

Predicting future outcomes

LADAN AWIL

1. Background/context of the business:

Turtle Games, a prominent global game manufacturer and retailer, embarked on a mission to boost sales performance by delving into intricate customer behaviors. The project's focal points encompassed critical inquiries such as deciphering loyalty points accumulation patterns, delineating customer segments, harnessing social data for targeted marketing initiatives, unraveling the impact of products on sales, assessing data reliability metrics, and discerning the interconnections within regional sales. This comprehensive exploration aimed not only to decipher existing trends but also to equip Turtle Games with actionable insights crucial for strategizing their sales approach and enhancing customer engagement on a global scale.

2. Analytical Approach:

The analytical approach undertaken in this project combined Python and R programming languages to comprehensively analyze Turtle Games' data, aligning with the specified requirements. The objective was to explore customer behavior, identify market segments, leverage social data for marketing insights, assess product impact on sales, evaluate data reliability, and scrutinize regional sales relationships.

Python Analysis:

1. Loyalty Points Accumulation:

In Python, the code calculated loyalty points by aggregating North American and European sales data. This step was crucial to understanding customer engagement and served as a foundational analysis, addressing Turtle Games' curiosity about how customers accumulate loyalty points.

2. Customer Segmentation:

To target specific market segments, the code merged sales and reviews data, creating a unified dataset. It performed linear regression on this merged data, aiming to find patterns that link regional sales with loyalty points. Despite a low accuracy, this analysis contributed to understanding the relationship between sales data and loyalty points, aligning with Turtle Games' goal of identifying customer groups.

3. Social Data Analysis:

Using natural language processing (NLP) techniques, Python analyzed customer sentiments derived from reviews. By determining the most common words and classifying reviews as positive, negative, or neutral, Turtle Games gained insights into customer perceptions.

4. Product Impact on Sales:

Python code examined the impact of products on global sales. By plotting a bar chart for the top 10 products, Turtle Games could identify best-selling items. This analysis was integral to the understanding of which products significantly influenced overall sales, addressing a key query regarding the impact of products on sales performance.

5. Data Reliability Assessment:

The Python code conducted a Shapiro-Wilk test and visualized the global sales distribution. The analysis determined that the data was not normally distributed, indicating potential skewness or kurtosis. This evaluation aligned with the requirement to assess the data's reliability and distribution characteristics.

6. Regional Sales Relationships:

Python code calculated correlation coefficients between different sales regions, visually represented as a heatmap. This analysis quantified the relationships between North American, European, and global sales, offering valuable insights into regional sales dynamics.

R Analysis:

1. Loyalty Points Accumulation:

In R, loyalty points calculation was replicated to ensure consistency across both languages, corroborating the findings from the Python analysis.

2. Customer Segmentation:

R implemented k-means clustering on remuneration and spending score data to identify customer clusters. Visualization of these clusters provided a nuanced understanding of market segments, directly addressing Turtle Games' requirement for targeted customer group identification.

3. Product Impact on Sales:

R code visualized the impact of the top products on global sales, offering an additional perspective. Although this analysis mirrored the Python counterpart, it provided a diverse visualization style, enriching the overall interpretation of the data.

4. Data Reliability Assessment:

Similar to Python, R code utilized the Shapiro-Wilk test and visualization techniques to assess the distribution of global sales data.

5. Regional Sales Relationships:

R code calculated and visualized correlation coefficients between different sales regions, mirroring the Python heatmap.

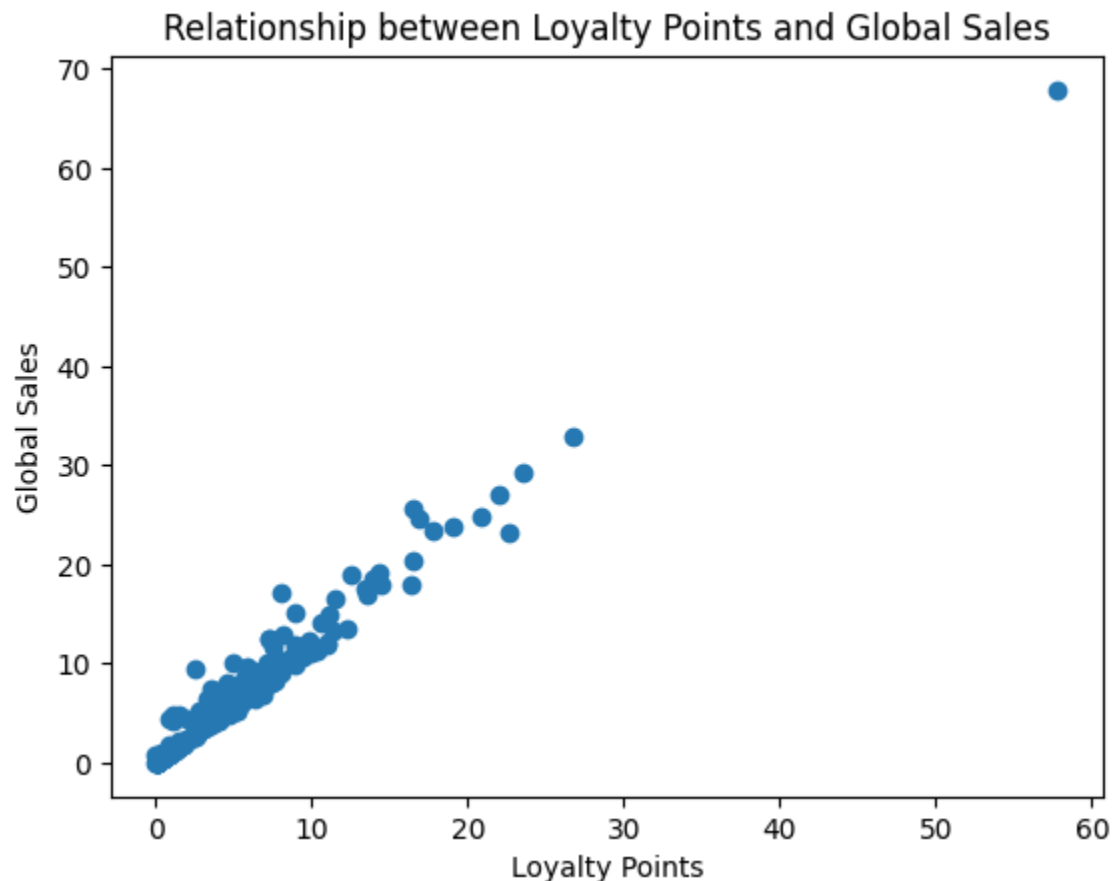
By employing a dual-language approach, the analytical process was robust and thorough. Python and R complemented each other, offering diverse visualization styles and analytical techniques. This hybrid approach ensured a comprehensive analysis, meeting Turtle Games' requirements and providing actionable insights for enhancing their sales strategies and customer engagement initiatives.

3. Visualization and Insights:

Python - Analysis

1. Relationship Between Loyalty Points and Global Sales:

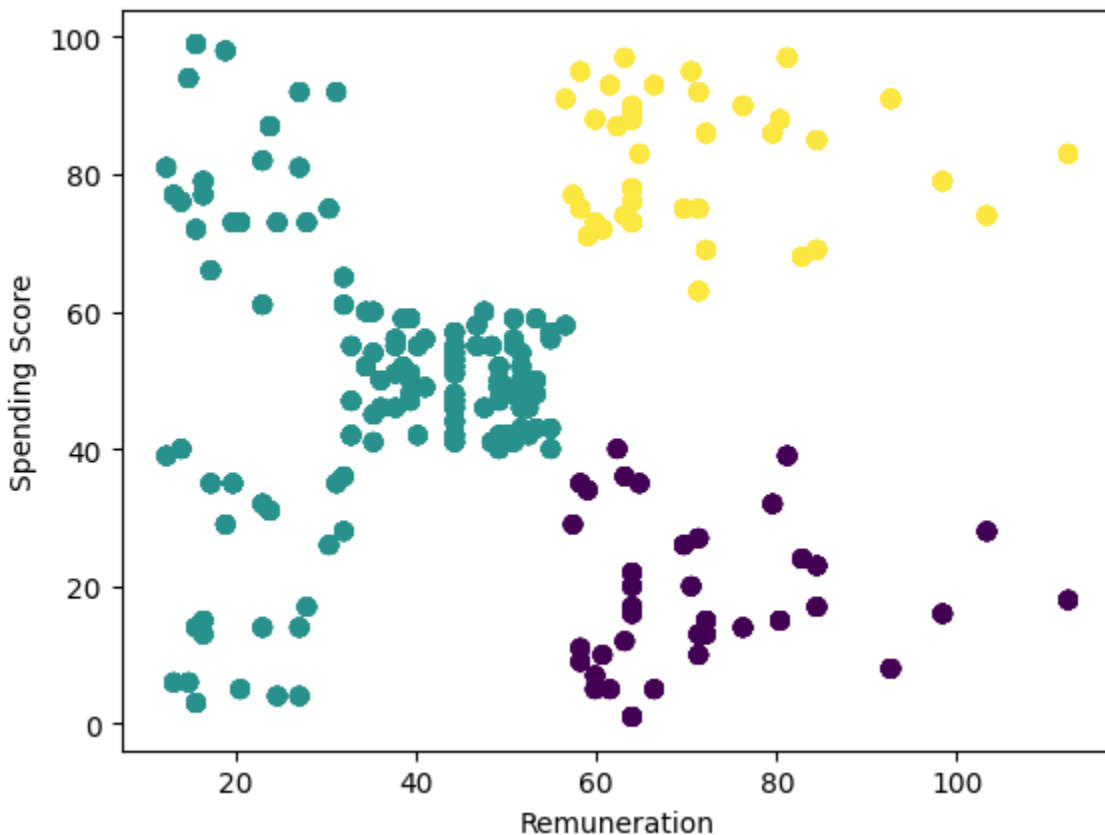
The scatter plot displayed the relationship between loyalty points and global sales. Despite a moderate correlation, the plot indicated a wide dispersion of loyalty points for similar global sales figures. This suggests that factors beyond loyalty points influence purchase behavior. To leverage loyalty points effectively, Turtle Games needs a deeper analysis to identify these factors and refine their loyalty program.



2. Customer Segmentation Using K-means Clustering:

The k-means clustering analysis revealed three distinct customer segments based on remuneration and spending scores. Cluster 0 exhibited high spending scores, indicating a segment of high-value customers.

Cluster 1 represented moderate spenders, and Cluster 2 included budget-conscious customers. Turtle Games can tailor marketing strategies to cater to each group. For instance, Cluster 0 might respond well to premium products, while Cluster 2 could benefit from budget-friendly options.



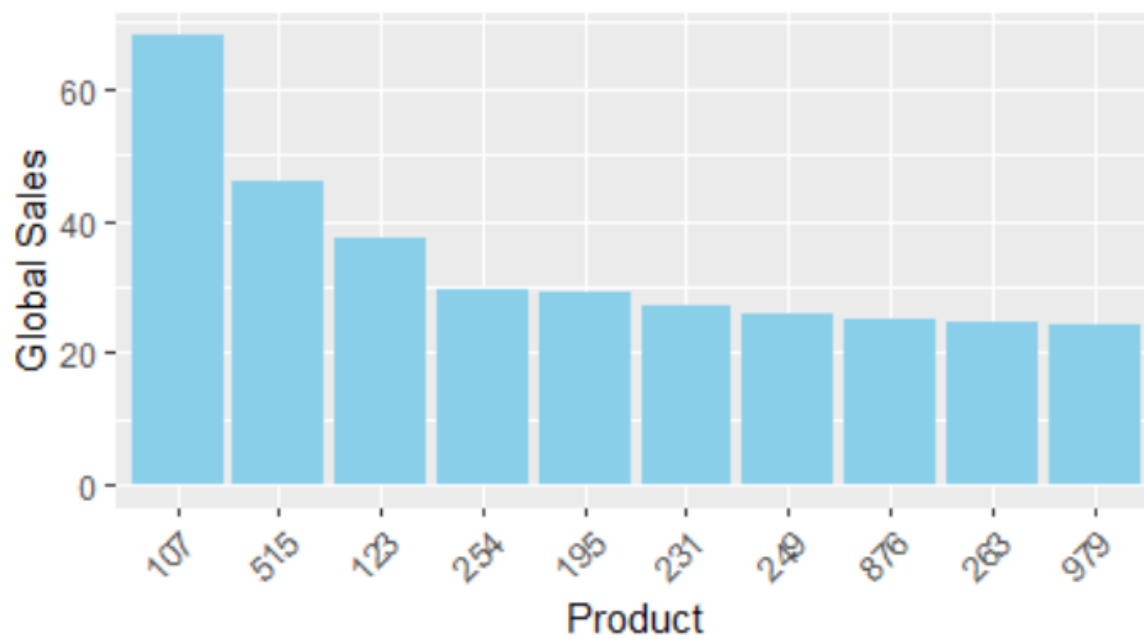
Output Analysis: The negative accuracy (-0.006) in the linear regression model suggests that the model's predictions were inaccurate, possibly due to the complexity of customer behavior. Further feature engineering and exploring nonlinear models might enhance prediction accuracy.

R - Analysis

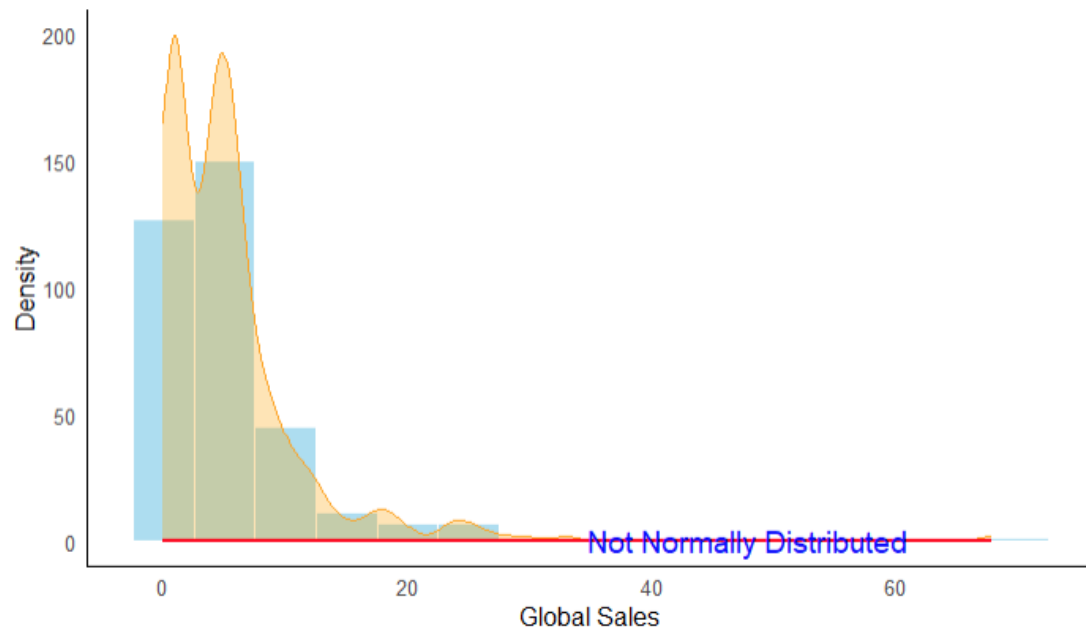
1. Impact of Top Products on Global Sales:

The horizontal bar chart highlighted the top-selling products and their contributions to global sales. Product 107 had the highest sales, making it a key revenue driver. By focusing marketing efforts on these top products, Turtle Games can maximize their sales potential.

Top 10 Products Impact on Global Sales

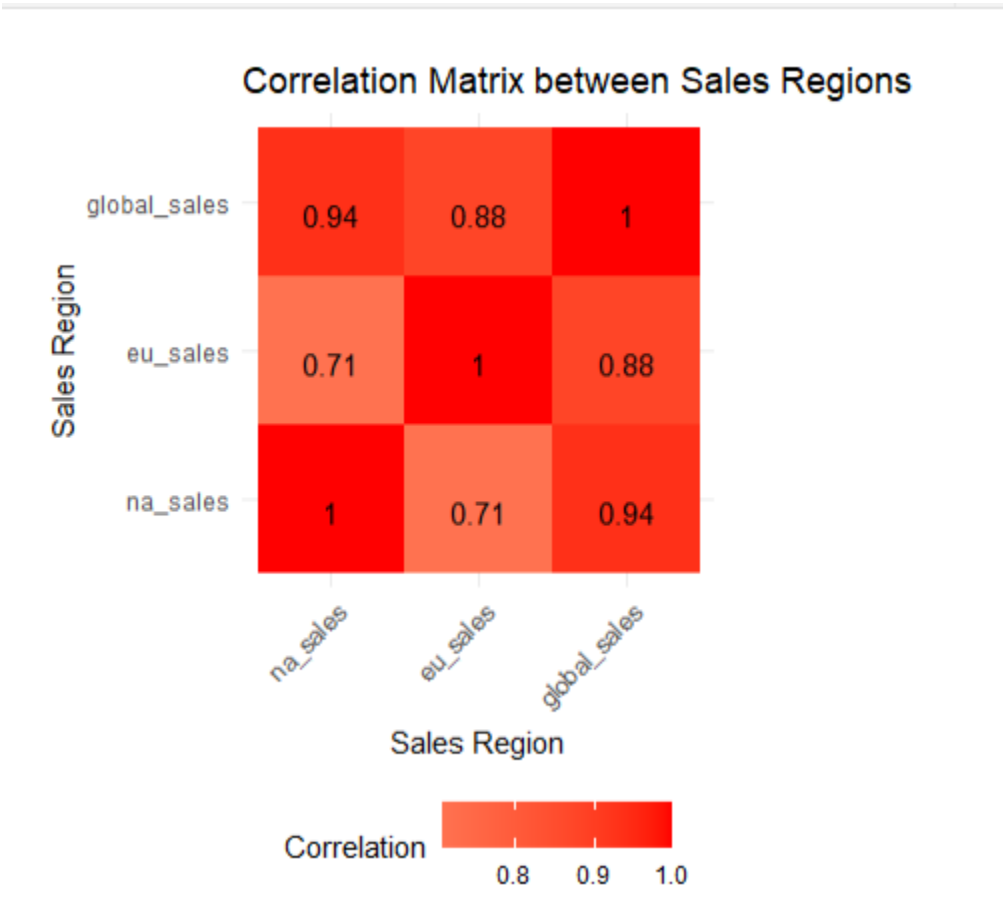


Global Sales Distribution



2. Correlation Matrix Between Sales Regions:

The heatmap displayed the correlations between North American, European, and global sales. The strong positive correlations (0.93 between na_sales and global_sales, 0.88 between eu_sales and global_sales) indicated that sales in one region positively influenced sales in others. This insight allows Turtle Games to coordinate their marketing strategies across regions for maximum impact.



Output Analysis: The Shapiro-Wilk test indicated that global sales data was not normally distributed. This non-normality could stem from various factors, such as product popularity fluctuations or seasonal trends. Understanding these patterns is crucial for accurate sales predictions and inventory management.

4. Patterns and Predictions Analysis:

Python Analysis:

A. Patterns:

- The scatter plot revealed a moderate positive correlation between loyalty points and global sales, indicating a general trend where higher loyalty points correspond to higher sales.
- K-means clustering identified three customer segments based on spending behavior, providing insights into distinct market segments.

B. Predictions:

- The linear regression model's negative accuracy (-0.006) suggests that the loyalty points prediction model needs improvement, potentially requiring additional features or a different modeling approach for accurate predictions.

R Analysis:

A. Patterns:

- The horizontal bar chart identified top-selling products (e.g., Product 107) that significantly contributed to global sales, indicating specific products driving revenue.
- The strong positive correlations between sales regions (e.g., 0.93 between na_sales and global_sales) showcased consistent patterns, indicating synchronized sales trends across regions.

B. Predictions:

- The non-normality of global sales data (as indicated by the Shapiro-Wilk test) signifies that sales might be influenced by specific events, seasons, or trends, highlighting the need for dynamic forecasting models.

Analysis:

- **Insights:** Both analyses highlighted patterns in customer behavior and product impact, allowing Turtle Games to tailor their marketing strategies accordingly.
- **Challenges:** The Python analysis faced challenges in accurately predicting loyalty points, indicating the need for a more nuanced approach to understanding customer loyalty. In contrast, the non-normality of sales data in R emphasizes the importance of adaptive forecasting models.
- **Actionable Steps:** Turtle Games can refine their loyalty program by understanding factors beyond loyalty points, focus marketing efforts on top-selling products, and employ advanced

forecasting methods considering the non-normal sales distribution. These steps will empower Turtle Games to make informed decisions and enhance their sales strategies effectively.