APRIL 24, 2023

DA301 ASSIGNMENT REPORT

Advanced Analytics for Organisational Impact

LOMBARD, LOIC

Introduction

The Turtle Games senior management asked us (BI team) to analyze how to **improve overall sales performance utilizing customer trends** and data insights. A good opportunity to apply **IDEAL**[1)], our structured corporate problem-solving framework. Leadership has **I**dentified the problem and broken it down into leading detail questions (**D**efined goals) to instruct our thinking. Here we are **E**xploring the data for solutions and strategies with an outlook to **A**nticipate outcomes and possible actions. In 6-12 months, we will **L**ook back and refine our approach (continuous improvement).

Leading questions:

- How customers accumulate loyalty points
- How groups within the customer base can be used to target specific market segments
- How social data (e.g. customer reviews) can be used to inform marketing campaigns
- The impact that each product has on sales
- How reliable the data is (e.g., normal distribution, skewness, or kurtosis)
- What the relationship(s) is/are (if any) between North American, European, and global sales?

Note: [1)] https://www.tntech.edu/cat/pdf/useful_links/idealproblemsolver.pdf

Analytical approach

We received two data sets (customer review and sales) to inform our thinking. After cleaning and transforming the data (invalid entries, duplicates, renaming columns, breaking down datasets) we explored them visually and with statistics. We ran several analyses which findings we visualize and share below. In 6 to 12 months' time, we would like to reconvene and assess if our findings had impact.

We used Python (Reviews) and R (Sales), while the capabilities largely overlap, R was chosen because of the sales department's preference for it and to continue the conversation later to generate additional insights.

Customer reviews:

In Python, libraries numpy, pandas, matplotlib, seaborn were most helpful. Statsmodels library facilitated creating linear regression models. Review comment (sentiment) analysis were performed with NLTK, Wordcloud, Textblob and vaderSentiment.

The linear regression models informed on statistically relevant correlations between investigated variables. Unfortunately, the data has limited predictive viability due to a non-normal distribution of the underlying data, however skewness and kurtosis weren't detected.

As sentiment analysis can yield ambiguous (even crude) results, hence we decide to perform it with two different libraries, because each of them was trained on different data (VaderSentiment – social, Textblob – lexicons).

Sales:

In R, sales data packages tidyverse, plotly, reshape2, moments, car and psych were most useful. The analysis yielded a list of top selling video games, the regional distribution of global sales and revealed the cyclical (or blockbuster) nature of our business. We also ran (multi) linear regression analysis. Here too, the model revealed significant correlation in the data, but proved unsuitable for predictive linear

analysis due to non-normal distribution, skewness, kurtosis, and heteroscedasticity. We speculate that because purchasing a video game is an individual and complex process, it generates noisy data. In addition, the blockbusters generate outsized revenue that skew data further (Outliers).

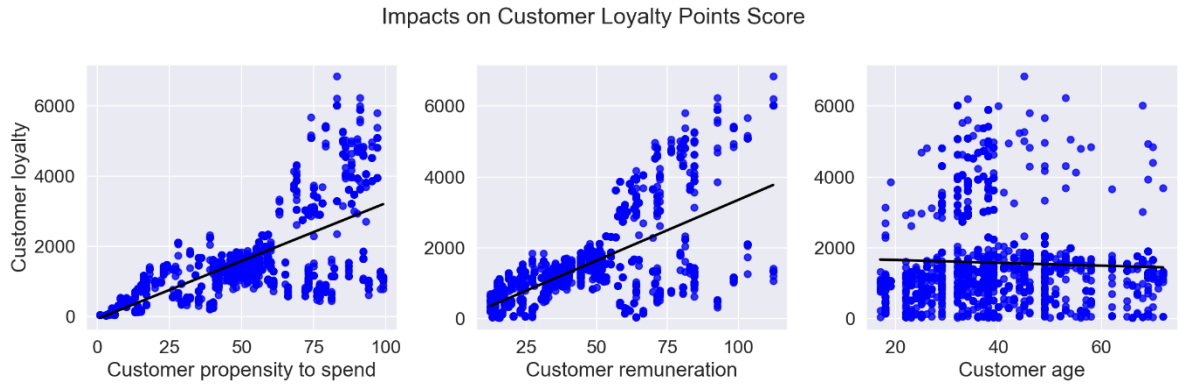
Data collection suggestions:

- Reviews: Segment by product & title
- Reviews: Add star rating (if available)
- Sales: Check rounding on EU and NA sales, two observations miss the “Year” field
- Sales: Increase granularity by country and region and add units sold/margin
- Sales: check platform affiliation of productid: 249 – looks wrong.
- Collect more data in general 😊

Visualization and insights

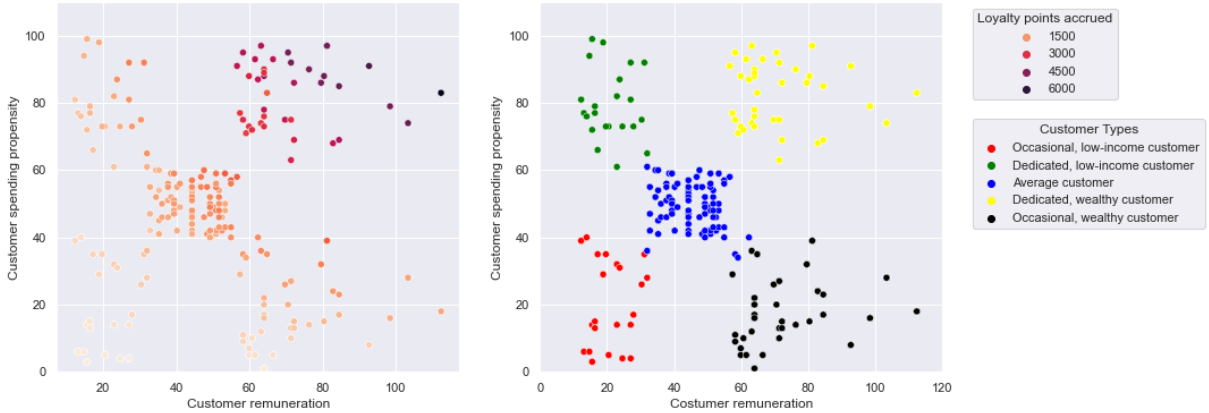
Loyalty:

Customers with a high earning and spending score tend spend more. No age wise correlation observed, this should be considered in age-targeted marketing campaigns.



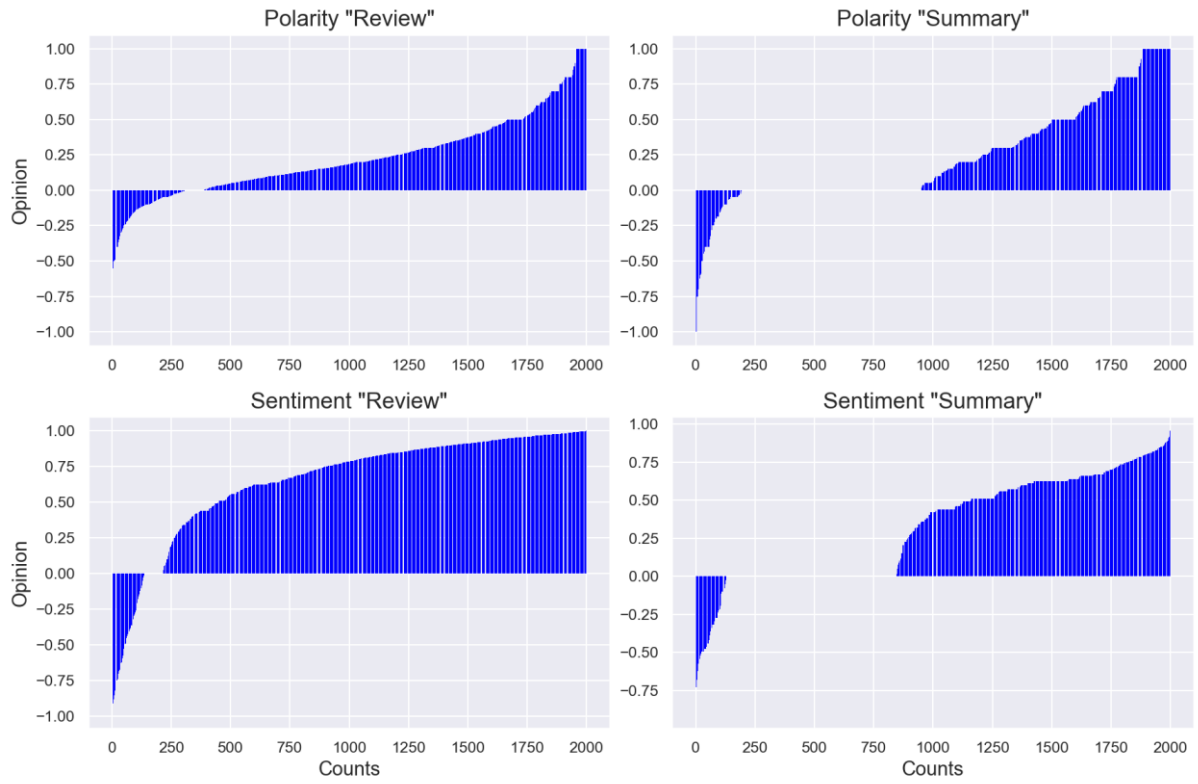
Five customer types

Comparing customer propensity to spend, renumeration and loyalty points, we see five, very clear clusters into which our customers can be segmented.



Review opinion

Even on the internet, where social media comments usually skew negative, our reviews are overwhelmingly favorable (80%+).



The top row “Polarity plots” have been analyzed with the textblob library, while the “Sentiment plots” with the vaderSentiment one. They align nicely. Summary comments, which are shorter, proved more ambiguous to both models than the - longer - review comments.

Top sellers

Aggregating the data by product Id reveals 3 categories of top sellers:

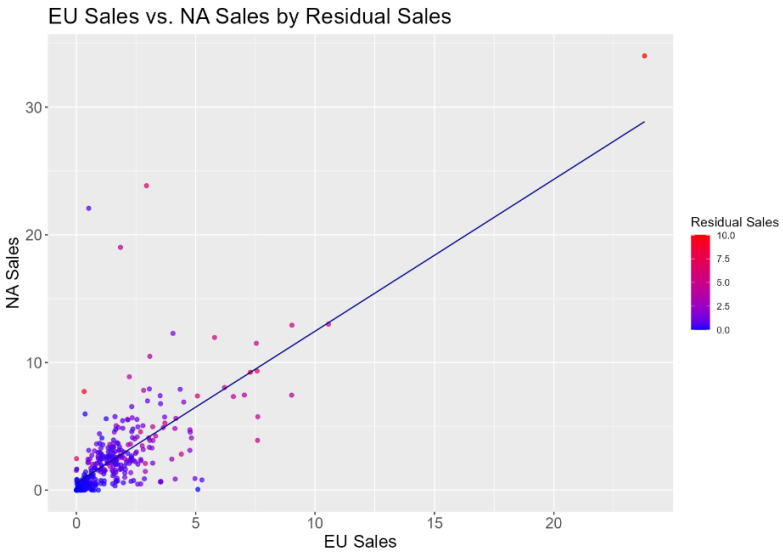
1. Titles marketed to a broad audience (Boys & Girls): Wii Sports - ID: 107

2. Well-made & successful multi-platform franchises:
    a. GTA: Grand Theft Auto - ID: 515
    b. Wii sports – ID: 107 & 231
    c. Mario kart racing – ID: 195 & 405
    d. Call of Duty – ID: 876
3. Highly anticipated consoles: NES or Gameboy color - D:123

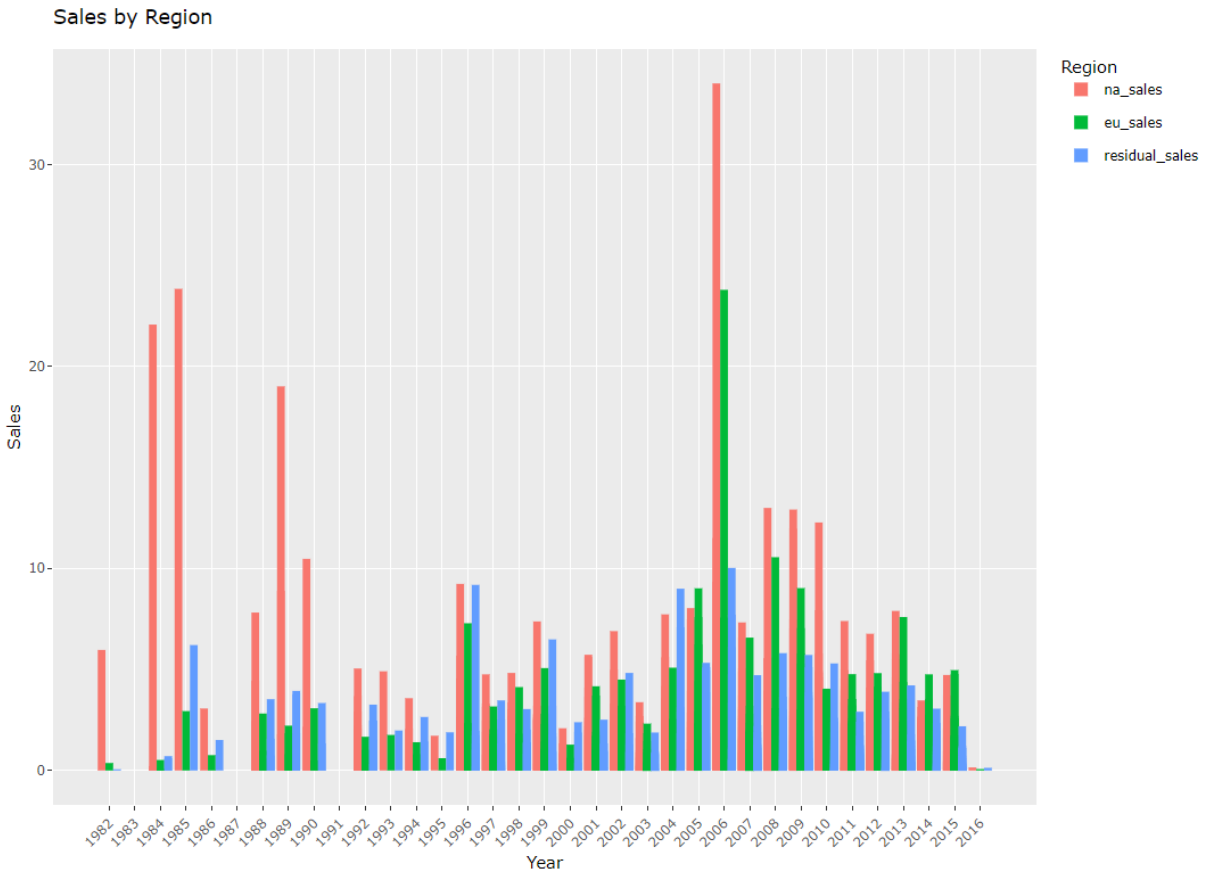
Any title falling into one or more categories does likely well. Key in categories 2 & 3 is title *anticipation*. One proxy for anticipation, could be buzz in niche online forums that can be quantified by NLP sentiment analysis.

Sales distribution

Two key figures describe Turtle Games sales well:



EU sales and NA sales show correlation: a hit in NA is likely one in EU. Color indicates residual sales, representing all sales outside of NA & EU markets. Most outliers are also spread roughly along the regression line. A third observation: Some titles were outliers in NA but not in Europe. Those are early titles in the 80ies when the NA market was more developed than the European one. As the next chart shows: Europe has caught up.



Long-term sales distribution shows NA leading in the 80ies and a low frequency of video game releases. As video games get more popular over time, releases increase, and sales become more reliable.

Patterns and predictions

Customer Type

The occasional customer groups have equivalents in remuneration that spend significantly more, the dedicated customer groups. By closely investigating the dedicated customer groups and identifying the difference to the occasional ones, we surely would find measures to convert occasional into dedicated customers.

1000 True Fans

Our reviews are very favorable. Dedicated customers have high conviction and the market rumors well. We propose selecting 1000 customers, characterized by multiple favorable reviews, high loyalty points score and impactful social media presence. Catering to them with special promotions/events tuning them into a community of “1000 true fans”. They could serve as amplifiers, sensitive market survey and brand ambassadors boosting Turtle Games popularity and sales.

Blockbusters

Sales show three types of video games that sell well: Titles of broad appeal, franchises, and anticipated consoles. We suggest searching for new titles that fill those criteria.

Cross sales

Dedicated customer types are engaged with our products. Sales data shows that these are likely franchises, which often have merchandise alongside the main product line on offer. They present a group that merchandise can be marketed to. Since we already have a broad product range in which the merchandise likely falls, we could add it with little effort.