TURTLE GAMES

Analysis & Insights

Rubin Analytics

Asaf Rubin

24 April 2023

Background: context of the business scenario

Turtle Games (TG) is a game manufacturer and retailer with a global customer base. The company manufactures and sells its own products, along with sourcing and selling products manufactured by other companies. Its product range includes books, board games, video games, and toys. The company collects data from sales as well as customer reviews. TG has a business objective of improving overall sales performance by utilising customer trends.

To improve overall sales performance, TG has come up with an initial set of questions aimed at uncovering trends, patterns and insights based on data collected on 1) its' customers and 2) its sales figures.

In terms of its customers, TG would like to understand:

1. How customers accumulate loyalty points
2. How groups within the customer base can be used to target specific market segments
3. How social data (e.g. customer reviews) can be used to inform marketing campaigns

In terms of its sales, TG would like to understand:

1. The impact that each product has on sales
2. The reliability of its sales data
3. The relationships between North American, European and Global sales

---

Analytical approach

This analysis comprised two distinct parts to help TG answer their initial questions. The approach undertaken to analyse data from TG's 1) marketing and 2) sales departments were analysed as follows.

TG's marketing department provided quantitative and qualitative data gathered from 2000 customers. The quantitative portion of this data comprised information on customers' remuneration, spending score, loyalty points, while the qualitative data comprised textual game reviews and summaries. Python was used to analyse the customer dataset, along with the libraries *Pandas*, used to import and clean the data, *Numpy* used for additional data manipulation capabilities, *Sklearn* for more advanced analytical techniques such as machine learning and *nltk* for processing qualitative data used natural language processing techniques. Data was imported into Python and cleaned to address missing values, duplicates, inconsistent data types and to remove unnecessary columns. Thereafter, three primary analyses were performed in Python:

1. **Simple and Multiple linear regression** was used to investigate how loyalty points might be impacted by spending scores, age and remuneration.
2. **K-Means clustering** was used to understand how the customer base might be segmented for market targeting.
3. **Natural Language Processing (NLP)** was used to analyse customer reviews to help inform marketing campaigns.

TGs sales data on 352 different games, the platforms on which they are played and their respective North American (NA), European (EU) and global sales was analysed using R. The *tidyverse* and *dplyr* libraries were used to import and clean the data by addressing missing values and removing unnecessary columns. Three primary analyses were conducted in R:

1. Initial **Exploratory Data Analysis** was conducted to examine the distribution of sales data across different regions and platforms and to determine the impact that each product has on sales.
2. **Normality testing** was used to assess the reliability of the sales data.
3. **Regression analysis** was used to investigate the relationships between NA, EU and global sales.

Following conclusions of the analysis, patterns, trends and predictions were summarized to answer TGs initial set of questions, and business recommendations were made, as detailed below.

Visualisation and Insights

Regression was performed to assess the impact of spending score, remuneration and age, individually and collectively, on loyalty points. Scatterplots were used to depict the strength and direction of the relationships, and to assess model accuracy. Spending score, remuneration and age all have significant positive relationships with loyalty points, meaning that customers who are older, have higher incomes or have higher spending scores are more likely to have accumulated higher numbers of loyalty points.

K-Means clustering was used to segment the customer base into five distinct groups (clusters), based on their remuneration and spending scores. The distribution of these clusters was visualized on a scatterplot to help inform targeted strategies for each segment. For example, TG can prioritize high-value customers with incentives and implement loyalty programs for more moderate spenders.

NLP techniques were used to analyse customer reviews, identifying frequently used words (visualised in word clouds), and sentiment and subjectivity scores. The results of the sentiment analysis were plotted on a histogram, showing that most reviews are more positive than negative. Reviews were also analysed to assess subjectivity, and plotted on histograms, showing that most reviews are a mix of subjective and objective.

An initial exploratory data analysis was performed to examine distribution of sales data across different regions and platforms. The top 10 highest and lowest selling products in each region were identified and visualized on bar plots.

Sales data's reliability was assessed with tests for normality and correlation analysis, revealing that there is a wide variation in sales performance for different products and that the data is not normally distributed. This suggests the presence of a few highly successful games, while the majority have modest sales.

Sales data was used to model regressions, and to make predictions based on sales values provided. NA sales were shown to be highly correlated with global sales. EU sales had a positive but weaker correlation with global sales. The regression models show reasonable predictions for global sales based on regional sales. However, further investigation showed that the model's residuals are not normally distributed, which means that the predictions become less accurate for larger sales values, and should therefore be used with caution.

Patterns and predictions

There are 6 key findings most relevant to TG's objectives:

1. Spending score, remuneration and age account for 84.4% of the variance in loyalty points. This means as a customer's age, remuneration or spending score increases, as do their loyalty points. However, the non-linear relationship between variables means that the results of this analysis should be used with caution.

2. TG's customer base has five distinct clusters, based on remuneration and spending scores. TG can optimize marketing efforts, drive business growth, and enhance customer satisfaction by tailoring strategies for each segment. For example, TG can prioritize high-value customers with incentives, create loyalty programs for moderate spenders, and explore cost-effective engagement for low-income segments.

3. Most customer reviews are generally positive. Some of the most positive reviews include words such as "awesome" and "excellent", while some of the most negative include "boring", "difficult" and "disappointed".

4. TG has several top selling games that significantly outperform others and a larger number that sell moderately or low.

5. TG's sales data is not normally distributed, with wide variety in sales performance. The sales data analysis should be treated with caution and it is recommended that more data be collected and analysed to produce more reliable results to inform decision-making.

6. 91.99% of the variance in global sales can be explained by NA and EU sales, meaning that as NA and EU sales increase, as do global sales.
