**Statistical Analysis Assignment Report**

**1. Business Questions:**

* **Sales Trends Across Branches:** How do total sales vary among different branches?
* **Impact of Customer Type on Sales:** Does the type of customer (Member vs. Normal) significantly affect total sales?
* **Average Rating by Gender:** Is there a significant difference in average ratings given by male and female customers?

**2. Hypotheses:**

* **Hypothesis 1:** There is a significant difference in total sales between different branches.
* **Hypothesis 2:** Customer type significantly influences total sales.
* **Hypothesis 3:** There is a significant difference in average ratings between genders.

**3. Data Preparation:**

The dataset was loaded and examined for missing values to ensure the integrity of the analysis.

# Checking for missing values

> missing\_values <- sum(is.na(data))

> print(missing\_values)

[1] 0

**4. Descriptive Statistics:**

Summary statistics provide insights into the overall performance of the supermarket.

summary\_stats <- data %>%

+ summarise(

+ Total\_Sales = sum(Total),

+ Average\_Rating = mean(Rating, na.rm = TRUE),

+ Total\_Quantity = sum(Quantity)

+ )

> print(summary\_stats)

Total\_Sales Average\_Rating Total\_Quantity

1 322966.7 6.9727 5510

**Key Metrics:**

* **Total Sales:** $**[insert total sales value]**
* **Average Rating:** **[insert average rating value]**
* **Total Quantity Sold:** **[insert total quantity]**

**5. Inferential Statistics:**

Statistical tests were conducted to validate the hypotheses.

**Sales Across Branches (ANOVA)**

Df Sum Sq Mean Sq F Value Pr(>F)

Branch 2 106988 53494 0.885 0.413

Residuals 997 60292151 60474

**Output**: ANOVA results revealed an F value of 0.885 and a p-value of 0.413.

**Interpretation**:

There are no statistically significant differences in total sales among the branches. The high p-value suggests that variations in total sales are likely due to random chance rather than branch-specific factors.

**Customer Type Impact on Sales (t-test)**

> t\_test\_results <- t.test(Total ~ Customer.type, data = data)

> print(t\_test\_results)

Welch Two Sample t-test

data: Total by Customer.type

t = 0.62155, df = 997.84, p-value = 0.5344

alternative hypothesis: true difference in means between group Member and group Normal is not equal to 0

95 percent confidence interval:

-20.85669 40.19359

sample estimates:

mean in group Member Mean in group Normal

327.7913 318.1229

**Output**: The t-test showed a t value of 0.62155 and a p-value of 0.5344.

**Interpretation**: The p-value exceeds 0.05, indicating no significant difference in average spending between Member ($327.79) and Normal customers ($318.12). The confidence interval for the mean difference (-20.86 to 40.19) includes zero.

**Ratings by Gender (t-test)**

> t\_test\_gender <- t.test(Rating ~ Gender, data = data)

> print(t\_test\_gender)

Welch Two Sample t-test

data: Rating by Gender

t = -0.15166, df = 996.5, p-value = 0.8795

alternative hypothesis: true difference in means between group Female and group Male is not equal to 0

95 percent confidence interval:

-0.2298699 0.1968882

sample estimates:

mean in group Female Mean in group Male

6.964471 6.980962

**Output**: The t-test for ratings produced a t value of -0.15166 and a p-value of 0.8795.

**Interpretation**: The p-value is well above 0.05, indicating no significant difference in ratings between genders. The average ratings are 6.96 for females and 6.98 for males, with a confidence interval of (-0.2299 to 0.1969) suggesting a lack of difference.

**6. Data Visualization:**

Visualizations enhance understanding of data trends.

Total Sales by Branch:

A graph of a sales by branch

Description automatically generated with medium confidence

Average Rating by Gender:

A graph with a couple of colored squares

Description automatically generated with medium confidence

**7. Conclusion**

1. **Sales Across Branches (ANOVA)**:
   * **Output**: F value of 0.885, p-value of 0.413.
   * **Interpretation**: No significant differences in total sales among branches. Branch identity does not significantly explain the variance in sales performance.
2. **Impact of Customer Type on Sales (t-test)**:
   * **Output**: t value of 0.62155, p-value of 0.5344.
   * **Interpretation**: No significant difference in total sales between Member and Normal customers. The confidence interval for the mean difference includes zero.
3. **Ratings by Gender (t-test)**:
   * **Output**: t value of -0.15166, p-value of 0.8795.
   * **Interpretation**: No significant difference in ratings between genders. The confidence interval suggests a lack of meaningful difference.

**Measures of Performance:**

* **ANOVA**: Indicates that branch identity does not significantly explain the variance in sales.
* **t-tests**: Measure the strength of evidence against the null hypothesis. The confidence intervals indicate uncertainty in the mean differences.

**Confidence Measures:**

* The confidence intervals derived from the t-tests show that we cannot confidently assert meaningful differences between the groups compared.

**8. Recommendations:**

1. **Branch Performance Strategies**: Analyze branch-specific operational efficiencies and share successful tactics across branches.
2. **Enhancing Customer Engagement**: Explore loyalty programs to convert Normal customers into Members through targeted promotions.
3. **Customer Satisfaction Improvement**: Conduct surveys to understand customer satisfaction across demographics, improving services and offerings.
4. **Further Analysis**: Consider analyzing additional factors, such as product lines or customer demographics, to gain deeper insights.
5. **Monitor Trends Over Time**: Implement a dashboard for real-time sales and rating tracking to respond quickly to emerging patterns.