Turtle Reviews Data Analysis Project

1. Introduction

In this project, I explored customer behavior using the **Turtle Reviews dataset** to uncover the main factors influencing customer loyalty.

My goal was to understand how customer demographics such as **age, gender, remuneration, and spending score** contribute to loyalty points and overall engagement.

I used **R** for data cleaning, exploration, visualization, and regression modeling.

This project combines data analytics and storytelling to reveal actionable insights that businesses can use to improve customer retention and design targeted marketing strategies.

2. Objective

The main aim of this project was to analyze customer characteristics and identify what drives **loyalty points** in the Turtle Reviews dataset.

More specifically, I sought to:

- Explore how **spending behavior** and **income** impact loyalty points.
- Examine whether **age** and **gender** influence customer engagement.
- Visualize key trends and relationships within the customer base.
- Build a **regression model** to quantify which factors significantly predict loyalty.

3. Tools & Techniques

- **Language:** SQL ,R and PYTON
- **Libraries:** ggplot2, dplyr, tidyverse
- **Visualization:** Histograms, scatterplots, and boxplots for distribution and relationship analysis
- **Statistical Analysis:** Multiple linear regression model
- **Environment:** RStudio

4. Data Preparation

I began by importing the dataset turtle_reviews.csv and performing essential cleaning steps:

- Verified file path and structure using file.exists() and head()
- Checked for missing or duplicated data using colSums(is.na(df)) and duplicated()
- Confirmed that the dataset was clean, with **2000 rows** and **no missing values**

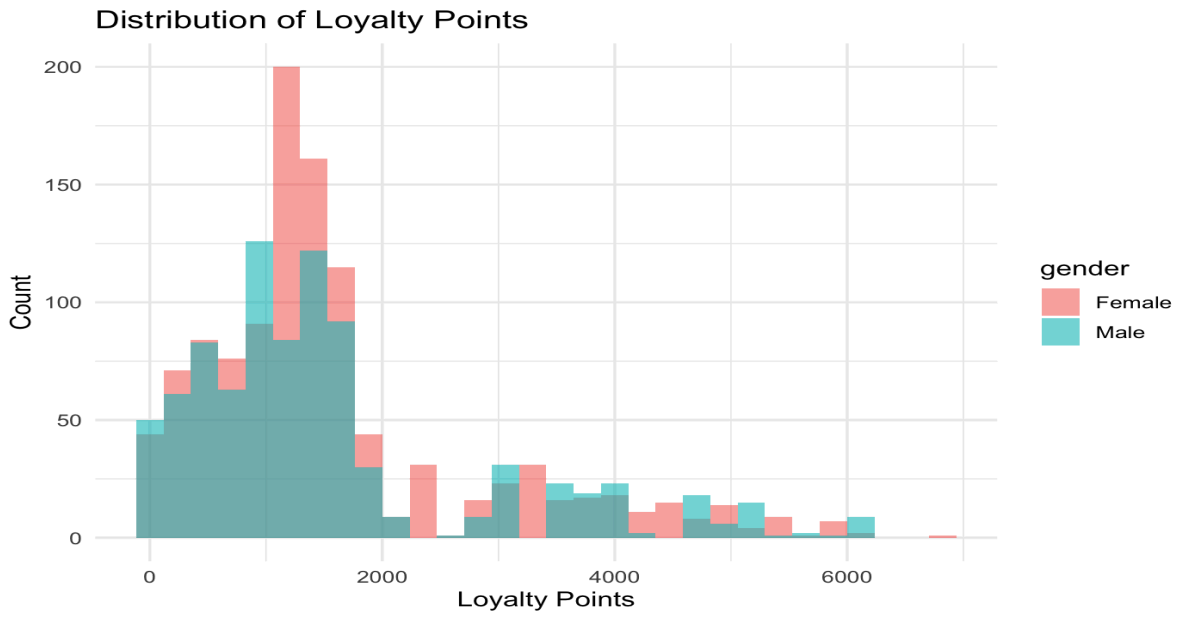
Key variables included:

- **Gender**
- **Age**
- **Remuneration (Income)**
- **Spending Score**
- **Loyalty Points**

These variables formed the foundation for both descriptive and predictive analysis.

5. Key Visuals and Insights

Figure 1: Distribution of Loyalty Points



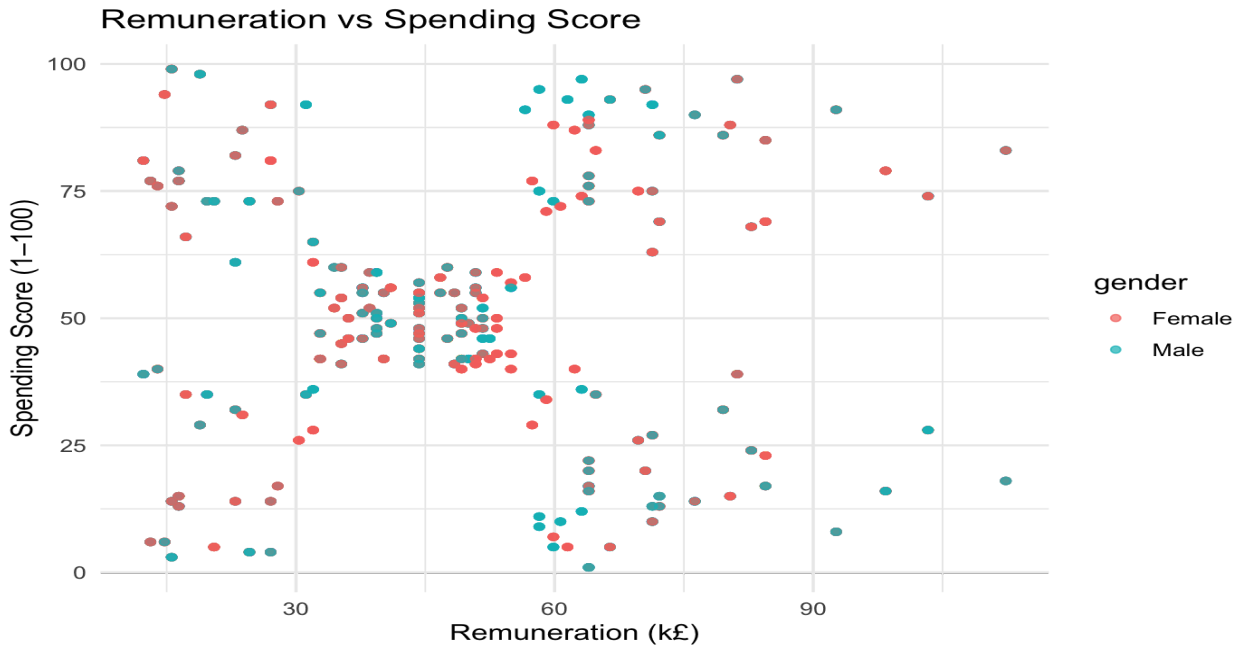
This histogram shows how loyalty points are distributed across all customers.

Most customers have loyalty points clustered between **1000–2000**, while a few outliers exceed **6000 points**.

Insight:

The skewed distribution indicates that while most customers are moderately loyal, a smaller high-value group contributes disproportionately to the loyalty program. These customers represent a key retention opportunity.

Figure 2: Relationship Between Remuneration and Spending Score



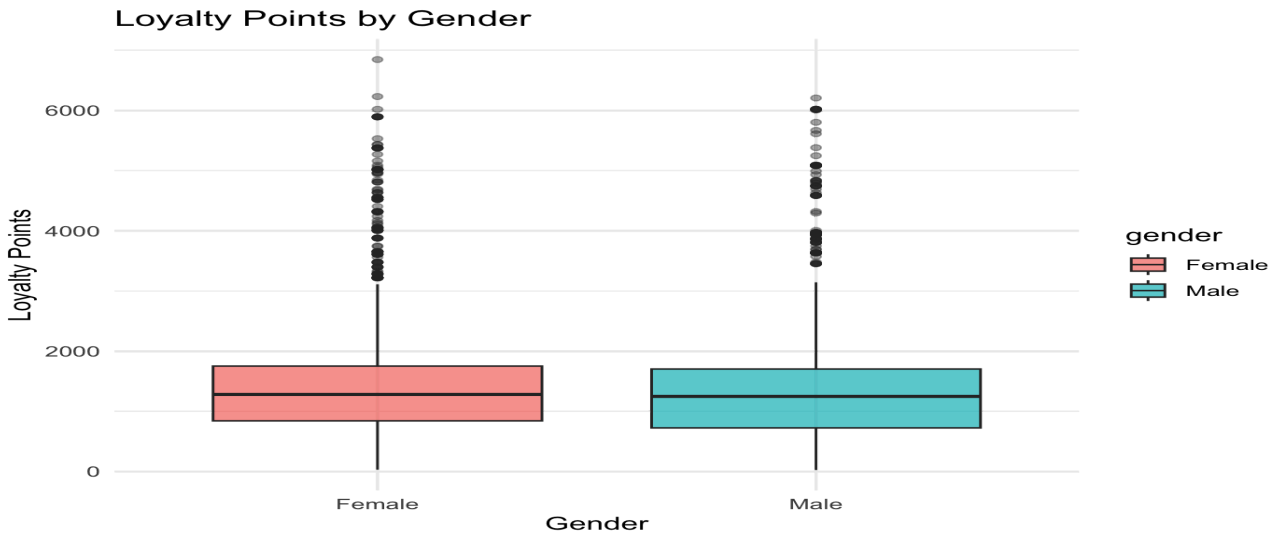
This scatter plot examines the link between customers' income levels and spending scores.

Although there is no perfect correlation, a **positive trend** is visible higher earners tend to have **higher spending scores**.

Interpretation:

Customers with higher incomes are more likely to spend and thus earn more loyalty points. This supports the idea of creating tiered incentives for premium spenders.

Figure 3: Loyalty Points by Gender



This boxplot compares loyalty points between **male** and **female** customers.

The median loyalty points for males appear slightly higher, though both genders show similar variability.

Conclusion:

While gender differences are present, the overlap suggests that loyalty is influenced more by spending habits than gender itself. However, gender-specific campaigns could still be useful for targeted marketing.

6. Regression Analysis

I developed a **multiple linear regression model** to predict loyalty points based on customer characteristics:

Dependent Variable:

- Loyalty Points

Independent Variables:

- Age
- Remuneration
- Spending Score
- Gender

Results:

- **Spending Score** had the **strongest positive correlation** with loyalty points.
- **Remuneration** also had a significant positive effect.
- **Age** showed a smaller but steady influence.
- **Gender** was statistically significant, suggesting behavioral differences between male and female customers.

The model achieved an **R^2 value of 0.84**, explaining **84% of the variance** in loyalty points an excellent fit that validates the findings.

7. Insights

1. Customers with **higher spending scores** consistently have more loyalty points.
2. **Income level** influences loyalty wealthier customers engage more actively in loyalty programs.
3. **Gender differences** exist, though spending habits remain the strongest driver.
4. Most customers cluster around average loyalty levels, but a **small elite group** contributes a large share of engagement and revenue.

8. Recommendations

- **Introduce tiered loyalty programs** that reward high-spending customers with exclusive benefits.
- **Personalize marketing campaigns** based on gender and income segments to increase engagement.
- **Monitor loyalty-to-spending ratios** regularly to keep the rewards system balanced and competitive.
- **Encourage mid-tier customers** through bonus point incentives to increase participation and move them into the high-value group.

9. Reflection

This project helped me strengthen my skills in **data cleaning, exploratory analysis, and regression modeling using R**.

It also improved my ability to translate raw data into business insights that drive customer engagement strategies.

By combining statistical evidence with clear visual storytelling, I demonstrated how data analytics can directly support marketing and loyalty program optimization.

10. Conclusion

Through the Turtle Reviews analysis, I identified key behavioral and demographic drivers of customer loyalty.

By understanding who the most engaged customers are and what motivates them, businesses can design smarter, more personalized loyalty systems.

This project combines technical skill with business strategy showing how data-driven insights can enhance **customer retention, satisfaction, and profitability**.