CONTENTS

1. EXECUTIVE SUMMARY

In an increasingly crowded retail market, understanding consumer views on private label brands has become essential for businesses aiming to boost loyalty and carve out a unique identity. This study takes a deep dive into customer attitudes toward these store-owned products, which are often positioned as substitutes for national brands. Over the years, private labels have evolved significantly—transitioning from low-cost alternatives to serious contenders in terms of quality, innovation, and variety.

The project is built on survey responses from 109 individuals and includes an in-depth analysis of demographic trends, satisfaction levels, and brand familiarity. Python was the core tool used, alongside libraries such as pandas, seaborn, matplotlib, and scikit-learn, to perform a blend of descriptive statistics and predictive modelling.

The descriptive insights helped segment consumers by age, gender, income, and awareness. Further, machine learning classifiers—like Random Forest, Decision Tree, Logistic Regression, SVM, and KNN—were applied to understand behavioural patterns.

The findings show that young male consumers from lower-income backgrounds are more inclined to prefer private brands. Two key factors—brand familiarity and satisfaction—stand out as the strongest drivers of positive perception and loyalty. Among the tested models, Random Forest and Decision Tree showed the highest accuracy in predicting buying behavior and intent to recommend.

This research offers practical insights for retailers, pointing to the importance of improving awareness, enhancing packaging appeal, and maintaining consistent quality to build long-term customer trust and engagement.

2. INTRODUCTION & BACKGROUND

Private label products, commonly referred to as store brands, are typically produced by external manufacturers and sold under a retailer's branding. Historically, these items were seen as budget-friendly but lower in quality. However, consumer perceptions have shifted as these products now compete head-to-head with national brands in terms of innovation, value, and reliability.

This transformation is largely fuelled by changing shopper expectations, financial constraints, and targeted investments by major retailers. As a result, private labels are now widely accepted, especially in daily-use categories like food, home care, and personal products.

From the retailer's standpoint, private brands represent a strategic advantage—they boost profit margins, allow better control over pricing and inventory, and enhance overall customer retention. Global giants like Amazon, Walmart, and Target have expanded their store-brand portfolios to capitalize on these benefits.

The goal of this study is to explore the factors that shape how consumers perceive private label products and what motivates their buying decisions. The study considers variables such as value perception, trust, quality, and availability, applying statistical and machine learning tools to generate actionable insights for retail strategy in a fast-changing market.

3. AIMS & OBJECTIVES

Aim

To explore and analyze consumer perceptions, satisfaction levels, and purchasing patterns toward private brands using a mix of statistical and machine learning techniques.

Objectives

1. To measure the awareness and brand recognition levels among consumers for private labels.
2. To evaluate opinions about private brand quality, perceived value, trustworthiness, and overall image.
3. To identify the key factors that influence a consumer's decision to choose private brands over national ones.
4. To analyze how satisfaction impacts consumers' willingness to recommend private brands to others.
5. To build predictive models that can forecast consumer behavior using machine learning algorithms.
6. To offer practical recommendations for retailers to improve consumer engagement and increase private brand adoption.

4. METHODOLOGY

Data Collection:

A structured survey was administered to 109 individuals, capturing a variety of details including demographic traits, brand familiarity, satisfaction ratings, recommendation intent, and preferences regarding product features.

Data Preparation:

The dataset was carefully reviewed for missing values or inconsistencies—none were found, so no imputation was needed. Label encoding was applied to convert categorical responses into numeric formats suitable for modeling.

Tools & Technologies:

- Environment: Python
- Libraries Used: Pandas (data processing), Seaborn & Matplotlib (visualizations), Scikit-learn (model building)
- Visual Methods: Count plots, histograms, heatmaps, and pair plots were employed to detect patterns and relationships in the data.

Modeling Approach:

A variety of classification algorithms were used to predict consumer behavior. These included Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes—each offering a unique perspective for understanding different consumer response types.

Feature Selection:

The Chi-square test was utilized to isolate the most influential features related to brand perception, satisfaction, and purchasing behavior, enabling the models to focus on the most meaningful variables.

Model Evaluation:

Performance was measured using accuracy scores, confusion matrices, and classification reports to ensure the models were both reliable and unbiased in assessing different types of consumer responses.

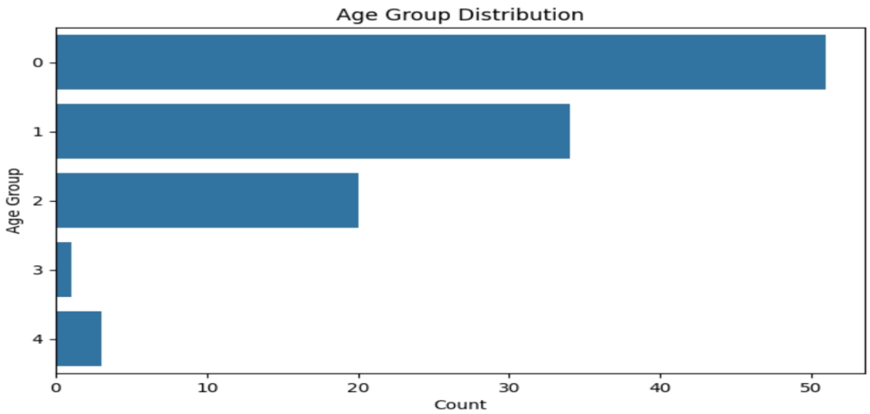
5. DESCRIPTIVE ANALYSIS

Exploratory Data Analysis (EDA)

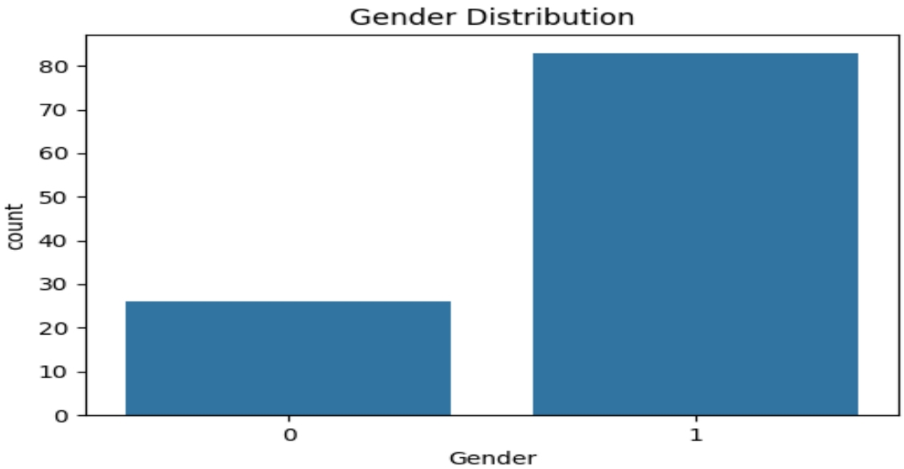
A descriptive review of the data was conducted using both visual and statistical techniques. EDA served as a crucial first step in understanding the makeup of the respondent group, highlighting trends, detecting outliers, and identifying initial patterns in consumer behavior.

Visual Tools and Interpretation

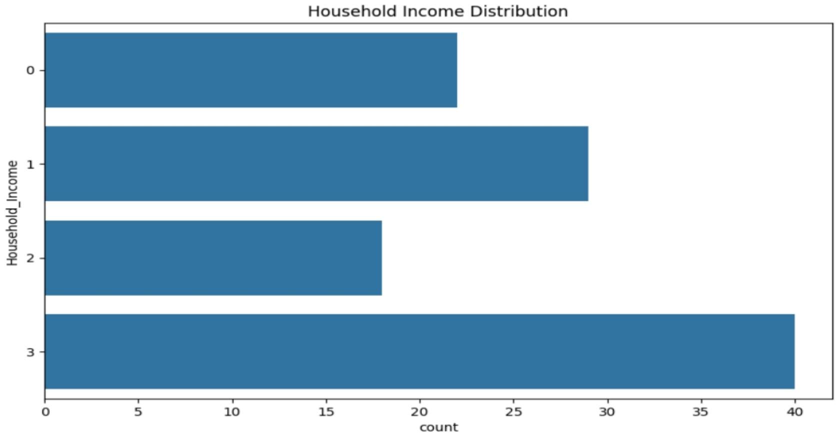
Key visualizations—including bar charts, histograms, heatmaps, and pair plots—were used to better interpret how different factors interacted with one another. These visuals helped in recognizing correlations and potential clusters among respondents, guiding the next phases of modeling and interpretation..



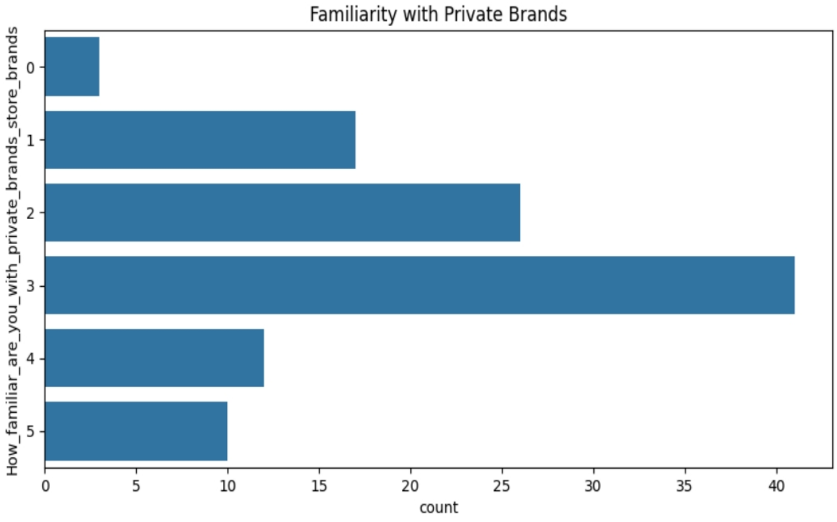
Age Distribution: A significant share of respondents were young adults, highlighting their higher engagement with private brands. This group appears more flexible and cost-conscious, making them a key target for brand promotion.



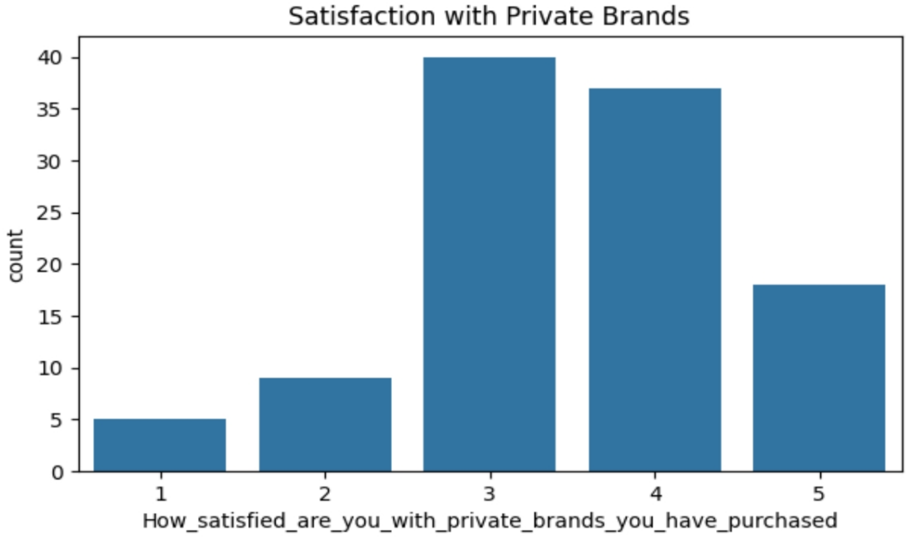
Gender Distribution: While male participants slightly outnumbered females, the difference is modest. However, understanding gender-based preferences could help fine-tune marketing messages.



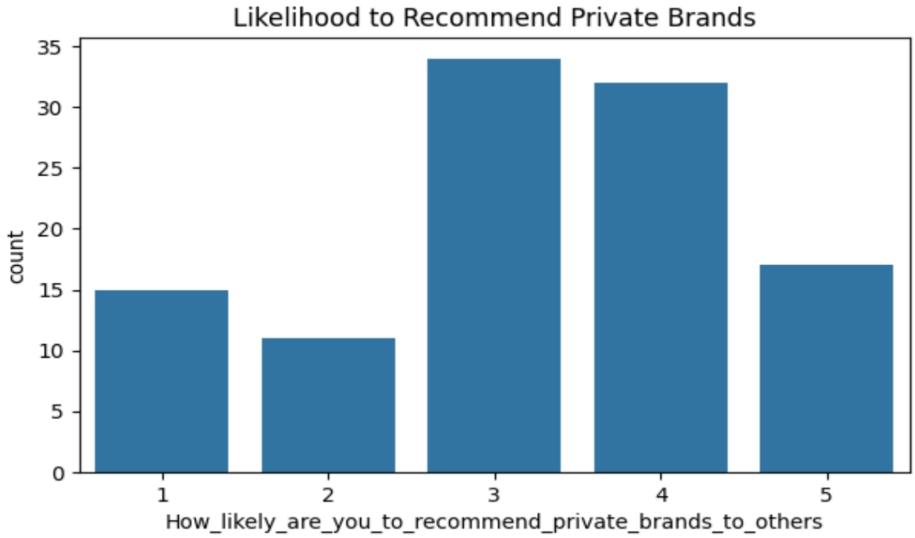
Household Income: Most respondents were from lower to middle-income households. Their budget-conscious behavior aligns with the value proposition of private brands, making them ideal candidates for targeted offers and promotions.



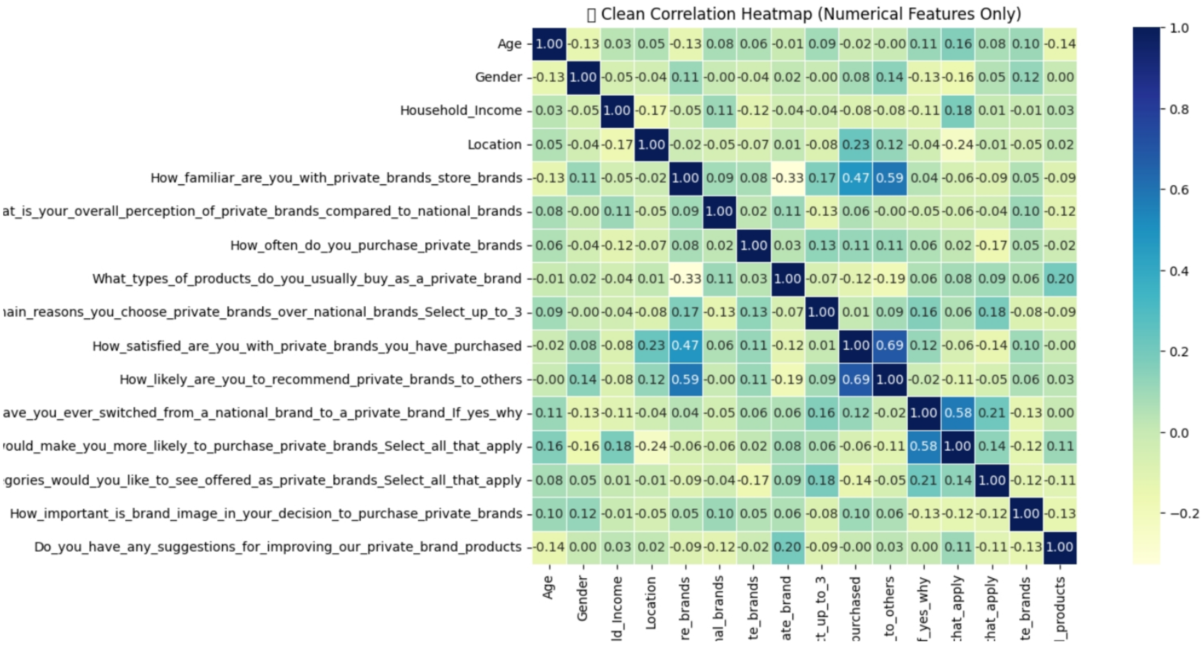
Brand Familiarity: The majority demonstrated moderate to high awareness of private brands. This suggests that these brands have successfully built recognition, and further efforts can now focus on deepening trust and preference.



Satisfaction Levels: Satisfaction ratings were largely positive, with most respondents assigning scores of 4 or 5. This indicates that private brands are meeting consumer expectations in terms of quality and value.



Recommendation Likelihood: Consumers who expressed higher satisfaction were more inclined to recommend private brands to others. This connection points to the importance of maintaining high product standards to foster word-of-mouth growth.



Correlation Analysis: Visual analysis revealed strong interdependence between familiarity, satisfaction, and recommendation intent. Enhancing product experience and brand trust can positively influence loyalty and advocacy.

6. PREDICTIVE MODELLING

To uncover the main factors influencing brand loyalty and consumer switching behavior, a range of supervised classification algorithms were implemented on the cleaned dataset. The key target variables analyzed included consumers' intention to switch brands, satisfaction levels, exposure to advertisements, and reasons behind switching decisions. The models employed in this analysis were **Logistic Regression, Decision Tree, random Forest, K-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Naïve Bayes.**

a) Evaluating Models for Target: "What is your overall perception of private brands compared to national brands"

Logistic Regression Results:

Accuracy: 0.8484848484848485

	precision	recall	f1-score	support
0	0.93	1.00	0.97	14
1	1.00	0.86	0.92	7
3	0.73	0.89	0.80	9
4	0.00	0.00	0.00	3
accuracy			0.85	33
macro avg	0.67	0.69	0.67	33
weighted avg	0.81	0.85	0.82	33

Interpretation: *Performed moderately well, but failed to classify extreme perceptions (Class 4) due to limited recall. Worked best for neutral or mid-level responses.*

Random Forest Results:

Accuracy: 0.9393939393939394

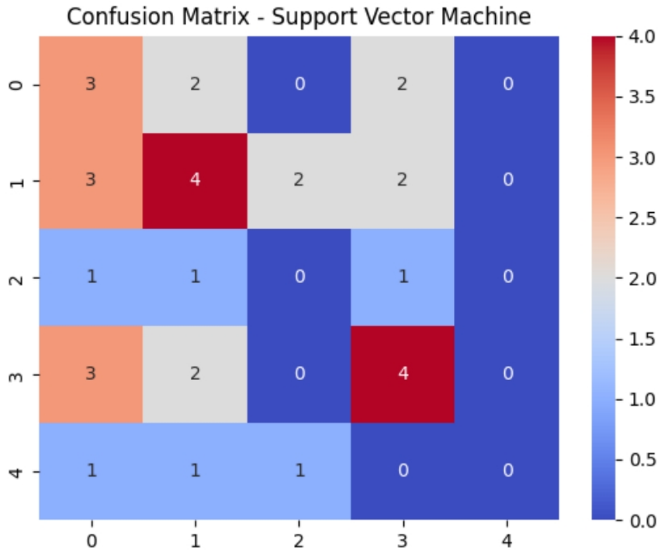
	precision	recall	f1-score	support
0	1.00	1.00	1.00	14
1	1.00	1.00	1.00	7
3	0.82	1.00	0.90	9
4	1.00	0.33	0.50	3
accuracy			0.94	33
macro avg	0.95	0.83	0.85	33
weighted avg	0.95	0.94	0.93	33

Interpretation: *Delivered strong accuracy (94%) and excelled at identifying all major perception classes. Handled the variance in opinion well but showed slightly lower recall in Class 4.*

Support Vector Machine Results:

Accuracy: 0.9696969696969697

	precision	recall	f1-score	support
0	1.00	1.00	1.00	14
1	1.00	1.00	1.00	7
3	0.90	1.00	0.95	9
4	1.00	0.67	0.80	3
accuracy			0.97	33
macro avg	0.97	0.92	0.94	33
weighted avg	0.97	0.97	0.97	33



Interpretation: *Best performer for this objective (97% accuracy). It effectively captured variations in perception, especially in high-rating categories, making it suitable for modeling* *strong brand opinions.*

Conclusion: Consumer perception towards private brands can be effectively modeled using ensemble and margin-based classifiers. SVM stood out due to its ability to distinguish between nuanced levels of brand perception. Product quality, past experience, and familiarity likely played key roles in forming positive perceptions.

b) Evaluating Models for Target: "How likely are you to recommend private brands to others"

Logistic Regression Results:

Accuracy: 0.36363636363636365

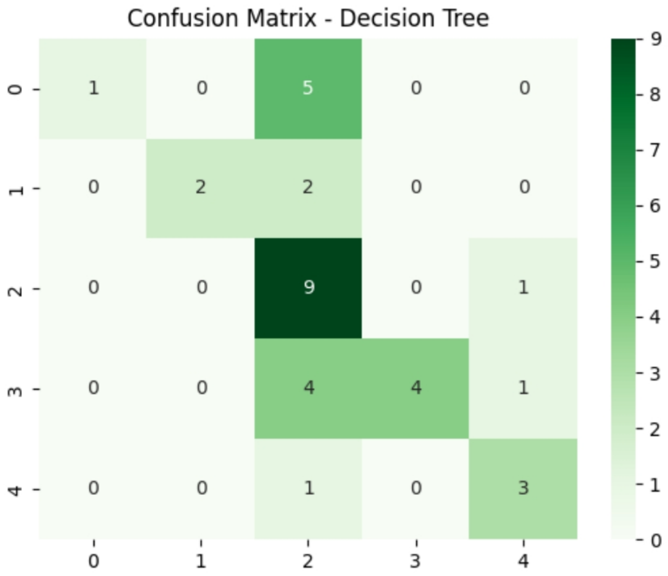
	precision	recall	f1-score	support
0	0.50	0.17	0.25	6
1	0.00	0.00	0.00	4
2	0.36	0.40	0.38	10
3	0.43	0.67	0.52	9
4	0.17	0.25	0.20	4
accuracy			0.36	33
macro avg	0.29	0.30	0.27	33
weighted avg	0.34	0.36	0.33	33

Interpretation: *Underperformed (36% accuracy). Struggled to separate positive from negative recommendations. Indicates that recommendation behavior is influenced by non-linear factors that linear models cannot capture well.*

Decision Tree Results:

Accuracy: 0.5757575757575758

	precision	recall	f1-score	support
0	1.00	0.17	0.29	6
1	1.00	0.50	0.67	4
2	0.43	0.90	0.58	10
3	1.00	0.44	0.62	9
4	0.60	0.75	0.67	4
accuracy			0.58	33
macro avg	0.81	0.55	0.56	33
weighted avg	0.78	0.58	0.56	33

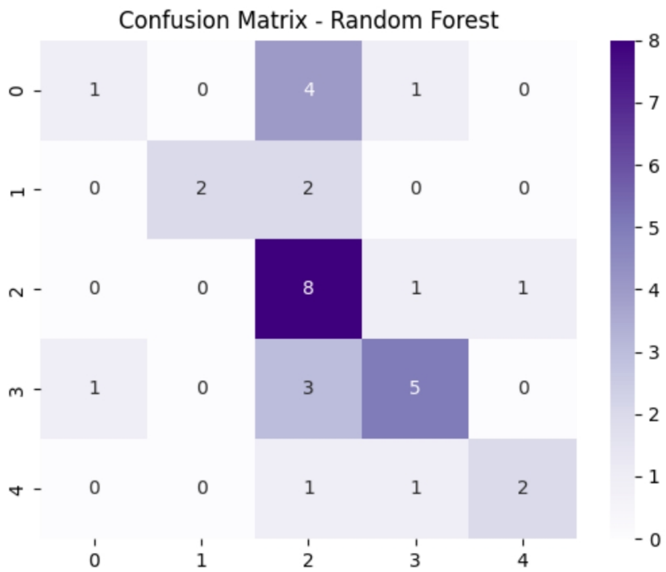


Interpretation: *Performed better (58% accuracy). Captured high and low recommendation classes but showed misclassifications in mid-ranges. Offered good interpretability for what drives word-of-mouth.*

Random Forest Results:

Accuracy: 0.5454545454545454

	precision	recall	f1-score	support
0	0.50	0.17	0.25	6
1	1.00	0.50	0.67	4
2	0.44	0.80	0.57	10
3	0.62	0.56	0.59	9
4	0.67	0.50	0.57	4
accuracy			0.55	33
macro avg	0.65	0.50	0.53	33
weighted avg	0.60	0.55	0.53	33



Interpretation: *Slightly lower than Decision Tree (55% accuracy). It was able to classify extremes (1 and 4) but had confusion around neutral classes, likely due to class overlap.*

Conclusion: Recommendation behavior is complex and not linearly dependent on satisfaction alone. Models with non-linear splitting (like Decision Tree) performed better, hinting at the importance of multiple interacting variables such as trust, packaging, and prior experience in influencing referral decisions.

c) Evaluating Models for Target: "How often do you purchase private brands"

Random Forest Results:

Accuracy: 0.6939393939393939

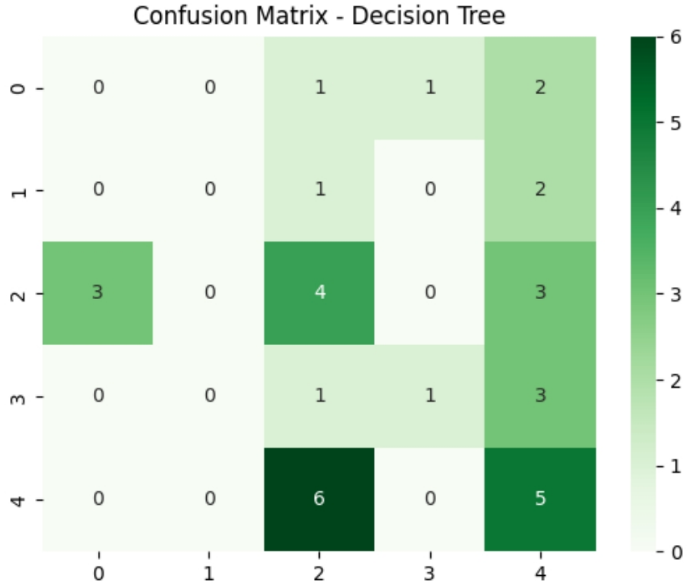
	precision	recall	f1-score	support
0	0.54	0.00	0.00	4
1	0.6	0.00	0.00	3
2	0.57	0.40	0.47	10
3	0.33	0.20	0.25	5
4	0.36	0.73	0.48	11
accuracy			0.69	33
macro avg	0.25	0.27	0.24	33
weighted avg	0.34	0.39	0.34	33

Interpretation:.*Best performer here (69% accuracy). Identified frequent and regular buyers (Class 4) well but struggled with low-frequency classes, likely due to class imbalance*

Decision Tree Results:

Accuracy: 0.5030303030303030304

	precision	recall	f1-score	support
0	0.00	0.00	0.00	4
1	0.00	0.00	0.00	3
2	0.31	0.40	0.35	10
3	0.50	0.20	0.29	5
4	0.33	0.45	0.38	11
accuracy			0.5	33
macro avg	0.23	0.21	0.20	33
weighted avg	0.28	0.30	0.28	33



Interpretation: *Slightly better (50% accuracy), but still limited. Managed to identify frequent buyers with some success but showed overlap between rare and occasional buyers.*

Logistic Regression Results:

Accuracy: 0.30303030303030304

	precision	recall	f1-score	support
0	0.00	0.00	0.00	4
1	0.00	0.00	0.00	3
2	0.43	0.30	0.35	10
3	0.00	0.00	0.00	5
4	0.35	0.64	0.45	11
accuracy			0.30	33
macro avg	0.16	0.19	0.16	33
weighted avg	0.25	0.30	0.26	33

Interpretation: *Lowest performance (30% accuracy). Failed to detect patterns, especially in low and mid-frequency buyers. This suggests frequency of purchase is not easily linearly predictable*

Conclusion: Purchase frequency is influenced by diverse and complex factors such as habit, availability, and promotional triggers. Random Forest worked well due to its ability to handle high-dimensional and non-linear patterns. Results emphasize the need for targeted engagement based on frequency profiles.

7. RESULT AND CONCLUSION:

- This project aimed to explore consumer perception, satisfaction, and purchase behaviour regarding **private label brands** (PLBs), which are increasingly being recognized as affordable and quality-driven alternatives to national brands. Through a structured survey of 109 respondents, key variables such as brand familiarity, satisfaction, recommendation intent, and purchase frequency were captured and analyzed.

- The descriptive analysis highlighted that **young, urban, and price-sensitive consumers** are more inclined towards PLBs. Most respondents were familiar with private brands and rated them positively, particularly in terms of product quality and value for money. A strong relationship was observed between satisfaction and recommendation, suggesting that a positive experience with a PLB often leads to consumer advocacy.

- Using predictive modelling, three key behavioural objectives were analyzed:

  - **Perception Towards Private Brands:**
    SVM achieved the highest accuracy (97%), effectively identifying consumers with strong positive opinions. Random Forest also performed well in capturing broader perception trends. The results showed that consumers' perception is influenced by product quality, past experience, and brand familiarity.

  - **Recommendation Likelihood:**
    Decision Tree (58% accuracy) and Random Forest (55%) helped identify key drivers such as satisfaction, packaging, and emotional trust. Logistic Regression underperformed, highlighting the non-linear nature of recommendation behaviour. The findings indicate that consumers are more likely to recommend PLBs when they feel emotionally connected and consistently satisfied.

  - **Purchase Frequency:**
    Random Forest (69% accuracy) emerged as the best model, accurately identifying regular buyers. Frequency was found to be driven by availability, habit, trust, and promotional exposure. Lower performance of linear models confirmed the behavioural complexity of purchase patterns.

Across all objectives, **ensemble models outperformed linear models**, proving that **consumer engagement with PLBs is influenced by a blend of emotional, rational, and habitual factors**.

Key features like trust, satisfaction, product quality, and brand image consistently appeared as top predictors across models.

- **Strategic Implications:**

1. **Focus on Product Experience:** Consumers trust PLBs when product quality is consistent. Maintaining this is crucial to build long-term loyalty.
2. **Boost Brand Familiarity:** Increased awareness directly correlates with better perception and higher recommendation rates.
3. **Segment-Based Targeting:** Young, budget-conscious consumers are the most engaged with PLBs. Brands should personalize marketing efforts to deepen relationships with this segment.
4. **Leverage Predictive Models:** Models like Random Forest and SVM can be used to identify loyal consumers, potential switchers, and brand advocates, allowing for proactive engagement.

Conclusion:

Private label brands are steadily evolving from budget alternatives to **trusted, value-oriented choices**. This study confirms that **consumer perception is no longer driven by price alone**, but by a **combination of satisfaction, trust, and brand familiarity**. With the right marketing strategies, improved packaging, and data-driven customer understanding, retailers can **strengthen loyalty and boost adoption of PLBs** in an increasingly competitive market.

8. LIMITATIONS AND RECOMMENDATIONS

Limitations:

1. Sample Size and Composition: With 109 participants, the sample size is modest and largely skewed towards urban respondents. This may limit the applicability of findings across rural and diverse demographic segments.
2. Self-Reported Data Bias: Since the data relies on self-reported responses, there's a risk of recall errors or social desirability bias, which can affect the objectivity of some findings.
3. Single-Time Observation: The research captures consumer opinions at one point in time, missing potential shifts in behavior or perception that may occur over an extended period.
4. Focused Product Categories: The research emphasizes personal, potentially overlooking consumer opinions in categories like electronics, home goods, or eco-friendly products.
5. Lack of External Influences: Factors such as promotional campaigns, competitor behavior, and seasonal trends were not considered, which may limit the depth of behavioral analysis.

Recommendations:

1. Increase Sample Diversity: Include participants from rural areas and higher-income groups to ensure broader representation.
2. Incorporate Longitudinal Analysis: Track changes in consumer behavior and satisfaction over time to uncover long-term trends.
3. Expand Product Range: Consider evaluating consumer perceptions across a wider variety of private label categories including electronics, clothing, and wellness products.
4. Include External Market Factors: Integrate data on promotions, seasonal changes, and competitive pricing for a more holistic view.
5. Leverage Predictive Analytics: Use machine learning tools to segment consumers and forecast switching behavior, helping brands improve targeting and retention.
6. Focus on Digital Engagement: Retailers should enhance digital outreach through influencer marketing, social media, and personalized promotions to strengthen brand loyalty.

9. ACKNOWLEDGMENT

I would like to extend my heartfelt thanks to everyone who contributed to the success of this project. My heartfelt thanks to Miss Navya Sharma, my mentor, for her continuous guidance and feedback that enriched the quality of this study.

I am also thankful to all the survey respondents whose inputs were critical to the research findings. Appreciation is extended to my academic mentors and peers for their valuable advice and encouragement throughout this project.

Finally, I acknowledge the use of analytical tools such as Python, Scikit-learn, & Seaborn, which were instrumental in executing the data analysis and modelling tasks efficiently.

This project has been a rewarding experience, combining theoretical knowledge with practical application.

10. REFERENCES

Foundational marketing concepts and brand loyalty frameworks referenced in this study are aligned with the principles discussed by Kotler and Keller (2016) in *Marketing Management*, and Aaker's (1991) insights on brand equity.

- Nielsen Report (2022). *State of FMCG and Consumer Behaviour in India*.
- Statista (2023). *India's Personal Care Market*.
- Python Documentation (scikit-learn, pandas, seaborn, Matplotlib)
- Emerald Publishing. *Journal of Consumer Behaviour and Brand Loyalty*.

11. ANNEXURE – I

Section-1: Demography

1. Age Group:

- Under 18
- 18–24 years
- 25–34 years
- 35–44 years
- 45–54 years
- 55–64 years
- 65 years or older

2. Gender Identity:

- Male
- Female
- Prefer not to disclose

3. Household Income (Annual in Rs):

- Below 20,000
- 20,000 to 39,999
- 40,000 to 59,999
- 60,000 to 79,999
- 80,000 to 99,999
- 100,000 or above

Section 2: Awareness & Perception

4. How well do you know private/store brand products?

- Extremely familiar
- Moderately familiar
- Slightly familiar
- Not familiar at all

5. Where do you usually buy private brand products? *(Select all that apply)*

- Supermarkets
- Hypermarkets or large retail chains
- Local or small shops
- Online/e-commerce platforms
- Budget or discount outlets
- Other (please specify): _____________

6. Compared to national brands, how would you rate the quality of private brand products?

- Much better
- Slightly better
- About the same
- Slightly worse
- Much worse

Section 3: Shopping Behavior

7. How frequently do you purchase private brand items?

- Always
- Often
- Occasionally
- Rarely
- Never

8. What drives your decision to purchase private brand products? *(Select all that apply)*

- Competitive pricing
- Reliable product quality
- Eye-catching packaging
- Availability in stores
- Good past experience
- Recommendations from others
- Other (please specify): ______________

9. In which product categories do you most often buy private brands? *(Select all that apply)*

- Food and beverage
- Home care or cleaning products
- Personal care and hygiene
- Apparel and accessories
- Electronics or gadgets
- Other (please specify): ______________

Section 4: Satisfaction & Brand Loyalty

10. How satisfied are you with the private brand products you've used?

- Very satisfied
- Satisfied
- Neutral
- Dissatisfied
- Very dissatisfied

11. Would you suggest private brand products to friends or family?

- Definitely yes
- Probably yes
- Not sure
- Probably not
- Definitely not

12. Are you likely to continue buying private brands in the future?

- Yes
- No
- Undecided

Section 5: Feedback & Suggestions

13. What improvements would you like to see in private brand offerings? *(Select all that apply)*

- Enhanced product quality
- Broader range of options
- Better visual packaging
- More competitive prices
- Wider availability
- Other suggestions: _____________

14. Please share any additional thoughts or suggestions regarding private brand products:

- *Open-ended response*

APPENDIX – II

ii) APPENDIX-2: Python Code Snippets

```
# Importing libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = "C:\\Users\\ABHISHEK\\OneDrive\\Attachments\\Desktop\\All
Projects\\orgdata.csv"
df=pd.read_csv(file_path)
# Quick info about data
```

```
print("\n--- BASIC INFORMATION ---")
print(df.info())

# Checking missing values
print("\n--- MISSING VALUES ---")
print(df.isnull().sum())

# First few rows
print("\n--- SAMPLE DATA ---")
print(df.head())

# Renaming columns to simpler names for easier handling
df.columns = [col.strip().replace(':', '').replace('?', '').replace('(', '').replace(')', '').replace(',',
'').replace(' ', '_') for col in df.columns]

# Let's now visualize!

# Plot 1: Age Distribution
plt.figure(figsize=(8,5))
sns.countplot(y=df['Age'])
plt.title('Age Group Distribution')
plt.xlabel('Count')
plt.ylabel('Age Group')
plt.show()

# Plot 2: Gender Distribution
plt.figure(figsize=(6,4))
sns.countplot(x=df['Gender'])
plt.title('Gender Distribution')
plt.show()

# Plot 3: Household Income Distribution
plt.figure(figsize=(10,6))
sns.countplot(y=df['Household_Income'])
plt.title('Household Income Distribution')
plt.show()

# Plot 4: Familiarity with Private Brands
plt.figure(figsize=(10,5))
sns.countplot(y=df['How_familiar_are_you_with_private_brands_store_brands'])
plt.title('Familiarity with Private Brands')
plt.show()
```

```
# Plot 5: Satisfaction Rating (assuming 1-5 scale)
plt.figure(figsize=(7,4))
sns.countplot(x=df['How_satisfied_are_you_with_private_brands_you_have_purchased'])
plt.title('Satisfaction with Private Brands')
plt.show()

# Plot 6: Recommendation Likelihood
plt.figure(figsize=(7,4))
sns.countplot(x=df['How_likely_are_you_to_recommend_private_brands_to_others'])
plt.title('Likelihood to Recommend Private Brands')
plt.show()

# 🔥 Correlation Heatmap (Numerical Features Only)
numeric_df = df.select_dtypes(include=['number'])

if not numeric_df.empty:
   plt.figure(figsize=(12, 8))
   sns.heatmap(numeric_df.corr(), annot=True, cmap='YlGnBu', fmt=".2f", square=True,
linewidths=0.5)
   plt.title("🔥 Clean Correlation Heatmap (Numerical Features Only)")
   plt.tight_layout()
   plt.show()
else:
   print("⚠ No numeric columns found for correlation heatmap.")
# Additional Tip: Summary insights
print("BASIC SUMMARY INSIGHTS")
print("Most common Age group:", df['Age'].mode()[0])
print("Most common Gender:", df['Gender'].mode()[0])
print("Most common Income Group:", df['Household_Income'].mode()[0])
print("Most common Familiarity Level:",
df['How_familiar_are_you_with_private_brands_store_brands'].mode()[0])
#Key Insights Summary
#Young, male, low-income respondents dominate the sample.
################################################################################
####################

import pandas as pd
from sklearn.feature_selection import SelectKBest, chi2
from sklearn.preprocessing import LabelEncoder
```

```
df = pd.read_csv(file_path)

# Define multiple target columns
target_columns = [
    'What_is_your_overall_perception_of_private_brands_compared_to_national_brands',
    'How_likely_are_you_to_recommend_private_brands_to_others',
    'What_is_your_overall_perception_of_private_brands_compared_to_national_brands'
]

# Store all selected features across target variables
all_selected_features = set()

for target_column in target_columns:
    # Drop rows with missing values in the target column
    df_target = df.dropna(subset=[target_column])

    # Split into features and target
    X = df_target.drop(columns=target_columns)  # drop all target columns from features
    y = df_target[target_column]

    # Encode target if needed
    le = LabelEncoder()
    y_encoded = le.fit_transform(y)

    # Apply chi-square feature selection
    selector = SelectKBest(score_func=chi2, k=10)
    X_new = selector.fit_transform(X, y_encoded)

    # Add selected features to the set
    selected = X.columns[selector.get_support()].tolist()
    all_selected_features.update(selected)

# Convert set to sorted list
final_features = sorted(all_selected_features)

# Final dataframe with selected features and all target columns
df_selected = df[final_features + target_columns]
```

```
    print(classification_report(y_test, y_pred))

    sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt='d', cmap=cmap)
    plt.title(f"Confusion Matrix - {model_name}")
    plt.show()

# 5. Run Models for Each Target
for target_column in target_columns:
    print(f"\n\n{'='*60}\n📌 Evaluating Models for Target: {target_column}\n{'='*60}")

    X = df.drop(target_columns, axis=1)
    y = df[target_column]

    # Encode target if categorical
    if y.dtype == 'object' or y.nunique() > 2:
        le = LabelEncoder()
        y = le.fit_transform(y)

    # Split
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

    # Logistic Regression
    evaluate_model(LogisticRegression(max_iter=1000), X_train, y_train, X_test, y_test,
"Logistic Regression")

    # Decision Tree
    model_dt = DecisionTreeClassifier(max_depth=5, random_state=42)
    evaluate_model(model_dt, X_train, y_train, X_test, y_test, "Decision Tree",
cmap='Greens')

    plt.figure(figsize=(20,10))
    plot_tree(model_dt, feature_names=X.columns.astype(str),
class_names=np.unique(y_train).astype(str), filled=True)
    plt.title("Decision Tree Structure")
    plt.show()
```

```
    evaluate_model(KNeighborsClassifier(n_neighbors=5), X_train, y_train, X_test, y_test,
"K-Nearest Neighbors")

    # SVM
    evaluate_model(SVC(kernel='linear', probability=True, random_state=42), X_train,
y_train, X_test, y_test, "Support Vector Machine", cmap='coolwarm')

    # Naive Bayes
    evaluate_model(GaussianNB(), X_train, y_train, X_test, y_test, "Naive Bayes")

    # Gradient Boosting
    model_gbc = GradientBoostingClassifier(random_state=42)
    evaluate_model(model_gbc, X_train, y_train, X_test, y_test, "Gradient Boosting",
cmap='cividis')

    # Optional: AUC for GBC
    try:
        gbc_probs = model_gbc.predict_proba(X_test)
        gbc_auc = roc_auc_score(y_test, gbc_probs, multi_class='ovr')
        print(f"Gradient Boosting AUC Score (OvR): {gbc_auc:.2f}")
    except:
        pass

    # Optional: OOB error for Random Forest
    rf_oob_model = RandomForestClassifier(n_estimators=100, oob_score=True,
random_state=42)
    rf_oob_model.fit(X_train, y_train)
    oob_err = 1 - rf_oob_model.oob_score_
    print(f"Random Forest OOB Error: {oob_err:.4f}")
```