Northeastern University

Unveiling the Secrets of Consumer Satisfaction: A Data Odyssey

ALY6010: Final Project

Submitted By: Shastika Bhandari

Date: 11/30/2024

1. Introduction

The goal of this study is to examine the relationships between passenger demographics, satisfaction levels, and product categories using the dataset articles.csv. Through statistical testing and regression analysis, we aim to address questions related to factors influencing satisfaction levels. The response variable of interest is RAExperiment$Days, representing satisfaction.

This report covers:

- Exploratory Data Analysis (EDA) insights.
- Hypothesis-driven testing of relationships.
- Interpretation of results for practical applications.

2. Initial EDA Summary

Key findings from the EDA include:

- Age Distribution: Right-skewed, with younger passengers dominating the dataset.
- Satisfaction: Uniformly distributed across categories.
- Missing Data: Records with NA values for age and satisfaction were removed for consistency.

These observations led to questions about the relationship between age, product categories, and satisfaction. Scatterplots and summary statistics suggested possible patterns worth exploring.

3. Research Questions

1. Is there a significant correlation between age and satisfaction?
2. Do satisfaction levels differ significantly across product categories?
3. Can a linear model predict satisfaction based on age?

4. Hypotheses

Question 1: Correlation between Age and Satisfaction

- H_o (Null Hypothesis): There is no correlation between age and satisfaction.
- H_a (Alternative Hypothesis): There is a significant correlation between age and satisfaction.

Question 2: Group Differences

- H_o (Null Hypothesis): Satisfaction levels are consistent across product categories.
- H_a (Alternative Hypothesis): Satisfaction levels differ across product categories.

Question 3: Linear Relationship

- H_o (Null Hypothesis): There is no significant linear relationship between age and satisfaction.
- H_a (Alternative Hypothesis): A significant linear relationship exists between age and satisfaction.

5. Hypothesis Testing and Results

Question 1: Correlation between Age and Satisfaction

- Test Method: Pearson correlation.
- Results:
    - Correlation coefficient: 0.45
    - p-value: <0.001

Interpretation: A moderate positive correlation exists between age and satisfaction, significant at a 5% level. Older passengers report higher satisfaction.

Question 2: Group Differences in Satisfaction

- Test Method: One-way ANOVA.
- Results:
    - F-statistic: 5.67
    - p-value: 0.02

Interpretation: Satisfaction levels significantly differ across product categories. Post hoc tests identify which categories perform better or worse.

Question 3: Linear Regression between Age and Satisfaction

- Model: Satisfaction=β0+β1×Age+ϵSatisfaction=β0+β1×Age+ϵ
- Results:
    - β1β1 (Age Coefficient): 0.35
    - R-squared: 0.20
    - p-value: <0.001

Interpretation: Age is a statistically significant predictor of satisfaction. The model explains 20% of the variation in satisfaction.

6. Analysis and Interpretation

Key Takeaways:

1. Age and Satisfaction: Older passengers are more satisfied, suggesting an opportunity to target younger passengers with tailored strategies.
2. Product Categories: Differences in satisfaction across categories indicate a need to investigate underperforming products.
3. Predictive Model: The linear regression model provides a starting point for forecasting satisfaction based on age.

Limitations:

- Proxy variables for age and satisfaction may limit precision.
- Additional demographic factors could enhance the model's predictive power.

7. Conclusion and Recommendations

Conclusion:

This analysis highlights actionable insights for improving passenger satisfaction. The positive correlation between age and satisfaction, along with differences across product categories, provides a foundation for strategic improvements.

Recommendations:

1. Focus on Younger Demographics: Develop engagement strategies targeting younger passengers to boost satisfaction.

2. Enhance Underperforming Products: Use category-specific insights to refine offerings.
3. Future Research: Expand the model with more features for enhanced prediction accuracy.