Turtle Games: Predicting Future Outcomes

By: Iulia-Diana Cristolovean

Background & context

TURTLE GAMES

A game manufacturer and retailer

OBJECTIVE

Improving overall sales performance by analysing and considering customer trends

QUESTIONS TO ANSWER

1. How do customers engage with and accumulate loyalty points?
2. How can customers be segmented into groups, and which groups can be targeted by the marketing department?
3. How can we use social data to understand customers?

Analysis Method

Predictive Models

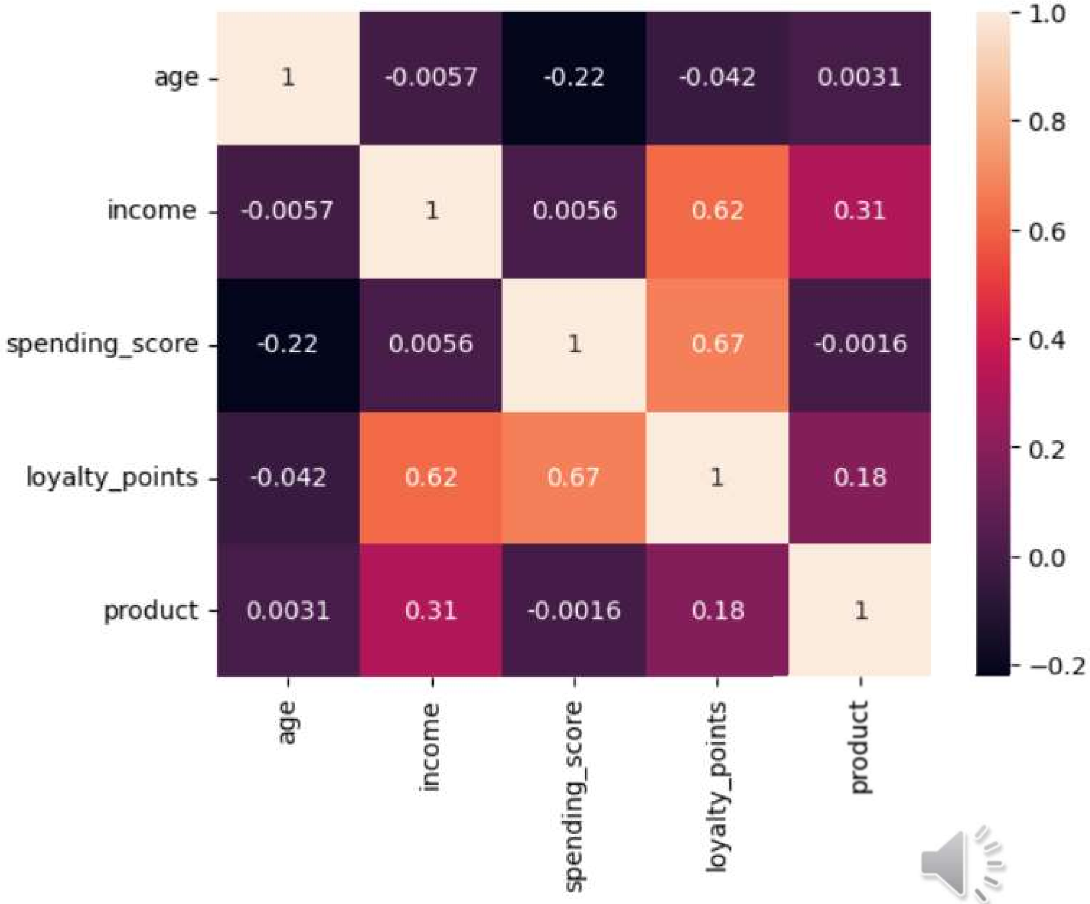
- Possible customer behaviors
- Customer segmentation

Natural Language Processing

- Customer opinions (reviews) on the website

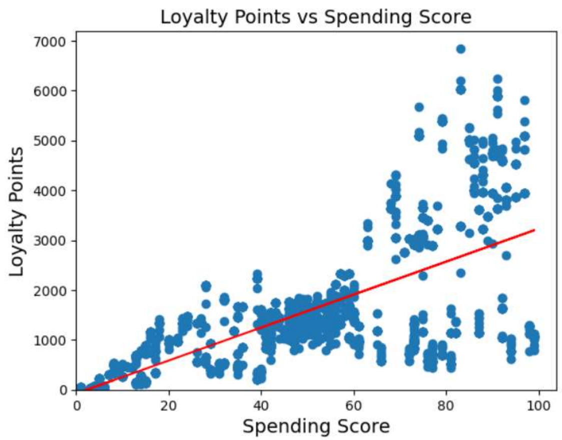
Tools used: Python , R

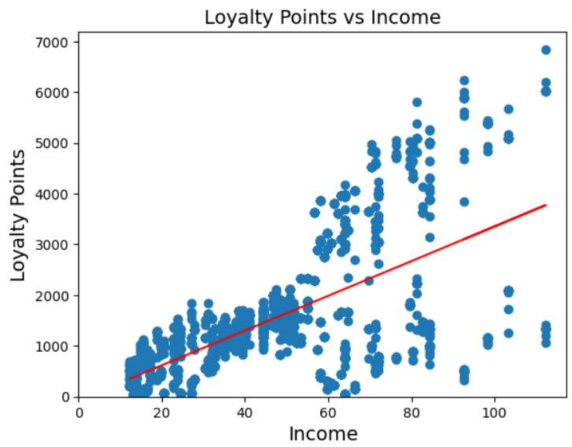
Factors influencing loyalty points

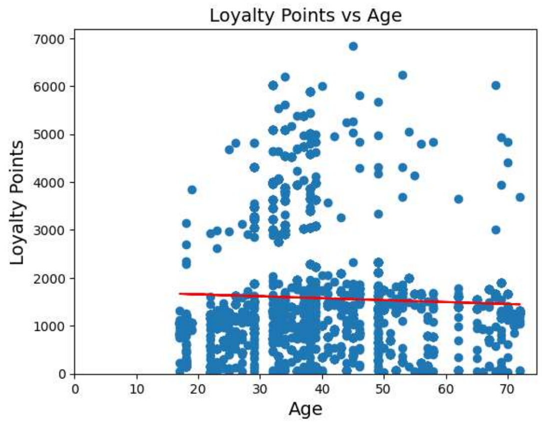


- Income moderately correlates with loyalty points (0.62).
- Spending score has a slightly stronger correlation (0.67).
- Age has no linear link to loyalty points.

To compute the statistical significance and explanatory power of these numerical variables on loyalty points accumulation, we used Simple Linear Regression Models.



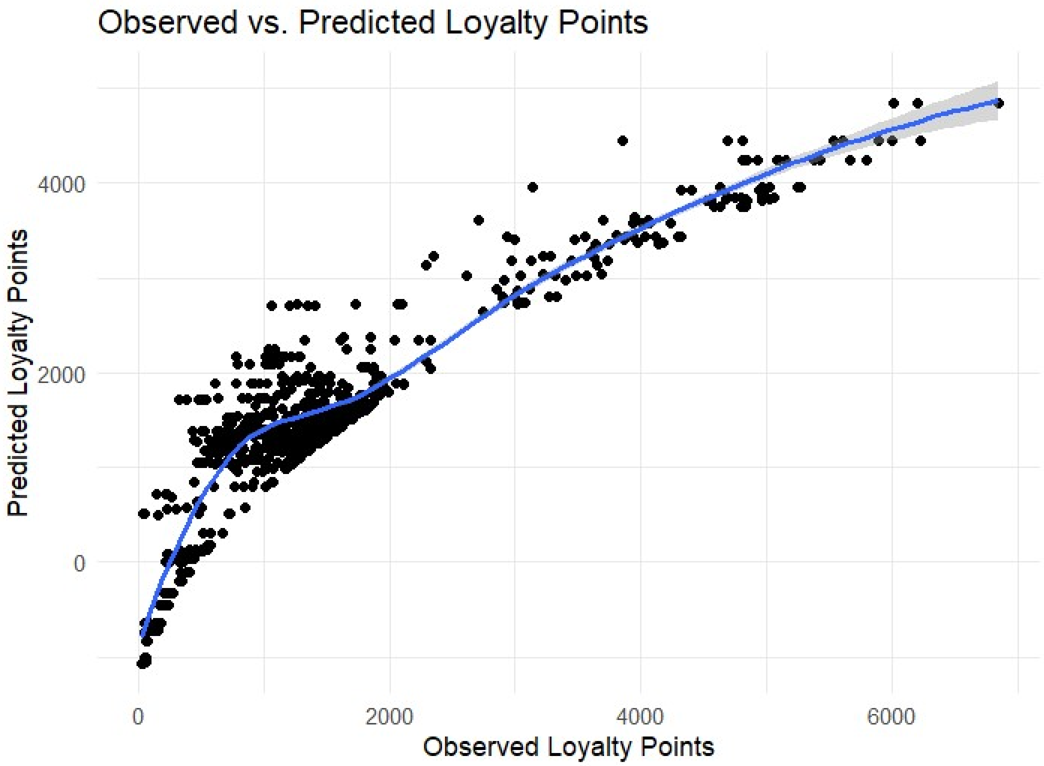




- ✓ *Spending Score vs Loyalty Points. [R2 = 0.452, β = 33.06, p = < 0.005]*
- ✓ *Income vs Loyalty Points. [R2 = 0.380, β 34.187, p < 0.005]*
- ✓ *Age vs Loyalty Points - No linear relationship.*

We used Multiple Linear Regression (MLR) with income and spending score

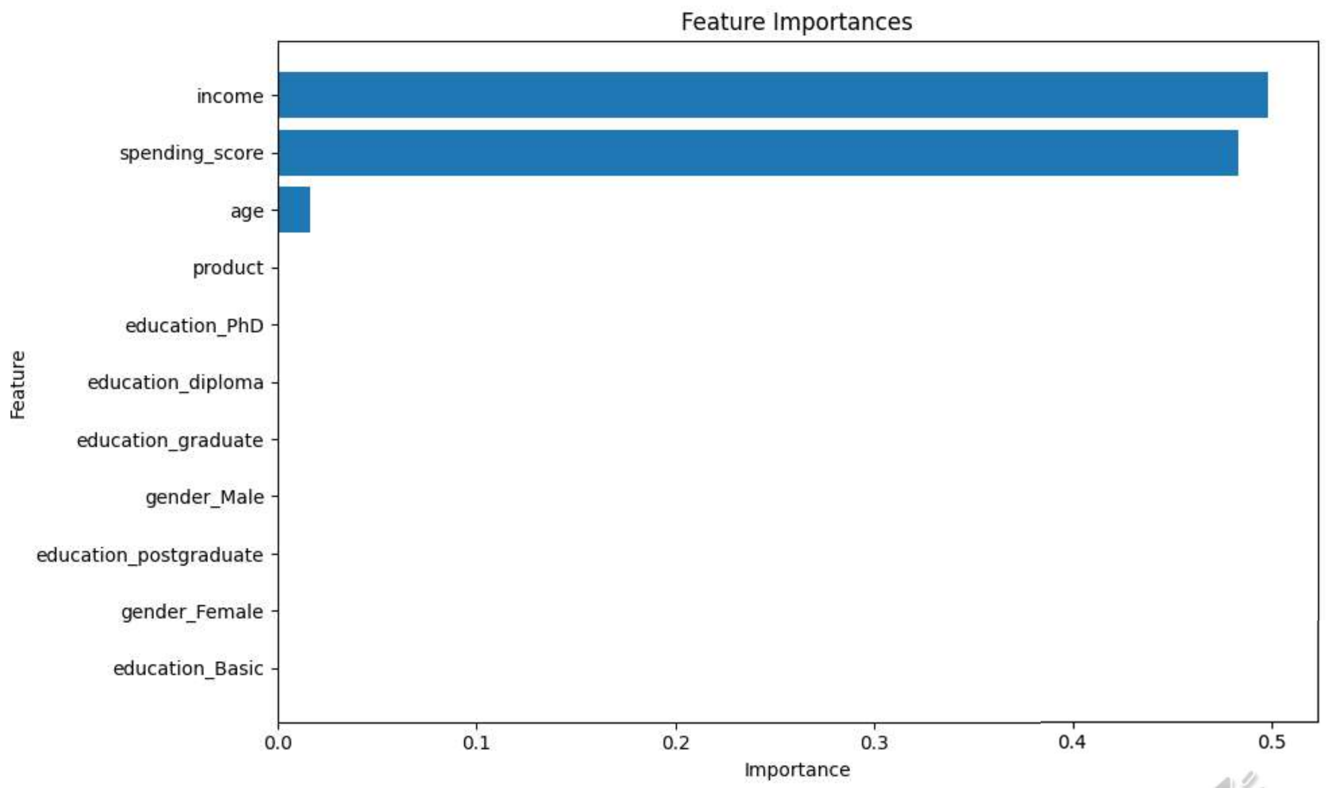
Multiple Linear Regression	
	Loyalty Points
R^2	0.83
Adjusted R^2	0.83
Mean Squared Error (MSE)	300944.1
Mean Absolute Error (MAE)	429.66



Income and Spending Score explained 83% of the variation in loyalty points.

To better understand which factor contributes to the loyalty points more, we used a Decision Tree Regressor because it manages non-linear relations well.

We determined the variables with the most contribution.

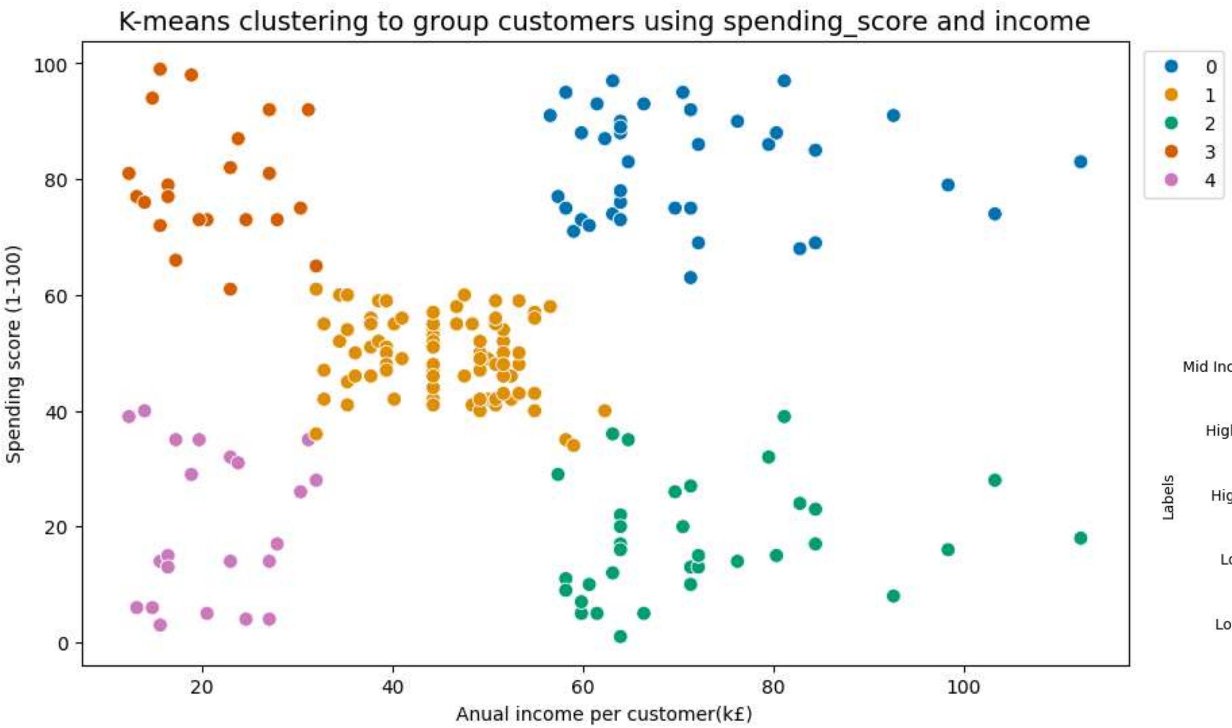


Key Insights on Predictors of Loyalty Points:

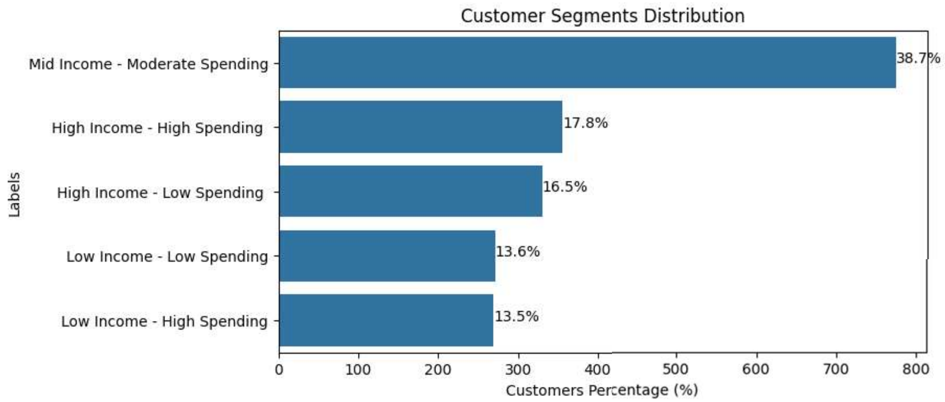
Random Forest — it's clearly the best-performing model in our lineup, both in terms of predictive power (R^2) and low error (MSE). If interpretability is more important than performance, the pruned Decision Tree or the Linear Regression models might still be worth considering.

Model	MSE	R^2
Multiple Linear Regression	300,944.09	0.83
MLR with log transformation	0.178	0.795
Random Forest	7,421.83	0.995
Decision Tree Regressor (pruned)	15,437.31	0.99
SVR with log transformation	1,441,148.25	0.111

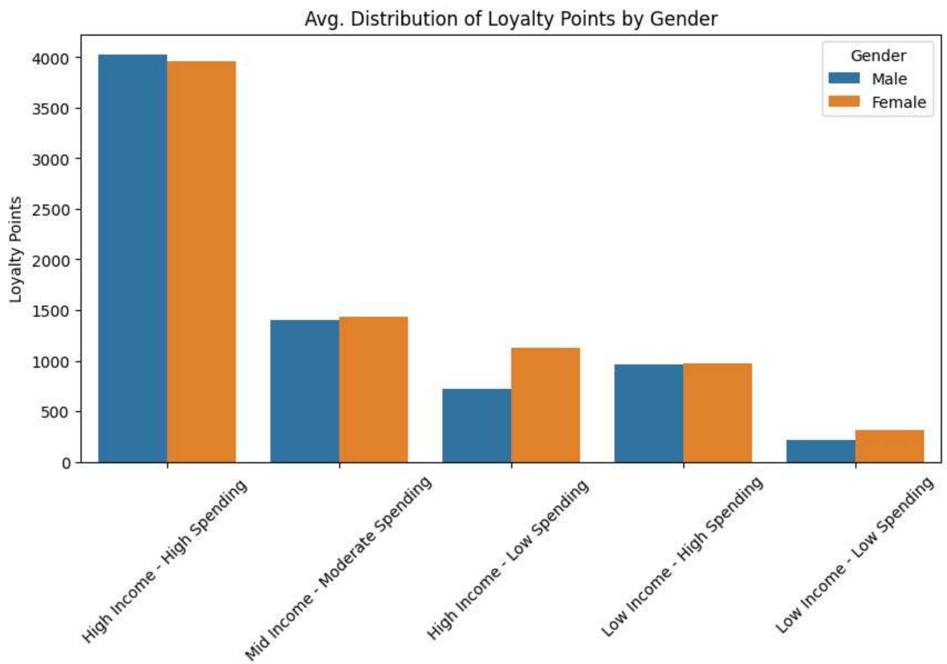
How can customers segmentation help drive targeted marketing?

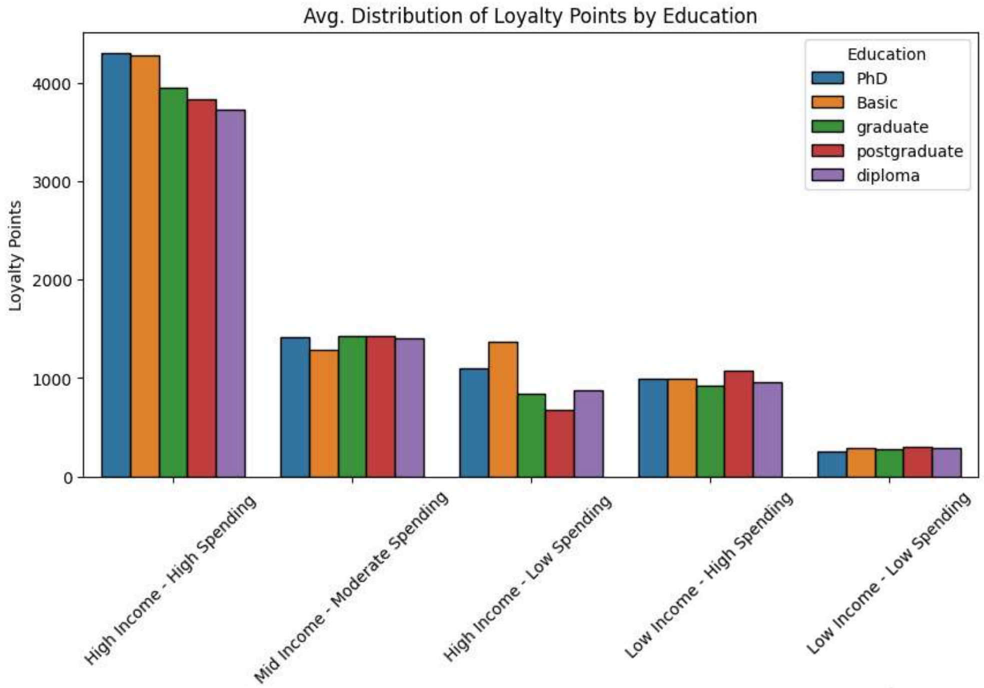


0	High Income - High Spending (Blue)
1	Mid Income - Moderate Spending (Orange)
2	High Income - Low Spending (Green)
3	Low Income - High Spending (Red)
4	Low Income - Low Spending (Purple)



Gender and education interactions with the loyalty program were analysed.





Education showed no clear pattern, but females engaged slightly more with loyalty points.

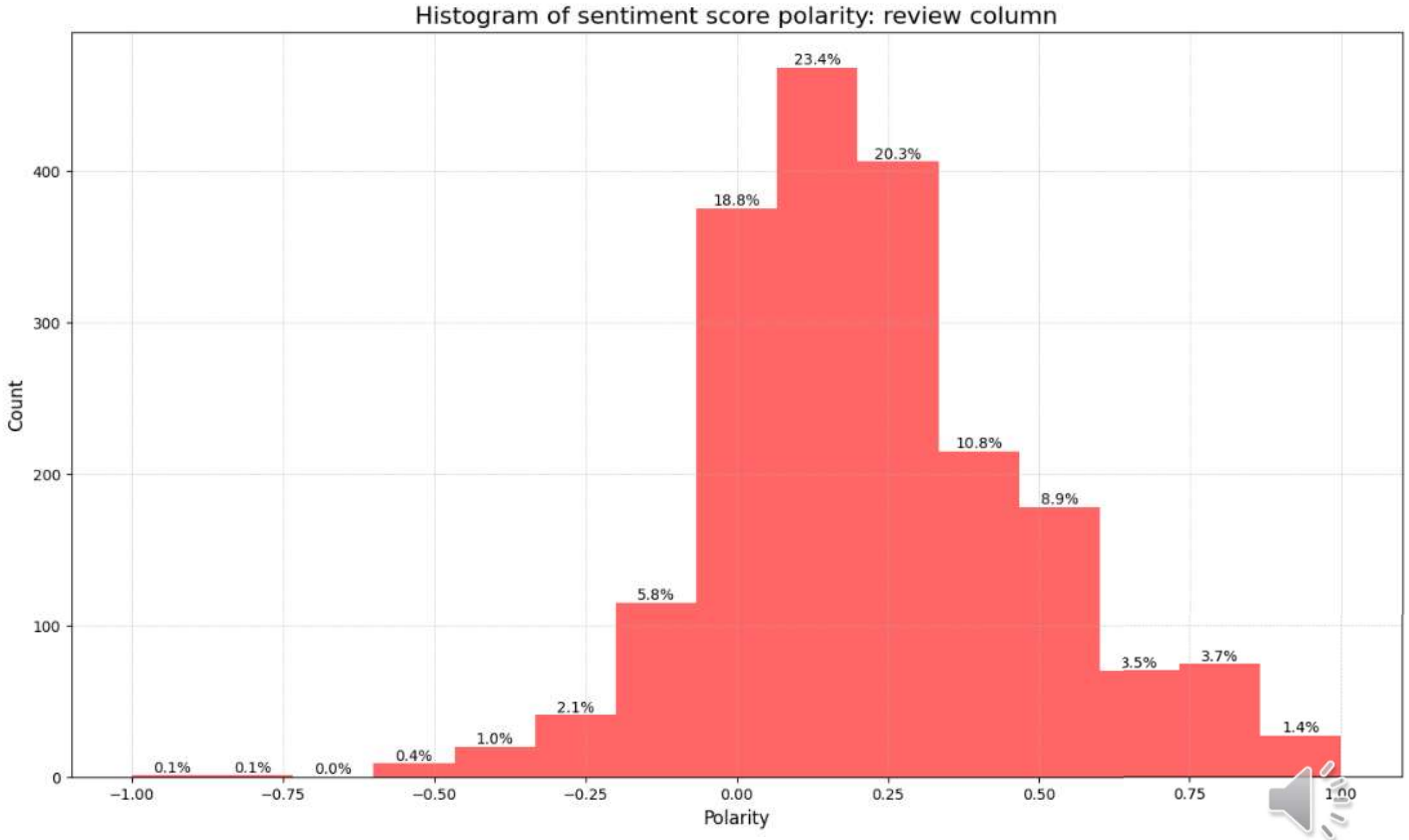
To consider language analysis we used Natural Language Processing

Word Cloud for ***review*** colmn

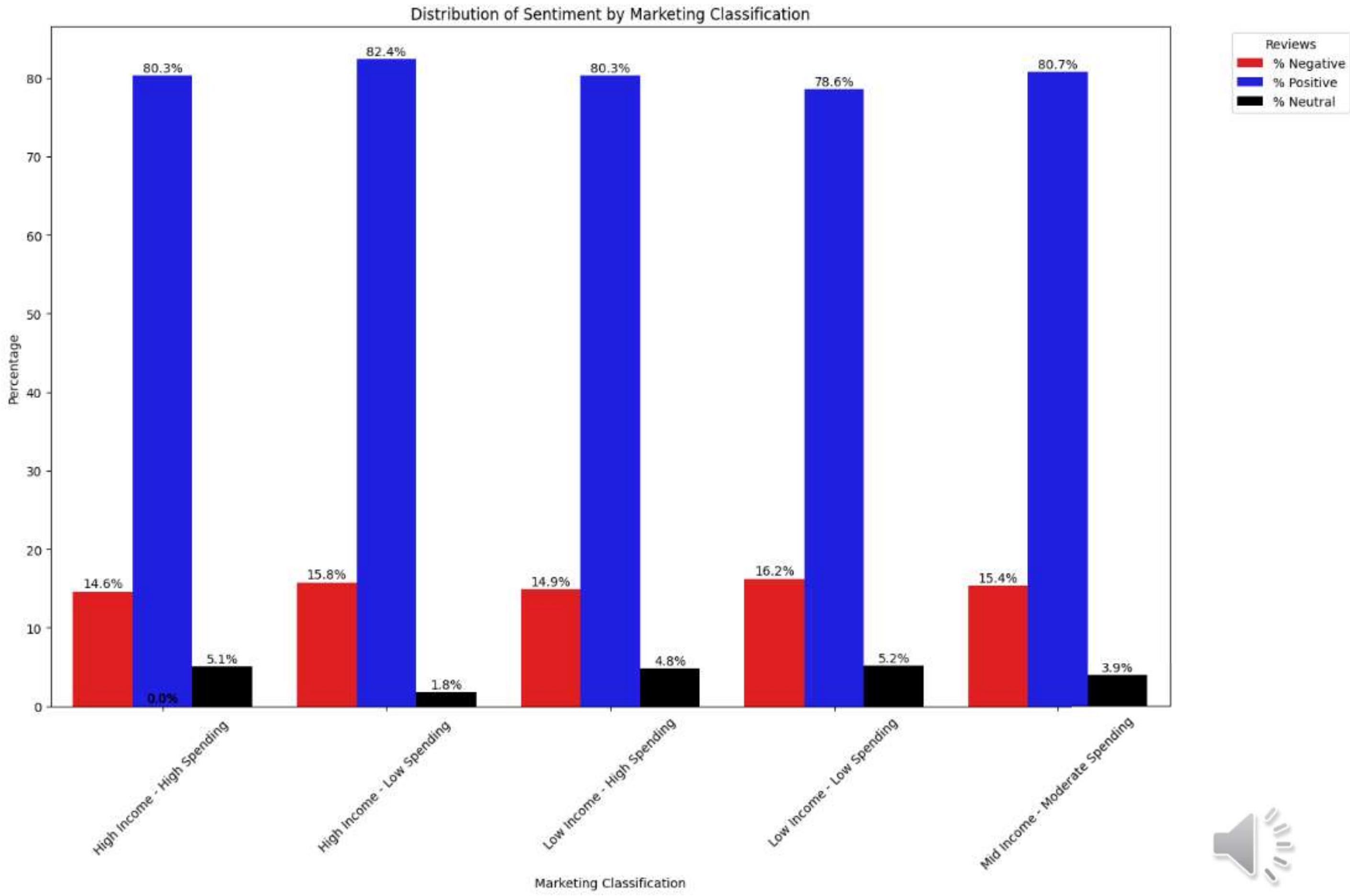
Word Cloud for ***summary*** colmn

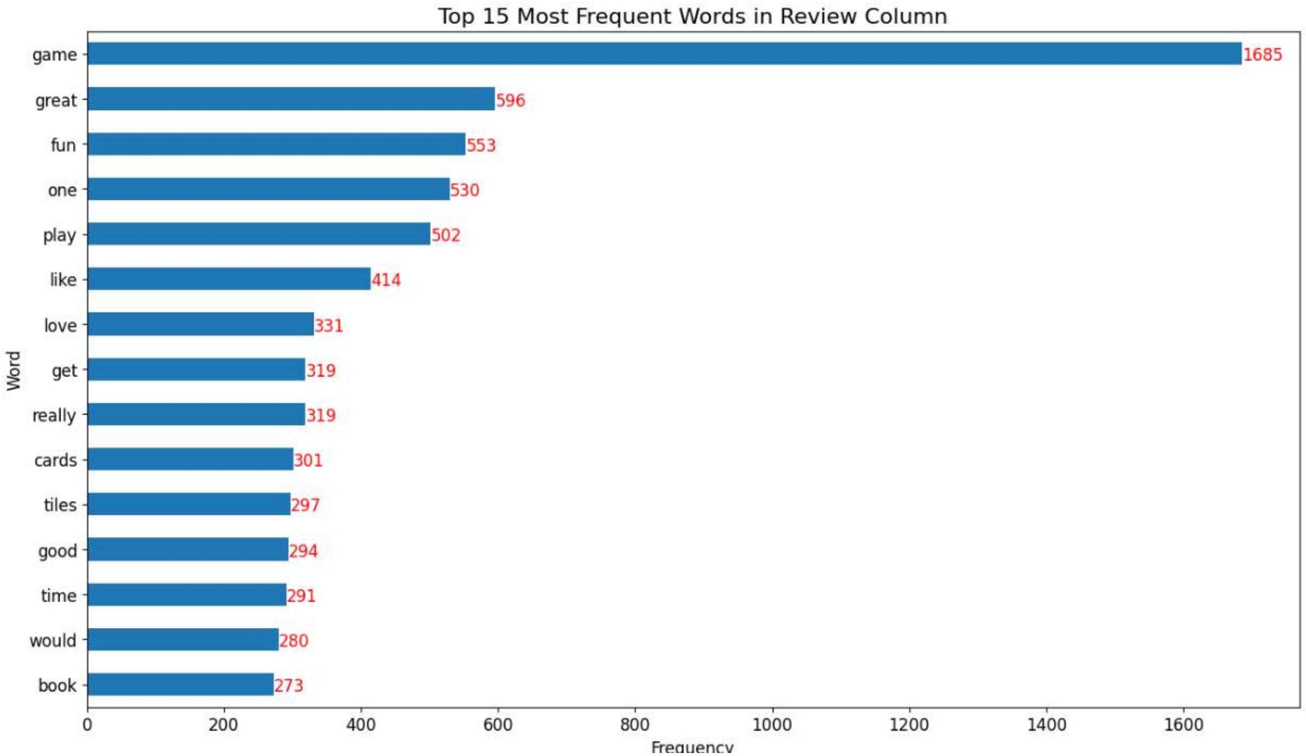
Social Data : what can we learn from customers reviews?

- Most review polarities are weakly positive (avg: 0.22), suggesting general satisfaction but not strong enthusiasm.
- Neutral reviews are over-represented—likely due to limitations in the sentiment model (e.g., rating "5 stars" as neutral).



- Most groups have a high rate of positive reviews.
- Conservative and practical spenders are the most satisfied.
- Occasional shoppers and conservative spenders are least satisfied—likely due to higher expectations—so targeting them could boost satisfaction.



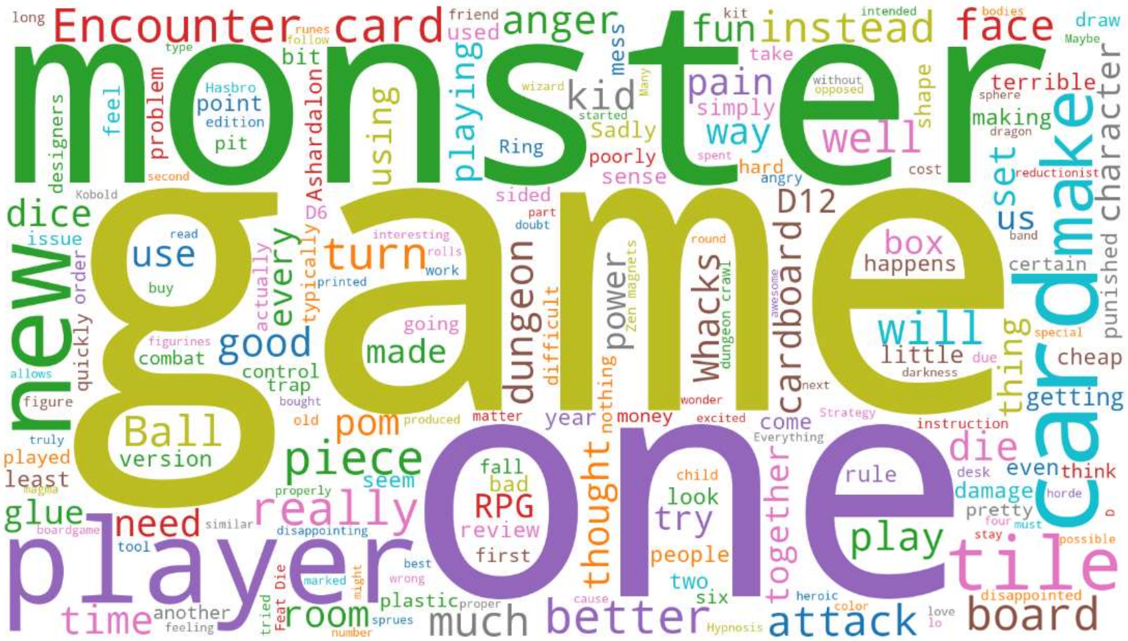


	word	polarity
0	game	-0.4
1	great	0.8
2	fun	0.3
3	one	0.0
4	play	0.0
5	like	0.0
6	love	0.5
7	really	0.2
8	get	0.0
9	cards	0.0
10	tiles	0.0
11	good	0.7
12	time	0.0
13	would	0.0
14	book	0.0

- Polarity of the top 15 review words shows “game” as negative (-0.4), while “great” (0.8), “good” (0.7), and “love” (0.5) are the most positive. Notably, 9 out of 15 words are neutral (0.0).

Many positive reviews centre around board games, cards, expansions and corporation.

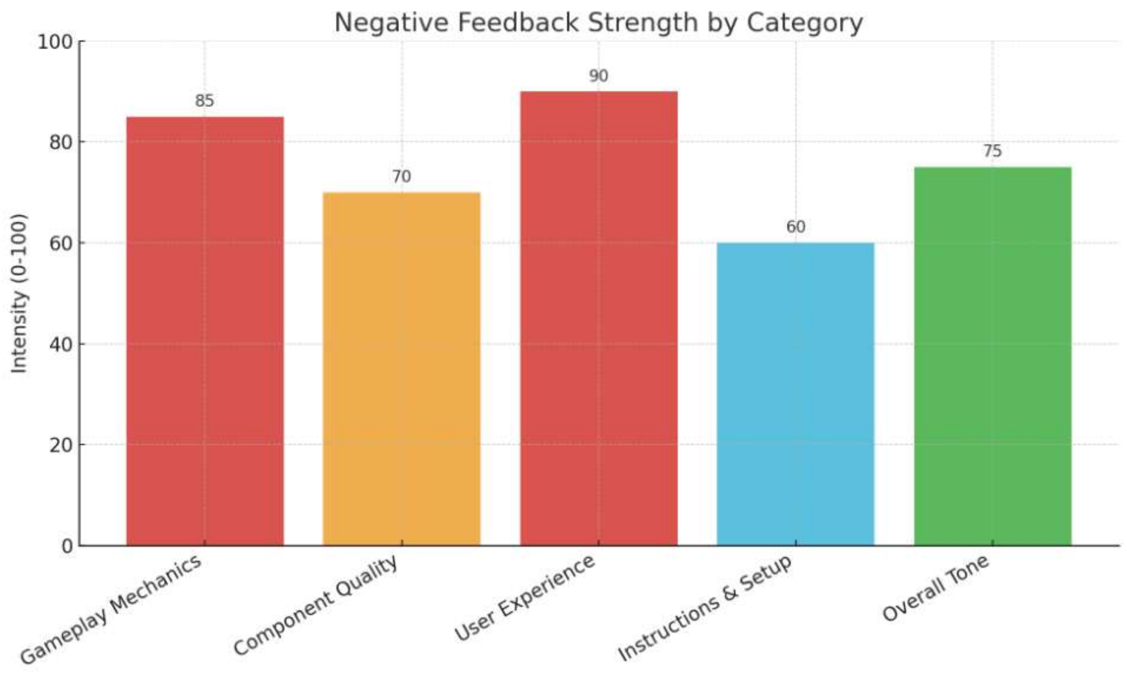
Given the higher risk posed by negative reviews for Turtle Games, we created a WordCloud revealing frequent complaints about board games, cards, quality, and usefulness.



Here's a summary table based on the word cloud analysis of negative feedback:

Category	Common Words/Phrases	Insights	Recommendations
Gameplay Mechanics	game, player, turn, attack, die, play, tile, card, monster, dungeon	Frustration with combat flow, randomness, or unfair game mechanics	Rebalance combat, reduce randomness, improve pacing and clarity of actions
Component Quality	glue, piece, cardboard, plastic, board, tile, magnet, fall, cheap	Complaints about fragile or poorly made pieces needing DIY fixes	Upgrade material quality, improve packaging, offer clearer assembly instructions
User Experience	pain, angry, terrible, disappointing, punished, much, really, difficult, feel, wrong	Emotional dissatisfaction, feelings of unfairness or tedium	Prioritize player enjoyment, reduce punitive elements, streamline gameplay
Instructions & Setup	instruction, rule, control, properly, tool, make, look, use	Confusing setup or poorly explained rules causing frustration	Rewrite rulebook with visuals, offer quick-start guides or tutorials
Overall Tone	try, new, version, better, thought, make, use, fun, instead, even, need	Desire for improvement or fixes; comparisons with other (better) versions or games	Take user feedback into account for updated editions or expansions

Negative Feedback Highlights:

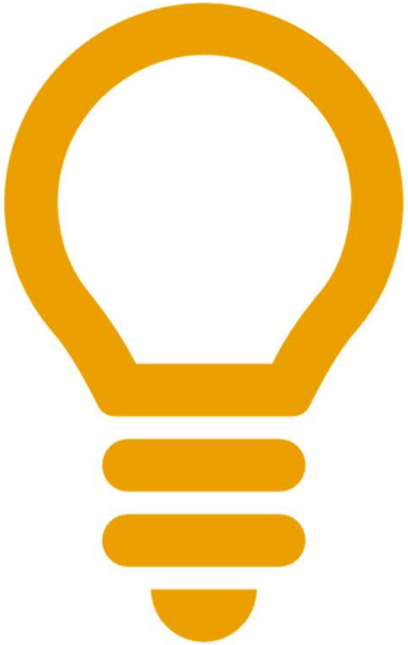


- Strongest Criticism: User Experience & Gameplay — frustration with feel and mechanics.
- Other Concerns: Component Quality (durability) and Overall Tone (disappointment).
- Moderate Issues: Instructions & Setup — clarity and ease of use.

To guide marketing strategies and product development, I made a list of products with the highest number of negative reviews.

	product	negative_review_count	total_review_count	negative_review_proportion
116	6431	5	10	0.5
90	4399	5	10	0.5
145	9597	5	10	0.5
106	5512	5	10	0.5
53	2387	5	10	0.5
10	486	4	10	0.4
3	231	4	10	0.4
58	2795	4	10	0.4
64	2870	4	10	0.4
72	3436	4	10	0.4

Insights and recommendations



- **Top Drivers:** Income & Spending Score explain 83% of Loyalty Points—useful for predicting and segmenting valuable customers.
- **Customer Segmentation:**
  - Target High-Income, High-Spending and Mid-Income, Moderate-Spending groups with tailored campaigns.
  - Females show higher loyalty engagement—consider gender-specific strategies.
- **Loyalty Engagement:** Wide variance in loyalty points suggests inconsistent engagement—opportunity for personalized rewards.
- **Sentiment Insights:**
  - Positive Reviews: Highlight in SEO, thank loyal users.
  - Negative Themes: Focus on product quality, age-fit, and clarity.
  - Improve Quality Control, simplify instructions, clarify age ratings, and respond to feedback.
- **Modeling Tips:**
  - Use linear regression for targeting.
  - Consider SVR for handling non-linear outliers.