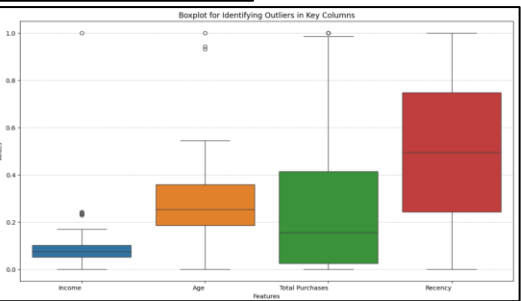# Dataset Overview & Data Quality
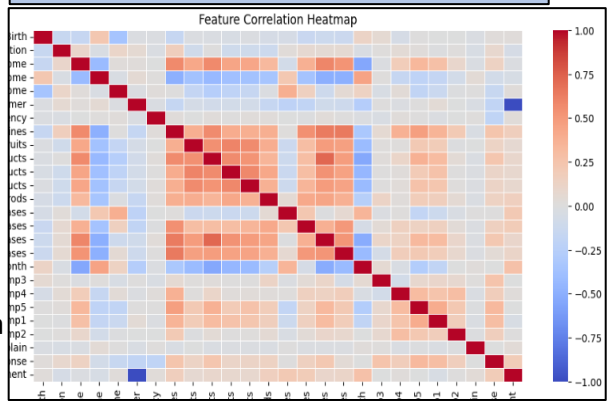
No. of features: 22   Null Values: 24 (income)

No. of Records: 2240 Duplicates: 182

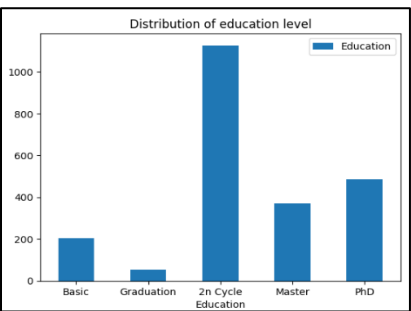**Outlier chart**

- Fairly clean data
- **Few outliers** in Income
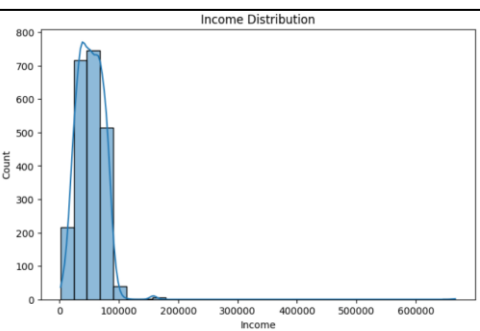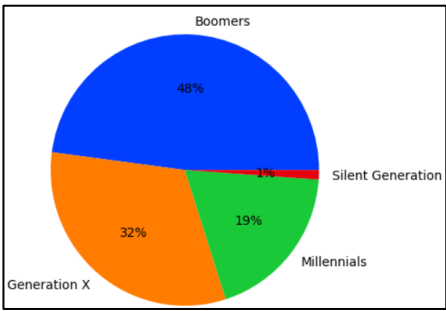- After removal we were left with **~98%** of the data
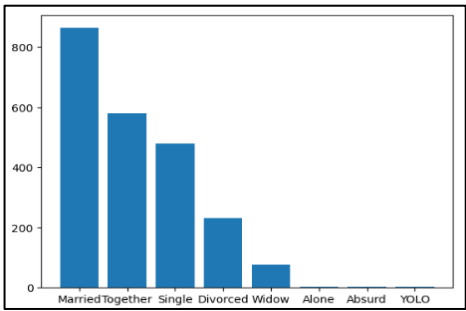
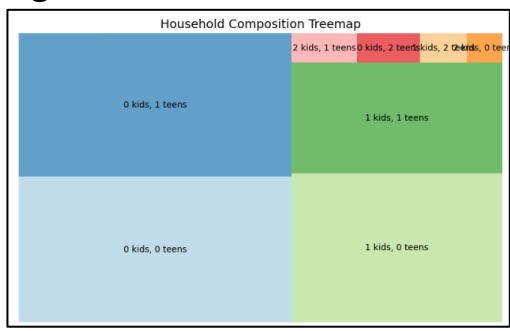**Correlation Heatmap**

Customer Demographics

Education Level

Age Distribution

Marital Status Distribution

Income Distribution

Household composition

- **Diverse Education Levels**
- **Older** Customer Base
- Household Composition & Income Variability
- Marital Status Influence

Purchasing Behaviour

Spend per Category by Generation (Age)
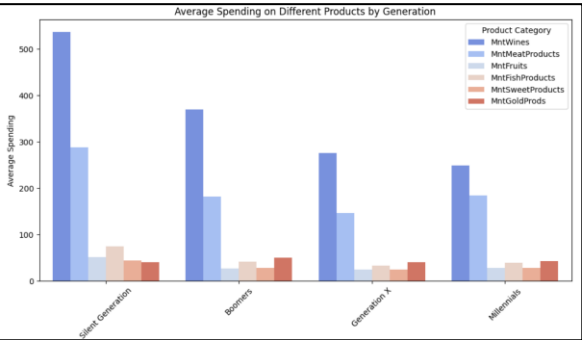
- **Millennials & Gen Z Are Digital Shoppers**
- Luxury & Bulk Purchases Increase with Age
- **Younger customers** prioritize **discounted purchases**

- **Store Purchases** Dominate for **Traditional** Shoppers
- Web & Catalog Sales Are Growing
- **Deals & Discounts** Drive Volume
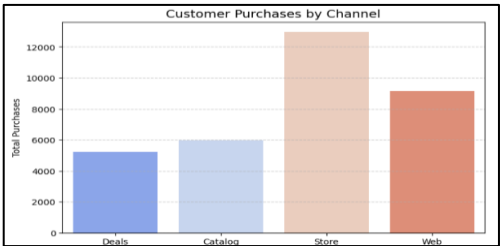
Channel analysis

Campaign Response

Overall Response Rate: **~15%**

Campaign Response Distribution

Campaign Acceptance by Generation

CUSTOMER SEGMENTATION AND TARGETED MARKETING

Scree Plot: Optimal no of clusters = 3

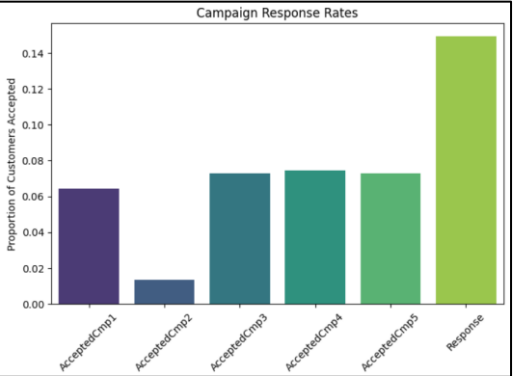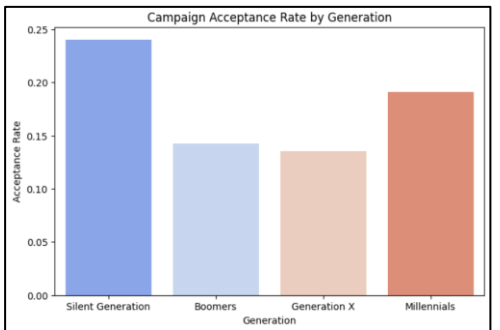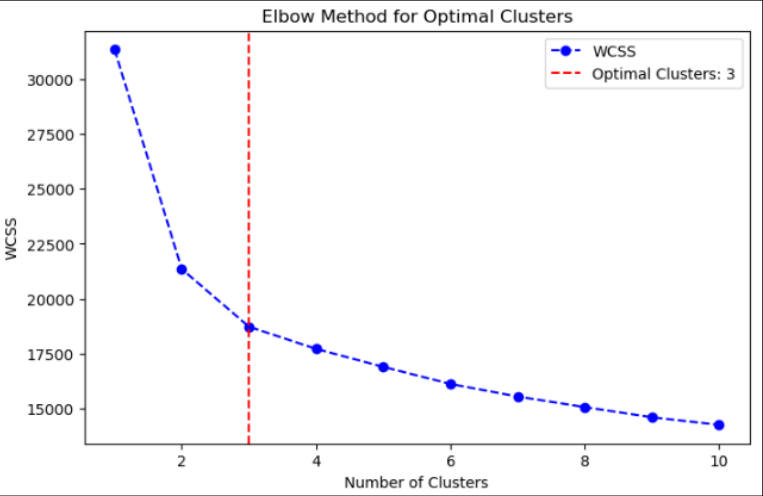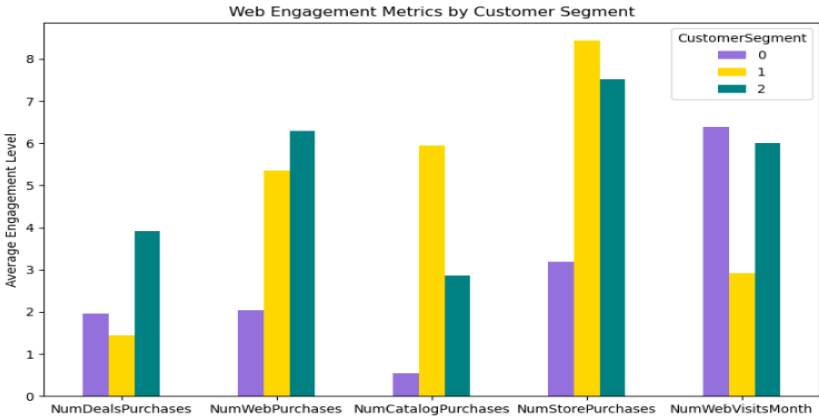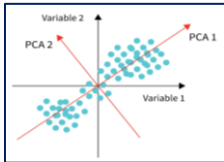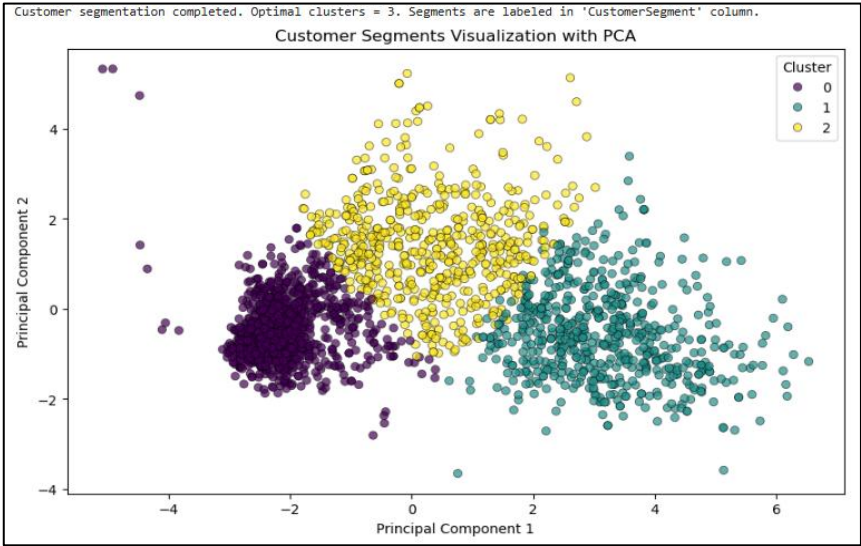Elbow Method for Optimal Clusters
- WCSS
- Optimal Clusters: 3

Clusters formed after PCA

Customer segmentation completed. Optimal clusters = 3. Segments are labeled in 'CustomerSegment' column.

Customer Segments Visualization with PCA

Web Engagement Metrics by Customer Segment

Business context based on clusters formed

Clusters	Customer Segment	Demographics	Purchase Behaviour	Online Acitvity	Business Label
Cluster 0: Purple	Low-Income | Price sensitive shoppers	Low Income	Minimal spending across all categories, particularly in wines, meat, and gold products.	Low engagement in web and catalog purchases. Moderate store purchases and web visits. Higher deal-based purchases.	Budget-Conscious Occasional Buyers
Cluster 1: Yellow	High Income | High spending Loyal customers	High Income	Significantly higher spending across all categories. Consistently engages in high-value purchases across multiple channels.	High web, catalog, and store purchases. Lower web visits, indicating they shop efficiently and with intent	Premium Loyalists
Cluster 2: Teal	Mid Income | Digital Savvy Shoppers	Mid Income	Moderate spending in all categories. Higher engagement in catalog purchases.	Highest engagement in web purchases. Frequent visits to the website, showing preference for digital shopping	Digital-First Shoppers

Engagement with Past Campaigns

Cluster 0:
Least engaged with past campaigns. Likely price-sensitive, cautious buyers.

Cluster 1:
Highly engaged in past campaigns. Likely to accept premium offers and exclusive deals.

Cluster 2:
Moderately engaged in campaigns. More likely to respond to digital promotions.
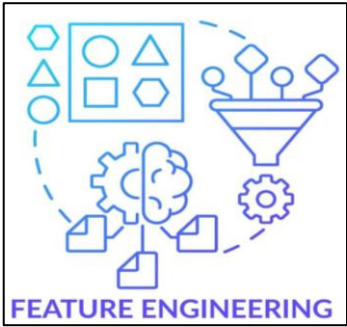
Predictive Modeling for Campaign Response

Feature Engineering Technique	Columns	Class used
Feature Extraction	Loyal customer flag etc	Pandas
Feature Trasformation	Age | Customer Tenure etc	Pandas
Feature Encoding	Education | Marital status	Label Encoder
Feature Aggregation	Kidhome & Teenhome	Childrenhome
Missing value Imputation	Income	Median Imputation
Outlier detection and removal	Income	Interquartile range(IQR)
Feature Scaling	Income | Age| Total_spend etc	Standard Scaler

FEATURE ENGINEERING

Algorithms Used	Accuracy	Recall		F1-Score	
		Class 0	Class 1	Class 0	Class 1
Logistic Regression	**0.8904**	0.96	0.49	0.94	0.58
Decision Trees	0.8501	0.93	0.42	0.91	0.46
Random Forests	0.8658	0.97	0.29	0.92	0.4
Gradient Boosting	0.877	0.98	0.32	0.93	0.44
XgBoost	0.8658	0.96	0.38	0.92	0.46

The ROC curve shows good model performance with an **AUC of 0.88**, indicating strong discrimination between classes.

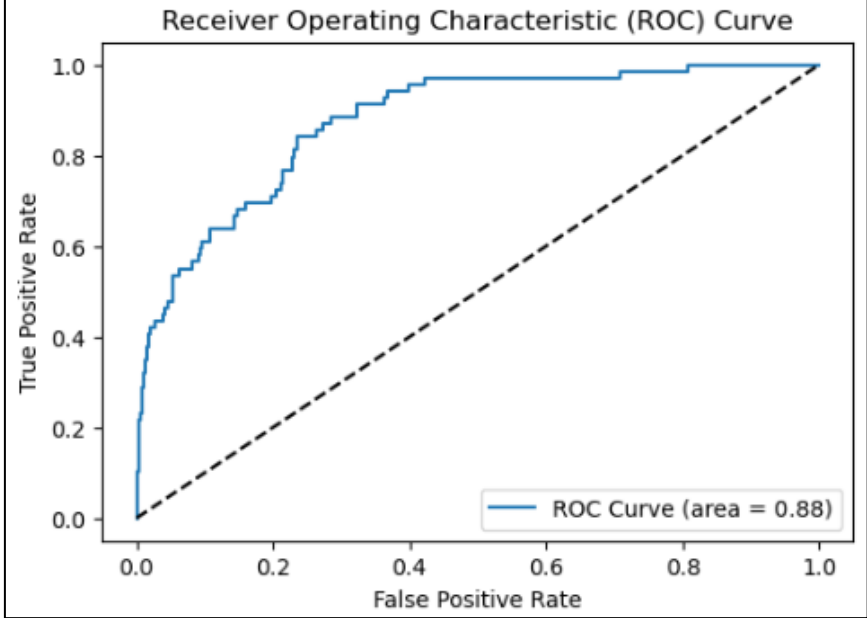Confusion Matrix for Logistic Regression Model

Binary Classification Model: Based on the evaluation metrics "Logistic Regression is the best performing model"

On performing GridSearchCV, Logistic Regression model accuracy improved a bit to 0.8949. With the best parameter of regularization strength as 0.1

```
Best parameters for Logistic Regression: {'C': 0.1}
Logistic Regression Accuracy: 0.8949
              precision    recall  f1-score   support

           0       0.91      0.97      0.94       378
           1       0.75      0.48      0.58        69

    accuracy                           0.89       447
   macro avg       0.83      0.72      0.76       447
weighted avg       0.89      0.89      0.88       447
```
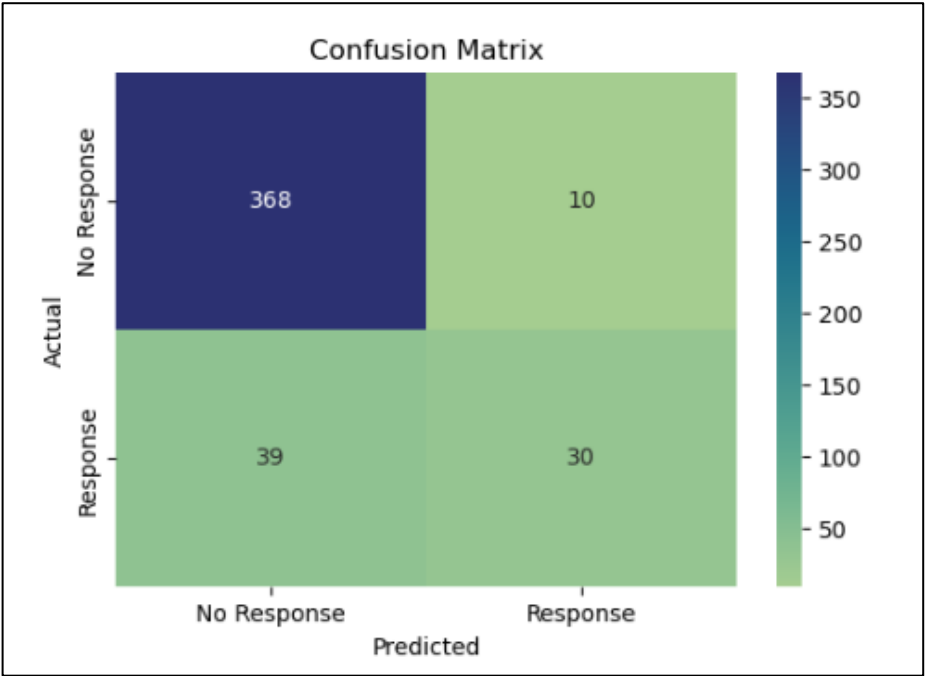
Key Drivers of Campaign Success:

Positive Impact on campaign success

No impact on campaign success

Feature Importance in Predicting Campaign Response (Logistic Regression)
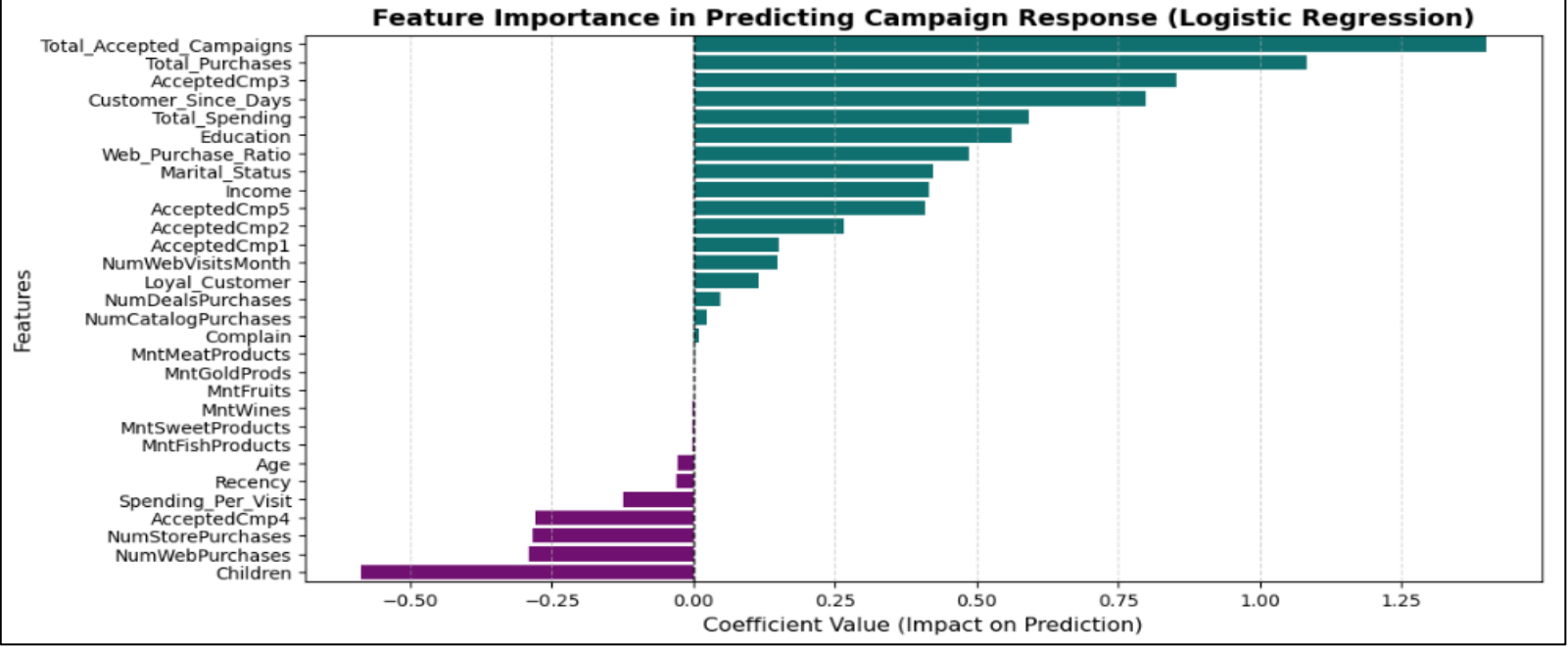
Total Accepted Campaigns (+1.40)
Customers who have previously accepted campaigns are highly likely to respond positively again.

Total Purchases (+1.08)
Customers with a higher total number of purchases are more likely to respond to campaigns

Customer Tenure (+0.80)
The number of days since a customer joined has a positive impact, indicating that long-term customers are more responsive

**Total Spending (+0.59)**
Higher spending customers are more engaged, likely because they perceive value in the promotions.

Children (-0.59)
Households with children are less likely to respond to campaigns, possibly due to financial constraints or time limitations

Recency (-0.03)
More recent buyers do not necessarily respond better, contradicting the assumption that recent activity signals engagement

Accepted Campaign 4 (-0.27) This specific campaign may have been less successful or targeted a less receptive audience.

Customers who spend more per visit are more likely to respond positively, suggesting TrendStyle Outfitters should target them with exclusive offers and loyalty rewards.

Distribution of Spending Per Visit by Response

Education Level vs Campaign Response
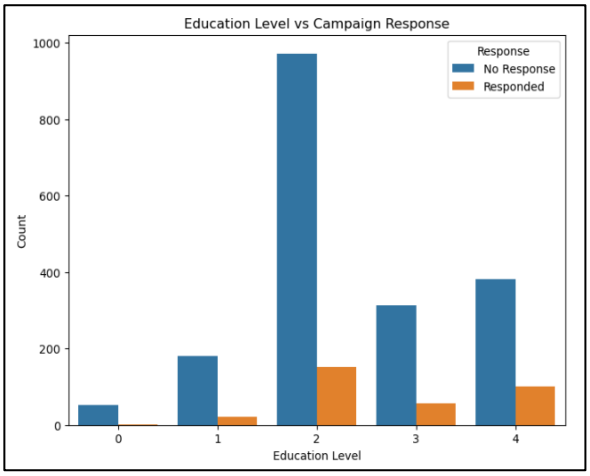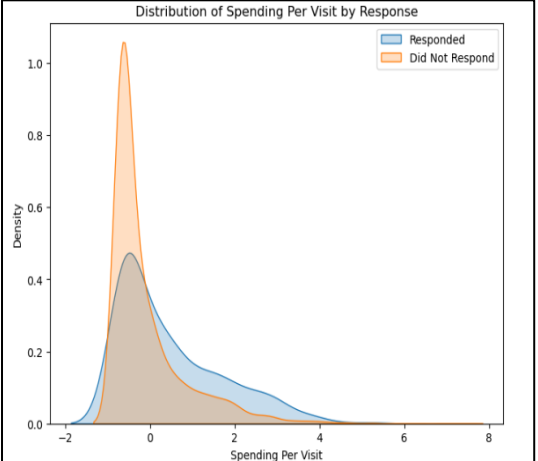
Higher education levels show better engagement. TrendStyle Outfitters should tailor messaging to highlight quality, brand prestige, and sustainability.

Customers with higher online purchase engagement respond better. TrendStyle Outfitters should boost digital marketing with personalized emails, online discounts, and retargeting ads.

Distribution of Web Purchase Ratio by Response
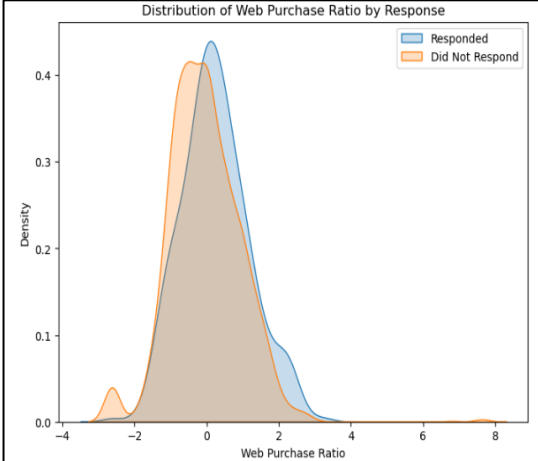
Actionable Recommendations for TrendStyle Outfitters

Target Audience Segmentation

- **High-Income Shoppers:** VIP programs, luxury branding, early access offers.
- **Moderate-Income Shoppers:** Seasonal promotions, personalized emails, social media ads.
- **Price-Sensitive Shoppers:** Flash sales, referral programs, app notifications.

$$\text{Increase in engagement} = \frac{R_{engaged} - R_{base}}{R_{base}} \times 100$$

Campaign Message & Offer Personalization

- **Luxury Buyers:** "Exclusive VIP Access to Premium Collections!"
- **Moderate Shoppers:** "Your Perfect Seasonal Picks – Special Offers for You!"
- **Budget Buyers:** "Big Savings Alert! Limited-Time Deals Just for You!"

$$\text{Increase in conversion} = \frac{C_{high} - C_{base}}{C_{base}} \times 100$$

Channel Optimization

- **Premium Shoppers:** High-end influencers, exclusive emails, digital ads.
- **Moderate Shoppers:** Facebook/Instagram ads, search ads, email promotions.
- **Price-Sensitive Shoppers:** SMS, flash sale notifications, TikTok ads.

$$\text{ROI Improvement} = \frac{S_{high} - S_{base}}{S_{base}} \times 100$$

Budget Allocation for Maximum ROI

- **60% in high-ROI channels** (social media, digital ads, personalized emails).
- **A/B testing to optimize** spending & adjust based on customer responses.

Expected Impact

- ✔ 15-30% increase in customer engagement.
- ✔ 10-20% higher conversion rates.
- ✔ 20-30% improvement in ROI through targeted promotions.