1. CONTEXT AND OBJECTIVES

Turtle Games wants to improve sales performance by understanding how customers earn loyalty points and which customer groups to target. This report has three goals:

I. Test whether spending score and income (remuneration kgbp) can reliably predict loyalty points with a simple, explainable model. Success target: R^2 above 0.75.

II. Identify practical customer segments that marketing can act on.

III. Turn findings into 3–5 specific actions with KPIs and a 30/60/90-day plan.

2. DATA AT A GLANCE

Dataset: 2,000 customers × 4 variables (loyalty points, spending, income, age).

Cleaning: removed language and platform, standardised names, kept high loyalty values.

Shape: loyalty is right skewed with a few very high values I report medians and IQR as well as means.

Working assumption: loyalty rises mainly with spending. Income adds extra lift and age provides little signal.

3. METHODS AND WHY I USED THEM

Exploratory data analysis: histograms, boxplots for distributions and outliers, scatterplots for loyalty relationships, and correlation matrix. To make findings easy for the business, I also grouped spending and income into Low, Mid, and High bands and summarised loyalty by these bands.

Guided by the exploratory work, I fitted a simple multiple linear regression model:
loyalty points = b0 + b1 * spending score + b2 * remuneration_kgbp

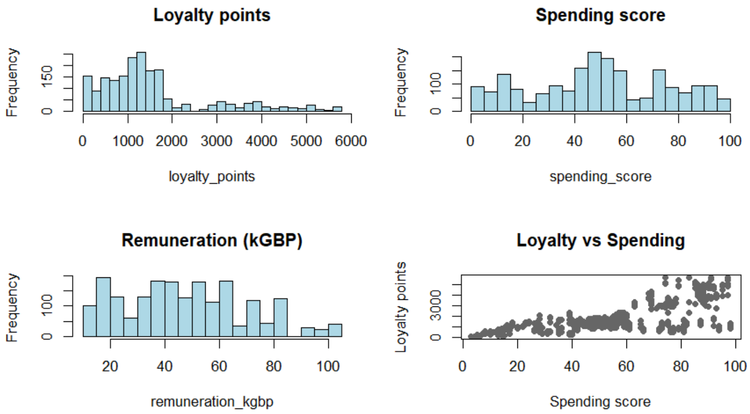
I checked model summaries and ran three diagnostic views: predicted versus actual, residuals versus fitted, and a normal Q–Q plot. I then created a few “what-if” scenarios with 95 percent prediction intervals to show how the model can support planning.

Outlier policy: kept high earners to reflect reality. Used medians and IQR to reduce their influence on summaries, plus residual diagnostics to check fit. I only remove a row if it is clearly erroneous otherwise, I prefer imputation.

4. KEY FINDINGS

4.1 Distributions and drivers

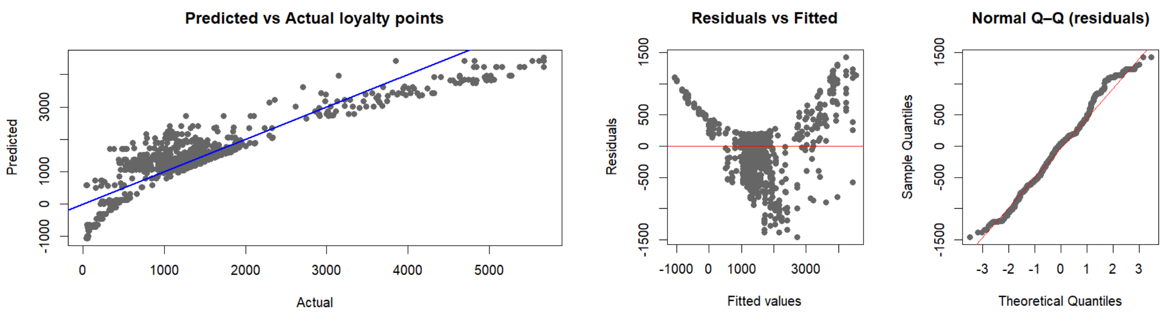
- Loyalty points are right skewed with a widespread. Most customers are in the middle, with a smaller group earning very high points.
- Spending score has the strongest positive relationship with loyalty. Income also adds independent lift. Age has minimal relationship with loyalty once spending and income are considered.
- Spending and income are almost uncorrelated with each other, so both add unique information for predicting loyalty. ($r=0.0064$)



Loyalty is skewed. loyalty rises with spending. Income is broadly spread.

4.2 Simple and explainable model

- Multiple linear regression using spending and income achieves R^2 of 0.83 which is strong for a two-variable model.
- Both coefficients are positive and highly significant ($p < 0.001$). Spending adds 32.8 loyalty points per unit (95% CI: 31.7-33.9). Income adds 33.9 points per £1k (95% CI: 31.5-36.4). VIF is 1.00, confirming no multicollinearity. The narrow confidence intervals indicate precise, reliable estimates for planning.
- Residual standard error is roughly 525 points. Residuals widen a bit at higher fitted values, which is normal given the skew and the presence of high earners.

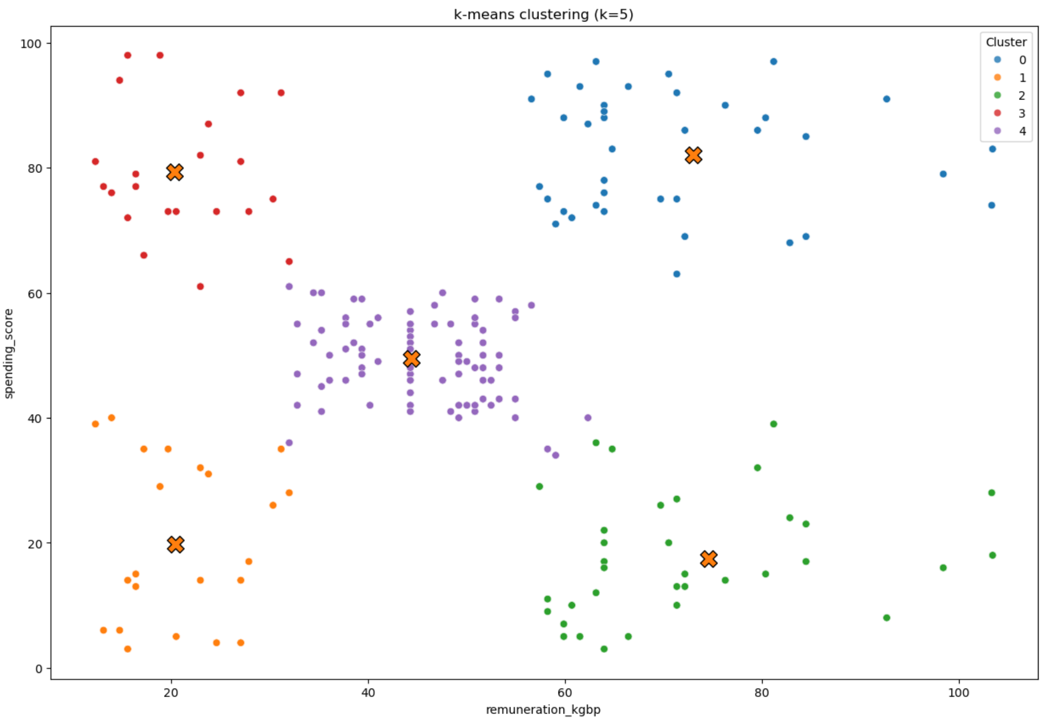


The fit is strong overall. Residuals show mild heavy tails and a little extra spread for very high predicted values.

4.3 Segments and thresholds

- By simple Low, Mid, High bands, loyalty increases steadily from Low to Mid to High spending. Within any spending band, higher income adds more loyalty. The highest loyalty is in High-spend and High-income.
- K-means clustering (k=5) on income and spending yields five practical segments. In short: a large Core Mid/Mid group, High-High VIPs, High-income but Low-spend (under-engaged potential), Low-income but High-spend (value enthusiasts), and Low-Low.
- K-means was chosen over hierarchical clustering for its computational efficiency and clear, interpretable segment boundaries suitable for operational use.

	spend	pay	n_customers	pct_customers	avg_loyalty	total_loyalty
1	Low	Low	392	19.6	528	206979
2	Mid	Low	175	8.8	1208	211351
3	High	Low	233	11.7	980	228349
4	Low	Mid	440	22.0	1176	517417
5	Mid	Mid	211	10.6	1702	359211
6	High	Mid	166	8.3	3384	561755
7	Low	High	193	9.7	1104	213057
8	Mid	High	47	2.4	3514	165158
9	High	High	143	7.2	4796	685786



Bands and clusters give straightforward targets for marketing and loyalty design. Mid pay clients represent 40.9% of customer base.

5. EVIDENCE INTO ACTIONS WITH KPIS

Actions are prioritized by expected impact: A and C target large, improvable segments (combined 60%+ of customers); B protects high-value customers; D and E optimize existing behaviour efficiently.

A. Move Core Mid/Mid up a tier. This is the largest group and the main growth lever.
Action: four-week (day 0-30) step-up bonus lifting spending from 60 to 70.
KPI: percentage of the group reaching spending 70 or above. A 10 point spending lift implies plus 328 loyalty points. Applied to 211 Core Mid/Mid customers, this generates approximately 69,208 additional points system-wide.
Measurement: A/B test within this segment; report proportions, not only counts.

B. Retain High-High VIPs.
Action: early access, premium multipliers with sensible caps, and VIP services.
KPI: VIP retention rate, average basket value, points per customer.

C. Activate High-income Low-spend (under-engaged potential).
Action: onboarding bundle, premium trials, cross-sell into higher margin lines.
KPI: plus 10 spending points in 30 days. Expected gain is roughly plus 328 loyalty points.

D. Protect margin for Low-income High-spend (value enthusiasts).
Action: non-cash perks or limited boosts, and cap discounts to protect margin.
KPI: points per £ and promo ROI versus control.

E. Nurture Low-Low efficiently.
Action: low-cost communications, bundles, and seasonal offers.
KPI: open and click rates, and modest but profitable uplift.

6. WHAT-IF PREDICTIONS FOR PLANNING

Five scenarios demonstrate planning applications with 95% prediction intervals:

Spend 60 and income £55k: about 2,140 points.

Spend 85 and income £80k: about 3,807 points (about 1,667 more than the previous scenario).

At spending 60, increasing income from £35k to £80k moves the prediction from about 1,461 to about 2,988 (about plus 1,527).

	spending_score	remuneration_kgbp	predicted	pi_low	pi_high
A) Low spend, low income	40	35	806	-225	1837
B) Mid spend, mid income	60	55	2140	1109	3170
C) High spend, high income	85	80	3807	2775	4838
D) Mid spend, low income	60	35	1461	430	2492
E) Mid spend, high income	60	80	2988	1957	4019

Use the predicted value for planning and the interval to set realistic expectations for individual outcomes.

7. LIMITATIONS AND HOW WE MANAGE THEM

- Causality: model identifies associations, not causal relationships. The step-up bonus (Action A) will test whether increasing spending causally drives loyalty.
- Skew and outliers: report medians and IQR in addition to means. Check residuals. If needed, test a log version of loyalty for sensitivity.
- Nonlinearity: if residuals show curvature, add a spending times income interaction or fit separate simple models by spending band.
- Generalisation: refit quarterly to watch for drift; validate on a simple hold-out split.

8. 30, 60, 90 DAY PLAN

Day 0 to 30: launch the step-up bonus to Core Mid/Mid and add a small dashboard showing conversion, average points, and points per £.

Day 31 to 60: roll out VIP retention to High-High. Publish uplift versus control.

Day 61 to 90: pilot the High-income Low spend activation, retrain the model, refresh clusters, and adjust thresholds if needed.

Days	**Action**	**Segment**	**Success Metric**
0-30	Step-up bonus	Core Mid/Mid	15% reach spending 70+
0-30	Dashboard setup	All	Real-time KPI tracking
31-60	VIP retention program	High-High	95% retention rate
61-90	High-income activation	High-income/Low-spend	+10 spending points
61-90	Model refresh	All	R^2 maintained >0.80

9. CONCLUSION

A simple model using spending and income explains 83% of variation in loyalty points, which is strong and easy to communicate. The segments and band tables give clear targets, and the what-if scenarios translate directly into planning numbers.

Next steps: move the large mid-tier group up, retain the VIPs, and activate the high-income low spenders, while monitoring results and retraining regularly.