

Turtle Games

Business Intelligence Report

**Overview**

Turtle Games is a global game manufacturer and retailer, which sells its own products as well as products manufactured by other companies.

This project report provides insight into how Turtle Games can improve sales performance; by outlining key insights into customer demographics and purchasing habits, their sentiment towards the current products and sales trends between regions and globally.

**Datasets**

Turtle Games provided two datasets for analysis of customer, product, and sales trends to inform the development of a new sales strategy. A metadata file was also provided with details of the data and the processing techniques already undertaken.

The ‘Turtles Reviews’ dataset provides details on the demographic of 2000 customers and their reviews of products. The ‘Turtle Sales’ dataset details the sales for each product in North America, Europe and globally along with the ranking, platform, year, genre, and publisher of the product.

Each dataset was checked for completeness and accuracy before being analysed and transformed into data visualisations, using either Python or R. The choice of programming language used for each dataset was as per the request of the department who supplied the data to align with existing business processes and facilitate further data investigation internally.

The following assumptions were made about the data provided:

* Data is assumed to be accurate as of December 2022, but no information provided on the timespan that either dataset covers
* All sales figures have been provided in £ British pounds and accurate conversion is assumed to have been undertaken
* All reviews were provided in English and accurate translation of the customer’s review is assumed.

An important note is that customers who choose to leave reviews are self-selecting and therefore may not be a true indication of customers’ thoughts.

**Analytical Approach**

Several analytical approaches were used to gather insights from the datasets into customer demographics and their sentiment towards the products offered, along with sales trends. The approaches used and reasoning are outlined below.

**Turtle Reviews**

The reviews dataset was analysed using Python to gain insights into Turtle Games’ customer base, how customers accumulate loyalty points and their reviews of products.

The describe () function was used to view the basic descriptive statistics for the Reviews dataset and gain exploratory insight into the customer demographics.

A good indication of each customer’s purchasing history in the dataset is their ‘Loyalty Score’, which is the monetary value of their purchases converted into points. Regression analysis was used to understand which factors influence the amount of loyalty points a customer has and therefore their purchasing history.

To ensure the data was suitable for regression analysis, the Breusch Pagan test was performed to check if the data was homoscedasticity. Different factors were analysed using the Ordinary Least Squares (OLS) regression technique to determine how they influence the number of loyalty points a customer holds.

To help Turtle Games’ segment their customer base and design targeted marketing and sales campaigns, a k-means clustering algorithm was applied to the dataset to group customers which share similar remuneration and spending scores. K-means clustering works by grouping data by similarities in distance from the centroid data point of each group.

Using k-means clustering to group customers four distinct groups were identified. Each customer is clearly within a market segmentation and allowing for the development of tailored marketing and sales strategies.

Sentiment analysis was performed using the NLTK library on customers’ reviews to gain insight into how customers feel about the products sold by Turtle Games.

**Turtle Sales Dataset**

The sales dataset was analysed to gain insight into sales trends of products and whether there is any relationship between North America, Europe, and Global sales.

Exploratory data analysis was undertaken on the dataset through plotting basic visualisations using R’s qplot function. These plots provided high-level insight into the relationship between North America and European Sales, and the top selling products and platforms which helped determine how to undertake further analysis.

The normality of the dataset was investigated by harnessing R’s statistical insight capabilities by performing a Shapiro-Wilk test, determining the skewness and kurtosis and the correlation between the sales data from all regions.

The relationship between North America, European and global sales was modelled using simple and multiple linear regression models.

**Insights**

**Customer Demographics**

Based on the ‘Sales’ dataset most Turtle Games customers are female (0.56%), with an average age of 39.49, are higher educated (88% with a graduate level of education or above) and average annual income of £48,079.60.

Chart, bar chart

Description automatically generated

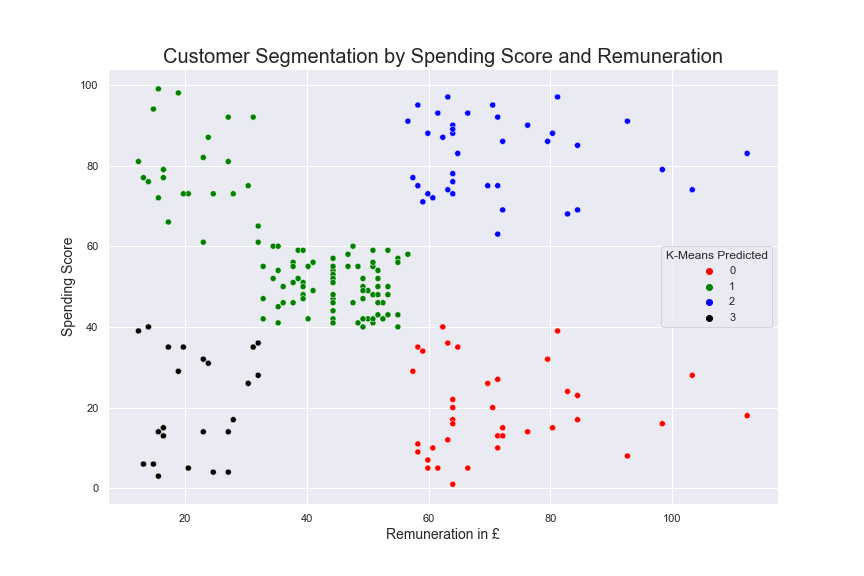
**Loyalty Points**

R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Using OLS regression analysis, to test how various factors influence the loyalty points (in the case y variable) we can see a low to medium correlation between customers' spending score and loyalty points with a R squared value of 0.452 and between remuneration and loyalty points a R squared value of 0.355.

The correlation between customers' age and the amount of loyalty points is almost non-existent. The results of the regression analysis have been plotted with a line of best fit to demonstrate the impact these variables have on customers’ loyalty points.

**Customer Segmentation**

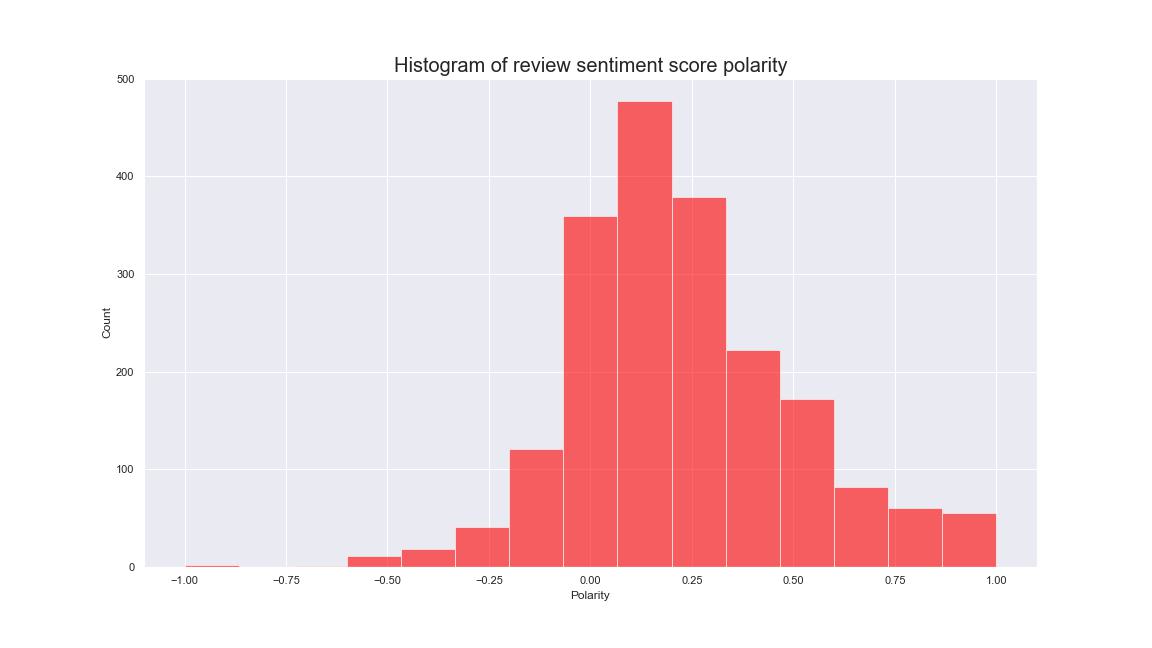
Using the k-means clustering technique there are four clear distinct customer groups within the Reviews dataset, as demonstrated by the visualisation below.

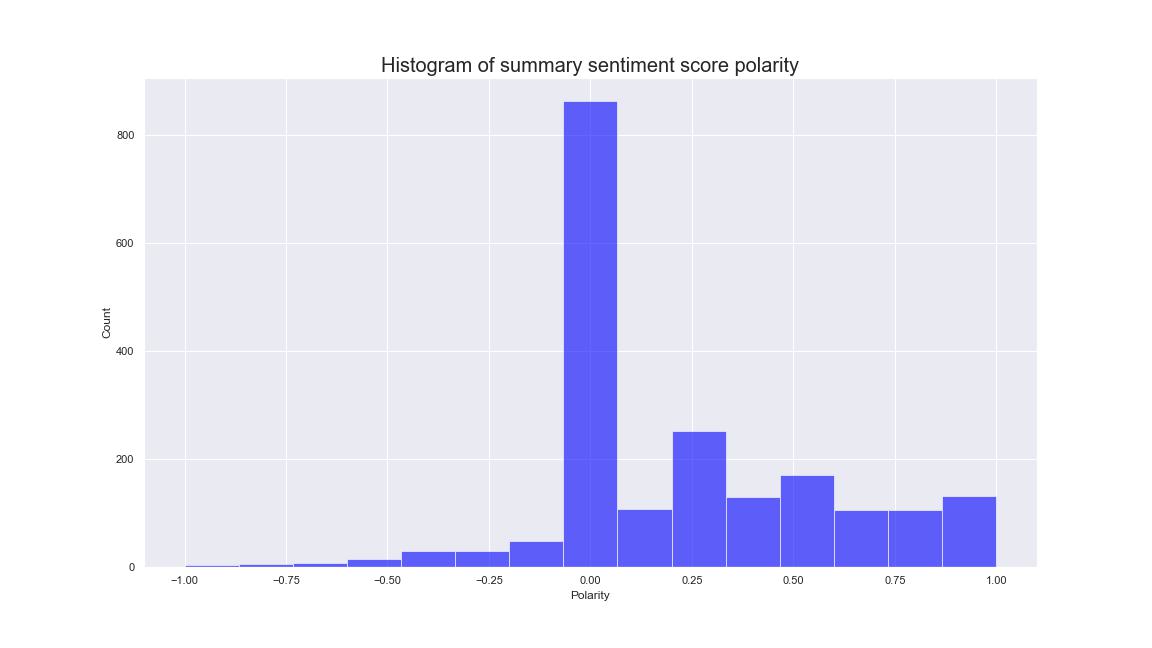


|  |  |  |
| --- | --- | --- |
| **Cluster** | **Number of Customers** | **Description of Remuneration and Spending Score** |
| 0 | 1013 Customers – 50.6 % | High Remuneration – Low Spending Score |
| 1 | 356 Customers – 17.8 % | Low Remuneration – High Spending Score |
| 2 | 351 Customers – 17.55 % | High Remuneration – High Spending Score |
| 3 | 280 Customers – 14 % | Low Remuneration – Low Spending Score |

**Customer Sentiment**

Sentiment Analysis was undertaken on the online review left by the customers. Sentiment polarity scores are assigned on key words left in the review and assigned a value between -1 (lowest negative sentiment) to +1 (highest positive sentiment). The analysis shows customers have an overall positively skewed sentiment towards Turtle Games products.

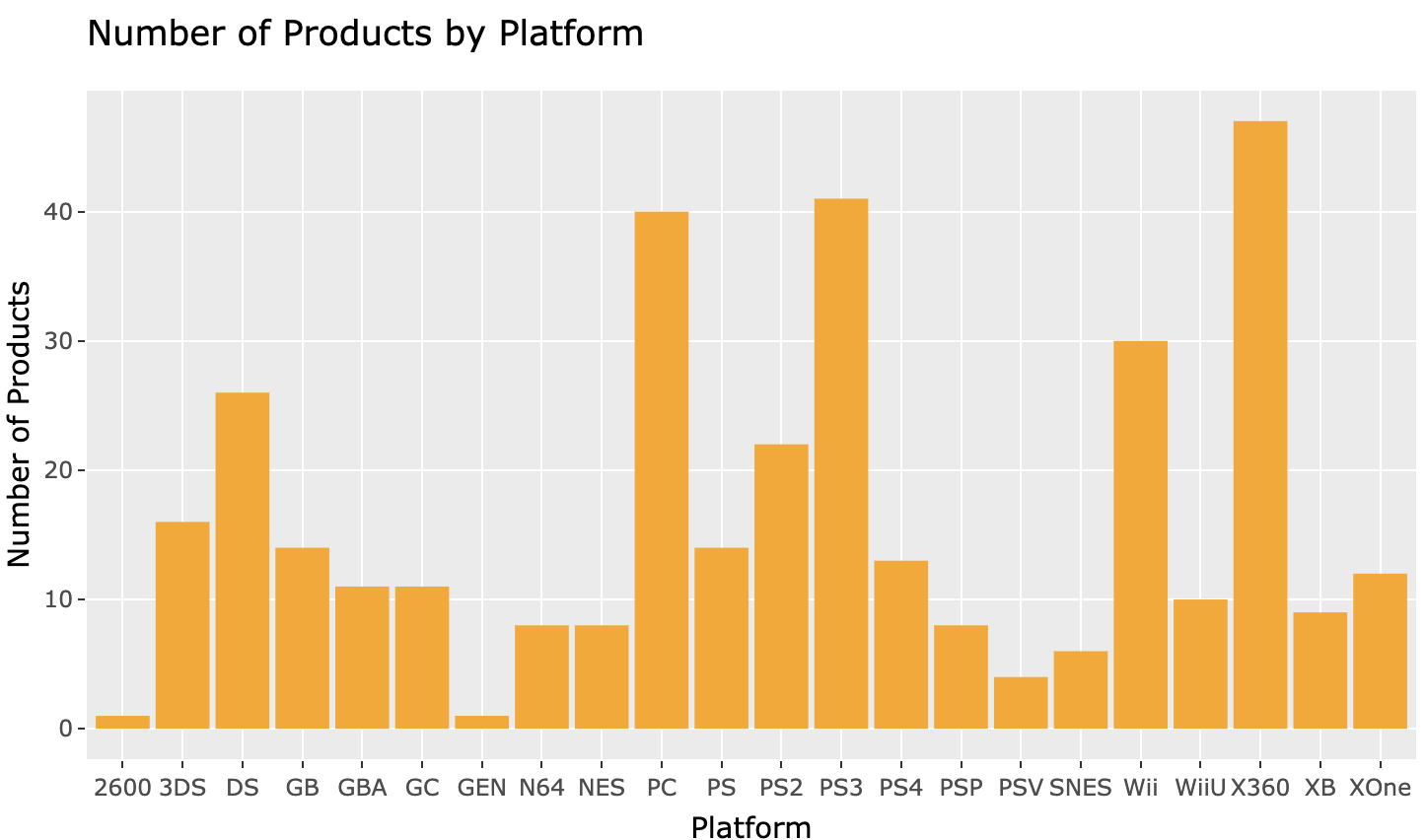
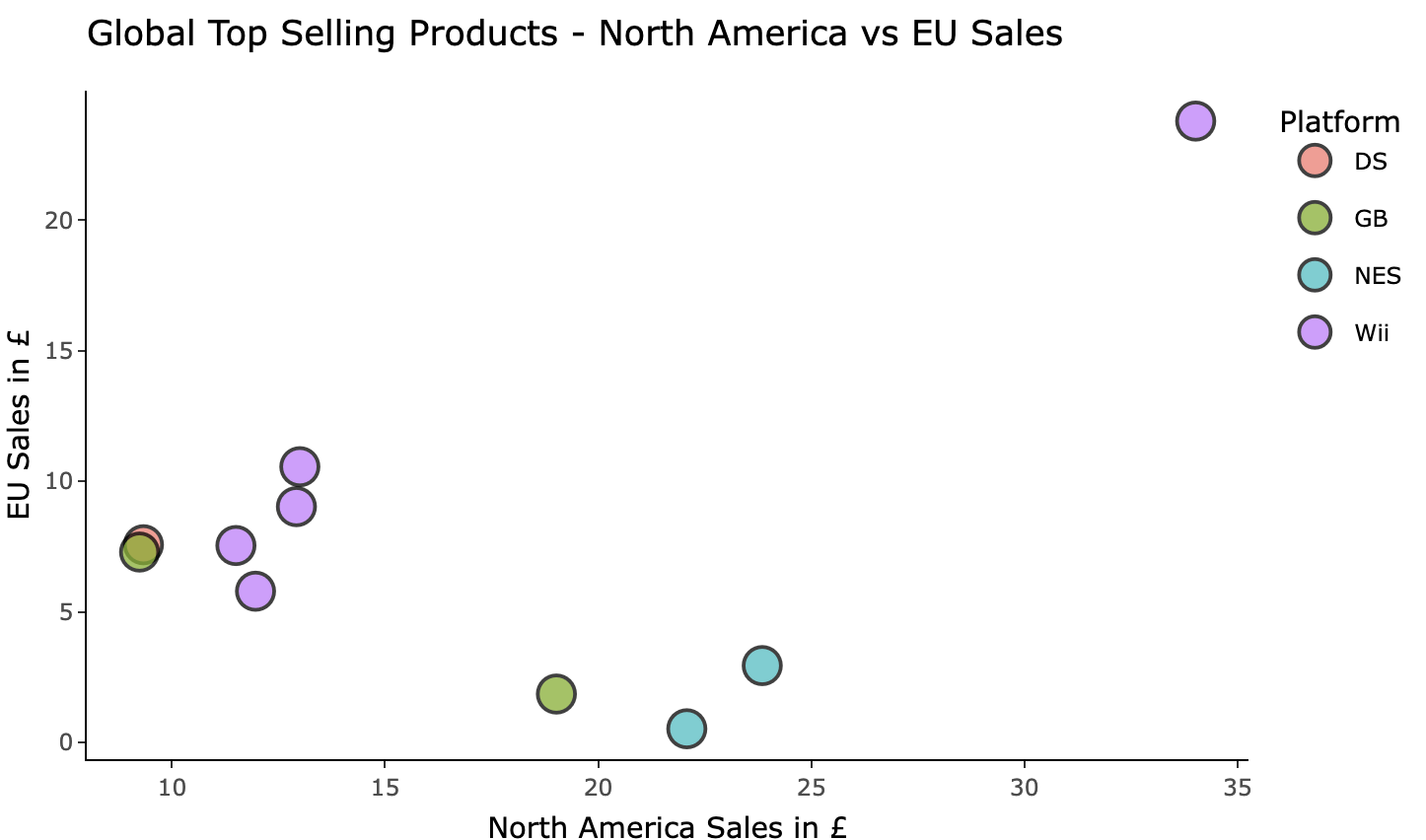




**Sales Trends**

North America and Europe represent most sales for Turtle Games, and there is one clear best-seller product offered by Wii. As demonstrated by the scatterplot below, there are differences in which products sell best in each region.

The products with the most sales are available on Wii, SNES and NES platforms, however Turtle Games offers the most products on X360, PS3 and PS4 platforms. There is an opportunity to provide more products from the best-selling platforms to increase sales.

****

**Key Insights & Recommendations**

* As there are differences in which products sell best between North America and Europe different sales strategies are recommended to optimise sales.
* Investigation into what products are selling best and not so well for Turtle Games, and whether the product offering can be improved
* There is an opportunity to improve customers’ sentiment of products. Further investigation is recommended into the negative reviews and whether this is product related and product offering could be changed.