IFOOD CRM CASE STUDY

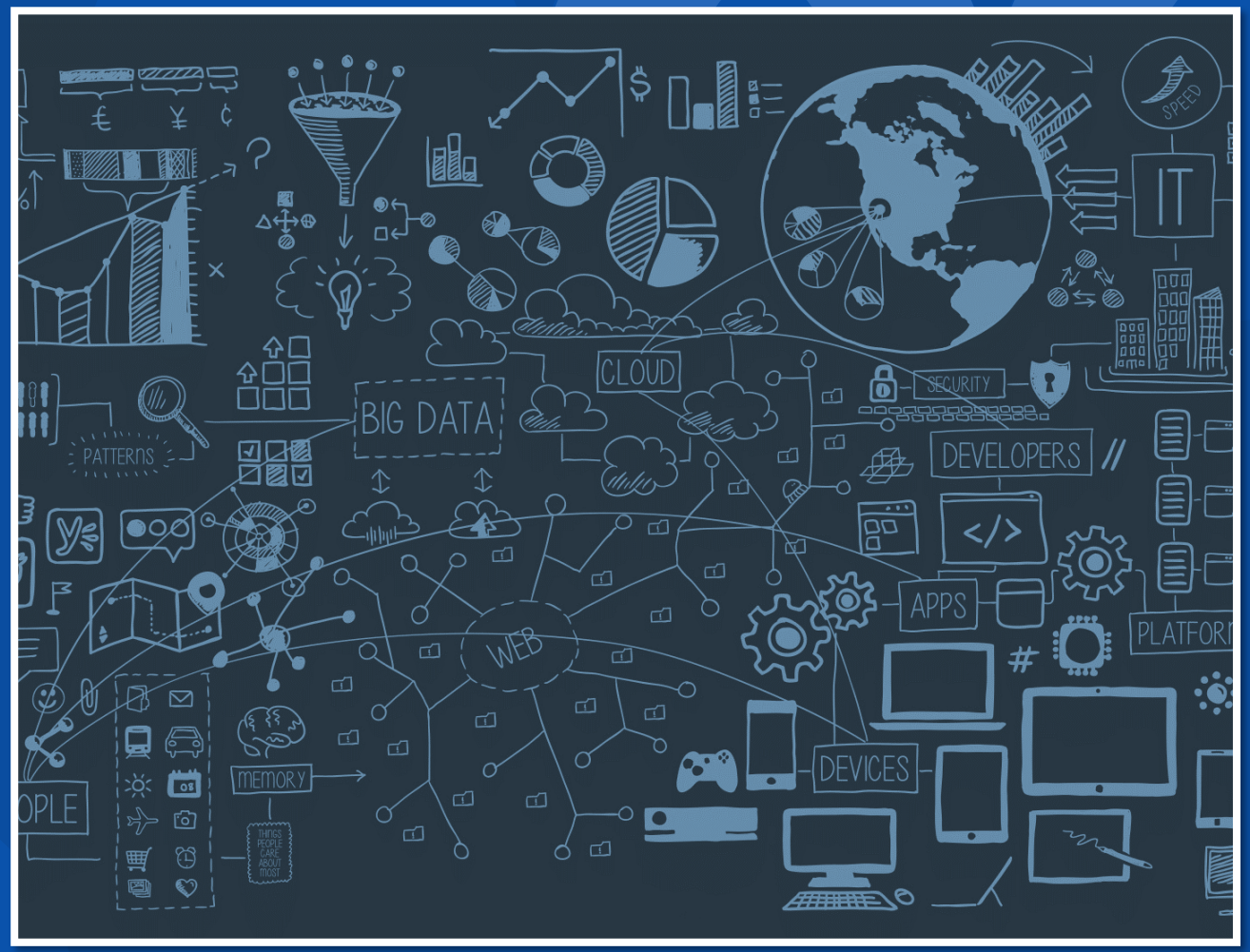
Exploratory Data Analysis Report

AGENDA

- Data Summary
- Exploration Plan
- EDA Results
- Data Cleaning & Feature Engineering
- Key Findings
- Hypotheses Proposals
- Significance Testing & Results
- Conclusion

STRUCTURE & ORGANIZATION

Data Summary



BIG PICTURE: A CRM SNAPSHOT

- **Origin:** The dataset was compiled by iFood, a leading Brazilian online food retailer, and represents customer-level data collected for internal marketing analysis from December 2014 to November 2016.
- **Purpose:** Identify data-driven solutions to isolate high-response customers to improve overall marketing ROI on future campaigns.
- **Content:** The data contains two socio-demographic data, two years of customer purchase history across six product categories, one month of digital engagement metrics, and past campaign interaction results from 2,240 customers enrolled between 2012 and 2014.
- **Structure:** The dataset contains the observations of all 2,240 customers (comprising the rows) among 29 features and includes the target variable, “Response” (comprising the columns).

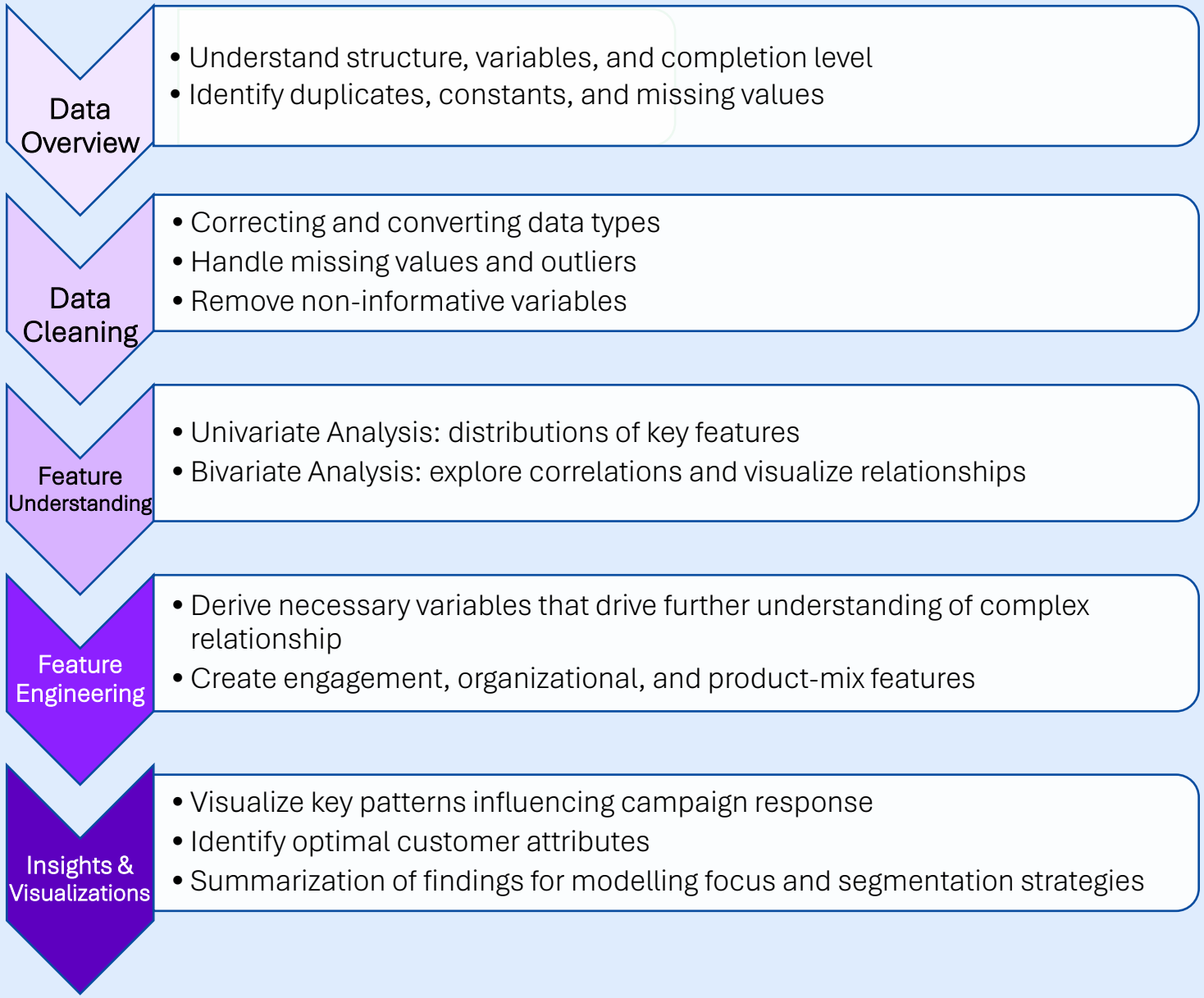
Feature	Description
ID	Unique ID to each customer
Year_Birth	customer's birth year
Education	customer's education level
Marital_Status	customer's marital status
Income	customer's yearly household income
Teenhome	number of teenagers in the customer's household
Kidhome	number of small children in the customer's household
DtCustomer	date of customer's enrollment with the company
Recency	number of days since the last purchase
MntWines	amount spent on wines in the last 2 years
MntFruits	amount spent on fruits in the last 2 years
MntMeatProducts	amount spent on meat products in the last 2 years
MntFishProducts	amount spent on fish products in the last 2 years
MntSweetProducts	amount spent on sweet products in the last 2 years
MntGoldProds	amount spend on *gold* products in the last 2 years
NumDealsPurchases	number of purchases made with discount
NumWebPurchases	number of purchases made through the company's web site
NumCatalogPurchases	number of purchases made using catalogue
NumStorePurchases	number of purchases made directly in stores
NumWebVisitsMonth	number of visits to company's web site in the last month
AcceptedCmp3	1 if the customer accepted the offer in the third campaign, 0 if otherwise
AcceptedCmp4	1 if the customer accepted the offer in the fourth campaign, 0 if otherwise
AcceptedCmp5	1 if the customer accepted the offer in the fifth campaign, 0 if otherwise
AcceptedCmp1	1 if customer accepted the offer in the first campaign, 0 if otherwise
AcceptedCmp2	1 if the customer accepted the offer in the second campaign, 0 if otherwise
Complain	1 if the customer complaineed in the last 2 years
Z_CostContact	the cost of each marketing contact
Z_Revenue	total revenue per campaign
Response (target)	1 if the customer accepted the offer in the last campaign, 0 if otherwise

THE PATHWAY TO INSIGHT

Exploration Plan

THE PROCESS

- Data Cleaning and Validation
- Feature Exploration & Engineering
- Insights & Visualizations

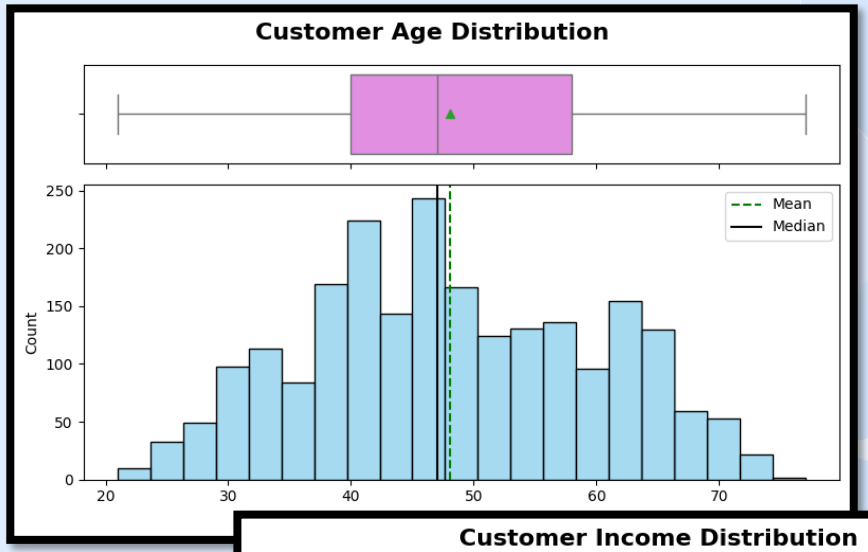


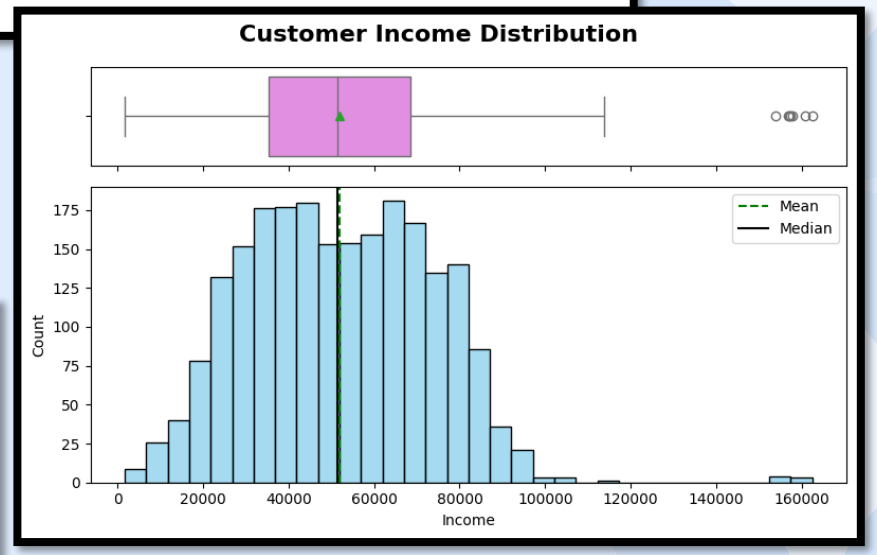
HIGH LEVEL INSIGHTS

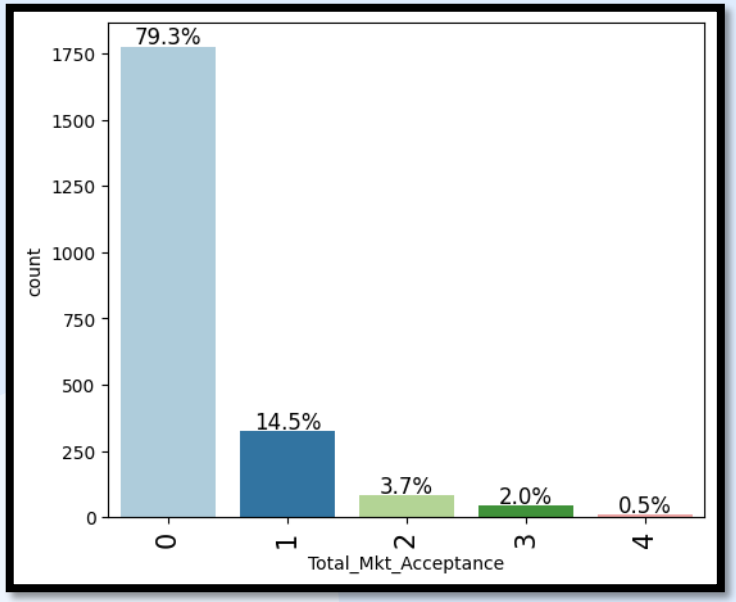
EDA Results

EDA RESULTS — CUSTOMER OVERVIEW

1. 2,240 unique customers; 15% overall campaign response rate.
2. Majority of customers are aged ~38–60, with middle to upper-middle income levels.
3. 79% of our customers did not respond to any of the previous marketing campaigns.

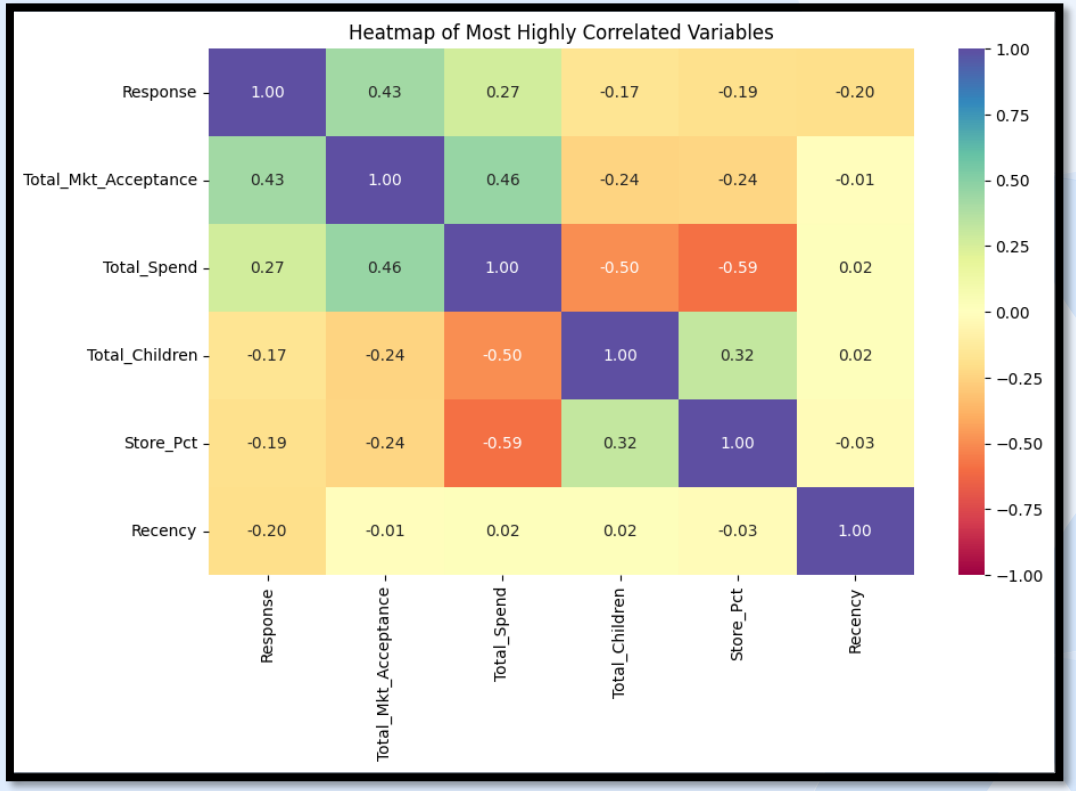


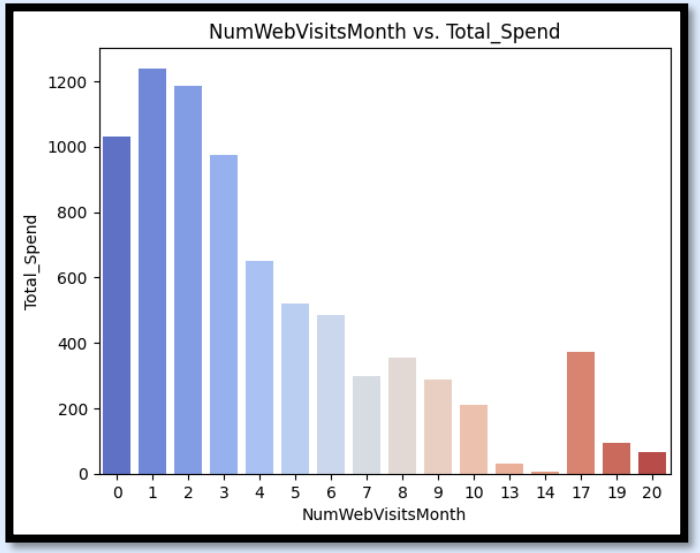


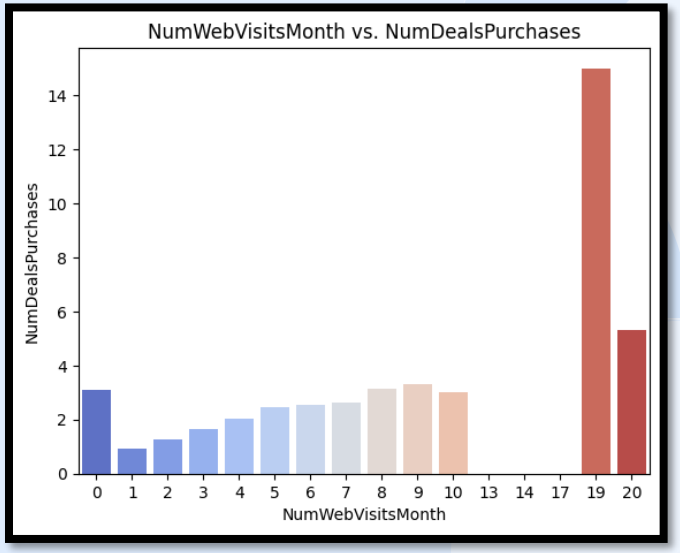


EDA INSIGHTS — BEHAVIORAL PATTERNS

- Past campaign acceptance and higher spending behavior (specifically in Wines and Meats) are the strongest indicators of future campaign acceptance.
- The long lapses in purchases and having children indicate a customer is less likely to respond.
- While not directly tied to marketing campaigns, frequent web visitors spend less unless items are on promotion.







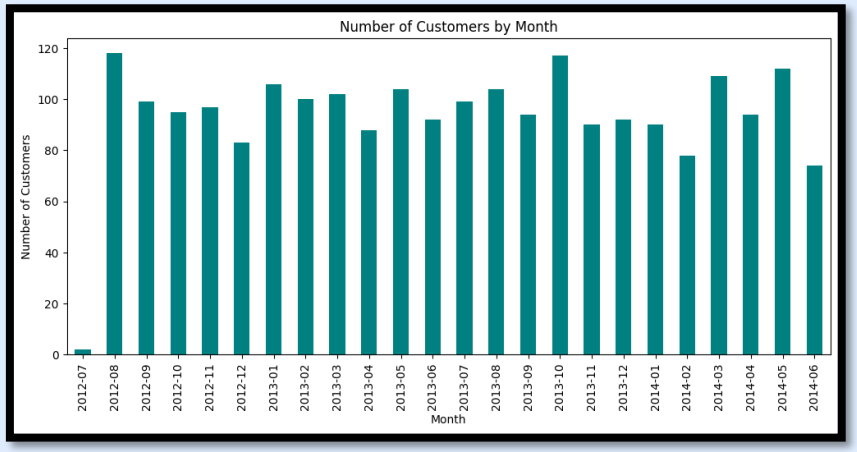
REFINEMENT & ENRICHMENT

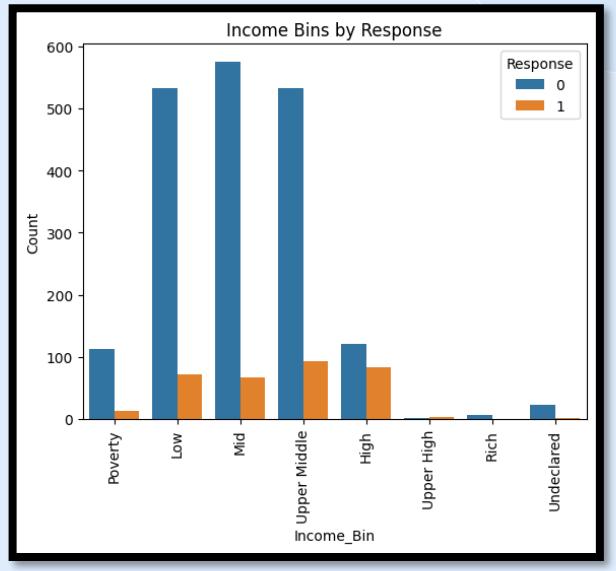
Data Cleaning & Feature Engineering

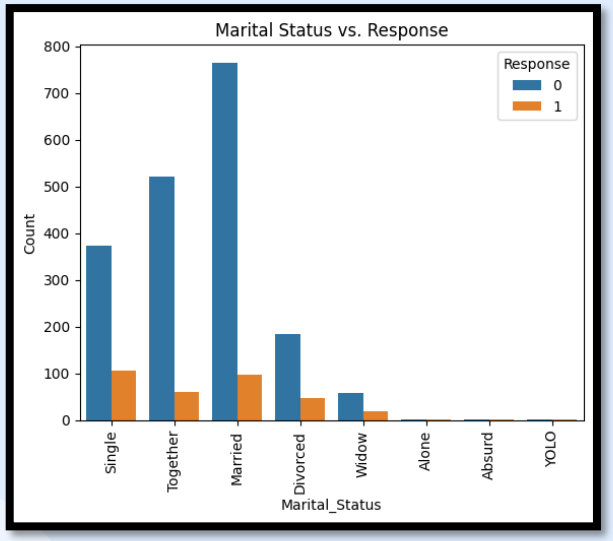
DATA CLEANING

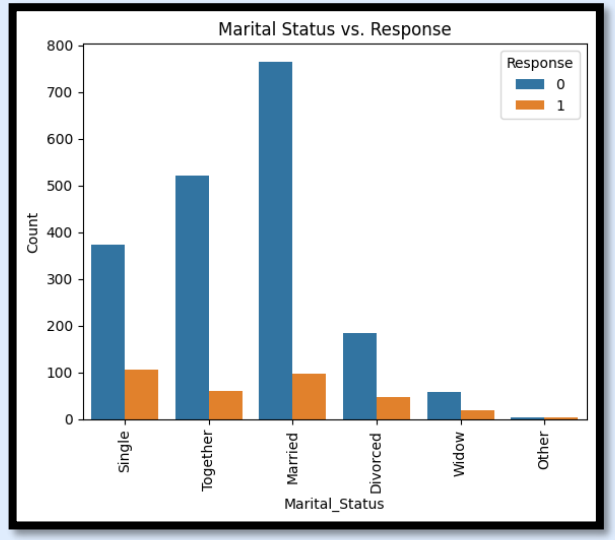
- Cleaning Steps:

    1. Datetime conversions
    2. Filled missing income with “Undeclared”
    3. Combined nonsensical Marital_Status values into “Other”
    4. Dropped constants and non-informative values. (Z_CostContact, Z_Revenue, ID)





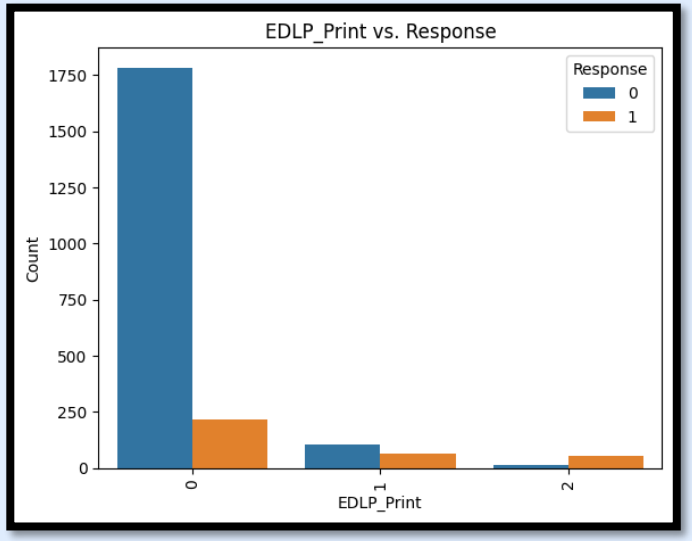




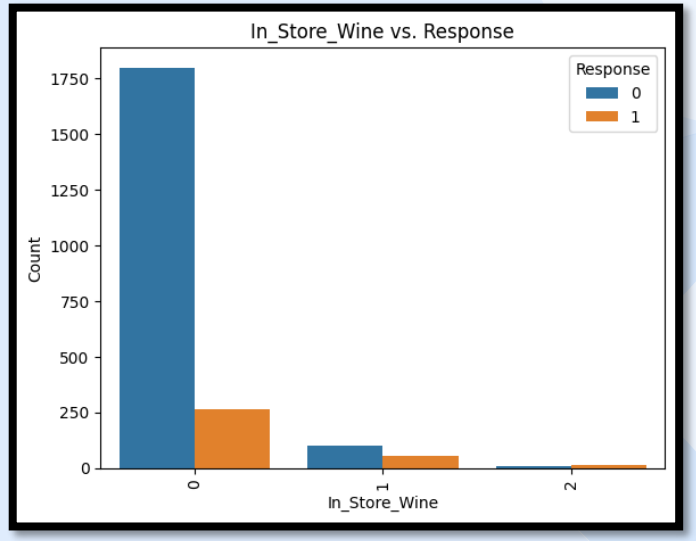
FEATURE ENGINEERING

- Combined similar campaigns into single variables based on products and channel performance
- Combined the separate campaign acceptances into one aggregated group – “Total_Mkt_Acceptance”
- Combined Kidshome and Teenshome into a single variable – “Total_Children”

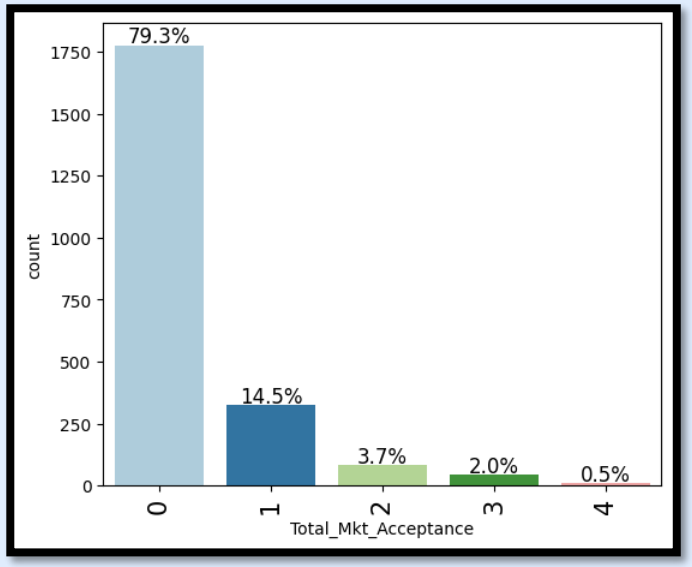
Campaign 1 + 5 Combined



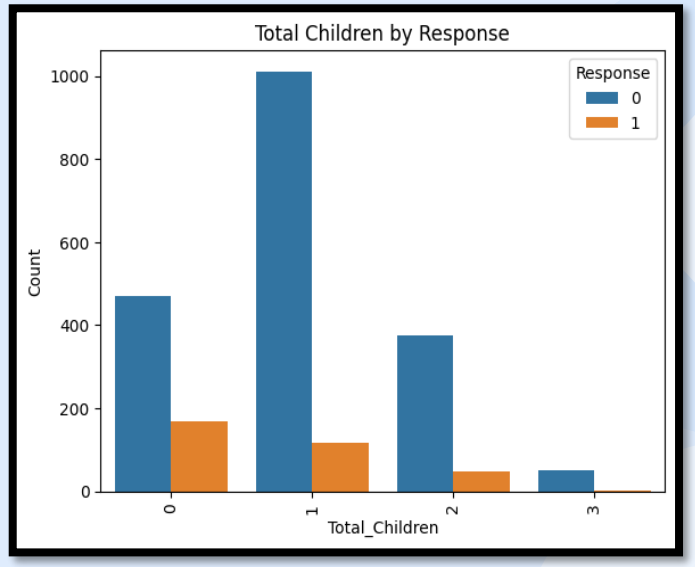
Campaign 2 + 4 Combined



Combined Campaign Acceptance

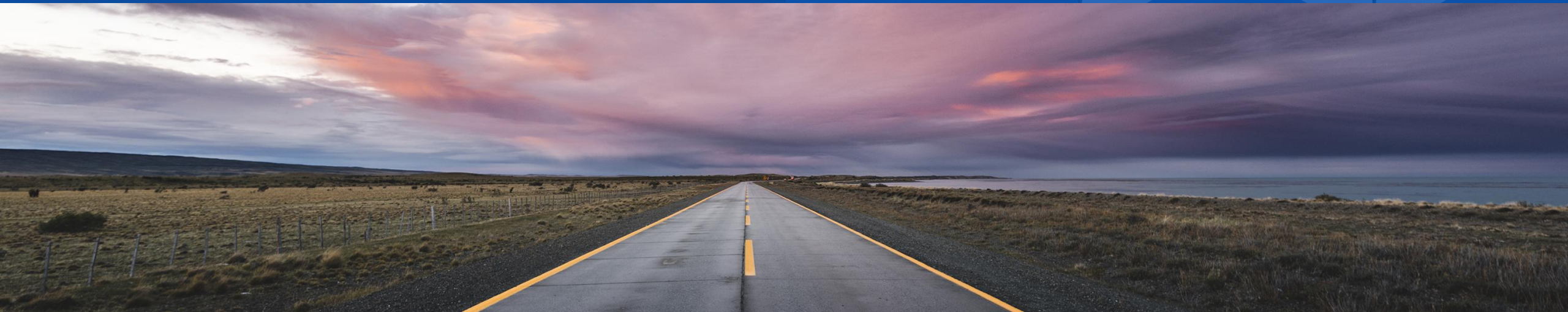


Combined Children Categories



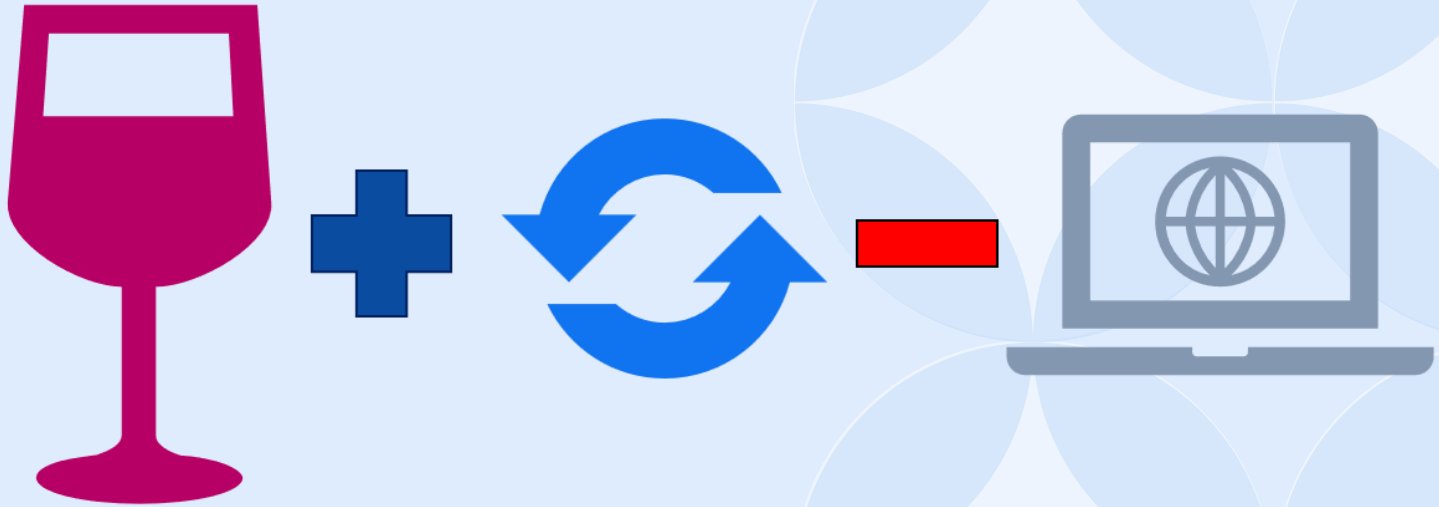
THE PATH FORWARD

Key Finding & Actionable Insights



KEY FINDINGS & ACTIONABLE INSIGHTS

- **High-value responders:** Heavy wine and meat buyers with recent purchases.
- **Repeat responders:** Past acceptance predicts future success.
- **Moderate web activity:** Converts better than high-frequency browsers.



Actionable Takeaways:

- Target high-spending wine customers.
- Focus on recent purchasers for time-sensitive offers.
- Refine digital strategy to balance promotion frequency.

UNCOVERING THE TRUTH

Hypotheses & Testing

HYPOTHESES TESTED

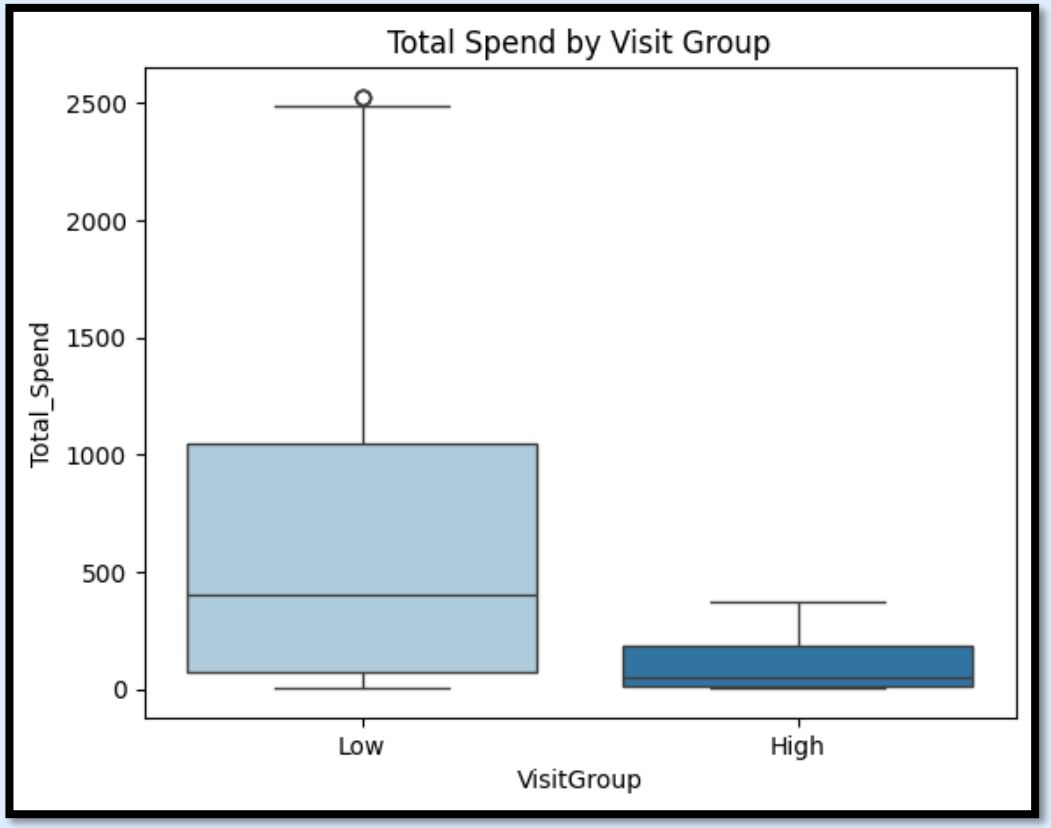
H_1: Customers that purchase above average amount of wine are more likely to respond.

H_2: Customers with children are less likely to respond to campaigns than those without.

H_3: Customers with ≥10 monthly web visits make fewer purchase than those with less visits.

*Each hypothesis includes both null (H_0) and alternative (H_1) statements for testing.

SIGNIFICANCE TESTING



Hypothesis: Customers with >= 10 website visits make *fewer* purchases than those <10 visits.

Test: One-tailed t-test.

Results:

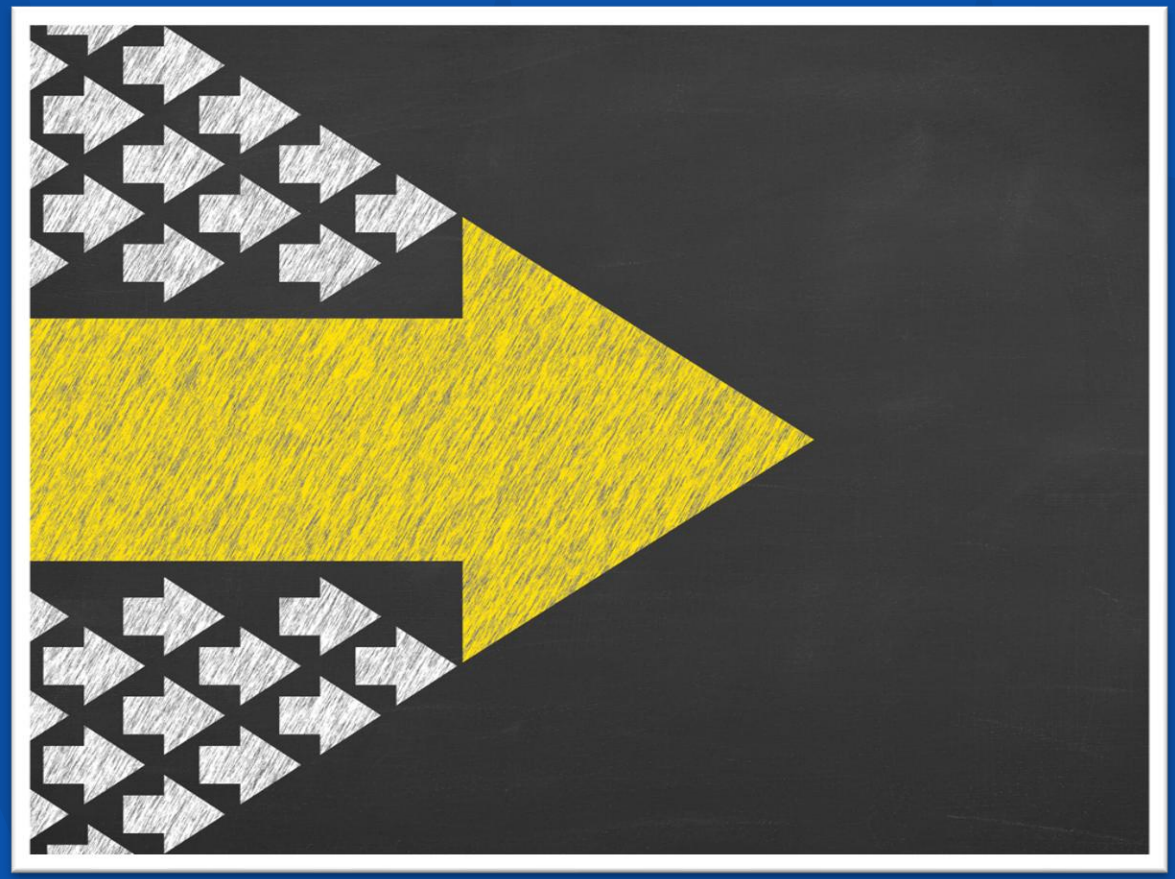
t-value = -2.808, p-value = 0.002 ($\alpha = 0.05$).

Decision: Reject H_0 — statistically significant difference.

Conclusion: There is evidence to support that customers with 10 or more visit make fewer purchases.

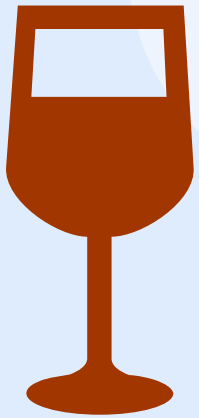
THE ANSWERS ALL OUR PROBLEMS

Conclusion & Next Steps



KEY TAKEAWAYS:

Behavior is a stronger predictor of response than demographics.



Wine spending, recency, and past engagement drive campaign success.

Average Responding Customer Spending Behaviors

Income	70253.962056
Kidhome	0.150609
Teenhome	0.372802
Recency	48.394697
MntWines	754.859670
MntFruits	35.678402
MntMeatProducts	303.864944
MntFishProducts	53.539335
MntSweetProducts	41.430719
MntGoldProds	66.831803
NumDealsPurchases	1.757398
NumWebPurchases	5.181328
NumCatalogPurchases	5.020113
NumStorePurchases	7.562433
NumWebVisitsMonth	4.621316
Customer_Age	48.297400

NEXT STEPS:

TRAIN PREDICTIVE MODEL (E.G., LOGISTIC REGRESSION, RANDOM FOREST).

EVALUATE MODEL METRICS (PRECISION, RECALL, F1-SCORE).

PILOT DATA-DRIVEN TARGETING TO VALIDATE ROI IMPROVEMENT.

Thank you

ADAM CUNNINGHAM
- BUDDING ANALYST AND MACHINE LEARNING ENGINEER