TURTLE GAMES

Report

Scenario Context

Turtle Games is a global game manufacturer and retailer who would like to use customer trends to improve overall sales performance.

Analysis of the provided customer and sales data will uncover and present findings with which the marketing manager can use to inform decision-making related to the project goal.

Questions for Turtle Games

- Is it possible to access data related to identifying which products Turtle Games manufacture and which they source from other companies?
- Is it possible to access customer data related to recency, and frequency, time on website?

Further Questions of the Data

- Could more information be gleaned from partitioning ‘Other’ sales data?
- Could advertising be streamlined further by breaking Europe sales data into individual countries?

Analytical Approach

Multiple Linear Regression (MLR)

- I initially created separate Simple Linear Regression (SLR) models, but none proved to be strong in their predictive power.
  - This is often the case with SLR as it is rarely apparent that a dependent variable is explained by just one other variable.
- MLR proved more comprehensive while using the same variables ('loyalty points' with 'age', 'renumeration', and 'spending score').
  - Variance inflation factor found there to be very little chance of multicollinearity.
  - The Breusch-Pagan test returned ***p = 1.14*** which indicates homoscedasticity.

K-Means Clustering

- Initial pair-plot highlighted potential outliers ('renumeration').
  - However, descriptive statistics of the reviews dataset showed there not to be any, and as 'renumeration' is a naturally wide-ranging variable, I did not remove any data.
- Both the Elbow and Silhouette methods showed the optimum number of clusters to be **5**.
  - However, *k = 6* separated the previously-highlighted potential outliers into a distinct cluster.
    - This may allow for increased accuracy in advertising efforts for these customers. Although, there are few of them and it may not be worth the expense to isolate them at this stage.
    - This model may also be overfit and might become inaccurate with a larger dataset.

Sentiment Analysis

- Although punctuation may keep certain sentiment (Sun, 2018), I chose to remove all punctuation so as to analytically treat each text equally (Rastogi, 2022).
- I chose to include the `product` code so as to identify the specific product that each review is referring to.
  - I also chose to include `age` and `renumeration` data as it may be useful to identify particular demographics of customers who leave reviews of this nature and for these products.
    - As the project continues, other variables could be included in this matrix to gain a deeper understanding of these customers.

- Initial visualisation highlighted a significant outlier in the Global sales data for the Wii platform.
  - There were some outliers in the North American (NA) and European (EU) sales data, but I decided that these were more likely due to strong-performing products and only removed the initial outlier as it may have significantly impacted statistical analysis.
    - This decision may have led to under-representing Turtle Games' sales performance.

- QQ-plots, Shapiro-Wilks, skewness, and kurtosis all indicated that sales data was asymmetrical and not normally distributed.
  - I took log square-root of the data but failed to normalise sufficiently, so conducted the more appropriate Mann-Whitney U Test which returned ***p = 0.0005501***.
    - This indicates that the difference between the NA and EU population medians is statistically significant.
      - MLR model A proved the strongest fit.
        - However, given the relatively unreliable nature of the data, I decided to pursue model B with transformed data.
        - Both models have limited reliability with their predictive accuracy, and this decision may result in more inaccurate findings.

Patterns, Predictions & Insights

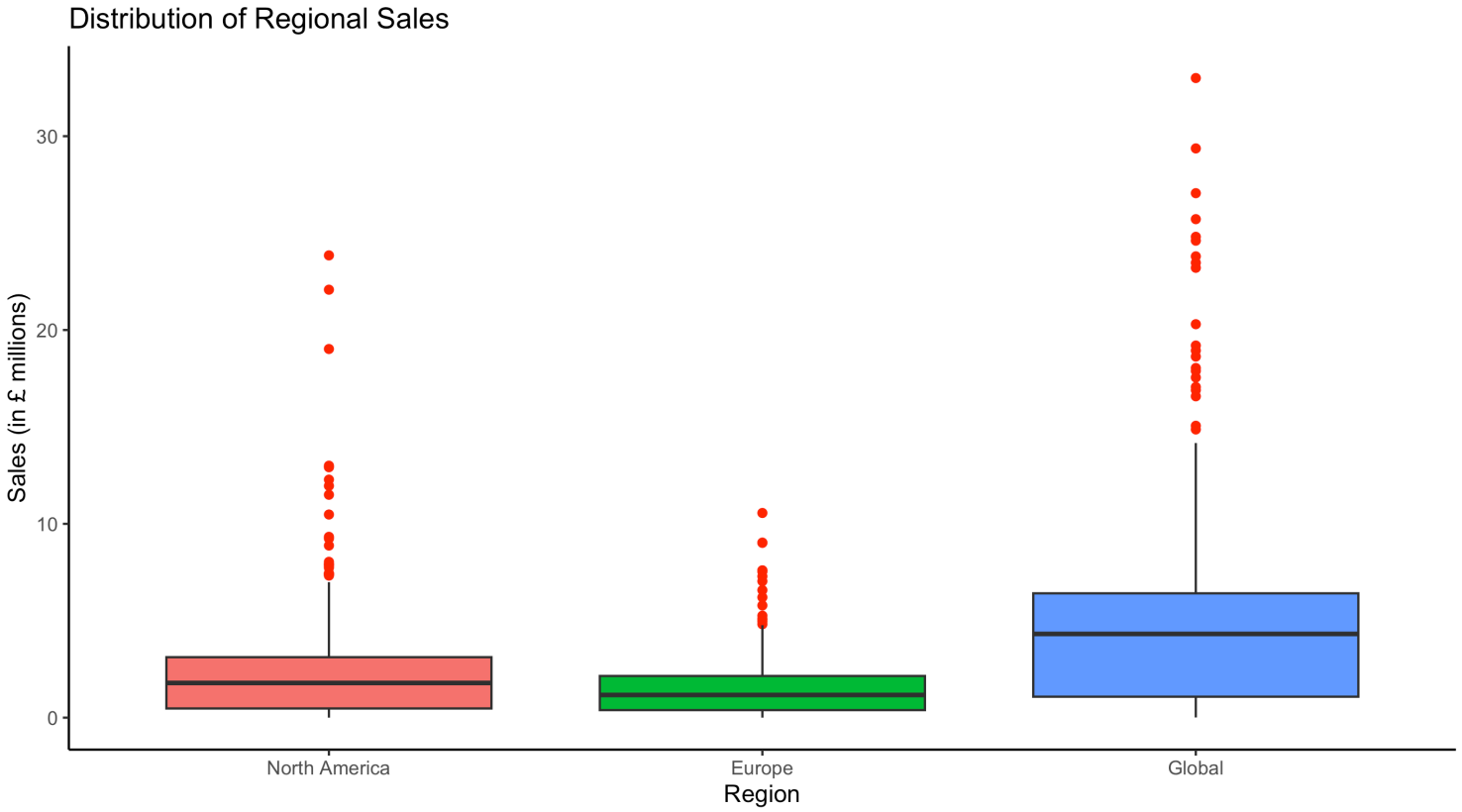
MLR

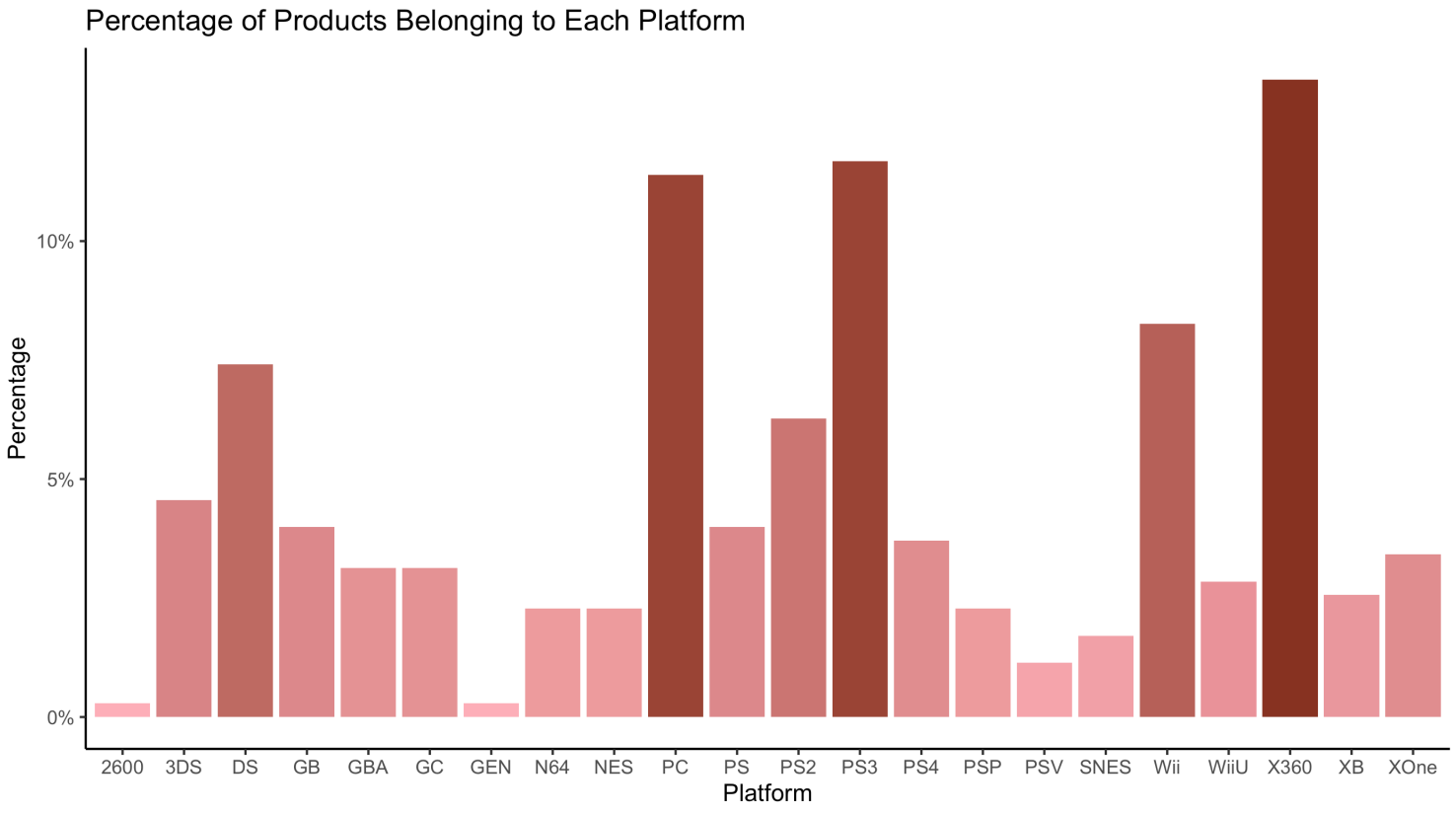
Customer Data

- Roughly **85%** of the variation in customers' loyalty points can be explained by the independent variables with a confidence of **95%**.
    - For example, provided customers have the same ‘age’ and ‘spending score’ values, then for every **£1,000** increase in their salary, the model estimates them to have a **33**-unit increase in ‘loyalty points’.
- **R^2 = 85%** is fairly strong, but a risk factor of **15%** is still significant and Turtle Games may want to assess the acceptability of the risk that a customer may not remain loyal under this model.
    - Accessing data related to recency, frequency, average basket size, etc. would be useful to strengthen the model and help determine a cohesive customer lifetime value.
    - As the project continues, I would recommend strengthening the model by evaluating misclassifications.

Sales Data

- Model B returned **Adj.R^2 = 0.9557**, indicating that Global sales are strongly predicted by EU and NA sales.
    - The model estimates that an increase in EU sales of **£1 million** would result in an increase of **£0.7 million** in Global Sales.
    - An increase of **£1 million** in NA sales would result in an increase of **£0.8 million** in Global sales.
        - These coefficients indicate that NA and EU sales are positively correlated with 'Other' sales from the rest of the world (which are not included in this data set).
        - It is perhaps worth taking into consideration that Europe, on average, has a lower cost-per-click rate than North America (Irvine, 2022).


Distribution of Regional Sales
Sales (in £ millions)
30
20
10
0
North America
Europe
Global
Region

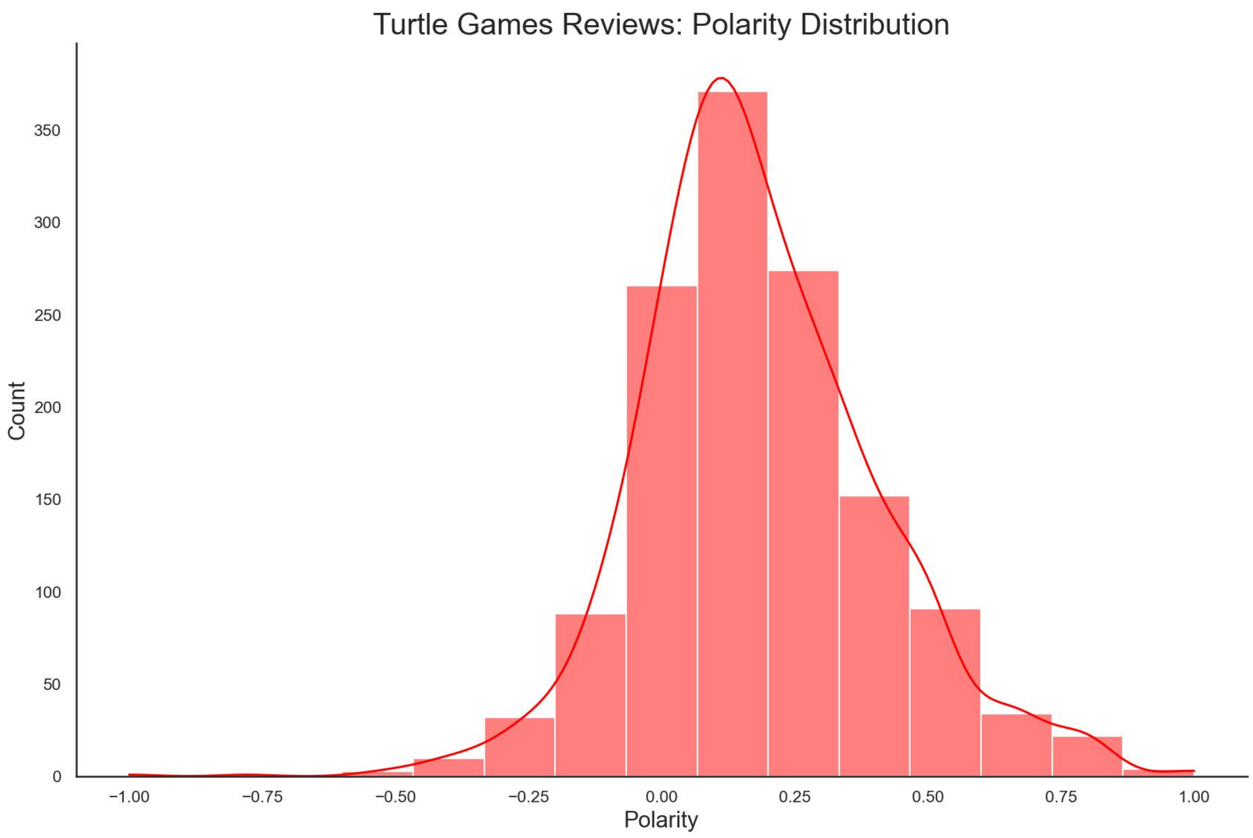

Percentage of Products Belonging to Each Platform
Percentage
10%
5%
0%
2600
3DS
DS
GB
GBA
GC
GEN
N64
NES
PC
PS
PS2
PS3
PS4
PSP
PSV
SNES
Wii
WiiU
X360
XB
XOne
Platform

- Most of Turtle Games' console products are sold for **Xbox 360** and **PS3**, which are no longer in production.
  - In 2021, Playstation 5, Nintendo Switch and the Xbox S were the highest-selling game consoles worldwide (Owens, 2022).
    - It may be worth looking to expand product catalogues to include games for these consoles and relinquishing products for more niche platforms, eg. GEN (Sega Genesis).
- In 2020, there were an estimated 1.75 billion PC gamers worldwide, an increase of 0.25 billion from 2019 (Clement, 2022).
  - As Turtle Games has a strong foothold in this platform's market, it should retain focus on such products.

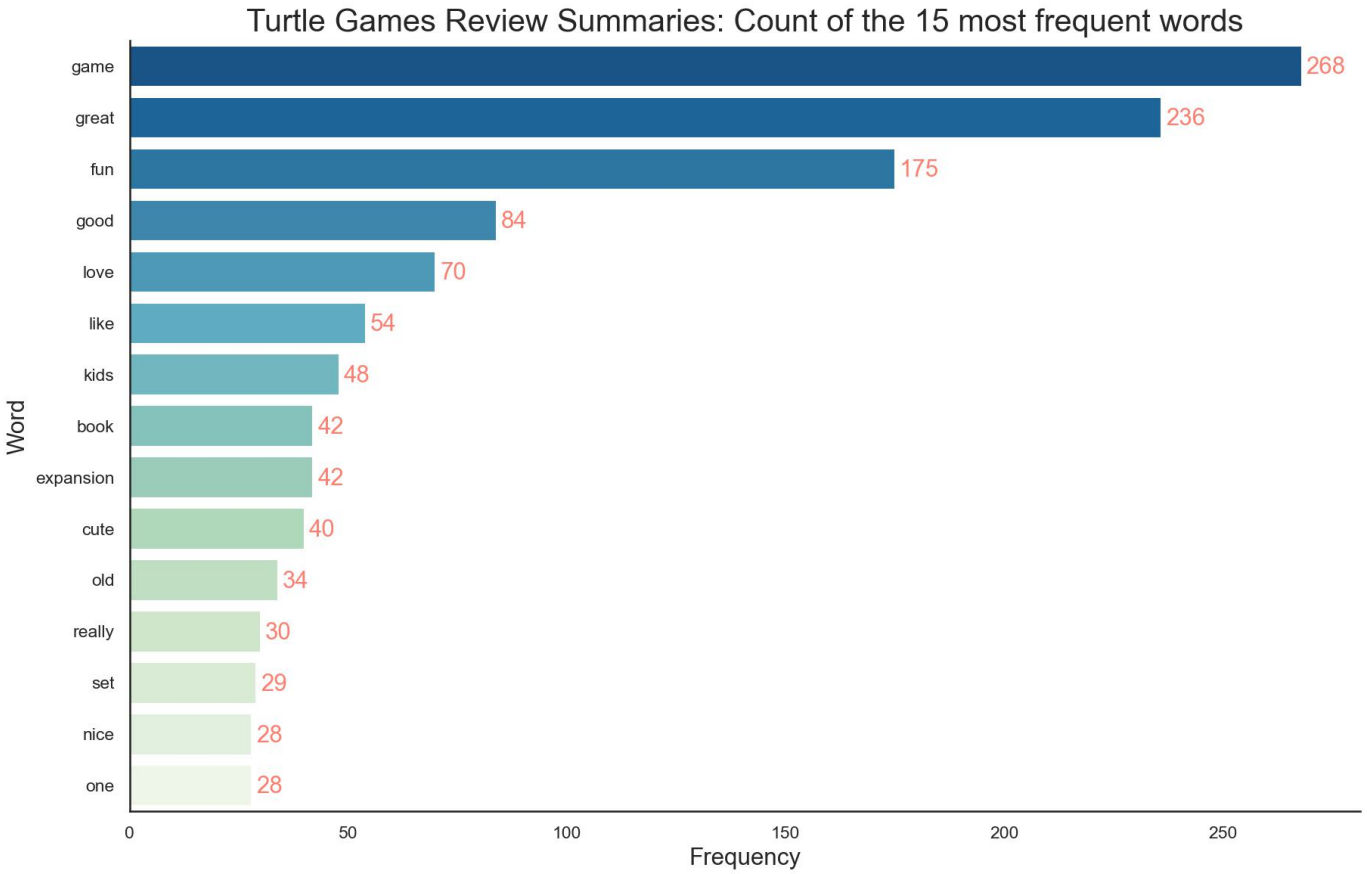
K-Means Clustering

- The 5 distinct spherical clusters can be used to pinpoint customers' demographic data collectively, allowing for a more efficient, effective advertising strategy for those in particular salary bands, for example.
  - As the project progresses, it may be useful to gain access to third-party website data so as to understand customers' preferences and habits in more detail, compound analysis, and further streamline advertising strategies.
    - However, this is dependent on the marketing budget.

Sentiment Analysis



- This above plot shows the reviews polarity data to be normally distributed with a slight skew to the left, meaning most reviews are generally more positive than negative.
  - This affirms the initial interpretation of the word clouds.



- "**Book**" and "**Expansion**" are mentioned often in a favourable manner. It may be beneficial to focus on these types of products as the project continues.

- Textblob subjectivity analysis determined that most reviews are neither particularly subjective or particularly objective, and range relatively evenly throughout the two.
  - However, limitations of the TextBlob sentiment analysis capabilities are apparent here:

" best gm screen ever "

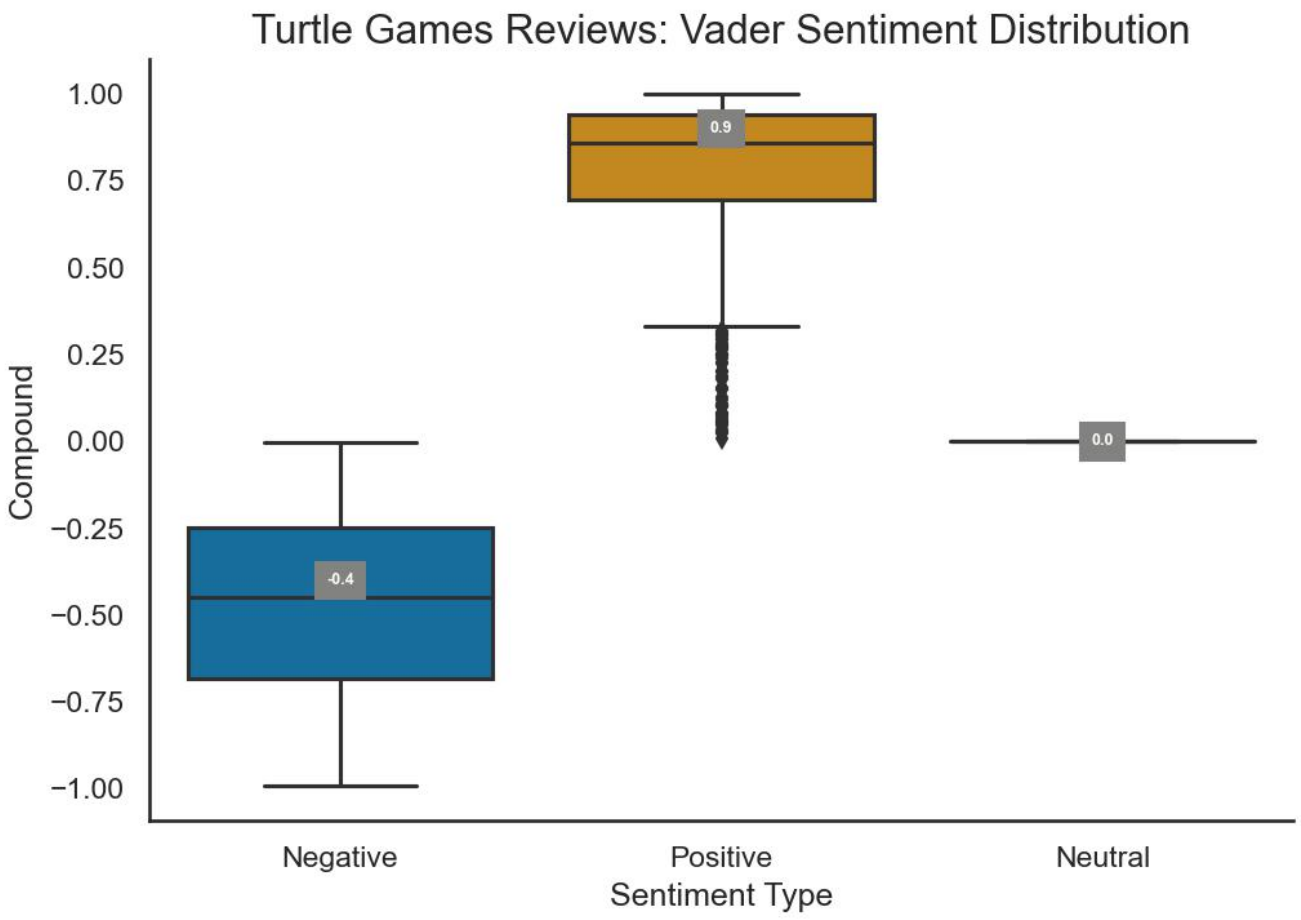
    - The subjectivity score is relatively low (**0.3**).
    - One would expect it to hold a much higher degree of subjectivity as it is clearly the opinion of this customer.

Product	Summary	Polarity	Subjectivity
10196	“ perfect for preschooler ”	1	1
10196	“ wonderful gift ”	1	1

- Product **10196** may be a star product as has been reviewed extremely positively by two customers.
  - Perhaps it is worth investigating this further so as to potentially feature this product in specific advertisement campaigns.

Product	Summary	Polarity	Subjectivity
1501	“ disappointing ”	-0.6	0.7
1501	“ very weak game ”	-0.4	0.6

- Turtle Games may want to investigate why these customers gave product **1501** such negative reviews and address the customers directly with assistance regarding their issues.



- Vader sentiment analysis showed there to be significantly more positive reviews (**n = 1208**) than negative (**n = 119**). This additional model provides affirmation.
  - The positive reviews (**median = 0.9**) have a smaller range than negative (**median = -0.4**), indicating that most positive reviews were very positive.
    - It may be worth looking to include quotes of positive reviews in advertisements.