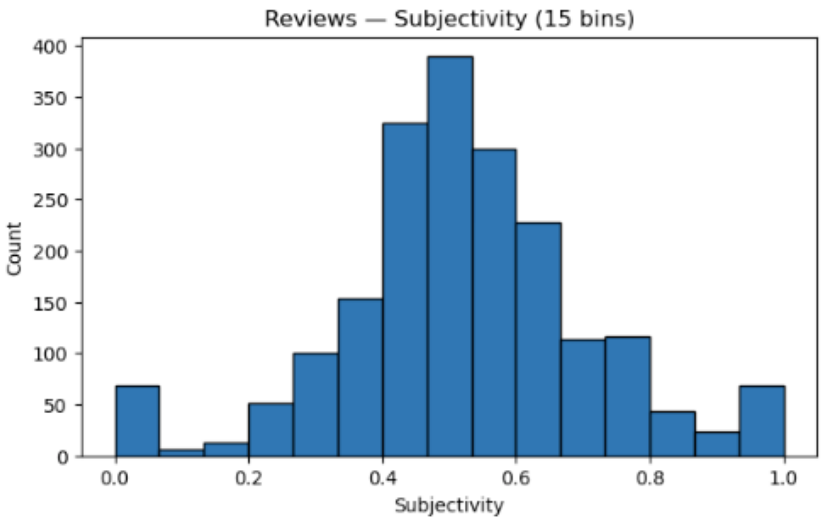
- Method: Used TextBlob to calculate polarity scores (-1 = negative, +1 = positive).
- Produced histograms (15 bins) for reviews and summaries.
- Why: Provides a simple, lexicon-based measure of emotional tone to establish a baseline (Loria, 2018).
- Results:
  - Reviews: Mean polarity = 0.217, Median = 0.18. Distribution spanned the full range -1 to +1, confirming that reviews captured both satisfied and dissatisfied customers.
  - Summaries: Mean = 0.220, but Median = 0.071 due to clustering at 0.0. This shows TextBlob treated many short summaries as neutral even when customers intended them as positive.

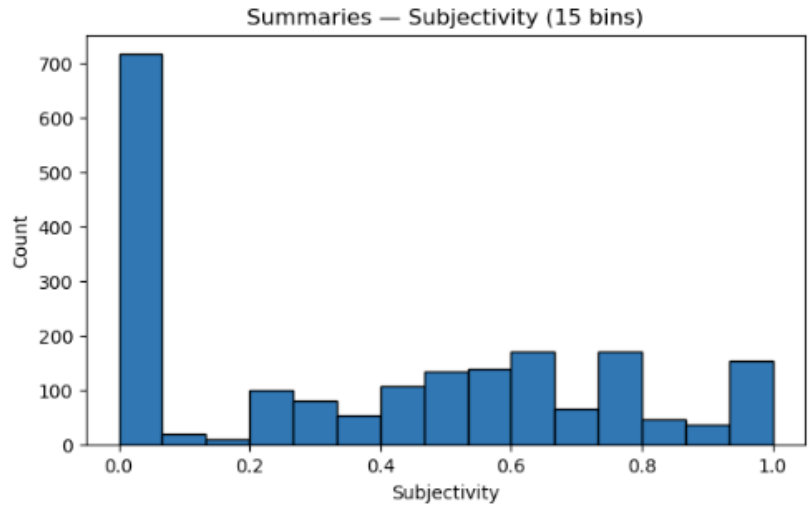




Step 3: Subjectivity (TextBlob)

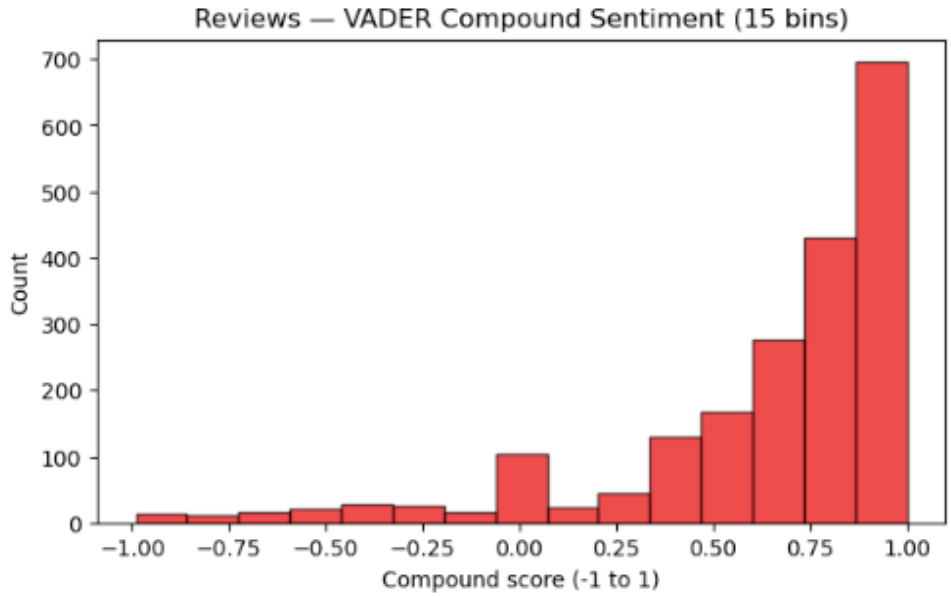
- Method: Calculated subjectivity scores (0 = factual, 1 = opinionated).
- Histograms (15 bins) created for both reviews and summaries.
- Why chosen: Helps distinguish between factual comments ("delivered on time") and opinion-rich comments ("amazing quality").
- Results:
    - Reviews: Average subjectivity = 0.518, clustered 0.4-0.6 → reviews were moderately subjective and therefore useful for insights.
    - Summaries: Average subjectivity = 0.379, with a large spike at 0.0. This reflects factual "labels" like *"five stars"*, offering little qualitative insight.

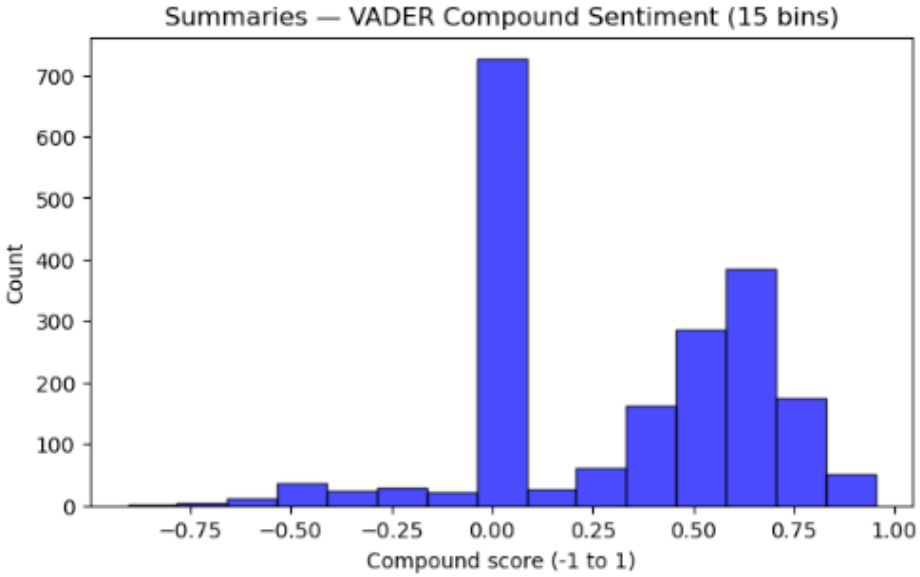




Step 4: Sentiment (VADER)

- Method: Applied VADER sentiment analyser, focusing on the compound score (-1 to +1).
- Histograms (15 bins) plotted for reviews and summaries.
- Why: VADER is tailored for short, informal text, making it suitable for customer comments. It captures intensity (e.g., “amazing!!!”), negation (“not bad”), and emojis, which TextBlob misses (Hutto & Gilbert, 2014)
- Results:
    - Reviews: Clear skew towards positive sentiment, majority clustered between 0.5-1.0. VADER also flagged strong negatives like *“money trap”*.
    - Summaries: Large spike around 0.0 (neutral). Many positive short forms (*“five stars”*) were misclassified as neutral because VADER does not recognise rating language without explicit sentiment words.





Step 5: Issue with Summaries Misclassification

- Observation: Both TextBlob and VADER misclassified short summaries such as *"five stars"* as neutral (0.0).
- Reason:
  - Neither model has "five stars" in its sentiment lexicon.
  - They rely on sentiment-bearing adjectives/adverbs. Since "five" and "stars" are not mapped to positivity, the phrase was treated as neutral.
- Impact:
  - Artificially inflated the neutral category in summaries.
  - Reduced calculated CSAT for summaries to 57.8%, compared with 88.7% for reviews.
- Mitigation Strategies:
  - Introduce explicit numerical rating system (1-5 stars) rather than free-text summaries.
  - Pair with guided review prompts (*"What did you like most?"*).

    - Extend lexicons to interpret phrases like *“five stars”* or *“10/10”* as highly positive.

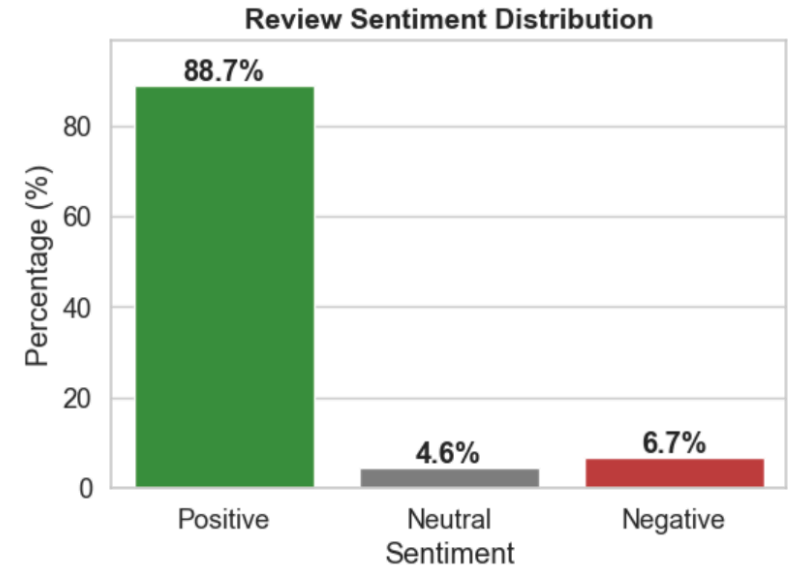
Step 6: Top 20 Positive and Negative Reviews

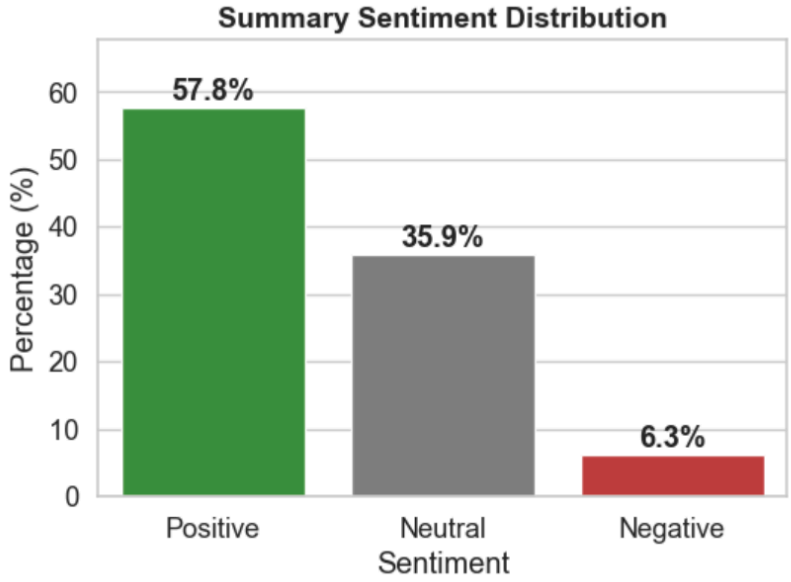
- Method: Extracted the 20 most positive and 20 most negative comments for both reviews and summaries using VADER compound scores.
- Why: Concrete examples highlight what customers value and where complaints cluster.
- Results:
    - Positive themes: Family fun, gameplay quality, enjoyment across age groups.
    - Negative themes: Product quality issues, misleading expectations, value-for-money complaints.
    - Summaries: Too short to add real depth; mainly repeated *“five stars”*.

These can be read in the notebook, not included here due to length.

Step 7: Sentiment Classification into Business KPIs

- Converted compound sentiment into percentages:
    - Positive: compound ≥ 0.05
    - Neutral: -0.05 < compound < 0.05
    - Negative: compound ≤ -0.05
- Aggregated into a barplot for business users.
- Results:
    - Review CSAT (Customer Satisfaction Score): 88.7%
    - Summary CSAT**:** 57.8%
    - Overall CSAT: 73.2%
- Interpretation: Reviews are a strong source of customer satisfaction signals, while summaries depress the score due to misclassification. The analysis also aligns with research on creating enduring customer value (Kumar & Reinartz, 2016).





Overall Insight

- Reviews provide rich, reliable, and nuanced feedback, suitable for sentiment analytics and actionable business insights.
- Summaries are shallow and often misleading, especially when algorithms misclassify shorthand expressions like *"five stars."*
- Recommendation: Turtle Games should prioritise structured ratings + guided reviews for future data collection, enabling both robust quantitative analysis (ratings/CSAT) and qualitative insight (reviews).

Appendix 13: Product Sales & Sentiment Proxy Analysis

Motivation & Approach

The dataset did not contain direct revenue or sales transaction data. Instead, I used spending_score as a proxy for product sales, since the metadata defined it as "a score between 1-100 assigned to customers based on spending behaviour." While imperfect, it provides a consistent, quantitative measure of relative sales performance across customers and products.

To complement this, I calculated average sentiment scores from customer reviews using the VADER compound sentiment score (range −1 to +1). This served as a proxy for customer satisfaction and reputation.

The aim was to combine financial (sales proxy) and qualitative (sentiment) measures to create a balanced framework for identifying products that are both profitable and well-regarded (or underperforming and at risk).

Why Reviews, Not Summaries: Although both reviews and summaries were available in the dataset, I based sentiment analysis exclusively on the full reviews:

- Reviews were longer and richer, providing stronger sentiment signals and context.
- Summaries were often short, generic, or neutral (e.g., "Five stars"), which reduced their discriminative power and caused misclassification.
- Reviews contained nuanced details (positive: "family fun for all ages"; negative: "misleading advertising"), making them more representative of actual satisfaction.

Thus, review sentiment is the more reliable input for assessing product performance.

Aggregation at Product Level
The analysis required moving from customer-level data to product-level metrics. I grouped the dataset by product code and calculated:

- Average spending_score as a proxy for sales.
- Average sentiment score (VADER compound) as a proxy for reputation.
- Number of reviews as a measure of reliability (plotted as bubble size).

This produced a structured dataset with one row per product, enabling comparison of products along financial and reputational dimensions.

Ranking Products Individually
Before combining sales and sentiment, I ranked products separately on each metric:

- Top 3 and Bottom 3 by Sales Proxy (average spending_score).
- Top 3 and Bottom 3 by Sentiment (average VADER compound).

This surfaced extremes in performance (best/worst products) and provided insight into which products performed consistently well across both measures versus those with discrepancies.

Top & Bottom Products by Sales and Sentiment

Category	Products (scores)
Top 3 by Sales	3153 (~67) 5510 (~66) 6466 (~65)
Bottom 3 by Sales	1031 (~31) 1175 (~29) 2173 (~21)
Top 3 by Sentiment	2874 (~0.86) 5429 (~0.85) 5493 (~0.84)
Bottom 3 by Sentiment	2253 (~0.27) 3165 (~0.26) 9597 (~0.24)

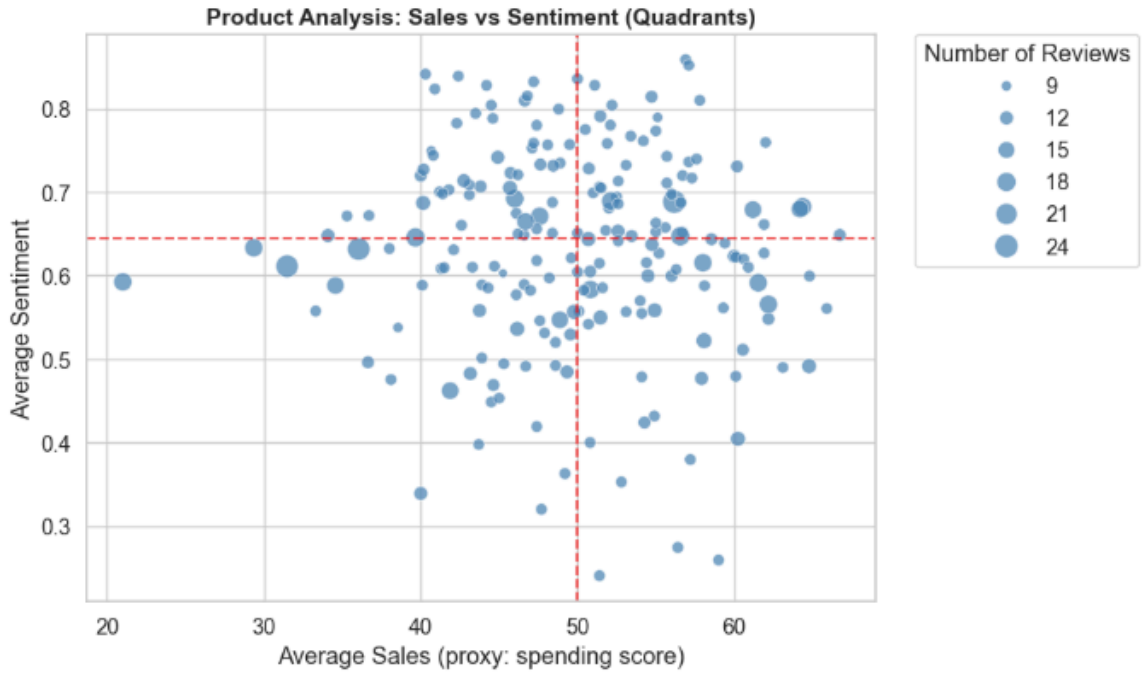
Scatterplot & Quadrant Analysis
To combine both measures visually, I plotted:

- X-axis: Average spending_score (proxy for sales).
- Y-axis: Average VADER sentiment (−1 to +1).
- Bubble size: Number of reviews per product.

Median values of both metrics were used as cut-offs to divide the plot into four quadrants:

- Top-Right (Cash Cows): High sales, high sentiment → maintain, protect, ensure supply reliability.
- Top-Left (Hidden Gems): Low sales, high sentiment → promote and expand distribution.
- Bottom-Right (Risks): High sales, low sentiment → address complaints, fix quality issues.
- Bottom-Left (Low Priority): Low sales, low sentiment → consider discontinuation or low-cost engagement only.

This quadrant framework allowed strategic interpretation.



Quadrant Interpretation: Sales vs Sentiment

Quadrant	Product Code (sales, sentiment)	Business Action
💰 **Cash Cows** (High Sales, High Sentiment)	3153 *(66.7, 0.65)* 4405 *(64.4, 0.68)* 5726 *(64.2, 0.68)*	Maintain investment, loyalty programs, ensure supply chain reliability
🌟 **Hidden Gems** (Low Sales, High Sentiment)	5493 *(40.3, 0.84)* 2371 *(42.4, 0.84)* 3427 *(47.2, 0.83)*	Increase promotion and distribution to scale successes
⚠️ **Risks** (High Sales, Low Sentiment)	5510 *(65.9, 0.56)* 6466 *(64.8, 0.60)* 2829 *(64.8, 0.49)*	Investigate quality issues, address complaints, improve perception
❌ **Low Priority** (Low Sales, Low Sentiment)	2173 *(21.0, 0.59)* 1175 *(29.4, 0.63)* 1031 *(31.5, 0.61)*	Consider rationalisation, low-cost engagement, or discontinuation

Interpretation & Business Value

- Cash Cows (e.g., Product 3153, 4405, 5726): Reliable performers. They sell strongly and are well-liked. They should be supported with loyalty programs and guaranteed supply.
- Hidden Gems (e.g., Product 5493, 2371, 3427): Loved by customers but low in sales. Marketing investment could unlock growth.
- Risks (e.g., Product 5510, 6466, 2829): These sell well but generate negative reviews. Urgent improvements are required.

- Low Priority (e.g., Product 2173, 1175, 1031): Weak across both measures. Rationalisation or deprioritising would release resources for higher-value products.

Why This Approach Is Appropriate

- Spending_score is a valid proxy for sales in the absence of actual revenue data.
- Full reviews provide higher-quality sentiment than short summaries.
- Combining both metrics creates a balanced view of financial and customer perception performance.
- Quadrant analysis translates complex product-level data into a business-friendly framework for decision-making.

Appendix 14: Further Exploratory Analysis in R

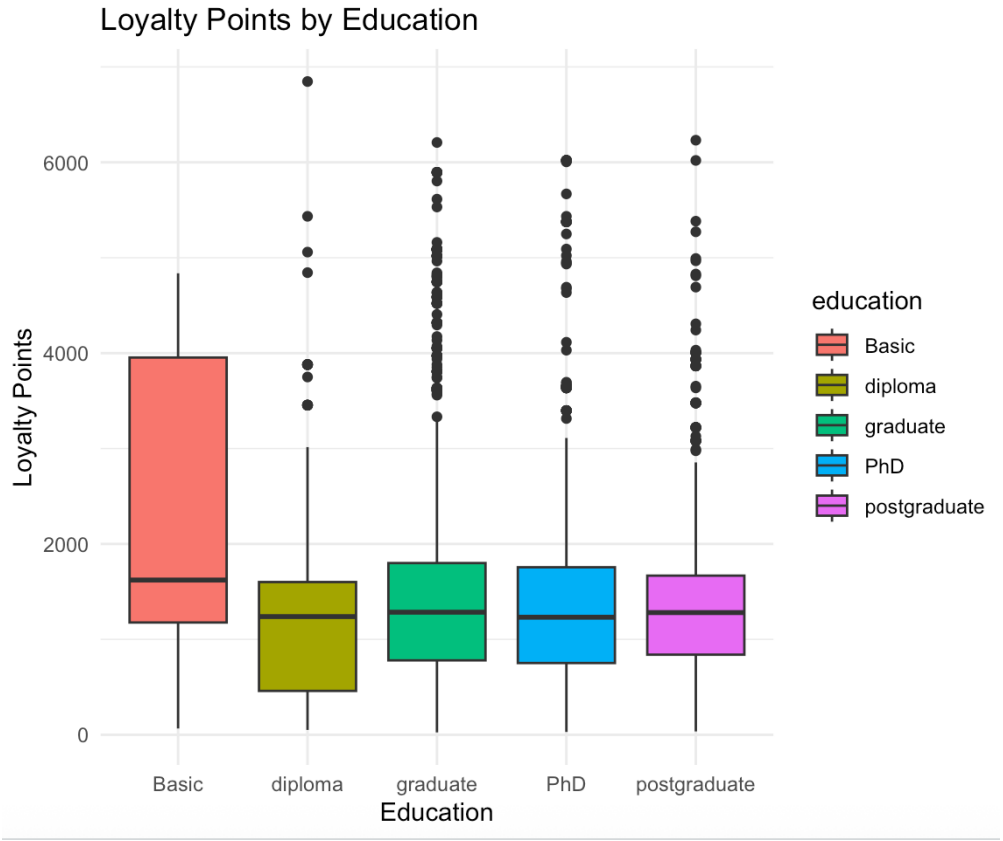
These exploratory plots were generated in R using ggplot2 to test whether demographic attributes (gender and education) had explanatory power for loyalty point accumulation. Several other plots were explored and are shown below. While visually useful, they did not reveal strong or systematic patterns and were not central to the main analysis.

Loyalty Points by Gender: Boxplot

- Created using boxplots in R (ggplot2) with loyalty points on the y-axis and gender on the x-axis.
- The median loyalty points for both male and female customers are very similar, just above 1,000.
- The interquartile range (IQR) is nearly identical, suggesting no meaningful gender effect.
- Both groups show high-value outliers (loyalty > 4,000), indicating that extreme behaviours occur in both genders.
- Interpretation: Gender does not appear to influence loyalty point accumulation.

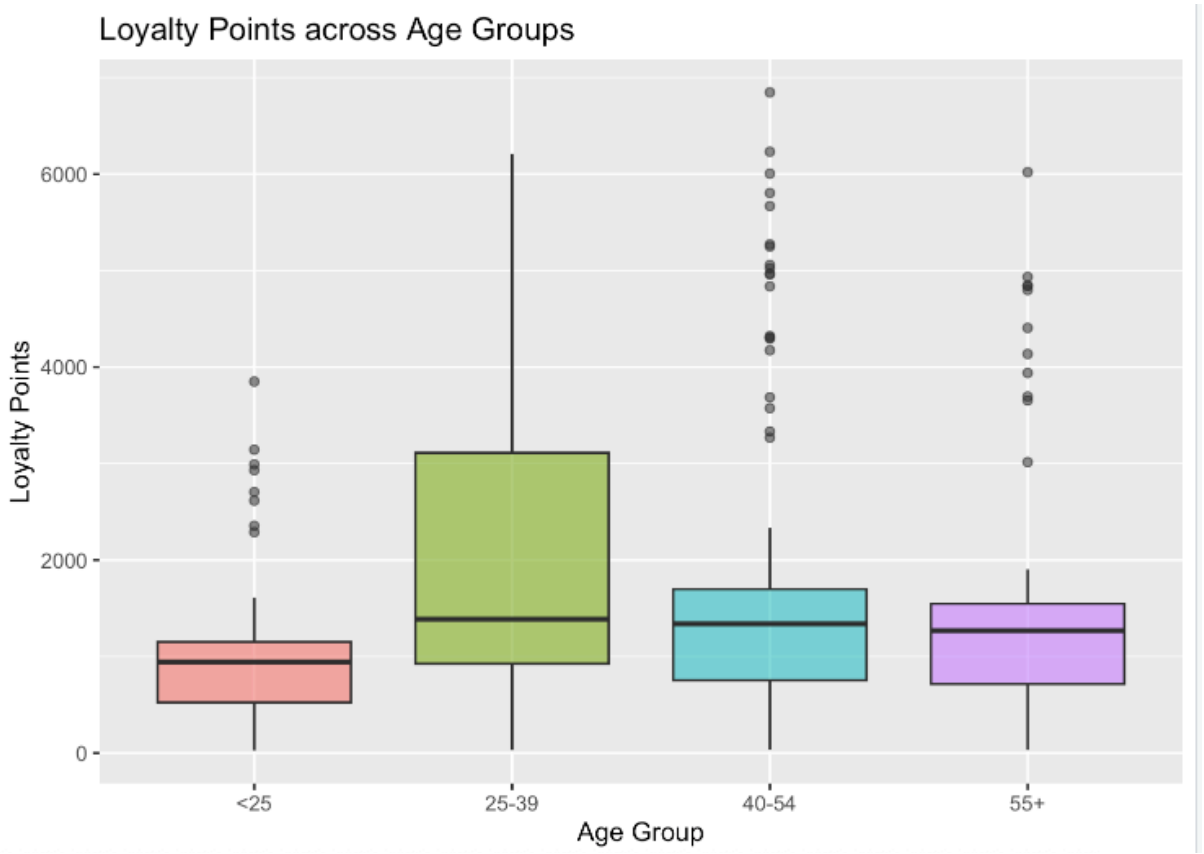
Loyalty Points by Education: Boxplot

- Created as a boxplot in R (ggplot2) with loyalty points on the y-axis and education category on the x-axis.
- Customers with basic education display the widest spread in loyalty points, with values ranging from near zero up to almost 6,000. Their median (~1,800) is slightly higher than other groups.
- Other education groups (diploma, graduate, postgraduate, PhD) have narrower spreads, with medians clustered between 1,000-1,500. Extreme outliers above 6,000 are only visible in the Graduate and PhD groups.
- Interpretation: While variability is greatest in the “basic education” group, no education level shows a consistent or systematic advantage in loyalty accumulation.



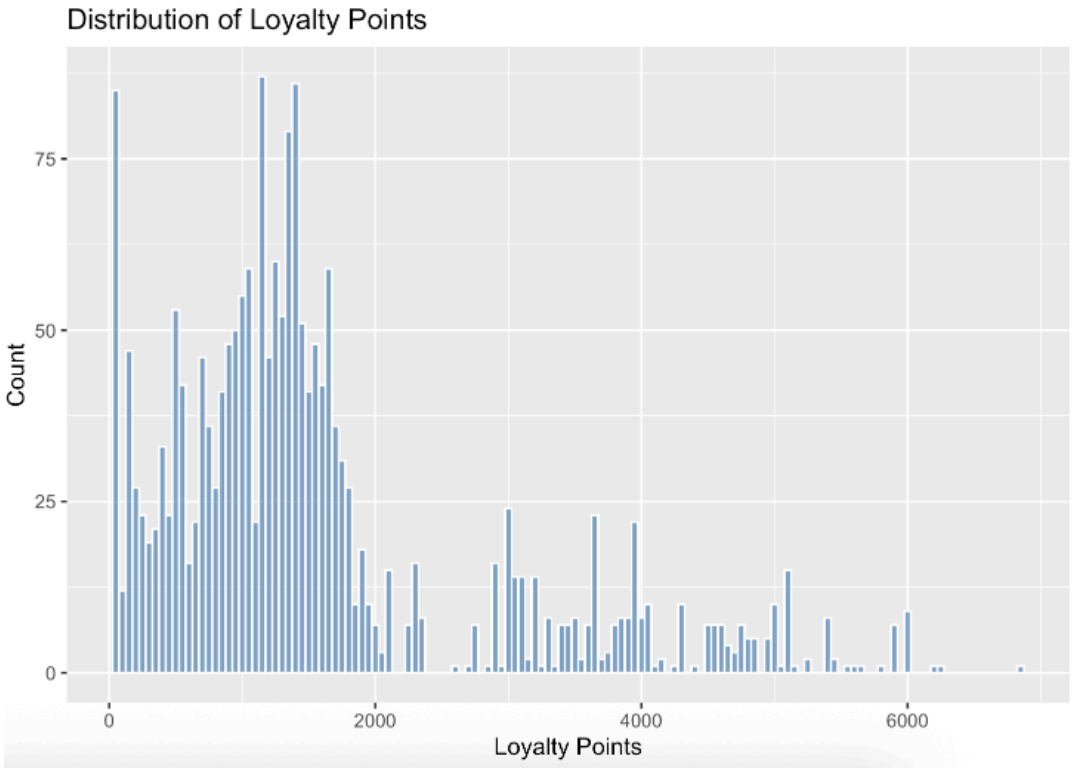
Loyalty Points by Age Groups: Boxplot

- Created as a boxplot in R (ggplot2) with loyalty points on the y-axis and age group on the x-axis.
- Customers aged 25-39 display the widest spread in loyalty points, ranging from near zero to over 6,000. Their median (~1,300-1,400) is slightly higher than other groups, though variability is substantial.
- The <25 group has a lower median (~1,000) and a tighter spread, suggesting younger customers tend to accumulate fewer loyalty points.
- The 40-54 and 55+ groups show similar medians (~1,400-1,500) and moderate spreads, with several high-value outliers above 5,000 points.
- Interpretation: Variability is highest among 25-39 year olds, but overall there is no consistent or systematic trend across age groups, supporting earlier regression findings that age has little predictive power for loyalty.



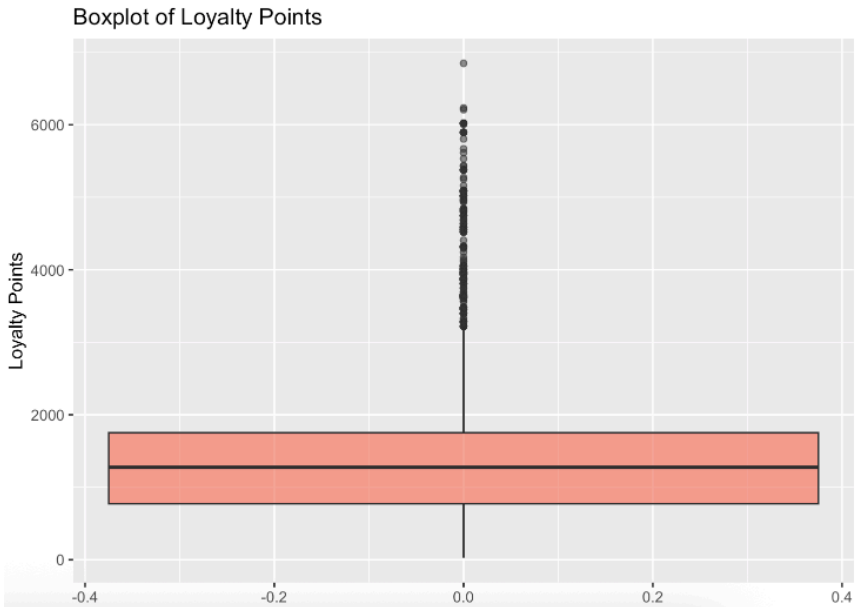
Initial Loyalty Points Distribution Plot:

Improved the interpretability for the purpose of the business presentation. Final plot in Appendix 4.

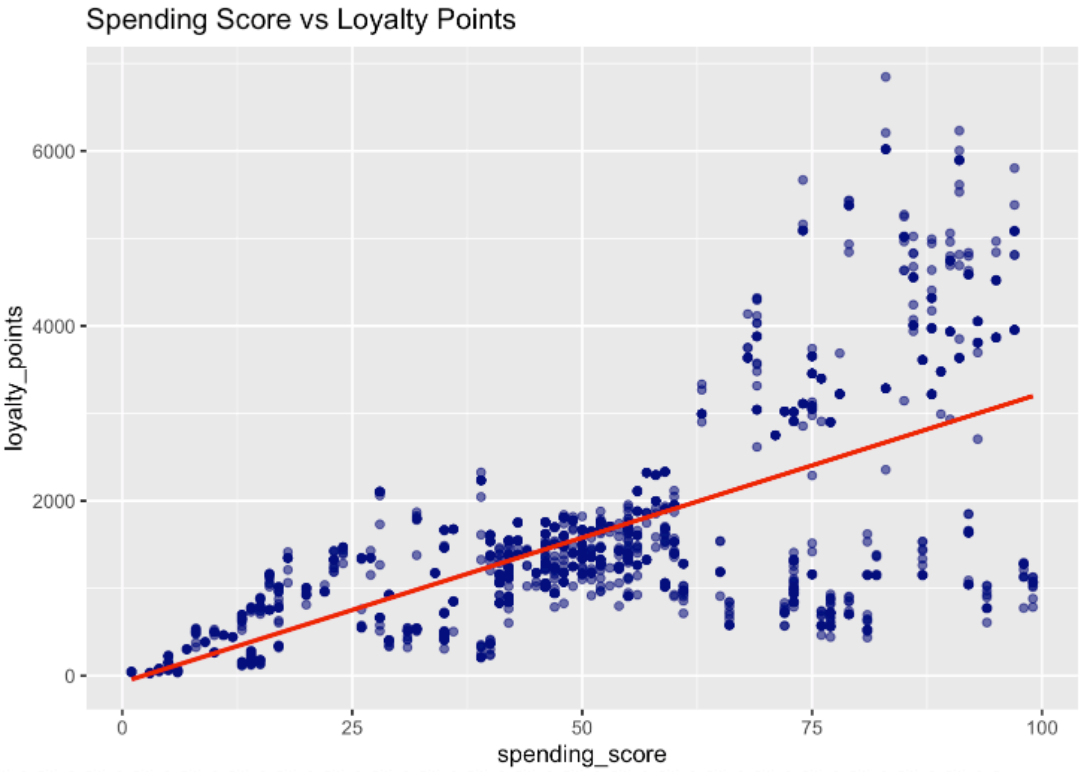


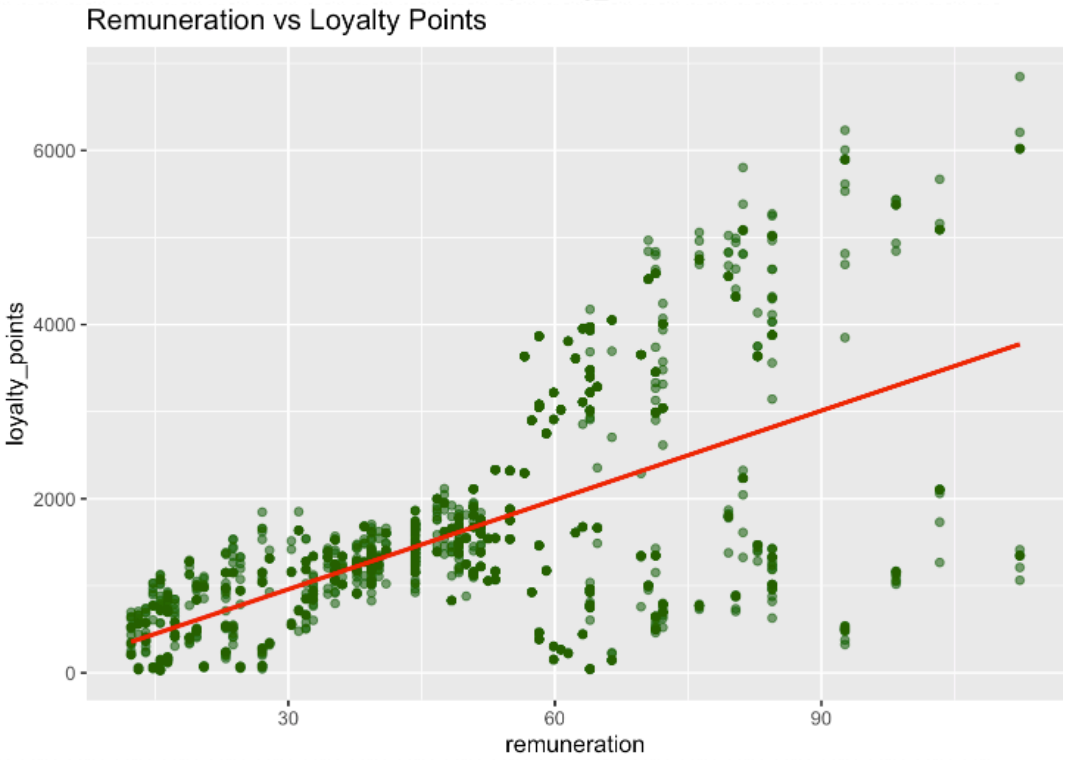
Boxplot of Loyalty Points:

- Used to view the distribution in another format.
- Shows a median just above 1,000, with many high-value outliers above 6,000, confirming the presence of a small but important group of extremely loyal customers.



Scatterplot of Key Variables:
As part of exploratory analysis, scatter plots were created in R to visualise the relationship between loyalty points and both remuneration and spending score. A linear regression line (red) was overlaid to assess direction and strength of association visually. These plots were used to confirm correlation patterns previously found in Python.





Appendix 15: Further Analysis Opportunities

1. **Enhance Data Collection**
   Collect more granular information on customer purchase history (e.g., frequency, product categories, transaction values), customer preferences, and cross-platform interactions. This would allow segmentation models to move beyond demographics and spending proxies, capturing richer behavioural and attitudinal insights.
2. **Adopt Advanced Predictive Models**
   Extend modelling beyond linear regression and decision trees by incorporating ensemble methods such as Random Forests or Gradient Boosting. These can capture complex, non-linear relationships, potentially improving prediction accuracy and business interpretability when feature sets expand.
3. **Link Loyalty and Sentiment to Actual Sales**
   Replace proxies (spending score, sentiment averages) with true financial metrics such as revenue per product and product category names. This would allow a direct evaluation of how loyalty points and customer sentiment drive commercial outcomes.
4. **Analyse Temporal and Regional Trends**
   Introduce time-series analysis to examine seasonal patterns in loyalty accumulation and product sentiment (e.g., holiday sales peaks, launch effects). Combine this with regional-level data to identify geographic variations in engagement, enabling more tailored marketing strategies.
5. **Connect Loyalty to Customer Lifetime Value (CLV)**
   Evaluate how loyalty point accumulation translates into long-term profitability. Linking loyalty balances with retention, churn, and repeat purchase behaviour would enable a full assessment of whether the loyalty program creates sustained customer value or short-term incentives.

Reference List:

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Hutto, C.J. and Gilbert, E.E., 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, pp.216-225.

Kaufman, L. and Rousseeuw, P.J., 2009. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons.

Kotler, P. and Keller, K.L., 2016. *Marketing Management*. 15th ed. Harlow: Pearson Education.

Kuhn, M. and Johnson, K., 2013. *Applied Predictive Modeling*. New York: Springer.

Kumar, V. and Reinartz, W., 2016. *Creating Enduring Customer Value*. Journal of Marketing, 80(6), pp.36-68.

Loria, S., 2018. *TextBlob Documentation*. [online] Available at: https://textblob.readthedocs.io/

Montgomery, D.C., Peck, E.A. and Vining, G.G., 2021. *Introduction to Linear Regression Analysis*. 6th ed. Hoboken, NJ: Wiley.