# **Final Report of Comprehensive Consumer Behavior Analysis Enriched by Google Trends**

## **Project Proposal**

## This project aims to analyze comprehensive consumer behavior data to uncover key insights into customer demographics, purchase behaviors, product preferences, and decision-making processes. The project will further enrich this analysis using external trend data from Google Trends to better understand how online search behaviors influence consumer purchasing decisions.

## **Motivation**

The primary motivation behind this project was to explore the relationship between online search behavior, represented by Google Trends data, and actual consumer purchase behavior. Understanding these relationships could potentially enable businesses to better predict consumer actions and enhance customer satisfaction through targeted strategies.

## **Data Source**

* Primary Data: The "Comprehensive Consumer Behavior" dataset was obtained from Kaggle, featuring detailed demographic information, purchase history, customer satisfaction ratings, and product feedback.
* Secondary Data: Google Trends search volume data across eight randomly selected product categories was collected using the pytrends Python library.

## **Data Collection and Preparation**

* The Kaggle dataset was downloaded in CSV format, and purchase data was aggregated weekly for analysis.
* Google Trends data was similarly downloaded, cleaned, and aligned temporally to match weekly aggregation.
* The datasets were merged based on corresponding weekly dates to allow joint analysis.

***Weekly Purchase Frequency vs Google Search Interest for Baby Products***

## 

## This time-series plot shows how consumer purchase frequency compares with Google search interest in "Baby Products" over time.

## 

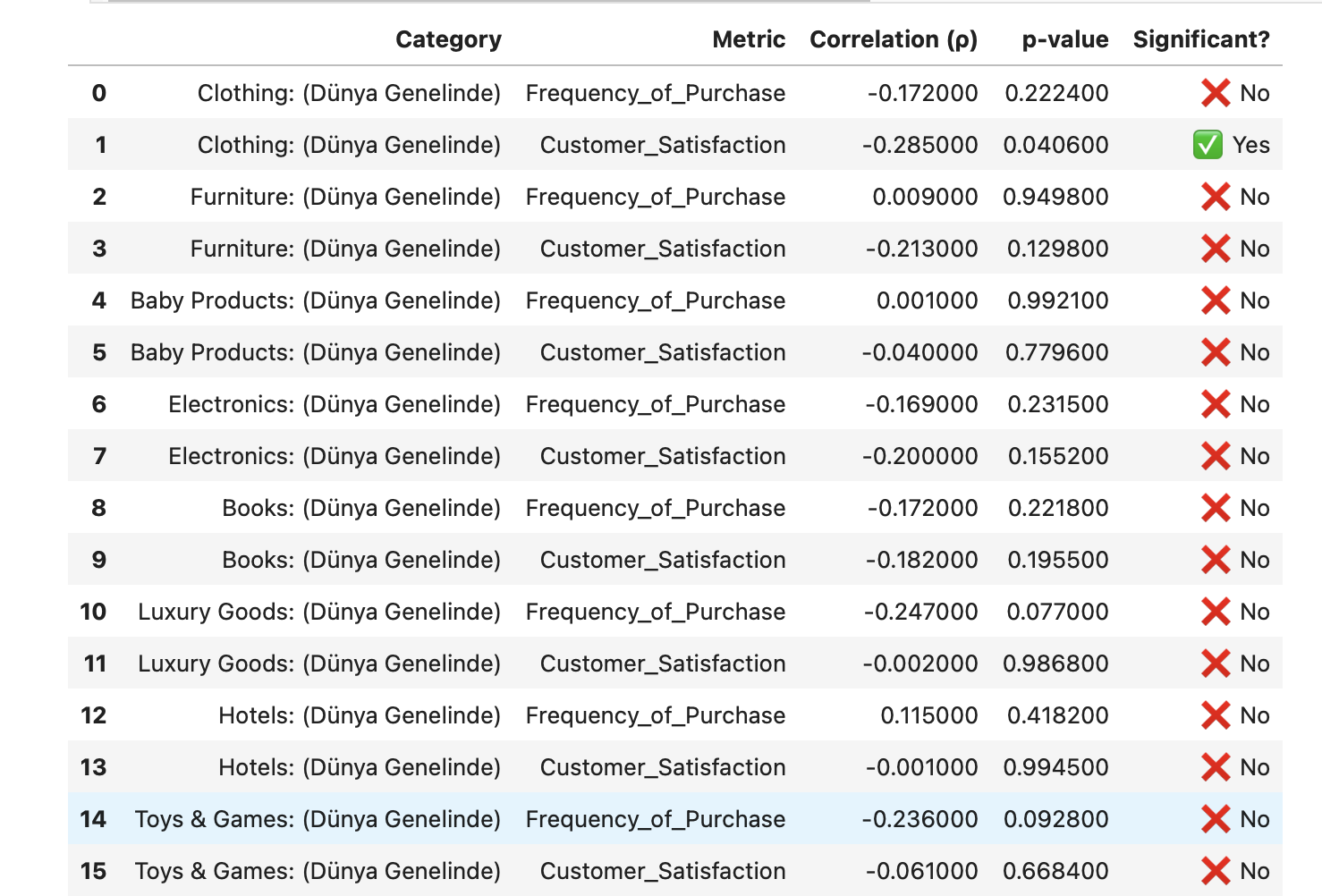
## **Data Analysis**

### **Exploratory Analysis and Hypothesis Testing**

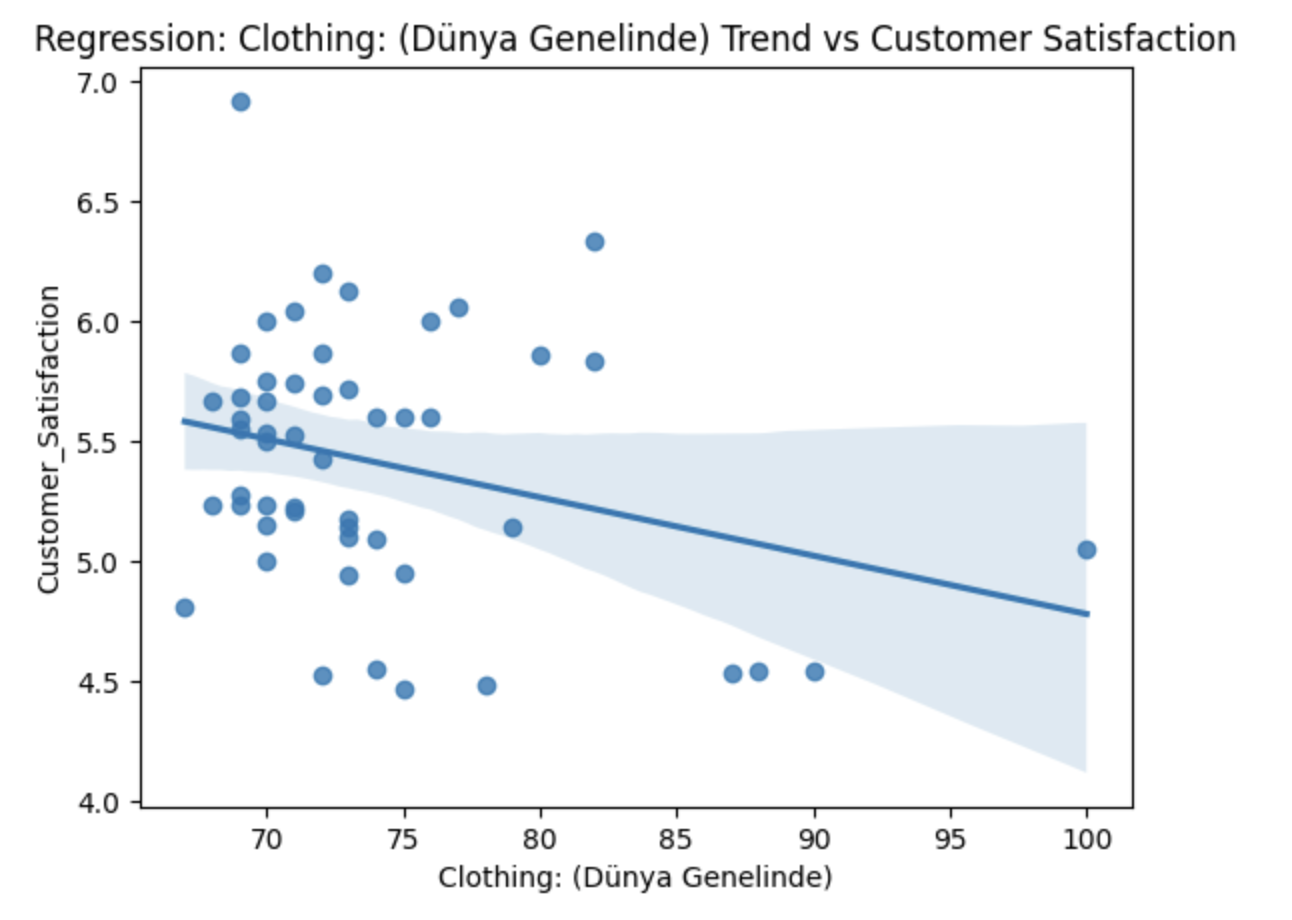
We investigated the following hypothesis:

* Null Hypothesis (H₀): There is no significant relationship between Google Trends search volume and consumer behavior metrics (frequency of purchase, customer satisfaction).
* Alternative Hypothesis (H₁): There is a significant relationship between Google Trends search volume and consumer behavior metrics.

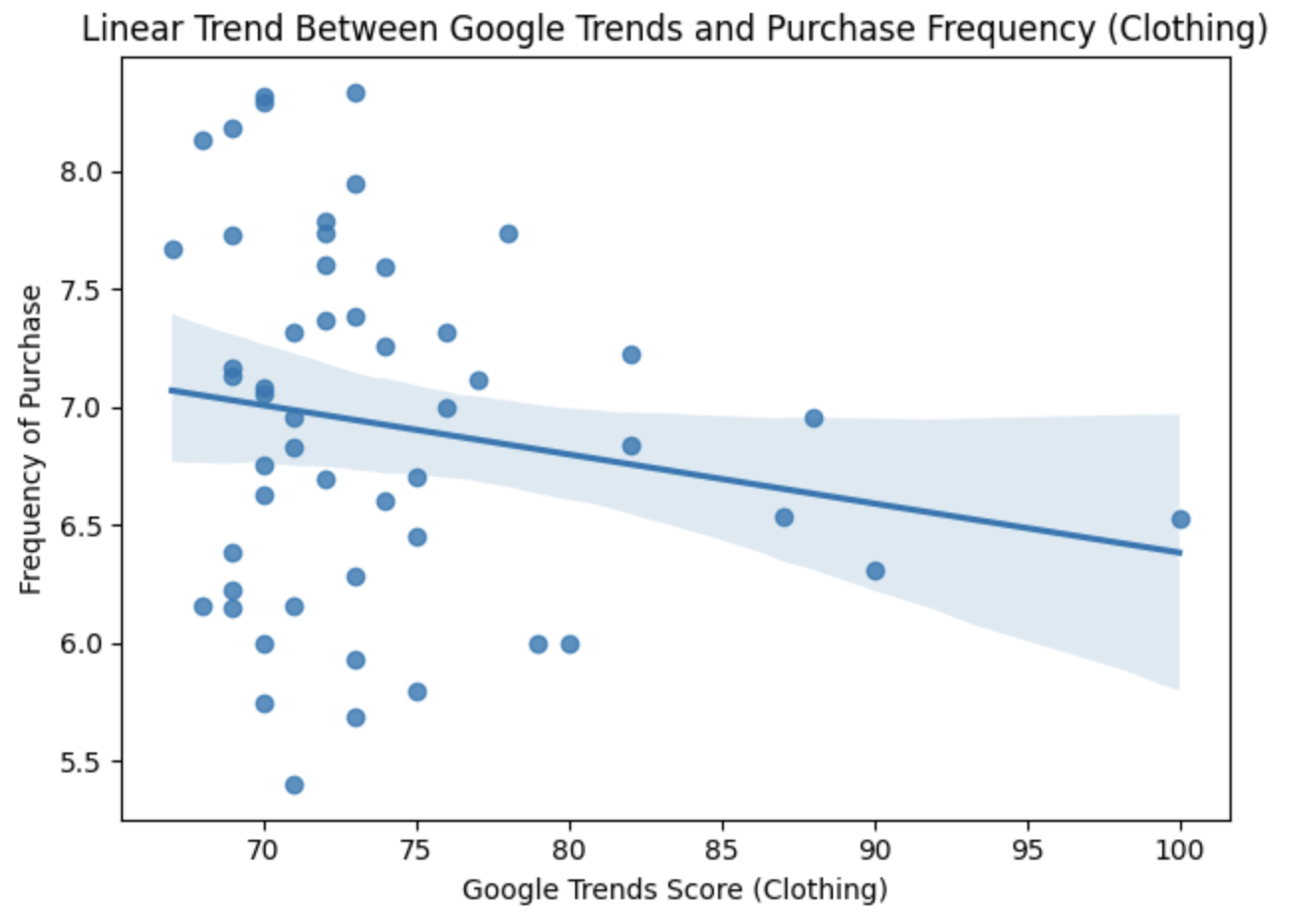
We performed Pearson correlation tests for each product category across two metrics: frequency of purchase and customer satisfaction.



This table summarizes the Pearson correlation and p-values for each product category and behavior metric.



*Linear Relationship between Clothing Search Trend and Customer Satisfaction  
Only this relationship was statistically significant (p = 0.0406), showing a modest negative correlation.*



This plot shows no significant correlation for frequency, supporting the earlier statistical result.

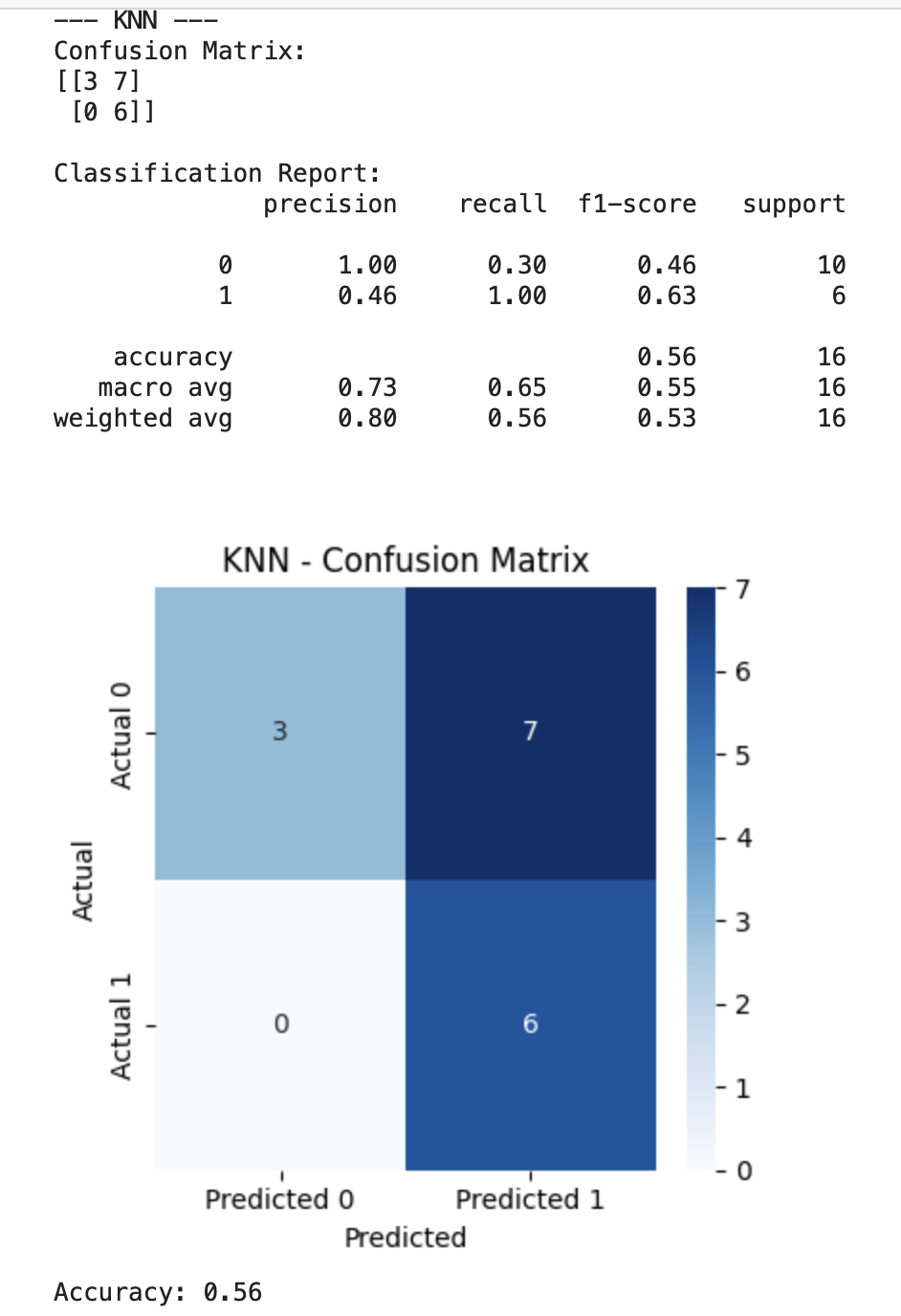
### **Key Findings from Correlation Analysis**

* Most categories showed no statistically significant correlation (p > 0.05), indicating weak predictive power of Google Trends alone.
* Clothing was the only category with a statistically significant negative correlation (ρ = -0.285, p = 0.0406) related to customer satisfaction, suggesting increased online interest in clothing correlates slightly with decreased customer satisfaction.
* Categories such as Luxury Goods (ρ = -0.247) and Toys & Games (ρ = -0.236) showed moderate, yet statistically non-significant correlations.

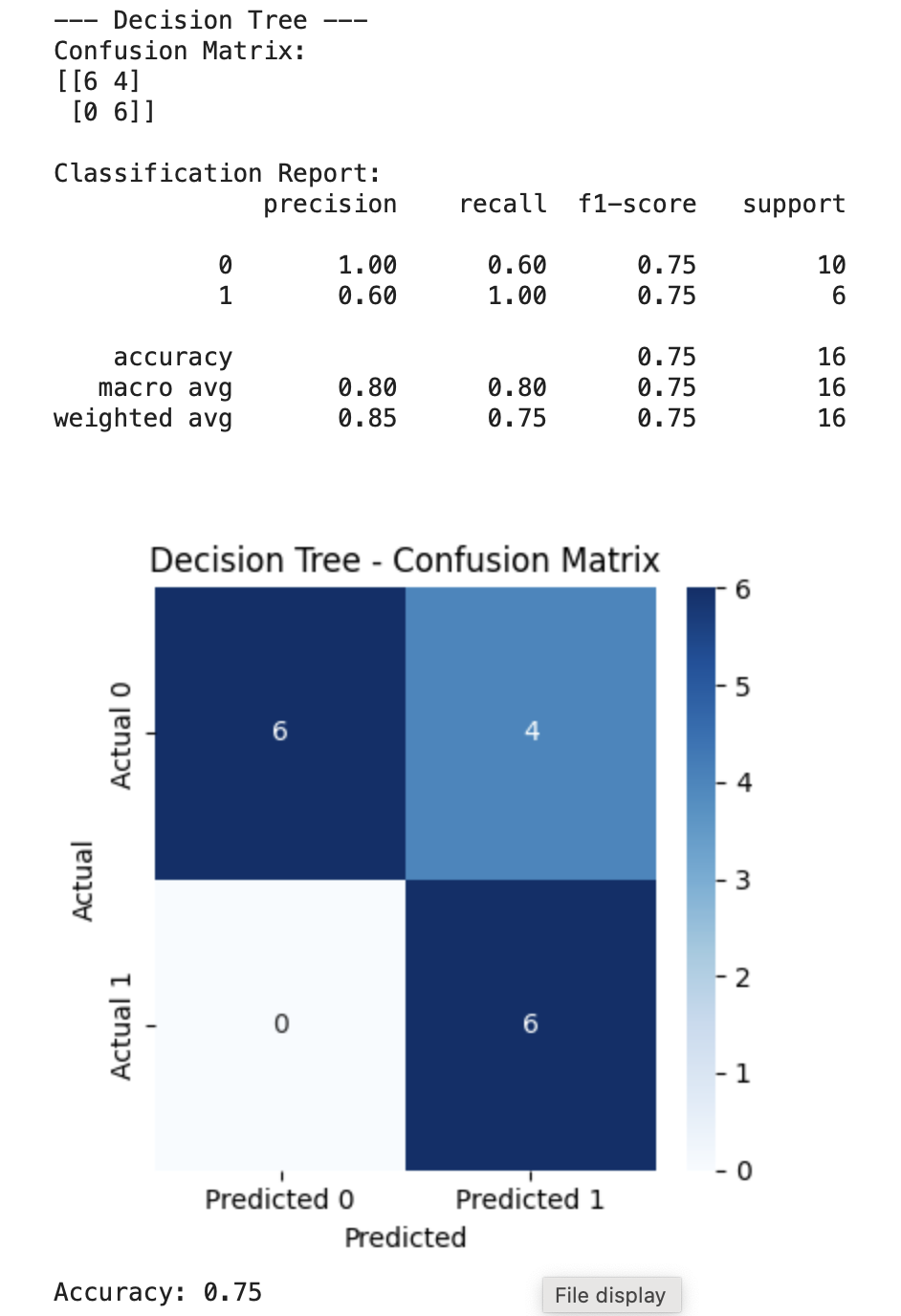
**Machine Learning Analysis**

We further investigated the significant category (Clothing) to predict customer satisfaction (converted to binary classification: high vs. low satisfaction) using three machine learning models:

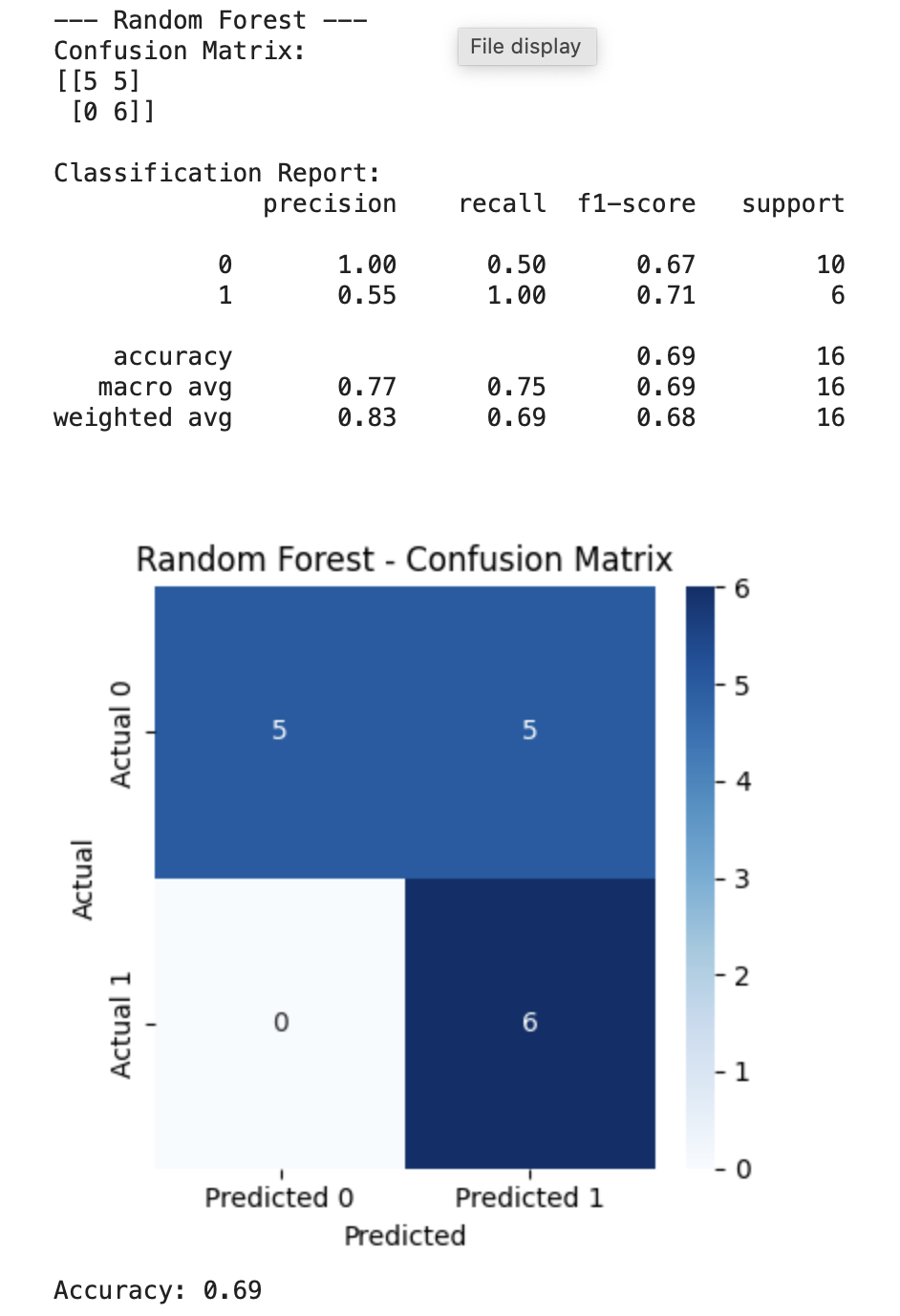
* K-Nearest Neighbors (KNN): Achieved an accuracy of 56%.



* Decision Tree: Achieved an accuracy of 75%.



* Random Forest: Achieved an accuracy of 69%.



The Decision Tree model performed best, indicating that while Google Trends data alone had limited predictive power, certain relationships could be leveraged with appropriate modeling techniques.

## **Findings**

* Limited Predictive Power: Google Trends data alone generally showed weak predictive relationships with consumer behavior metrics.
* Notable Exception (Clothing): There was a meaningful, although modest, inverse relationship between clothing-related searches and customer satisfaction.
* Moderate Correlations: Categories like Luxury Goods and Toys & Games exhibited moderate correlations worth further exploration, potentially driven by seasonal or contextual factors.

## **Limitations and Future Work**

* Data Limitations: The primary dataset and Google Trends data were aggregated weekly, possibly masking finer temporal dynamics.
* Feature Enrichment: Future analysis could benefit from additional data sources like social media sentiment, economic indicators, or demographic-specific online trends.
* Expanded Modeling: Employing more sophisticated machine learning models and feature engineering could improve predictive accuracy.

## **Conclusion**

This analysis provides an initial exploration into the relationship between consumer behavior and online search trends. While the findings indicate limited predictive capacity for Google Trends data alone, the modest yet significant relationship identified in the clothing category highlights opportunities for deeper investigation, especially with richer data sources and advanced analytical techniques.

Aysu Yazdıçutan

31945