

Predicting future outcomes : Turtle Games

Predictive analysis using Python & R

Background

Turtle Games, a global game manufacturer and retailer, is determined to boost overall sales and revenue using data-driven insights. They've engaged a skilled Data Analyst team to explore sales data and customer reviews. The focus is on understanding microeconomic factors influencing sales and product success. By analysing trends, correlations, and regional preferences, the team aims to guide strategic decisions. Armed with actionable intelligence, Turtle Games' leadership aims to optimize marketing, distribution, and product offerings, ensuring they stay ahead in the competitive gaming industry. With data as their compass, the company is poised for sustainable growth, customer loyalty, and continued success in the gaming market..

Analytical Approach

To achieve business objective, our team of data analysts has devised a structured and comprehensive analytical approach. We will address initial set of questions using Python and R :

1. **Data Ingestion and Wrangling:** To start, we imported the relevant .csv datasets into Python and R, ensuring accuracy and consistency by standardizing the data format and handling missing & redundant values. This process laid a strong foundation for subsequent analysis.

1. Load and explore the data

```
In [1]: # Install the statsmodels package.  
!pip install statsmodels  
  
# Import the necessary libraries.  
import numpy as np  
import pandas as pd  
  
# Visualisation  
import seaborn as sns  
import matplotlib.pyplot as plt  
  
# The statsmodels  
import statsmodels.api as sm  
from statsmodels.formula.api import ols  
  
# Import and read the data file (turtle_reviews.csv file)  
reviews = pd.read_csv('turtle_reviews.csv')  
  
print (reviews.head())  
print (reviews.info())  
  
reviews.describe ()  
  
print(reviews.dtypes) # Print the data types of each column  
print("\nMetadata:")  
print(reviews.info()) # Print metadata about the DataFrame
```

2. Drop columns

```
In [4]: # Remove redundant columns (language and platform)

reviews = reviews.drop(['language', 'platform'], axis=1)

print (reviews.head())

reviews

   gender age  remuneration (k€)  spending_score (1-100)  loyalty_points \
0   Male   18            12.30                  39                210
1   Male   23            12.30                  81                524
2 Female   22            13.12                  6                40
3 Female   25            13.12                  77                562
4 Female   33            13.94                  40                366

   education product                                         review \
0 graduate      453 When it comes to a DN's screen, the space on t...
1 graduate      466 An Open Letter to GaleForce9*:\\nYour unpaint...
2 graduate      254 Nice art, nice printing. Why two panels are f...
3 graduate      263 Amazing buy! Bought it as a gift for our new d...
4 graduate      291 As my review of GF9's previous screens these w...
```

3. Rename columns

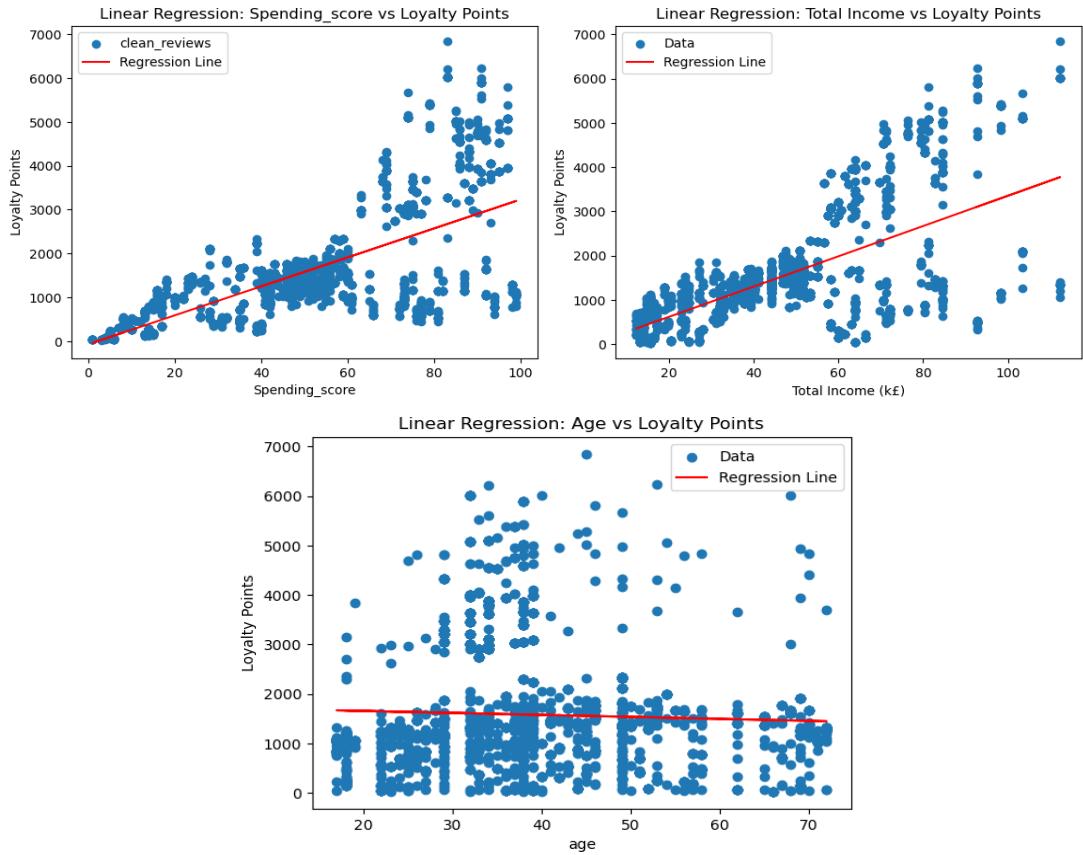
In [5]:	# Change column headings to names that are easier to reference (e.g. renumeration and spending_score).																																																						
	reviews = reviews.rename(columns={'remuneration (k€)': 'Total_Income_(k€)', 'spending_score (1-100)': 'Spending_Score'})																																																						
reviews.head ()																																																							
Out[5]:	<table border="1"> <thead> <tr> <th>gender</th><th>age</th><th>Total_Income_(k€)</th><th>Spending_Score</th><th>Loyalty_points</th><th>Education</th><th>product</th><th>review</th><th>summary</th></tr> </thead> <tbody> <tr> <td>0</td><td>Male</td><td>18</td><td>12.30</td><td>39</td><td>210</td><td>graduate</td><td>453</td><td>When it comes to a DM's screen, the space on ...</td></tr> <tr> <td>1</td><td>Male</td><td>23</td><td>12.30</td><td>81</td><td>524</td><td>graduate</td><td>466</td><td>An Open Letter to GaleForce9:\n\nYour unpaid... Another worthless Dungeon Master's screen from...</td></tr> <tr> <td>2</td><td>Female</td><td>22</td><td>13.12</td><td>6</td><td>40</td><td>graduate</td><td>254</td><td>Nice art, nice printing. Why two panels are ... pretty, but also pretty useless</td></tr> <tr> <td>3</td><td>Female</td><td>25</td><td>13.12</td><td>77</td><td>562</td><td>graduate</td><td>263</td><td>Amazing buy! Bought it as a gift for our new ... Five Stars</td></tr> <tr> <td>4</td><td>Female</td><td>33</td><td>13.94</td><td>40</td><td>366</td><td>graduate</td><td>291</td><td>As my review of GF9's previous screens these w... Money trap</td></tr> </tbody> </table>	gender	age	Total_Income_(k€)	Spending_Score	Loyalty_points	Education	product	review	summary	0	Male	18	12.30	39	210	graduate	453	When it comes to a DM's screen, the space on ...	1	Male	23	12.30	81	524	graduate	466	An Open Letter to GaleForce9:\n\nYour unpaid... Another worthless Dungeon Master's screen from...	2	Female	22	13.12	6	40	graduate	254	Nice art, nice printing. Why two panels are ... pretty, but also pretty useless	3	Female	25	13.12	77	562	graduate	263	Amazing buy! Bought it as a gift for our new ... Five Stars	4	Female	33	13.94	40	366	graduate	291	As my review of GF9's previous screens these w... Money trap
gender	age	Total_Income_(k€)	Spending_Score	Loyalty_points	Education	product	review	summary																																															
0	Male	18	12.30	39	210	graduate	453	When it comes to a DM's screen, the space on ...																																															
1	Male	23	12.30	81	524	graduate	466	An Open Letter to GaleForce9:\n\nYour unpaid... Another worthless Dungeon Master's screen from...																																															
2	Female	22	13.12	6	40	graduate	254	Nice art, nice printing. Why two panels are ... pretty, but also pretty useless																																															
3	Female	25	13.12	77	562	graduate	263	Amazing buy! Bought it as a gift for our new ... Five Stars																																															
4	Female	33	13.94	40	366	graduate	291	As my review of GF9's previous screens these w... Money trap																																															

The screenshot shows an RStudio interface with several tabs open. The code editor contains R code for reading a CSV file, creating a new data frame from a subset, and performing multiple regression analysis. The console tab displays the results of the regression, including predicted values and observed values.

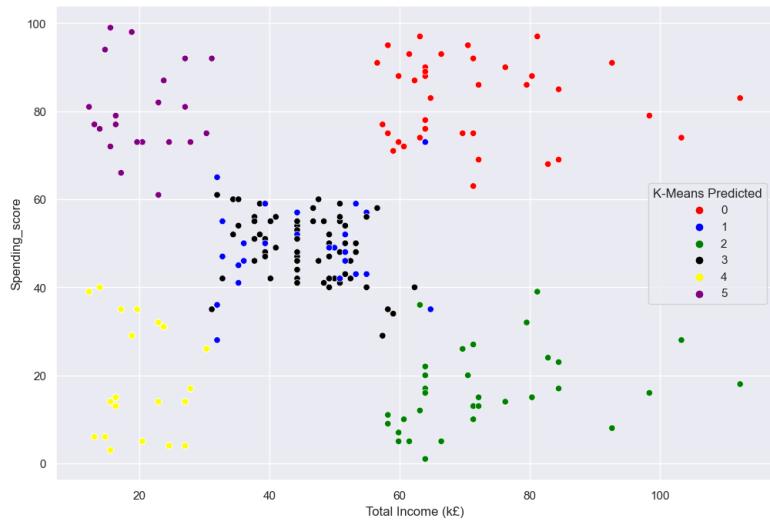
```
23 # Import a CSV file
24 Sales <- read.csv('turtle_sales.csv', header=TRUE)
25
26 # Print the data frame.
27 View(Sales)
28 as_tibble(Sales)
29
30 ## View a summary of the data frame.
31 summary(Sales)
32
33 # Create a new data frame from a subset of the sales data frame.
34 # Remove unnecessary columns
35 Sales_1 <- select(Sales, -Ranking, -Year, -Genre, -Publisher)
36
37 # Convert product into factors (categorical variable) # Create new data frame with Clean data
38
39 Sales_clean <- mutate(Sales_1, Product = as.factor(Product))
40
41
42 # View the data frame.
43 as_tibble(Sales_clean)
44
45 # View the descriptive statistics.
46 summary(Sales_clean)
47
48 #####
49
50 #####
```

```
R 4.2.3 : ~
> predict_global_sales_multiple <- predict(lm_multiple, newdata = new_data)
> # Compare the predictions to the observed values
> obs_sales <- c(50.2, 40.5, 30.8, 55.1, 48.6) # Replace this with your actual observed values
> # Print the results
> print("Predicted Global Sales (NA_Sales):")
[1] "Predicted Global Sales (NA_Sales):"
> print(predict_global_sales_multiple)
      1        2        3        4        5 
59.457821 7.470625 3.943404 4.884945 38.945263
> print("Predicted Global Sales (EU_Sales):")
[1] "Predicted Global Sales (EU_Sales):"
> print(predict_global_sales_eu)
      1        2        3        4        5 
65.466667 5.107314 2.637387 3.506082 2.284768
> print("Predicted Global Sales (Multiple Regression):")
[1] "Predicted Global Sales (Multiple Regression):"
> print(predicted_global_sales_multiple)
      1        2        3        4        5 
71.468572 6.856883 4.248367 4.134744 26.431567
> print("Observed Global Sales Values:")
[1] "Observed Global Sales Values:"
> print(observed_values)
[1] 50.2 40.5 30.8 55.1 48.6
> # Determine your working directory
> setwd("C:/")
[1] "C:/Users/shubhrabhanot"
> |
```

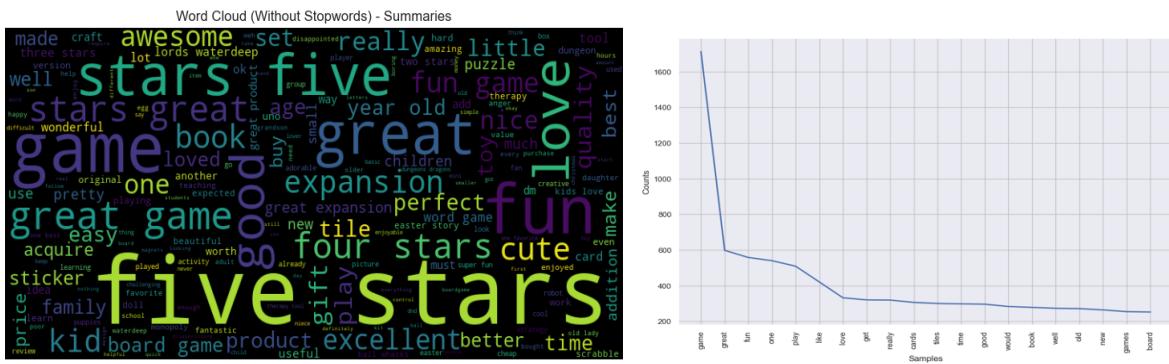
2. **Exploratory Data Analysis** : With reliable & clean datasets, we focused on customer loyalty points and product impact on sales using predictive modeling. We found strong positive correlations between loyalty points, spending score & total income and weaker associations with age. Targeting high-spending and high-income customers can boost loyalty and sales.



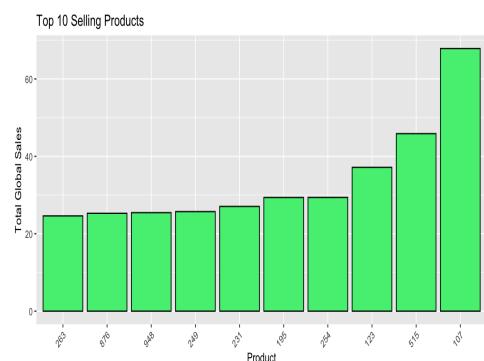
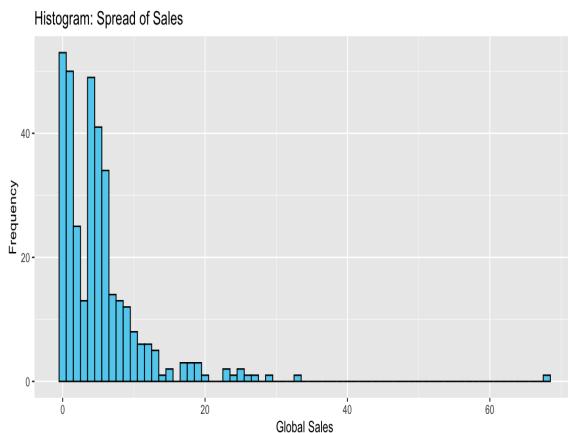
Utilizing K-Means clustering, we identified distinct customer groups based on spending and income levels. Personalized marketing and loyalty strategies tailored to each group can drive growth and enhance customer satisfaction, fostering long-term loyalty and increasing sales performance.



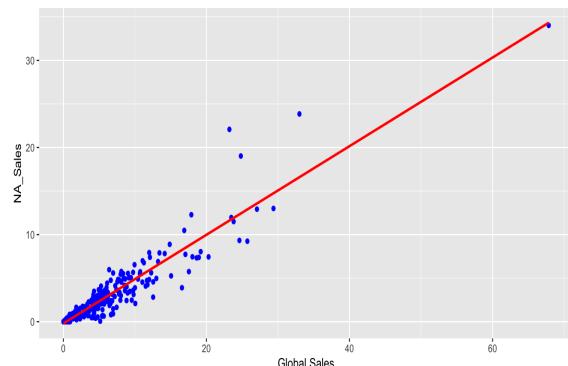
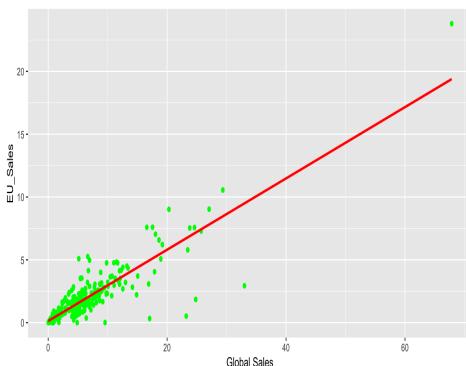
Leveraging NLP and sentiment analysis, we identified positive & negative sentiments which can be worked upon to improve customer experience. Utilize NLP insights to guide marketing strategies and enhance customer satisfaction, leading to increased loyalty and overall sales growth.



By analysing top-selling and underperforming products, Turtle Games can refine its product offerings and improve sales outcomes. Understanding distribution patterns and extreme sales values allows the team to focus on high-performing products and explore factors contributing to their success.



The histogram and QQ plot showed slight skewness towards higher sales values, indicating some products performed exceptionally well. The kurtosis value indicated heavier tails than a normal distribution, suggesting extreme sales values for certain products. To make data-driven decisions, the business team should focus on analysing top-selling products and identifying factors contributing to their success.



We observed significant positive correlations between North America, European union and global sales, indicating a clear pattern of sales alignment across regions.

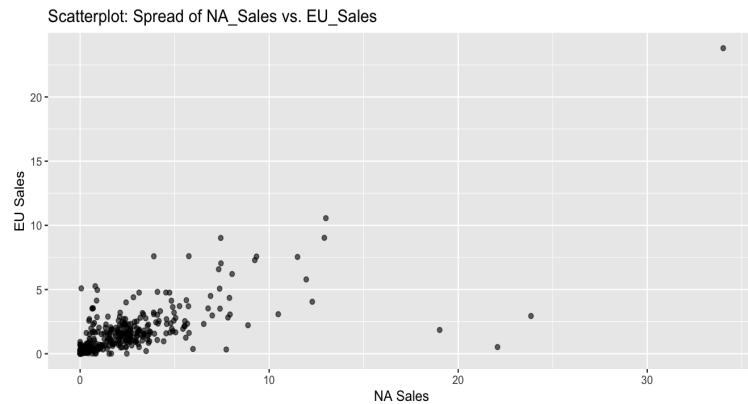
Understanding these trends allows Turtle Games to leverage regional strengths, optimize marketing strategies, and make informed decisions to maximize revenue and profitability.

By leveraging data and refining strategies based on analytical insights, Turtle Games can maintain its competitive edge, enhance overall sales success, and achieve long-term growth and profitability in the gaming market.

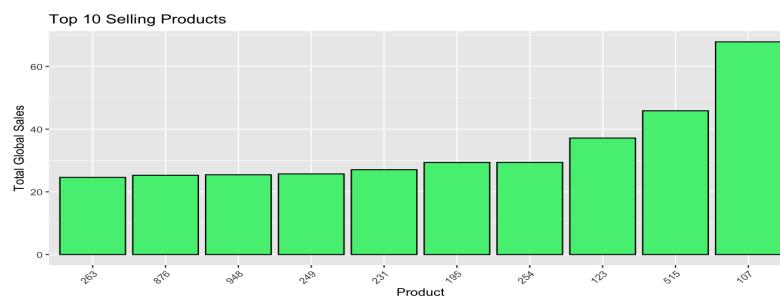
Visualization and Insights

In our analysis, we carefully selected visualizations that align with Turtle Games' business objectives, enabling data-driven decision-making and strategic planning.

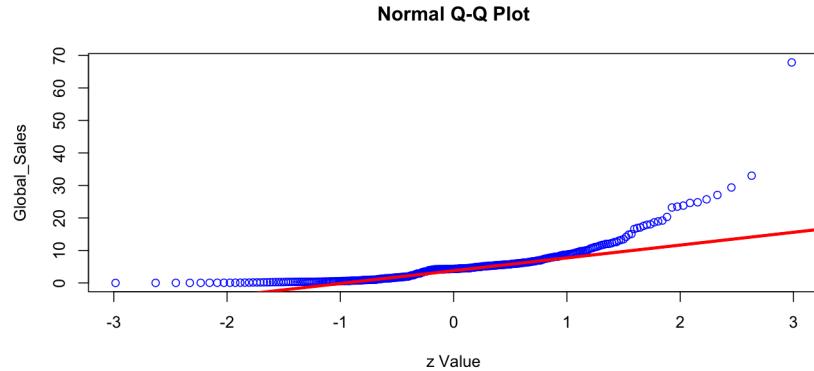
Scatterplots were utilized to show the strong correlation between regional sales North America and EU sales with global sales, helping the business team identify regions with the highest impact on overall sales. This insight guides targeted marketing and resource allocation for maximum revenue generation.



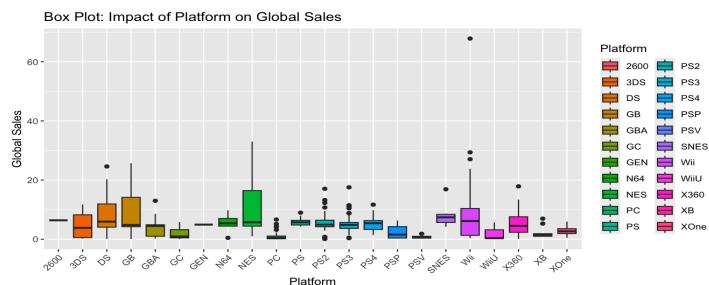
Histograms provided valuable information on the distribution of global sales data, revealing top-performing products and patterns of exceptional sales performance. Identifying these high-performing products allows the team to prioritize marketing efforts and optimize resource allocation, driving overall sales growth.



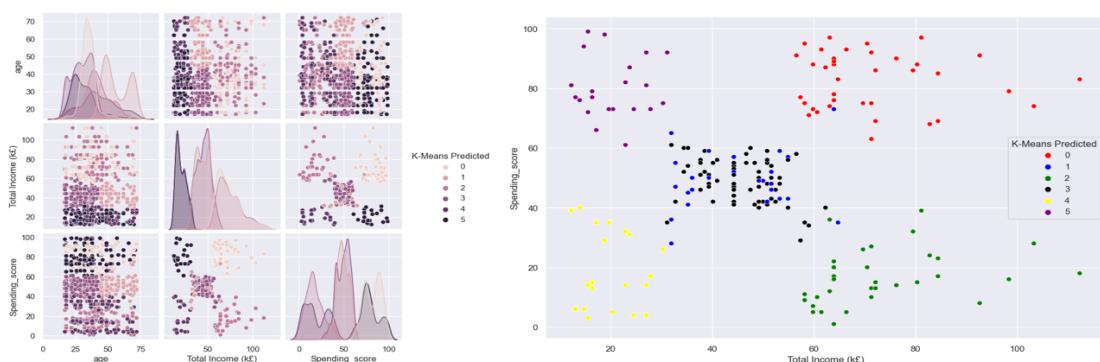
QQ Plots were employed to assess the normality of global sales data, detecting deviations from the reference line and indicating non-normal distribution with extreme sales values for certain products. Understanding these outliers helps the business team make informed decisions to address underperforming products, optimize pricing, and adjust marketing strategies accordingly.



Boxplots were used to compare global sales across different gaming platforms, uncovering potential platform-specific trends. This insight guides the team to refine platform-specific marketing strategies, optimize partnerships, and make informed decisions for better sales outcomes.



Cluster Analysis, specifically K-Means clustering, was applied to segment customers based on spending and income levels, creating distinct customer groups with unique preferences and behaviours. These clusters provide valuable insights for tailoring marketing and loyalty programs, enhancing customer satisfaction, driving repeat purchases, and fostering brand loyalty.



Overall, these visualizations offer actionable insights that inform data-driven decision-making and shape effective marketing strategies. The business team can prioritize marketing efforts, optimize product offerings, and foster customer loyalty by understanding regional impacts, identifying top-performing products, addressing outliers, and segmenting customers.

Armed with these valuable insights, Turtle Games can navigate the competitive gaming market, drive revenue growth, and achieve sustained success in the industry. By leveraging the power of data visualization, the company gains a competitive edge in the gaming industry, enabling effective strategies to increase customer satisfaction, loyalty, and overall sales performance.

Recommendations

Based on the analysis conducted, we recommend the following strategies to Turtle Sales Company for increasing their sales and revenue:

1. Target High-Impact Regions: Focus marketing efforts on regions with the strongest correlation to overall sales, such as North America and the European Union. Tailoring strategies to these regions can maximize revenue generation and market penetration.
2. Promote Top-Selling Products: Identify top-performing products using histograms and prioritize marketing campaigns to highlight their unique selling points. Emphasizing these products can attract more customers and boost overall sales.
3. Address Underperforming Products: Use QQ plots to identify products with extreme sales values or outliers. Analyse the reasons for underperformance and take corrective actions, such as pricing adjustments or improved marketing, to enhance sales performance.
4. Personalize Marketing and Loyalty Programs: Utilize cluster analysis to segment customers based on spending and income levels. Implement personalized marketing and loyalty programs to enhance customer satisfaction and loyalty.
5. Leverage Customer Feedback: Use sentiment analysis on customer reviews to understand positive and negative feedback. Strengthening positive aspects while addressing negative sentiments can improve overall customer experience and drive sales.

By implementing these strategies, Turtle Sales Company can optimize sales performance, enhance customer satisfaction, and achieve sustained revenue growth. Data-driven decision-making and tailored marketing efforts will position the company for success in the competitive gaming industry.