Analytical report

Turtle Games: Customer segmentation and spending behaviour

Prepared by: Yuliya Pauzunova

Background and context

Client: Turtle Games

Business problem: Improving overall sales performance by analysing and considering customer trends

Analytical problem: Identify statistically significant features and segment customers to predict loyalty point balances

Analytical objectives/questions:

- Identify relevant features that best explain the variance in customer spending behaviour and evaluate possible relationships between features and loyalty point balances to develop predictive models
- Group customers based on models' predictions to increase marketing strategy efficiency by implementing tailored marketing campaigns and promotions based on the identified segments
- Identify and summarise common themes and opinions in product reviews and evaluate sentiment to improve overall sales performance by addressing customers' needs related to customer service or products

Dataset Details:

Provider: Turtle Games
File Name: turtle_reviews.csv
Size: 2000 rows, 11 columns

Key data fields:

- Online Reviews: full review text submitted by customers and review summary
- Customer Demographics:
    - Gender
    - Age
    - Education
    - Remuneration
- Spending Score: A score assigned to a customer based on the customer's spending nature and behaviour. The value ranges between 1 and 100
- Loyalty Points: A score based on the point value of the purchase, converting the monetary value to point value, and the point value of an action (purchase)
- Product ID: Identifies the product associated with the review
- Language of the review and platform source

Methodology and Approach

Regression Analysis

Linear Regression

Decision Tree

We employed regression analysis techniques to identify relevant features that best explain the variance in customer spending behaviour and evaluate possible relationships between features and loyalty point balances to develop predictive models

Segmentation

Clustering

We utilised centroid-based clustering models to group customers into clusters based on similarities

Exploratory Data Analysis (EDA)

Visualisation

Summary statistics

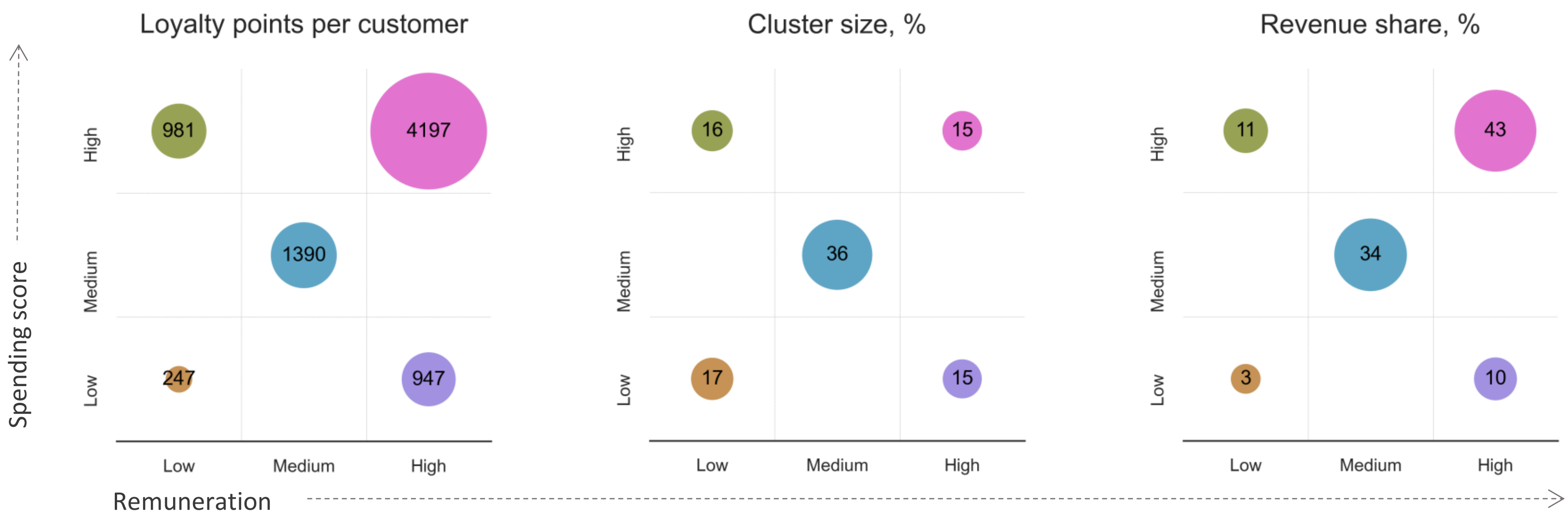
Visualisation techniques and summary statistics were utilised to identify patterns and interpret customer clusters, providing valuable insights to inform decision-making

Sentiment Analysis

NLP

We utilised a lexicon-based sentiment analysis tool (VADER), which is specifically designed for social media text and short sentences and Word Clouds to assess sentiment and summarise most common themes in product reviews

Clustering analysis



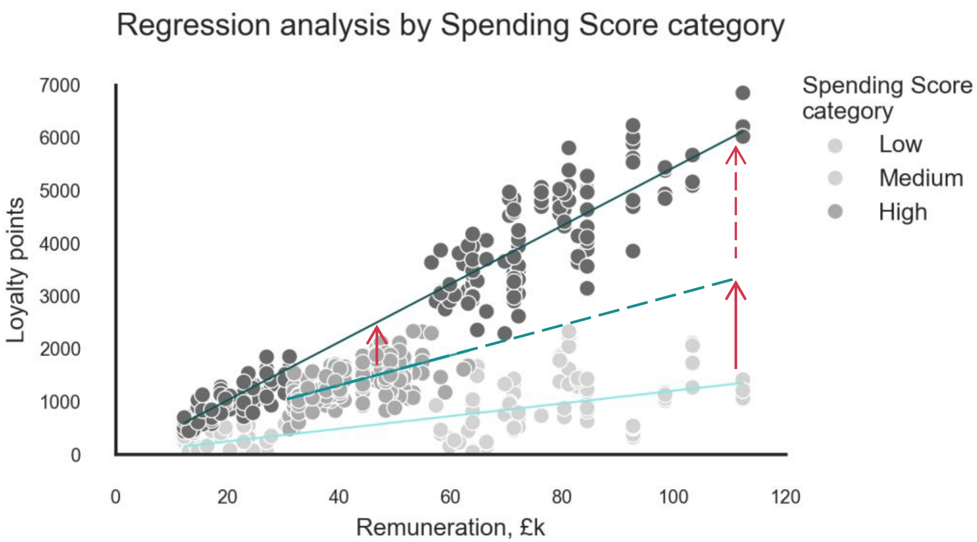
Customers can be grouped into five clusters based on spending score & remuneration, and three groups by spending score category:

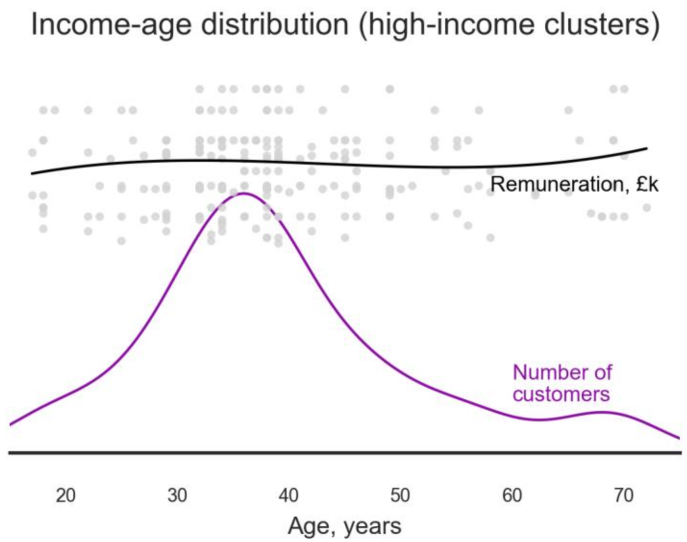
- <u>'High' spending score:</u>
  - Cluster 4: High-Value (high-income)
  - Cluster 2: Limited-Potential (low-income)
- <u>'Moderate' spending score:</u>
  - Cluster 3: Moderate (average income)
- <u>'Low' spending score:</u>
  - Cluster 1: High-Potential (high-income)
  - Cluster 0: Low-Value (low-income)

Assumptions and Limitations:

- When developing a marketing strategy by customer cluster, it is crucial to consider both cluster characteristics and cluster size in terms of customer count and revenue share. Since loyalty points represent the value of purchases, we extrapolate these results to revenue.
- We assume that the sample is representative of the overall population of customers. However, we note that the sampling approach appears non-random, potentially introducing bias. With 10 reviews per product, the sample may not reflect the actual sales structure and could misrepresent certain demographic groups or products.

Key patterns and observations





Regression analysis

Loyalty points have a positive correlation with remuneration and spending score, i.e. the higher the income and spending score the higher the loyalty points balance.

This indicates that increasing spending scores can drive loyalty points balances up.

Moreover, loyalty points demonstrate greater income elasticity in clusters with higher spending score categories.

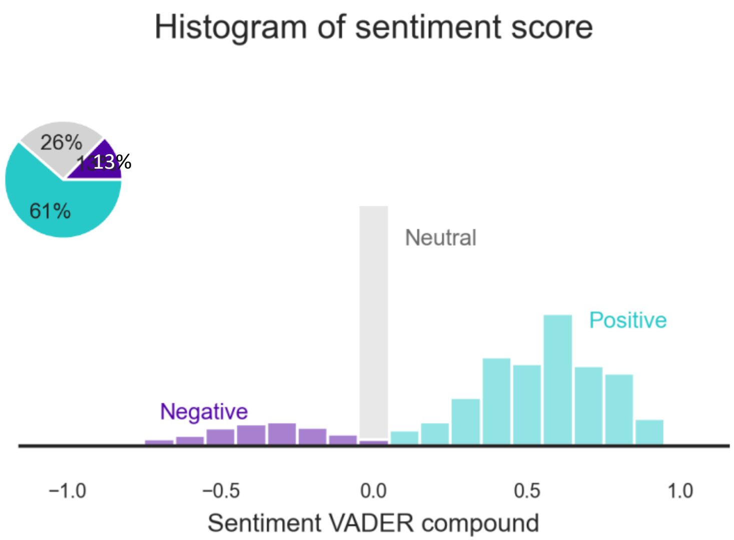
This suggests that the optimal strategy would be to focus on higher-income customers as an identical increase in spending score will yield a stronger response in loyalty points compared to lower-income customers.

Customer demographics

Remuneration shows little variation by age. At the same time, customers over 40 appear underrepresented in clusters 2 and 4, presenting a potential opportunity to grow the high-value segment.

Note: Income by age does not represent income over an individual customer's lifecycle but rather a snapshot of income distribution across turtle games customers across different age groups (the line is not adjusted for differences in customer mix across age groups). Though there are some limitations it must reflect the overall relationship between age and income

Sentiment analysis



Average score

-0.36 | 0 | 0.55

Note:
Compound sentiment score (ranges between -1.0 and 1.0):
- Score > 0.05: positive sentiment
- Score < -0.05: negative sentiment
- -0.05 <= Score <= 0.05: neutral sentiment

Sentiment score

- Sentiment analysis shows a wide range of scores, from -0.93 to 0.99, but positive sentiment dominates at 61%, with an average score of 0.55 in this group
- Customer sentiment scores and distribution patterns seem to be consistent across clusters and spending score categories

Predicting future outcomes

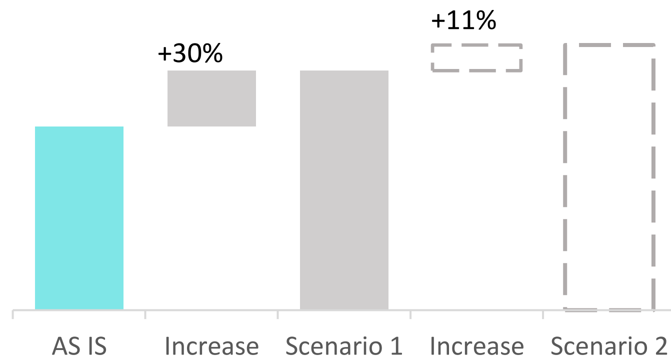
Scenarios

Average spending score per category

	AS IS	Scenario 1 "Base-case"	Scenario 2 "Best-case"
High-Potential (Cluster 1)	18	50	80
Moderate (Cluster 3)	50	80	80

"Medium"
"High"

Predicted total revenue



If the target spending score levels specific to each scenario are achieved, the total loyalty points balance is projected to increase by 30.2% in Scenario 1 and 10.7% in Scenario 2.

Recommendations summary

Existing Customer Campaigns

- Regression analysis suggests that the optimal strategy would be to focus on higher- and average-income clusters with lower spending scores targeting an increase in engagement:
    - Cluster 1: High growth potential with opportunities to increase loyalty points balances by 3–4 times
    - Cluster 3: Moderate potential with an expected 30% increase in loyalty points balances. This segment accounts for approximately one-third of the total customer base and revenue, making it strategically important despite smaller growth opportunities
- Cluster 4 (High-Value) Use premium loyalty programmes and personalised engagement to sustain retention and maximise lifetime value.
- Cluster 2: Limited potential for growth, with a price-sensitive audience that may respond to cost-saving promotions.
- Cluster 0: Low-Value segment with limited purchasing power and minimal growth potential. Marketing efforts for this group should be scaled back.

New Customer Acquisition

- Incentivise referral campaigns via Cluster 4 (High-Value) and Cluster 3 (Moderate) to encourage existing customers to recommend new clients. Leveraging customer networks may help to expand reach cost-effectively

Customer Sentiment & Engagement:

- Leverage customer surveys to address negative feedback and recurring concerns including product quality, unclear rules, and complexity and to guide product improvements.

Tracking and Metrics:

- Monitor sentiment and Net Promoter Score (NPS) to evaluate campaign effectiveness and loyalty trends over time.

Data Collection & Analysis Improvements:

- Address sampling bias in reviews and ensure representation across demographics.
- Introduce time-stamped reviews to measure campaign impact over time.