Context

Turtle Games manufactures and sells an extensive range of products, including its own products and products manufactured by other companies. In this analysis report, we focused on the video game product line and gathered sales data as well as customer feedback to evaluate Turtle Games' performance. The analysis is conducted based on 2 datasets and we will examine from 3 perspectives:

1. Customers segmentation and trends
2. Sentiment analysis
3. Market sales performance

With the findings and insights, recommendations will be provided to inform the commercial and marketing approach to improve the overall sales performance.

Analytical Approach

There are 2 datasets that are analysed for the report:
1. **turtle_reviews:** analysis with Python
2. **turtle_sales:** analysis with R

To start with, we imported the data into Python and R and view the data frame to check and see if there is a null value. The null value and unnecessary columns are removed from the database to better bring them together into valuable insight.

The function of .describe() is commonly used as a quick glimpse on the database prior to further analysis. We did not remove the outliner in the data frame considering the possibility of the great difference in sales between different games.

Identify the Customer Trends & Segmentations

Reviews are gathered from 2,000 clients with 56% of them being women and the remaining being men (See Fig 1). Customers are between the ages of 17 and 72 and average at 39 years old.
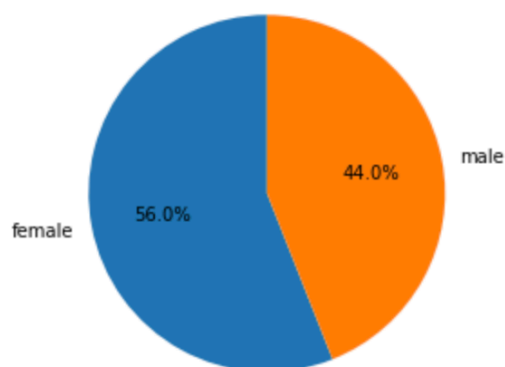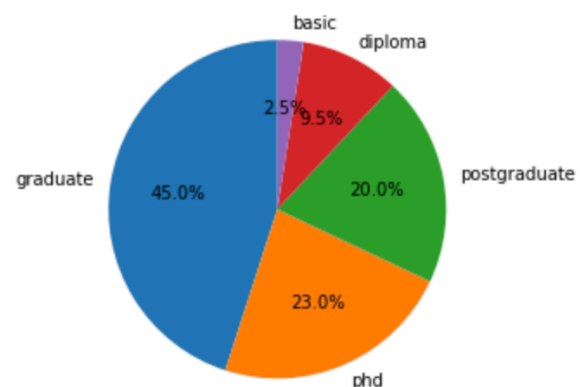

Fig 1. Gender

Fig 2. Education

To examine the customer trend and understand the relationship between variables, we use the Seaborn library to examine the relationship and dissemination of data with the pair chart. (See Fig 3)
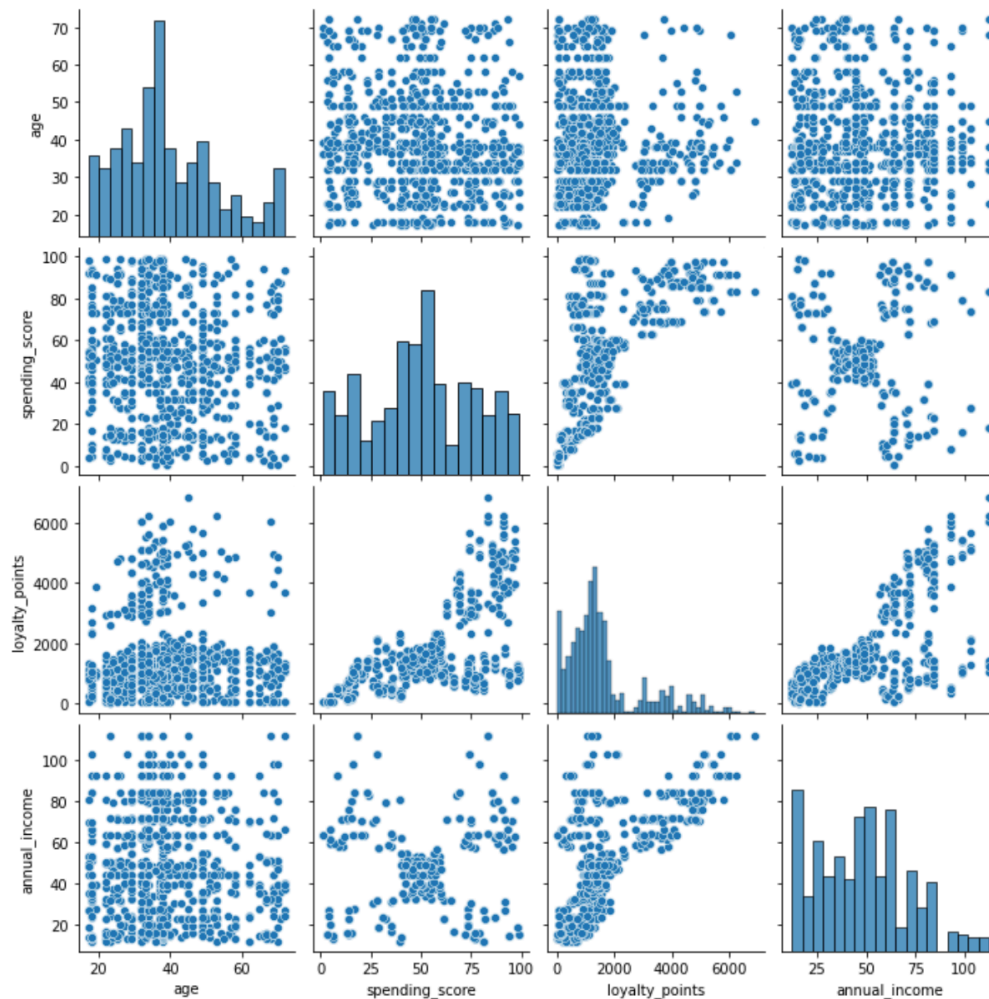

Fig 3. Pair Plot

Based on the pair plot, we conducted a further deep dive into the correlation between spending and loyalty, remuneration and loyalty, age and loyalty using linear regression.
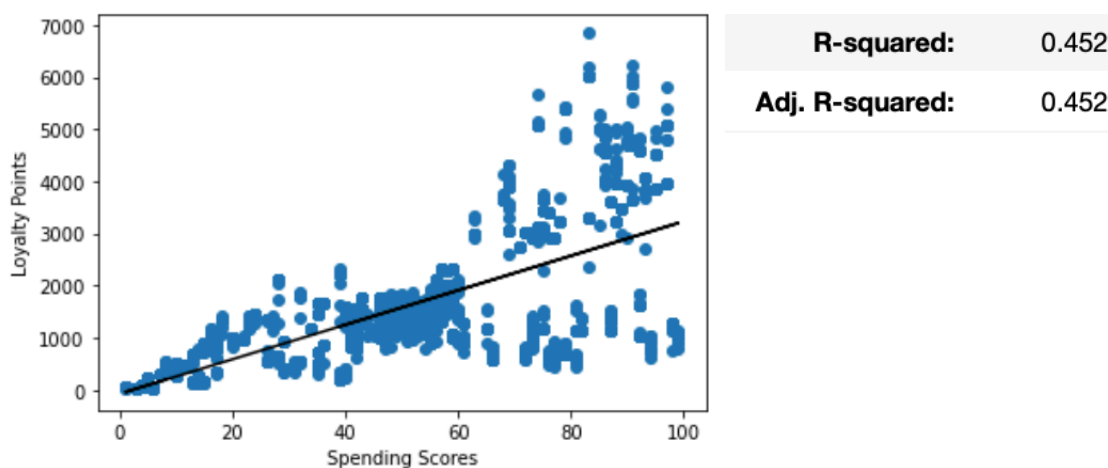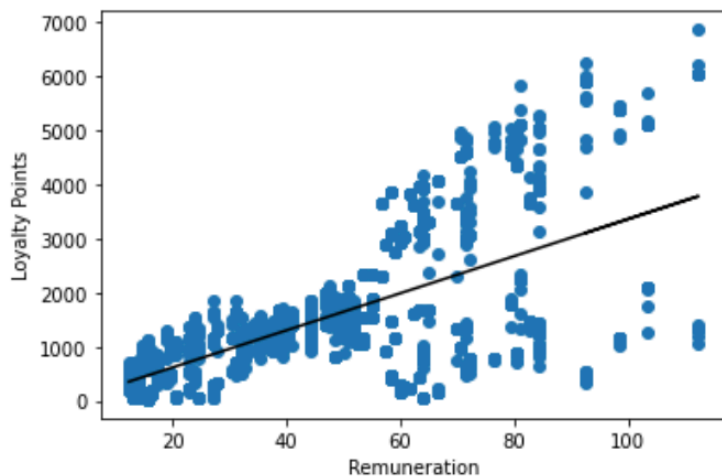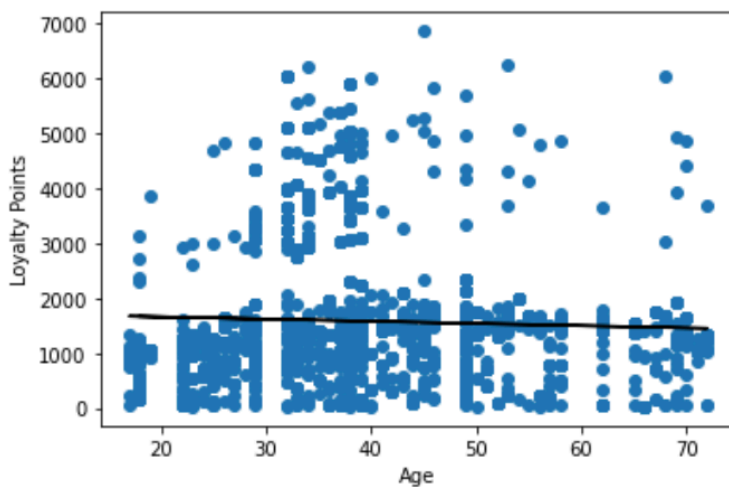

| R-squared: | 0.452 |
| Adj. R-squared: | 0.452 |

Fig 4. Spending v.s. Loyalty   Page 2 of 10

| R-squared: | 0.380 |
| Adj. R-squared: | 0.379 |

Fig 5. Renumeration v.s. Loyalty

| R-squared: | 0.002 |
| Adj. R-squared: | 0.001 |

Fig 6. Age v.s. Loyalty

There is a positive correlation between spending scores and renumeration with loyalty points. The relationship between renumeration and loyalty points is slightly stronger in relation to the coefficient. However, we can see the megaphone pattern from the linear regression model that indicate the model gets more erratic at higher fitted values. The low Adj. R-squared is also an indicator of the inaccuracy of this model for this data set. On the other hand, in the regression model for age and loyalty, it shows no correlation between variables.
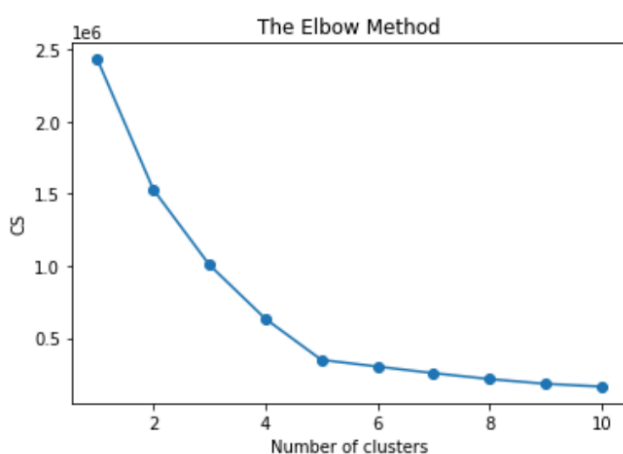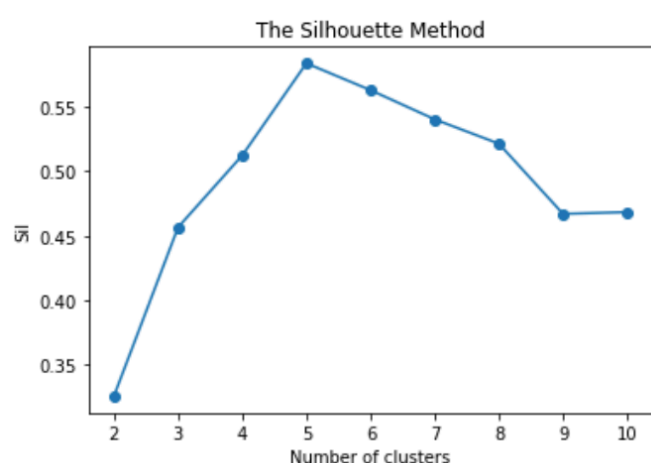

Fig 7. The Elbow Method

Fig 8. The Silhouette Method

To segment our customers for business planning, the Elbow method and the Silhouette method are used to determine the optimal number of clusters and split our data into groups accordingly.
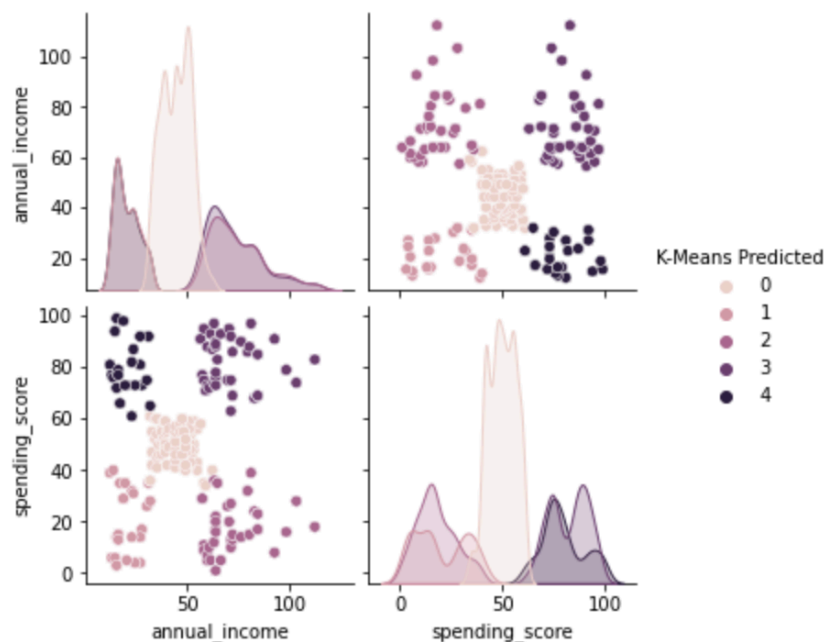

Fig 9. K-Means Clustering

We can see the potential among the 5 customer segments and different marketing approaches can be considered to best target these customers.

- **Cluster 0 & Cluster 3:** High value customers with the greatest potential for generating sales. The marketing campaign objective is to develop the customers' shopping habits and increase their loyalty to maximize sales.
- **Cluster 2:** High potential customers segment with high purchasing power. Brand awareness campaign is considered to develop the instinctive preference towards Turtle Games and motivate them to purchase from the company.
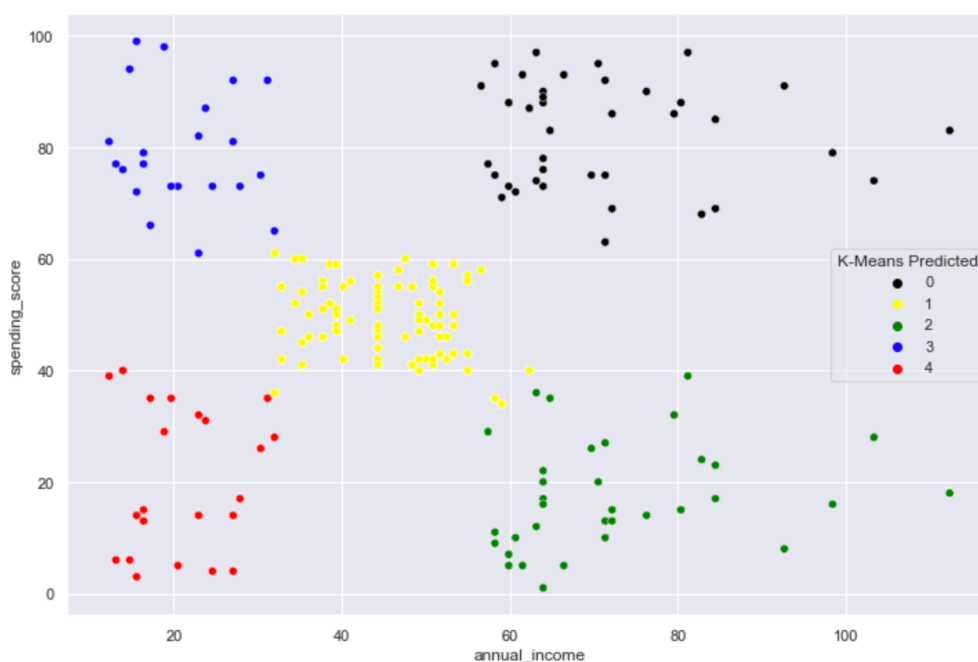

Fig 10. Clustering Model (k = 5)

Sentiment Analysis to Explore Customers Feedback

We use the function of generate_polarity to perform Natural Language Processing on the collected data. In general, we are seeing a positive sentiment on both summary and review for the Turtle Games.

Over 80% of the reviews and 54% of the summary are positive. With reference to the histogram below (See Fig 11 & Fig 13), the data are skewed towards the positive side with high volume of the feedback are close to neutral.
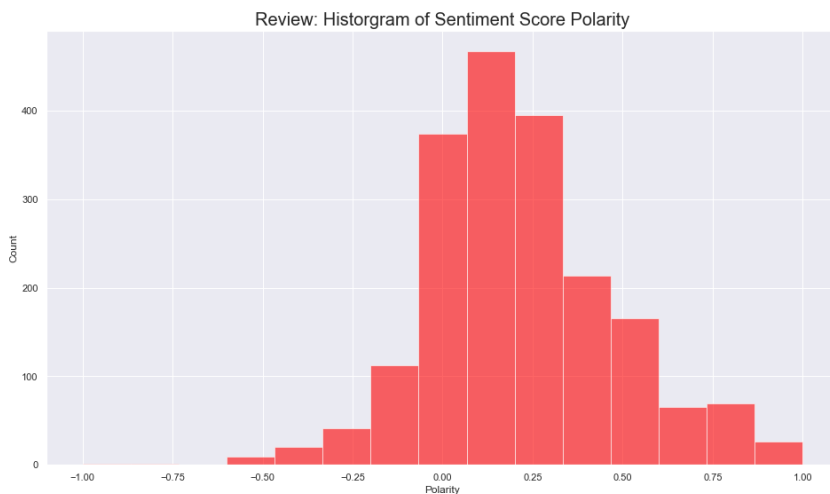

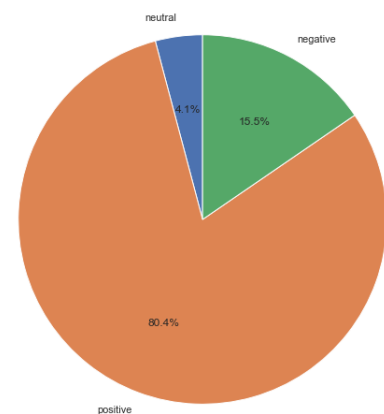Fig 11. Histogram of Sentiment Score Polarity on Review
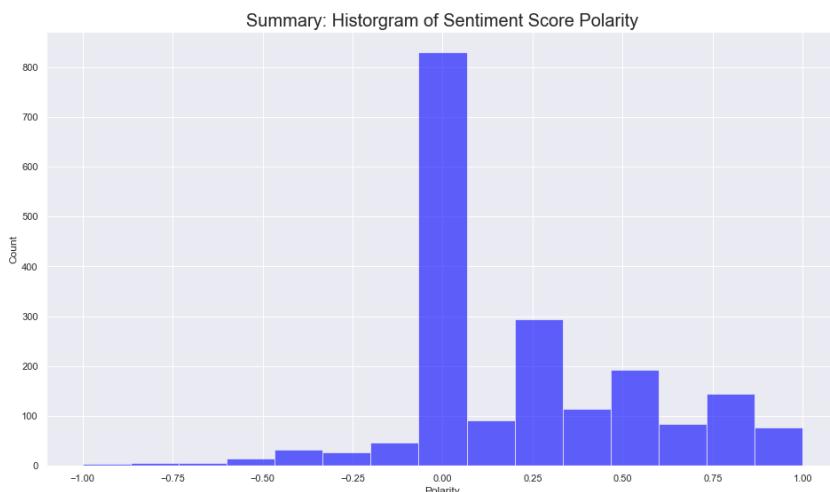

Fig 12. Review Sentiment Analysis

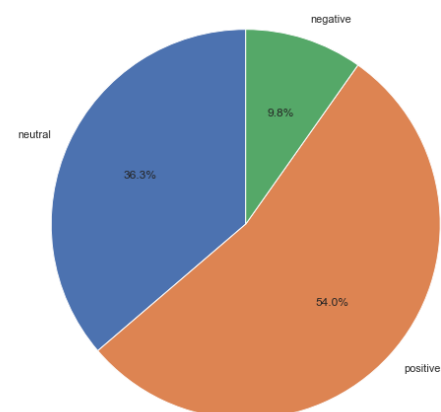Fig 13. Histogram of Sentiment Score Polarity on Summary

Fig 14. Summary Sentiment Analysis

To further explore the audience sentiment towards the Turtle Game, we dive into the top 20 positive and negative review and summary. In general, we can see that the game is commonly used as a gift, people also appreciated the design and found the game great. On the contrast, the negative sentiment stated that the game is boring, disappointing, difficult and complicated to follow.

All in all, Turtle Games received positive feedback from the customers.

Market Analysis to Reveal the Market Dynamic

In the sales data, we reviewed 165 games that are available on 22 platforms, some are available on multiple platforms while others are only available on one. 12 different game genres are included in the data. Majority of the games are Shooter and Action for X360 and PS3.
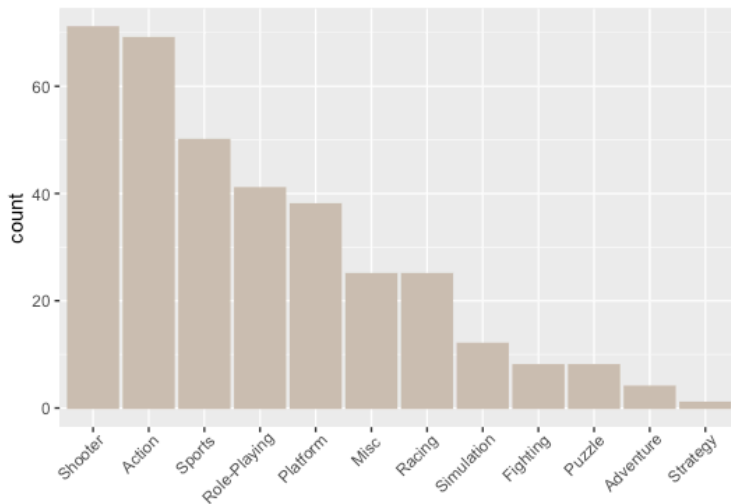
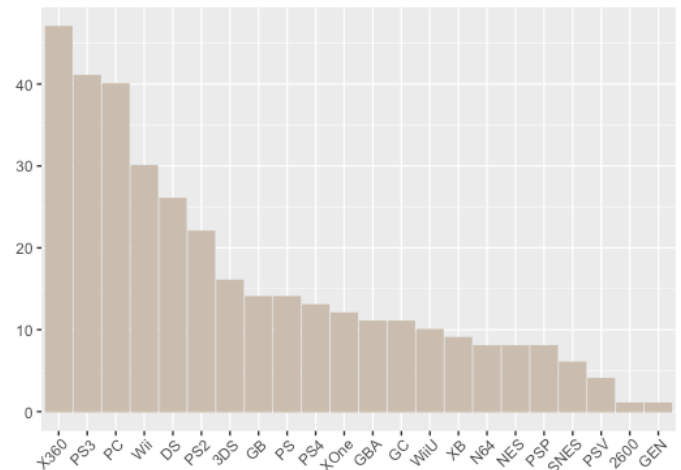

Fig 15. Count of Game Genre

Fig 16. Count of Platform

The 3 market sales share similar trend yet there are difference between market that indicate the difference in customers behaviour including the preferred platform and genre preference. The Q-Q Plot (See Fig 17) and Shapiro-Walk test both indicate the dataset is not normally distributed with positive skewness. This is likely to be caused by the best seller product 107 launched in 2006 by Nintendo with a total global sales of GBP 67.85M that skewed the data.
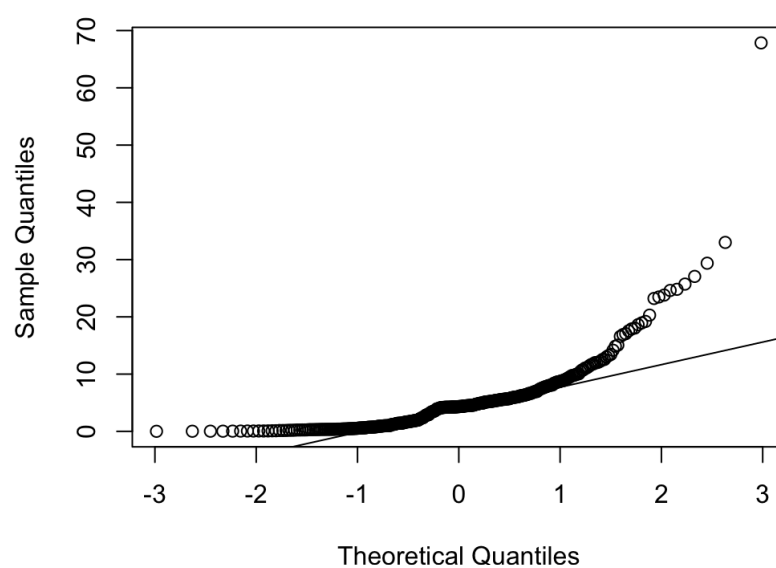
Global Sales Q-Q Plot

Fig 17. Global Sales Q-Q Plot

Shooter, Action and Sports are the top 3 genre with large volume of game available. (See Fig 18, 19 & 20). Generally, X360 and Wii are the most common gaming platform for the market, PS3 remains to be the top 2 gaming platform in the EU market. (See Fig 21, 22 & 23)
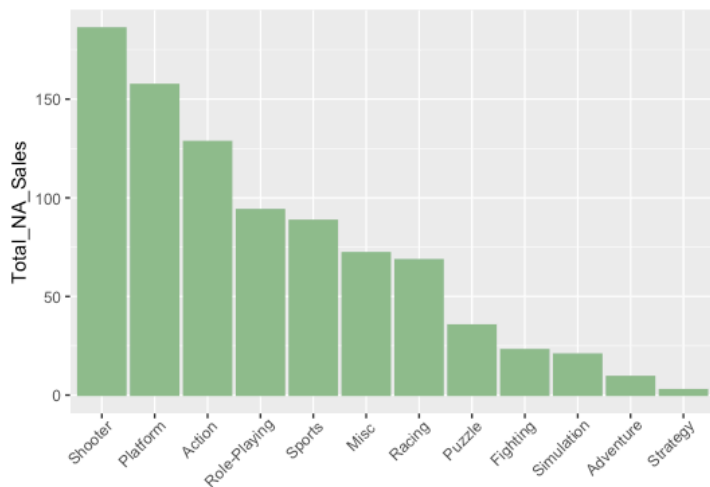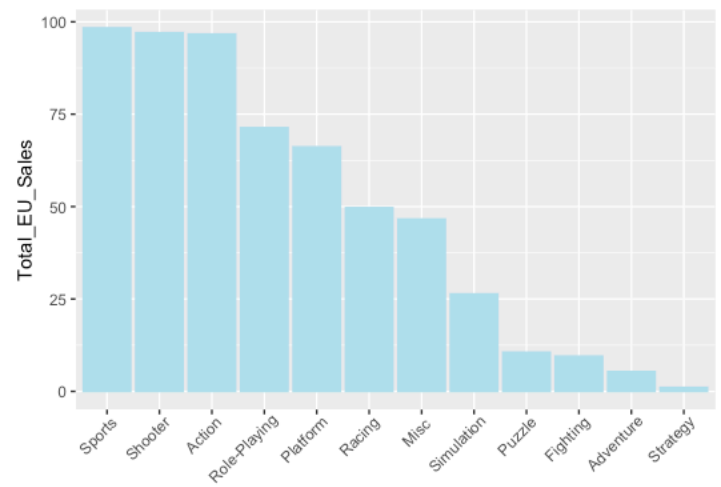

Fig 18. North America Sales by Genre
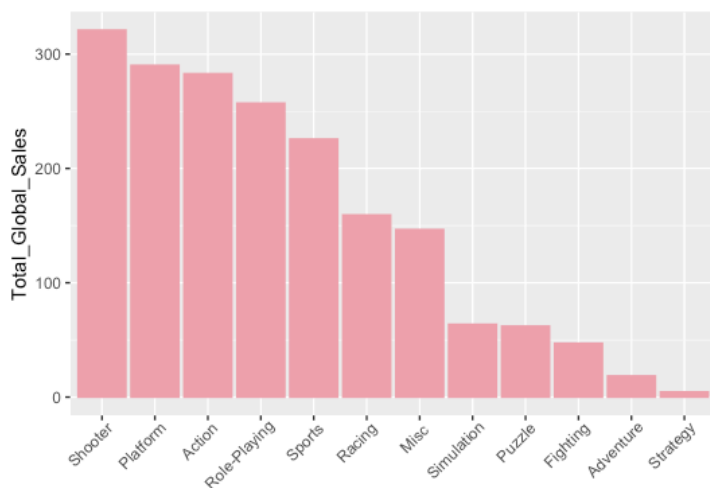

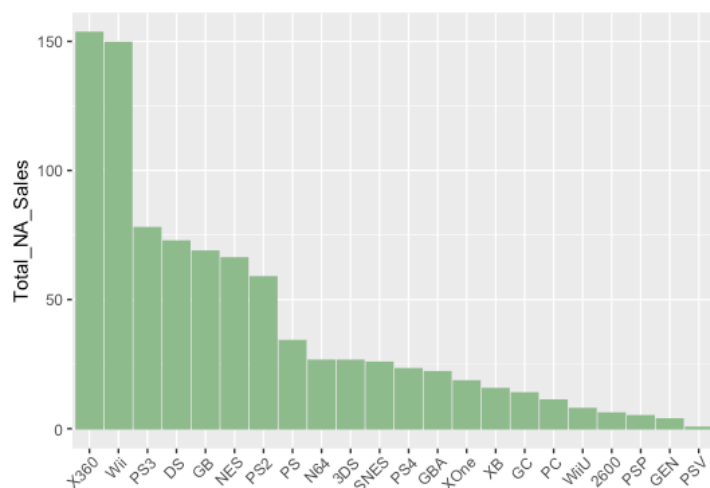Fig 19. Europe Sales by Genre

Fig 20. Global Sales by Genre

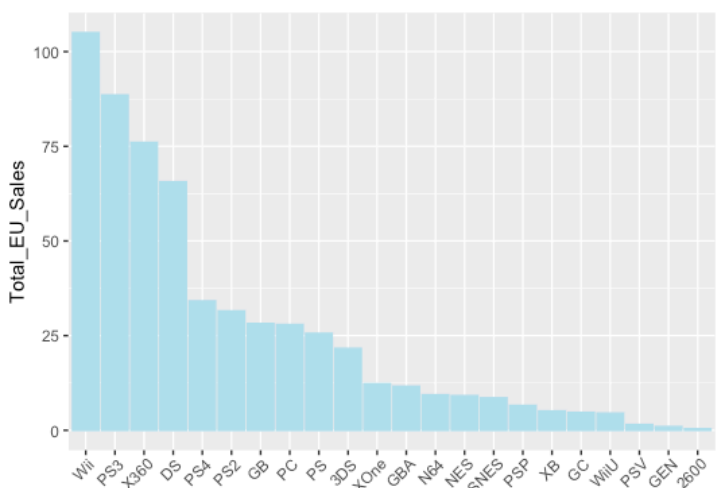Fig 21. North America Sales by Platform
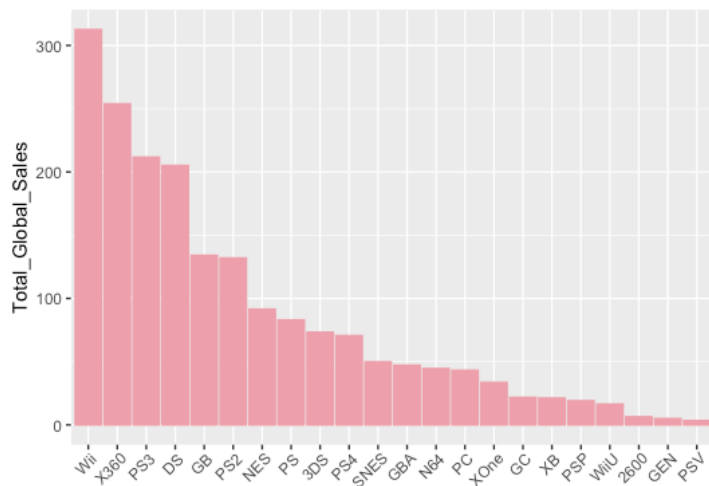

Fig 22. Europe Sales by Platform

Fig 23. Global Sales by Platform

There seems to be a slightly negative relationship between the Product ID and sales that the smaller Product ID tend to have higher sales. (See Fig 24). It is interesting to further investigate and understand the rationale of product ID to identify any pattern behind.
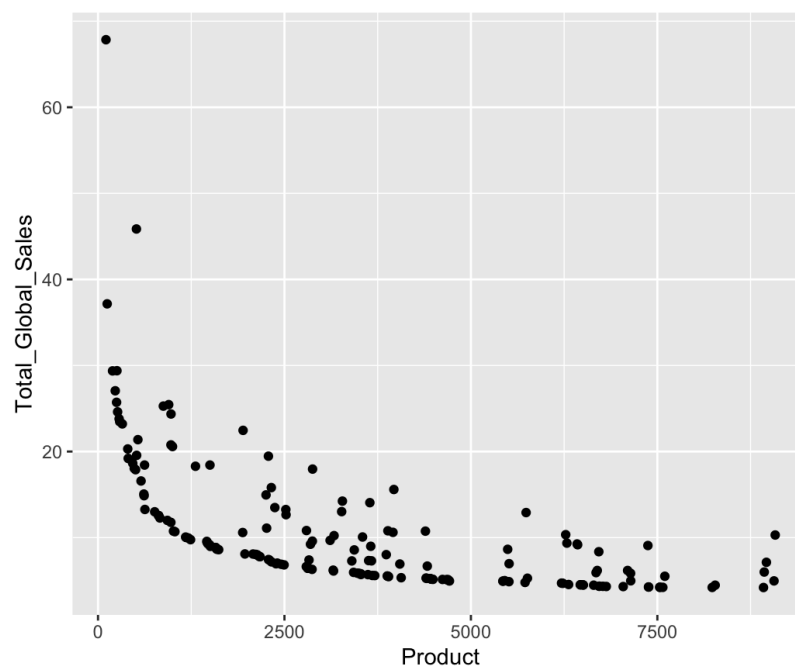

Fig 24. Correlation between Global Sales & Products

Discover Pattern and Create Prediction Model

NA Sales has a high positive correlation with Global Sales with an index above 93%. EU sales also show a strong positive correlation with global sales with an index of approximately 88%. The relationship between NA and EU sales is relatively weak with an index of approximately 70%.
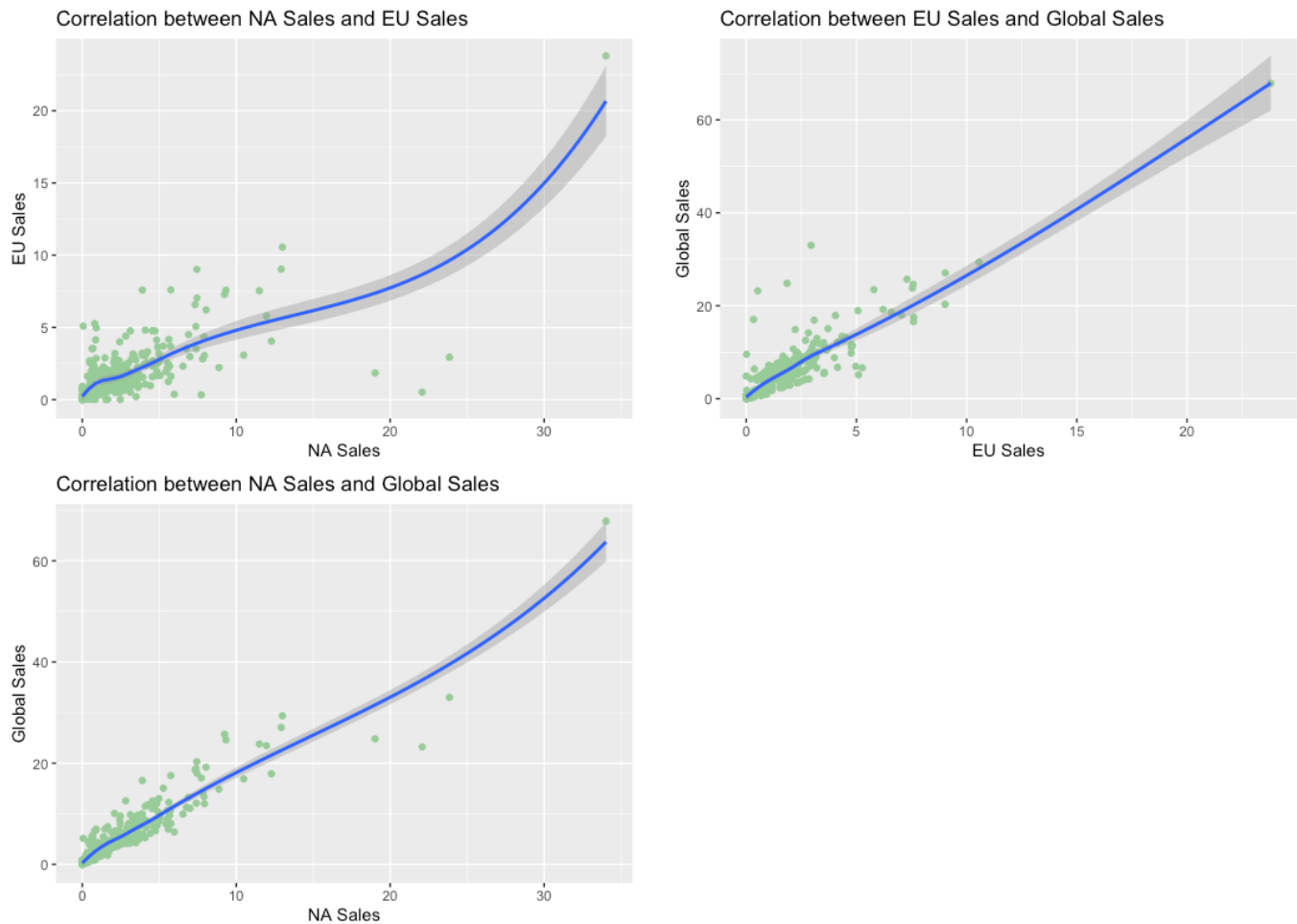
To further explore the best model that can be integrated into the data, we have tried with the multiple linear regression model to review NA, EU Sales and Global Sales sales. The adjusted R-squared is 0.9685 and the residual standard error is 1.112, which is considered an accurate pattern.
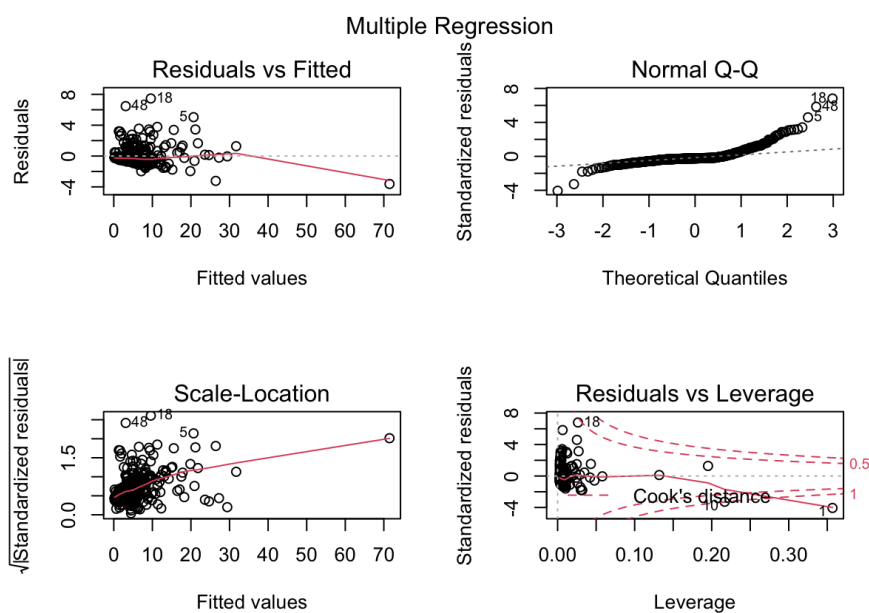

Fig 25. Multiple Linear Regression

To take forward the analysis, a prediction model is created based on the provided data. This model can then help to predict the global sales for game that is first launch in the North America and Europe market.

```
  Product NA_Sales EU_Sales Predict_Global_Sales Actual_Global_Sales
1     107    34.02    23.80            71.468572               67.85
2    3267     3.93     1.56             6.856083                6.04
3    6815     2.73     0.65             4.248367                4.32
4    2877     2.26     0.97             4.134744                3.53
5     326    22.08     0.52            26.431567               23.21
```

Recommendations

1. **Marketing Campaign:** Different tactics to target different customer segments

2. **Procurement:** Taking the genre preference, platform popularity and customers feedback into consideration

3. **Market Strategy:** Evaluate the investment and approach based on the market potential

Continuous Improvement

1. **Data Accuracy:** Based on research (**ISFE Report**, 2021), the average age of a video game player in Europe is 31.3 years old and women make up 47% of the gaming population. This is a different breakdown from what we collected. We should investigate further how we collected the data and make sure that it is the representation of Turtle Games customers.

2. **Data Completeness:** The current data include only one of the product line for Turtle Games and is unable to reveal the full picture of the company sales performance.

3. **Data Source:** There is limited information on how the data is collected and when is this collected. More data is required to improve the accuracy of the model and analysis.