Turtle Games

Objective

Turtle Games has a business objective of improving overall sales performance by utilising customer trends.

Data Sets

turtle_reviews.csv

turtle_sales.csv

Methodology

- Python and R were utilised to clean, explore, manipulate, and visualise the data to answer the key questions from Turtle Games

- DataFrames were sense-checked, errors were checked
- Descriptive Statistics was utilised to describe the data in statistical context
- Data was explored to determine unique identifiers for more accurate analysis
- String variables were cleaned to ensure uniformity.
- K-Means clustering was used to determine loyalty points accumulation.
- NLP was used to determine how social data can be used to inform marketing campaigns
- Data visualisations were used to explore the data and determine distribution.
- Exploratory data analysis (EDA) techniques were performed to clean and manipulate the data so that we can determine the reliability of the data.
- Regression techniques were used to determine possible relationships between variables within the sales data.

Data Analysis by:
Florida_Phoebe
LSE Data Analytics Career Accelerator

Findings

(1) How customers accumulate loyalty points:
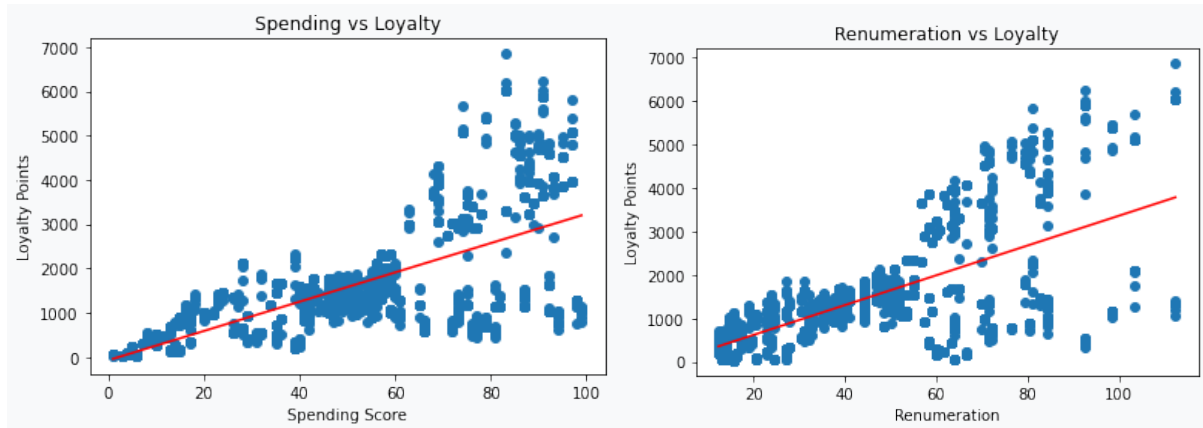
Renumeration - Total income per customer per year in pounds, where k=1000

Spending Score (1-100) - A score is assigned to the customer by Turtle Games based on the customer's spending nature and behaviour. The value ranges between 1 and 100.

Linear regression techniques show that renumeration and spending score can be used to know how customer accumulate loyalty points.

The higher the renumeration and spending score, the higher the likelihood of the customer accumulating loyalty points.

(2) How social data (e.g. customer reviews) can be used to inform marketing campaigns

Cluster 2 (BLACK) – high renumeration but low spending score

This customer base can be targeted for marketing, they got higher income but have low spending score therefore this customer base is not spending as much.

Marketing campaigns for higher range products can be implemented for this group to try and increase potential profit from this group.

Cluster 4 (GREEN) – low renumeration but high spending score

This customer base can be investigated as to which products this group spends on. This group can be targeted to increase product reviews through incentives such as discounts per every product review. Which can help future product review analysis.

(3) How social data (e.g. customer reviews) can be used to inform marketing campaigns

Histogram of sentiment score polarity
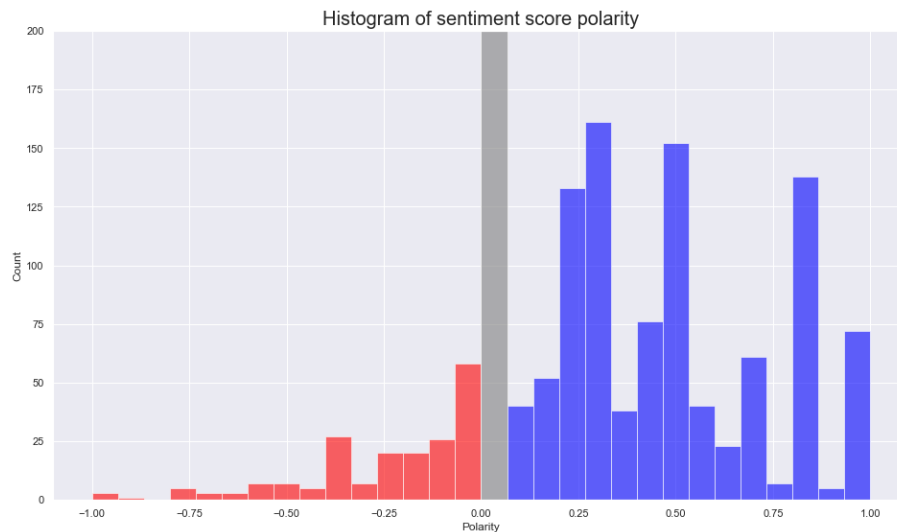
The histogram of the overall polarity score shows a distribution that is right skewed which can mean that the general sentiment of the customers is positive.

Most Common Words in Bad Reviews
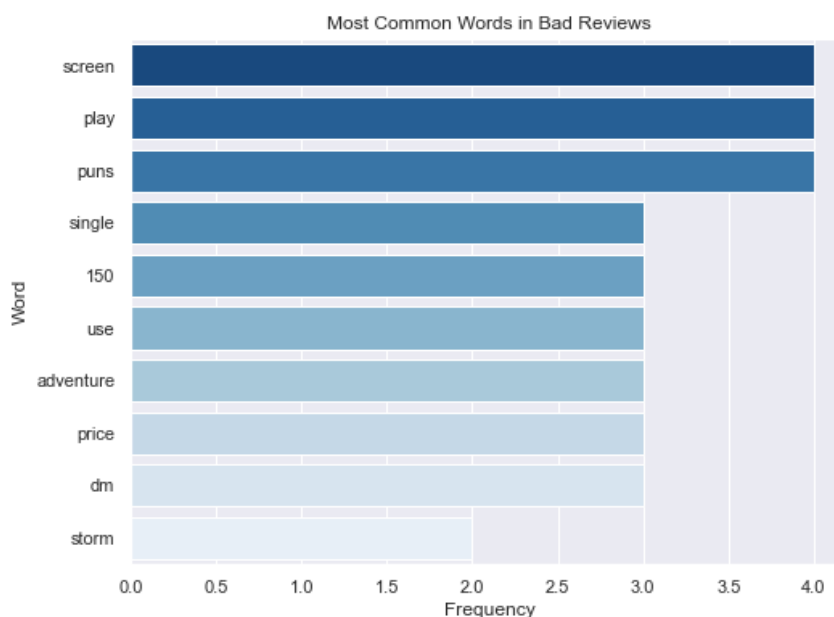
There are some negative sentiments towards the words "screen", "play" and "puns". We cannot deduce what this analysis can imply but we would suggest further investigation on products that contain screens to determine what negative sentiments are being shared towards those products.

Most Common Words in Good Reviews
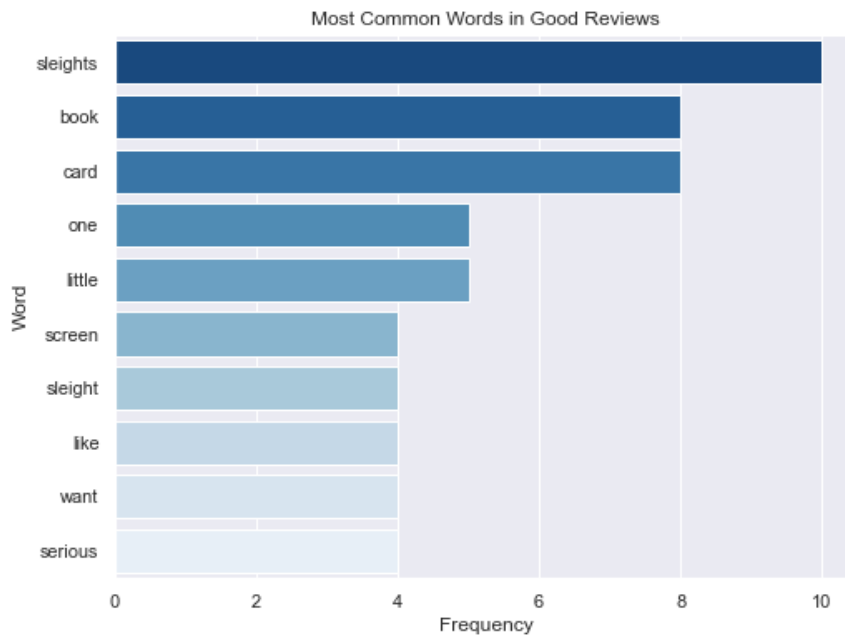
There are some positive sentiments towards the words "sleights", "book", and "card". Further investigation is recommended to find out which products these sentiments were referring to and maybe those products can be used to target marketing campaigns for an increase in positive sentiments on the web.

(4) The impact that each product has on sales

According to the metadata, "Product" is a numerical value which is a unique code allocated to the product based on the item description.

Upon investigation, the product number is not unique at all.

We have used "Ranking" as a unique identifier instead, to determine how certain products by genre, publisher, platform and release year could impact global sales.
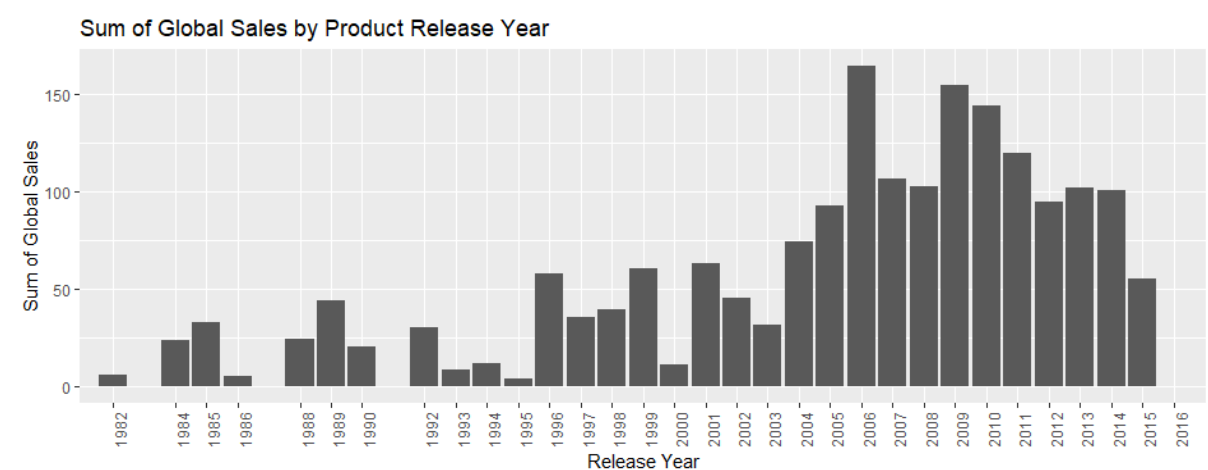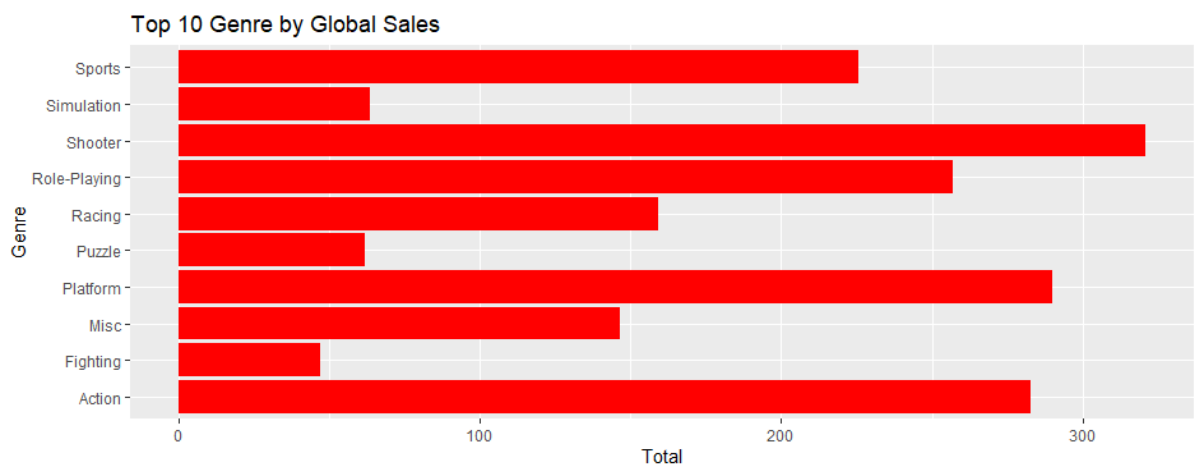

It seems that games that were released in year 2006 accounts for the highest sum of Global_Sales, followed by year 2009 and year 2010.
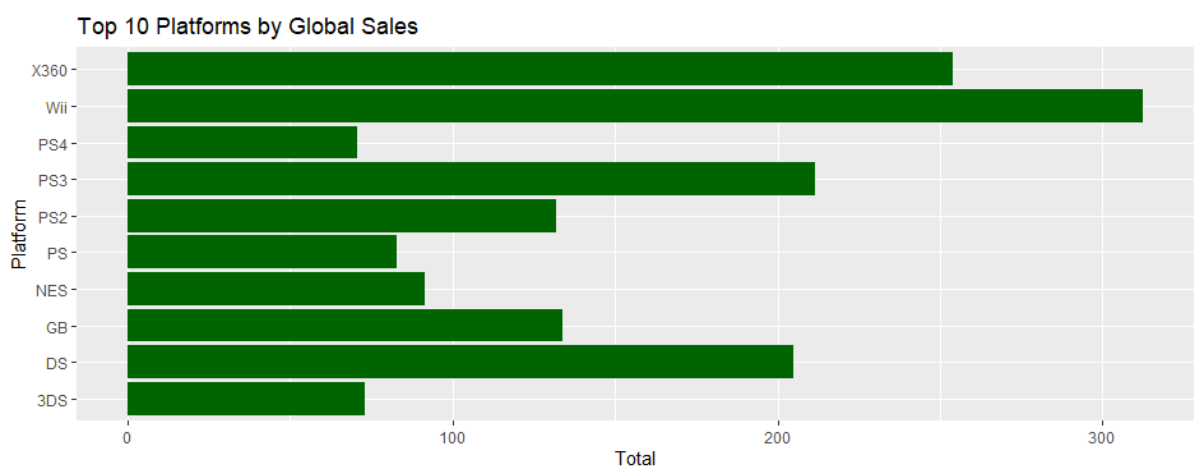
The top 5 best-selling videogames from 2006 are all published by Nintendo.

Nintendo accounts for the highest global sales across all publishers as shown below.

Top 10 Genre by Global Sales
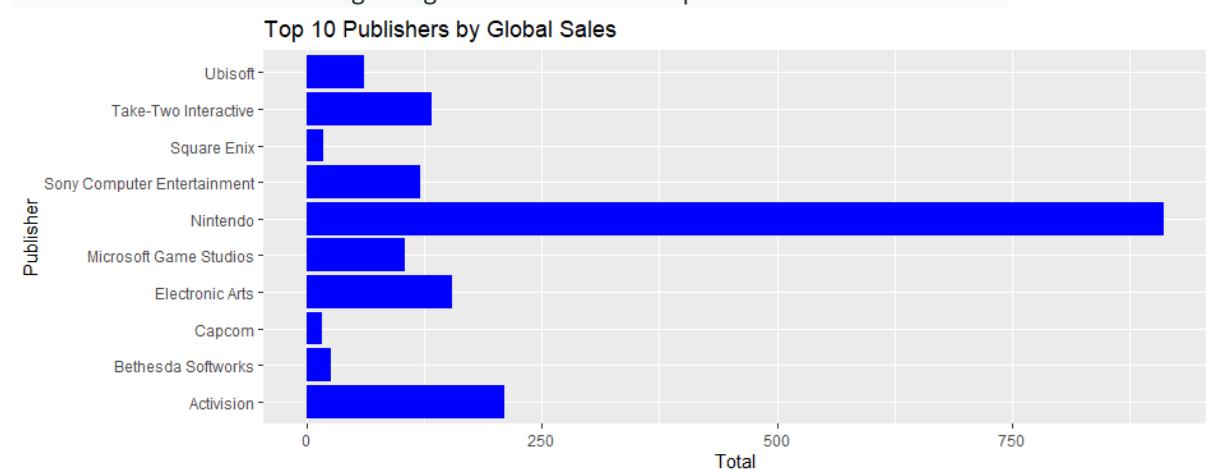
Video games in the "Shooter" genre accounts for the highest global sales per genre.
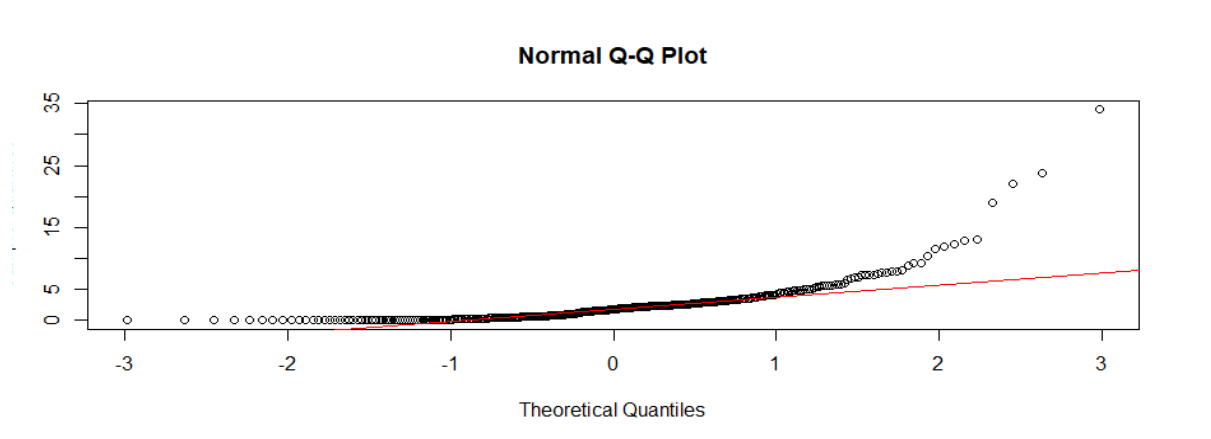

Top 10 Platforms by Global Sales

Video games on the Wii platform contributes the most to the global sales.

(5) How reliable the data is (e.g. normal distribution, skewness, or kurtosis)
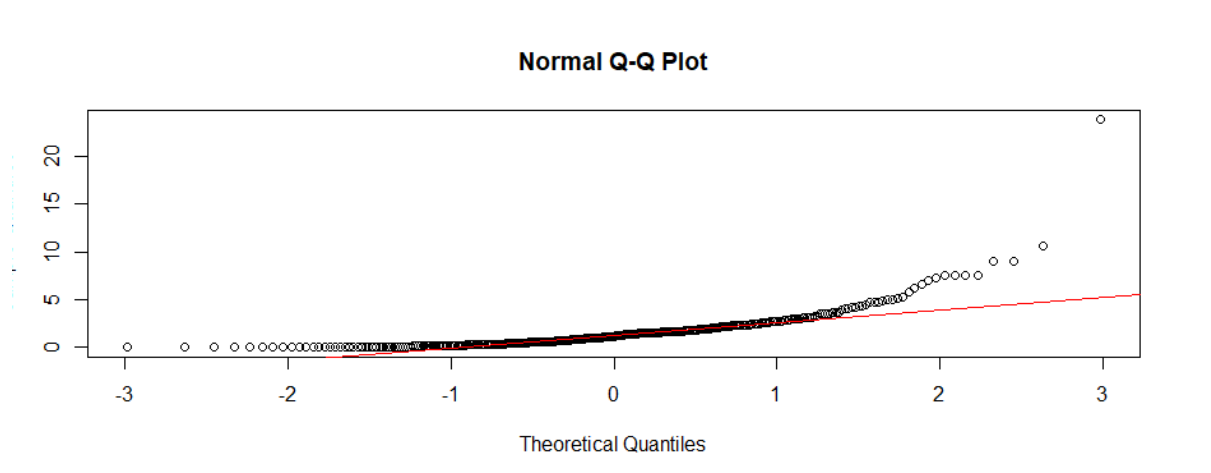
All three sales data show strong statistical evidence against the null hypothesis and although these are highly peaked with heavier tails compared to a normal distribution, the data indicates a highly positive skew across all three sales.

The data is reliable and can be used to build and fit a MLR model to predict future sales.
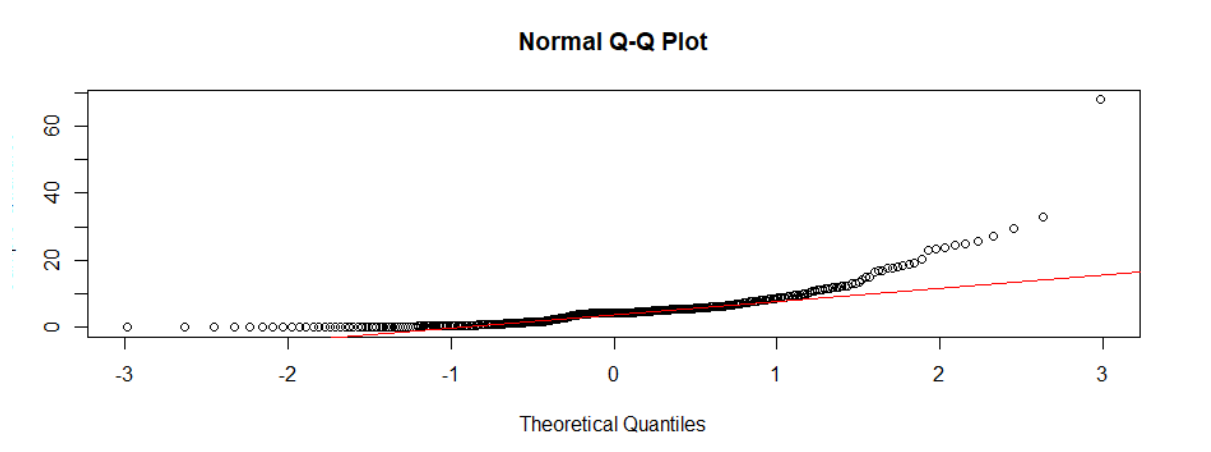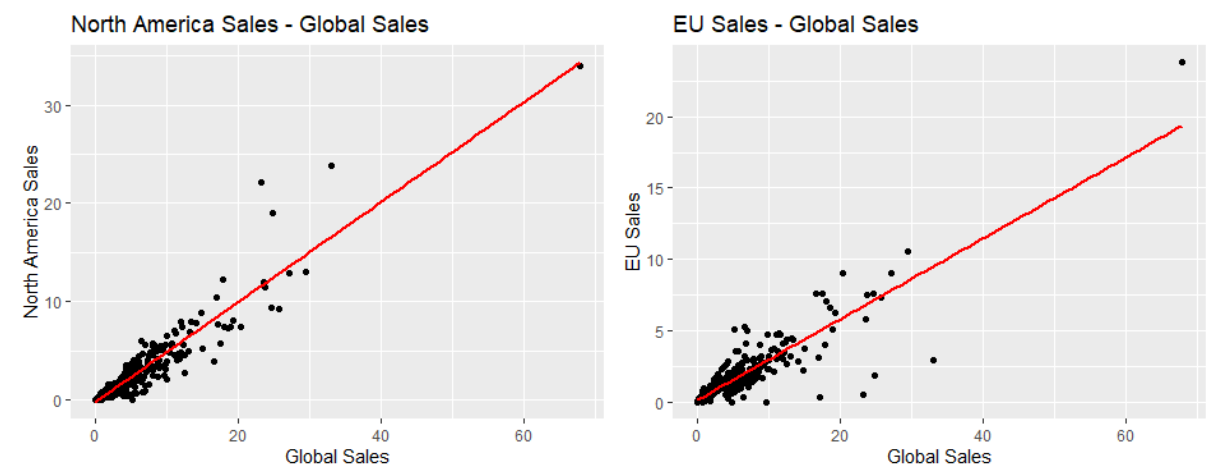
North Americal Sales - Q-Q Plot

Normal Q-Q Plot

Theoretical Quantiles

EU Sales – Q-Q Plot

Normal Q-Q Plot

Theoretical Quantiles

Global Sales – Q-Q Plot

Normal Q-Q Plot

Theoretical Quantiles

(6) what the relationship(s) is/are (if any) between North American, European, and global sales?

The data was then explored to ensure a linear relationship exists between the two predictor variables (North America Sales and EU Sales) and the observed values (Global Sales).

The figure below shows this relationship.

A multiple linear regression (MLR) model was then built and fitted using the North America Sales and EU Sales to predict future Global Sales. (next page).

Summary

Marketing

Loyalty points accumulation can be determined by customer renumeration and spending scores. The higher the renumeration and spending score, the higher the likelihood of the customer accumulating loyalty points.

Customers with 'high renumeration, low spending score' can be targeted for marketing campaigns of products with higher value to try and increase their loyalty points.

Customers with 'low renumeration, high spending score' can be targeted with incentives if they are to leave product reviews on social media. This could potentially add value for future product feedback analysis and also increases the overall positive sentiment towards Turtle Games.

More investigation is recommended for the customer product reviews to determine which specific products the negative sentiments are referring to.

Supply Chain Forecasting

The data shows that if a product is newly released, it does not necessarily equate to sales.

The top 10 selling products were originally released between 1984 and 2009.

Mostly in the Shooter, Platform and Action genres.

Mostly published by Nintendo.

And products that are on the Wii and X360 platforms.

This information can assist on supply chain forecasting to help anticipate customer purchase and prevent stock unavailability which can impede on potential profits.

Sales Predictions

An MLR model was built that is highly accurate and the predicted values will be right 96.85% of the time.

Global Sales predictions for the Top 10 best-selling Turtle Games products:

Ranking	Global_Sales	Predicted_Sales
1	67.85	71.47
2	33	31.72
3	29.37	29.41
4	27.06	27.27
5	25.72	20.68
6	24.81	24.68
7	24.61	21.16
8	23.8	23.63
9	23.47	21.81
10	23.21	26.43

We are confident that the predicted figures above are derived from reliable data. We recommend that further analysis and forecasting can be made through the utilisation of similar predictive models.