Superstore Marketing Campaign

94-881 MANAGING ANALYTICS PROJECTS
APRIL 23

Jingchen Fu, Jiuyuan Xie, Minkang Li, Yi Liu

STAGES AND MILESTONES

Define project context and decision to be improved

Stage 1:
Problem Framing

Stage 2:
Data Preparation

Get the dataset from Kaggle and perform data cleaning.

Univariate / Bivariate analysis
Customer segmentation

Stage 3:
EDA

Stage 4:
Modeling

Initial models to test and evaluate the model

Develop resource allocation, risk assessment and mitigation plans

Stage 5:
Risk and Resource

Stage 6:
Reporting and Action

Develop final reports
Make decisions

1

Problem Framing

Deliverables:

Proposal detailing stakeholders and decision to be improved for the end-of-year gold membership sale

Superstore: Gold Membership Promotion

- **Objective:** Introduce discounted Gold membership
- **Target:** Existing customer base
- **Offer Details:** 20% off all purchases, priced at $499 (down from $999)

Decision to Be Improved

Current decision process: Random or broad-spectrum customer contact

Improved approach: Use predictive modeling to target likely buyers

Goal: Improve campaign efficiency and success rate

Decision Makers

Primary decision-makers: Marketing team

Support by: Data science team

Responsibilities: Utilize model predictions, strategize and execute phone campaign

Value of Improved Decision

Higher conversion rates, more memberships sold

Reduced marketing costs

Enhanced customer experience and increased loyalty

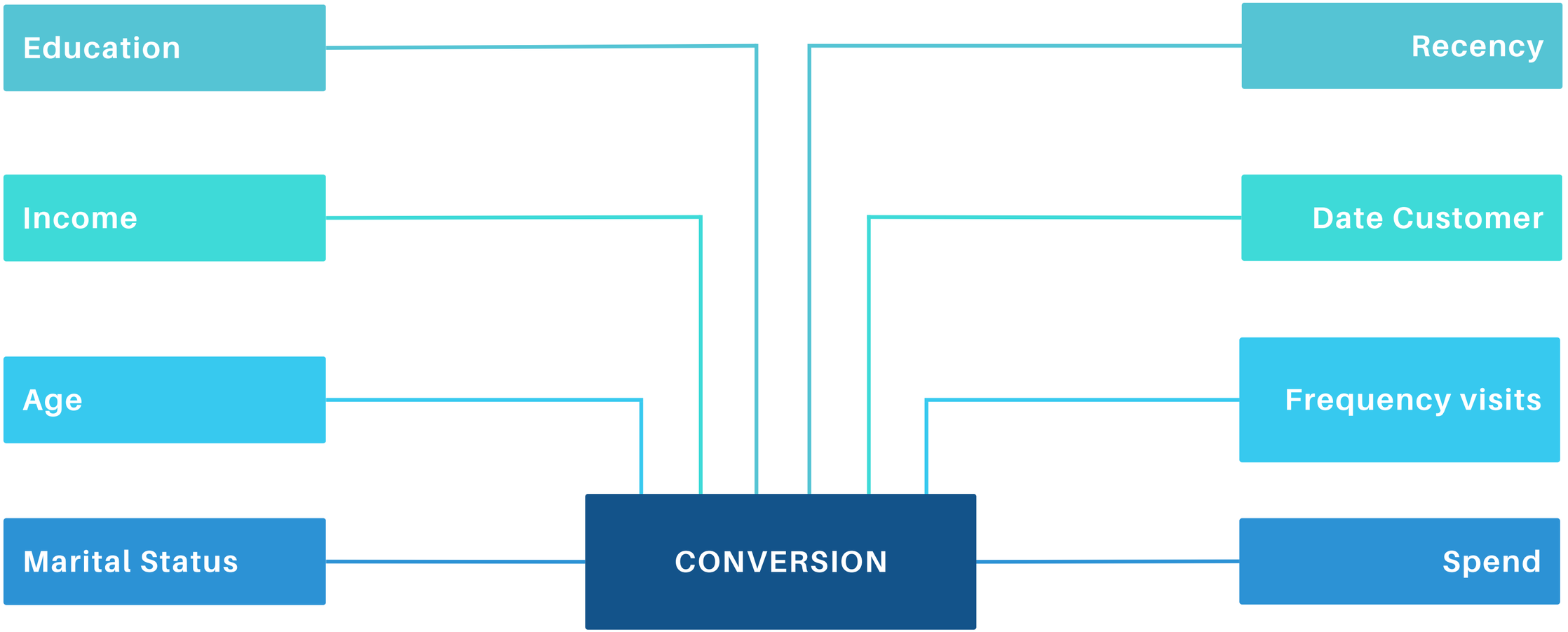
2

Data Preparation

Deliverables:

Cleaned and transformed data for analysis.

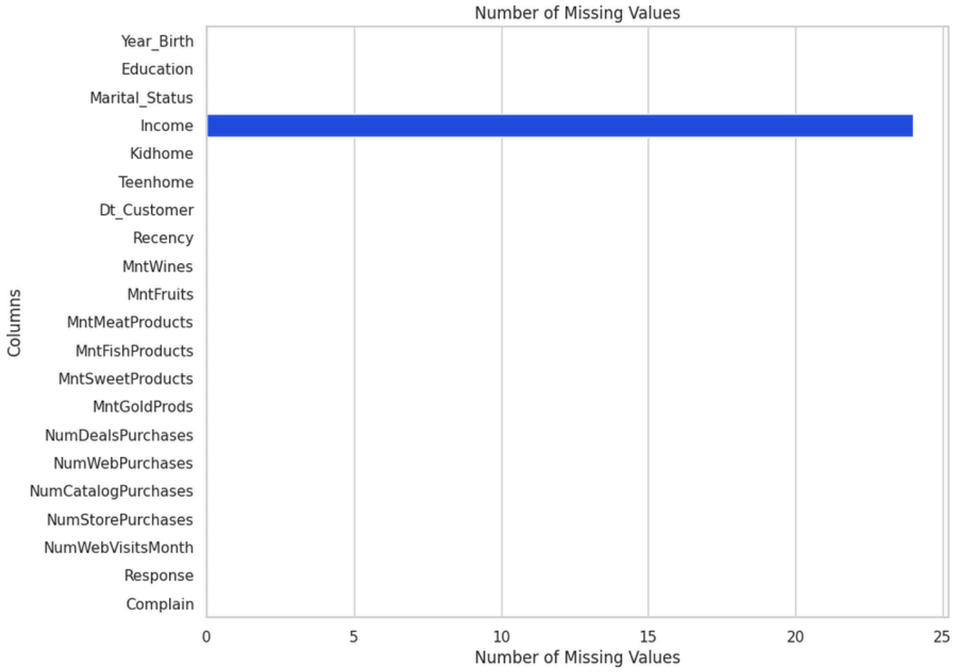
DATA - PILOT CAMPAIGN

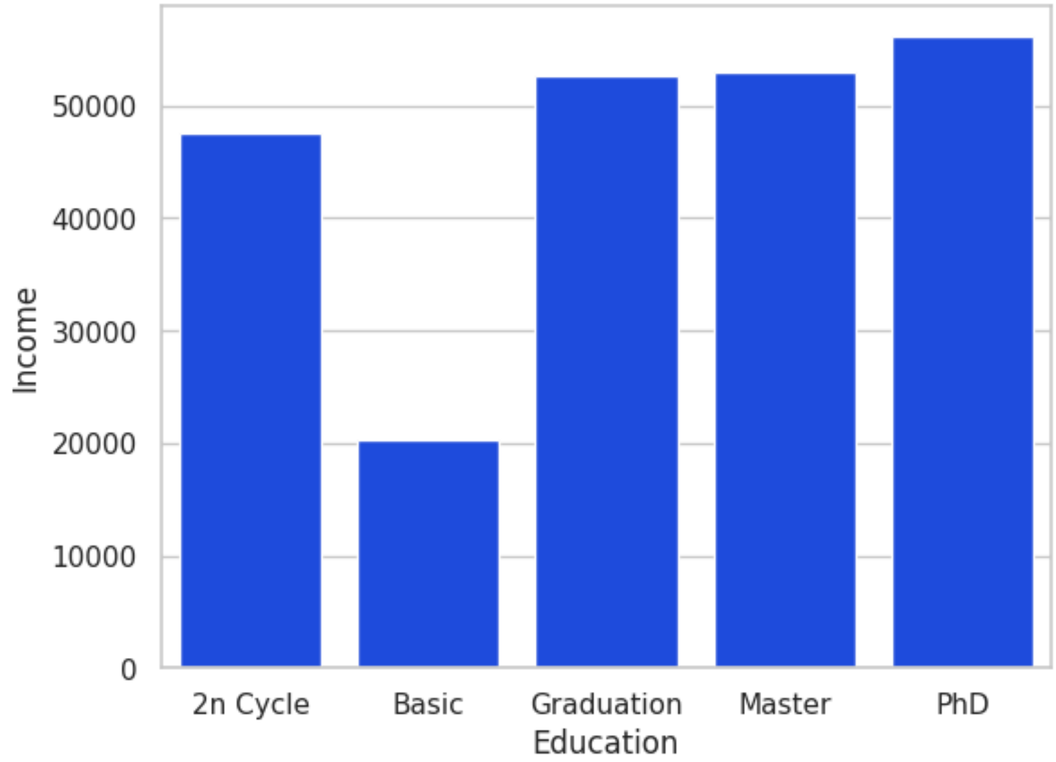


Demographics

Behaviors

Null Values





Fill with mean income in the education group

The Power of Visual Charts.

3

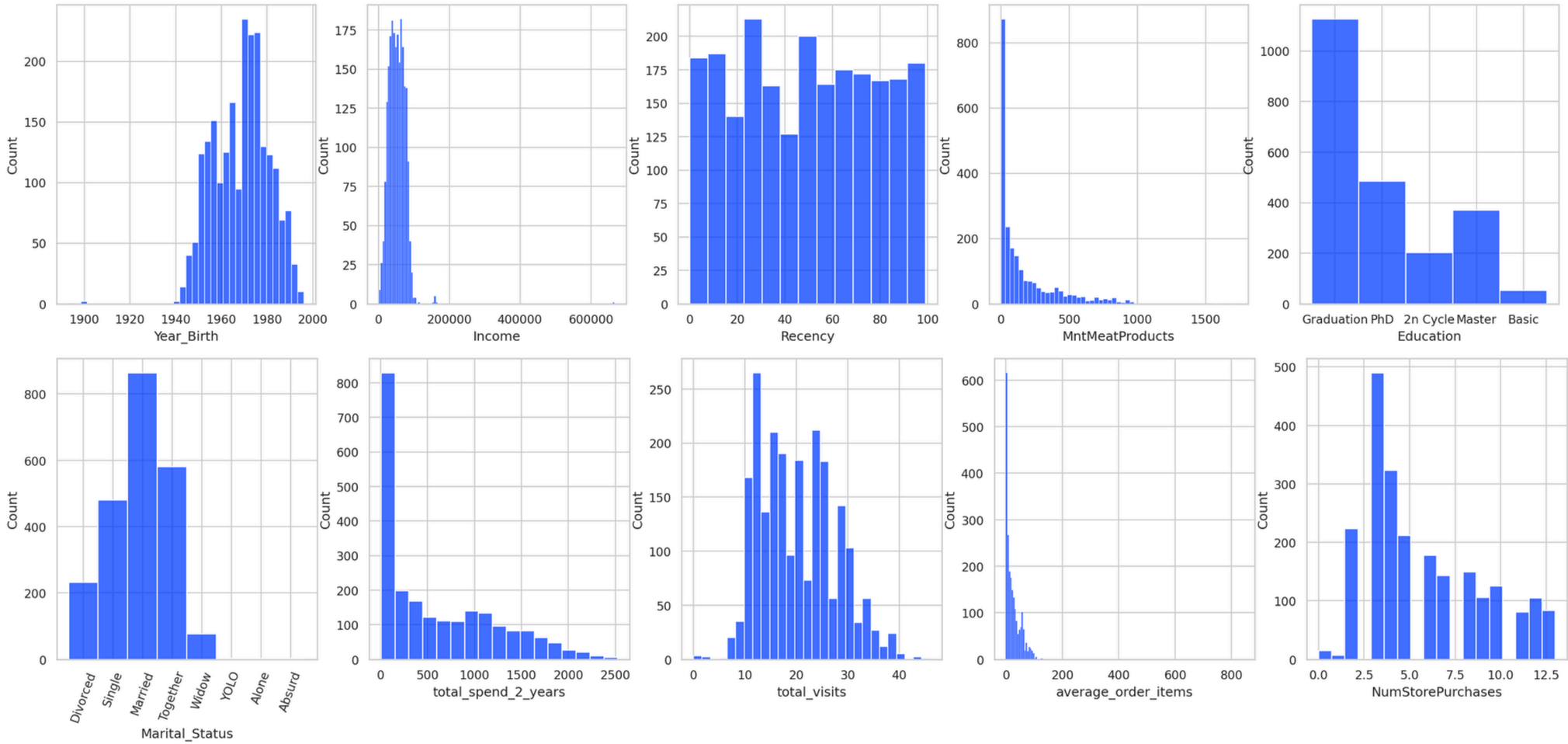
Exploratory Data Analysis

Deliverables:

Comprehensive EDA report including univariate and bivariate Analysis, and customer segmentation insights

Univariate Analysis

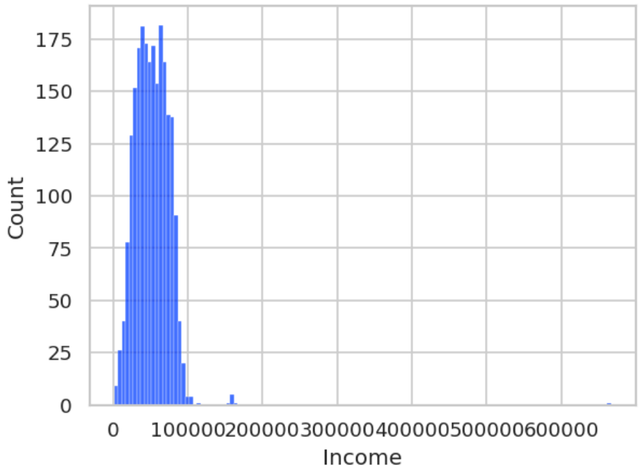
Conversion Rate: 14.9%

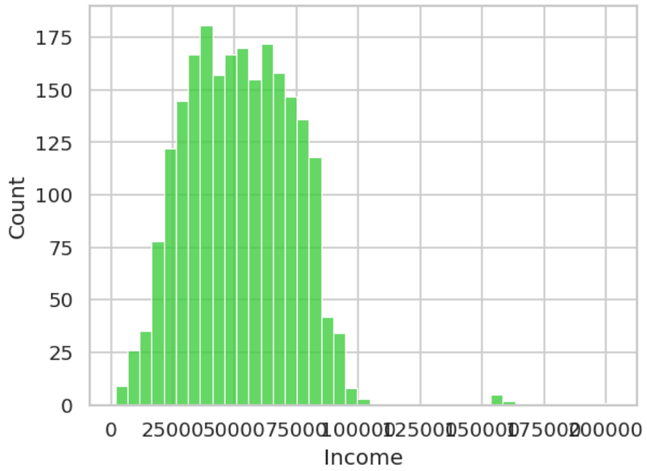


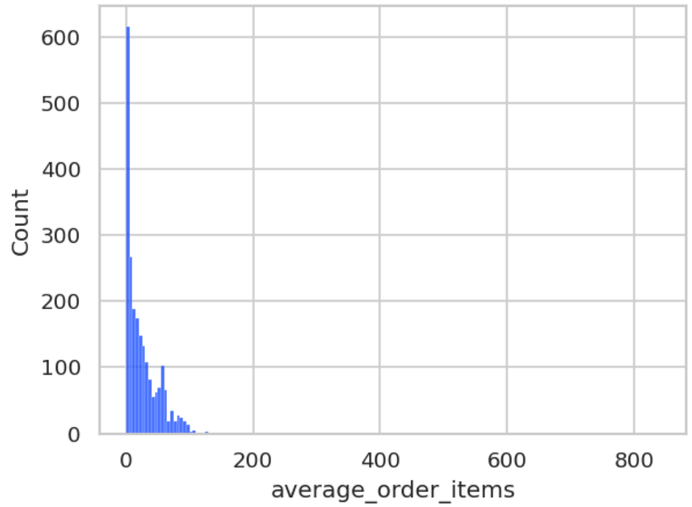
The Power of Visual Charts.

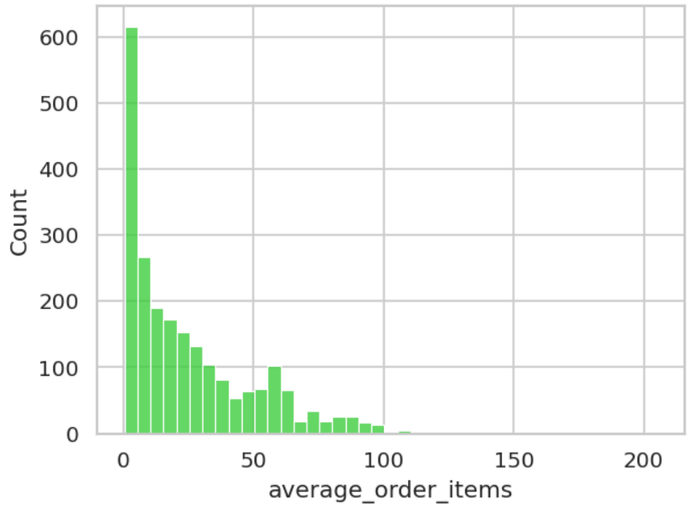
Extreme Value

6 Sigma Clipping



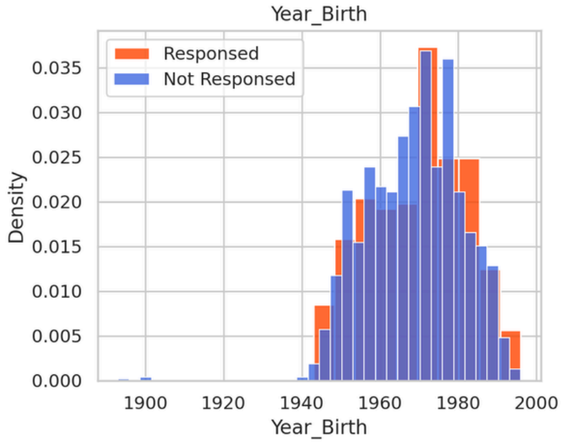


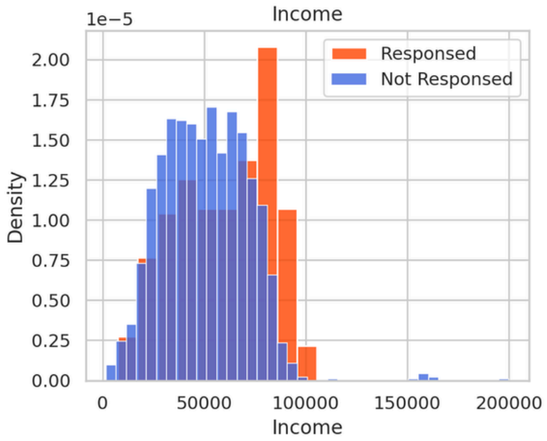


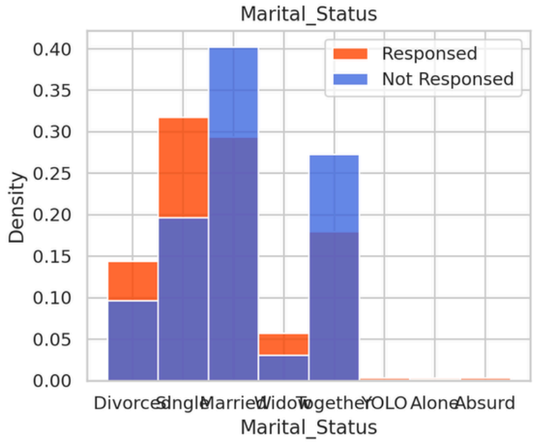


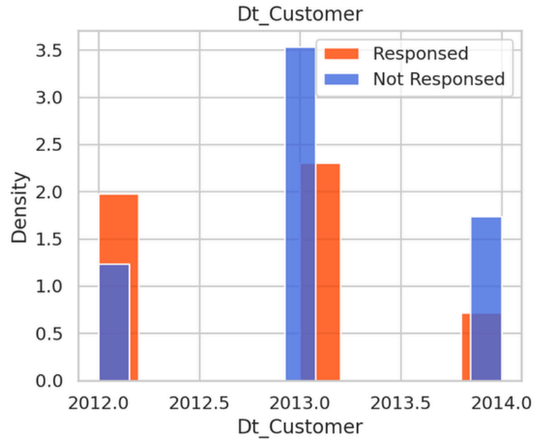
The Power of Visual Charts.

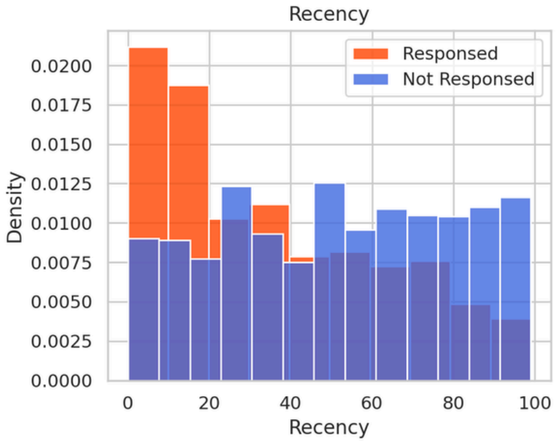
Bivariate Analysis

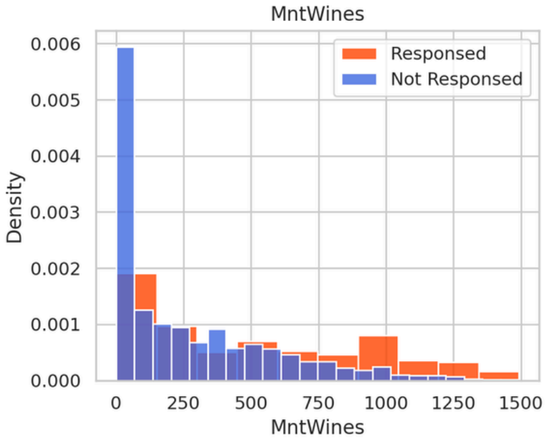


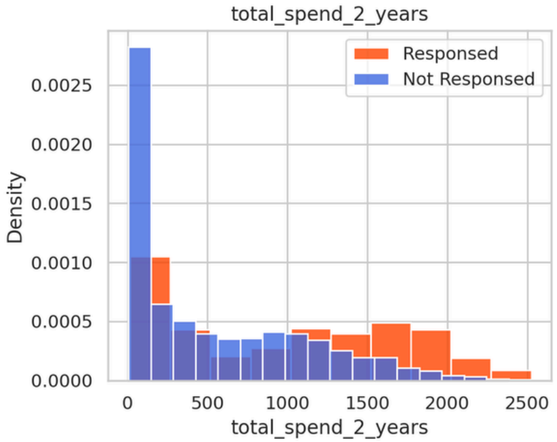


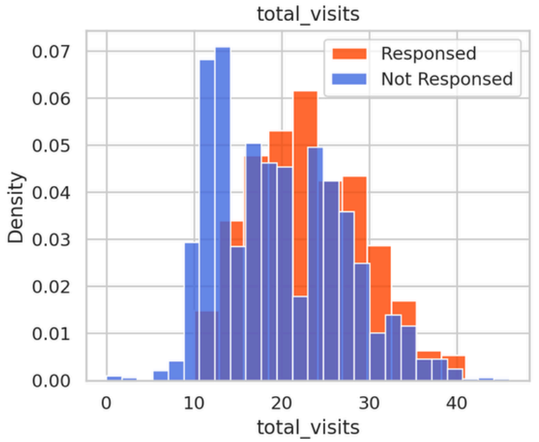






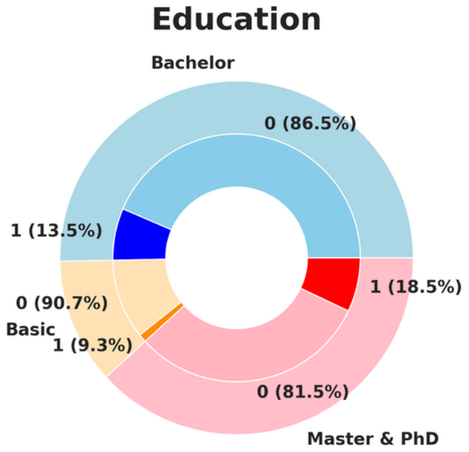


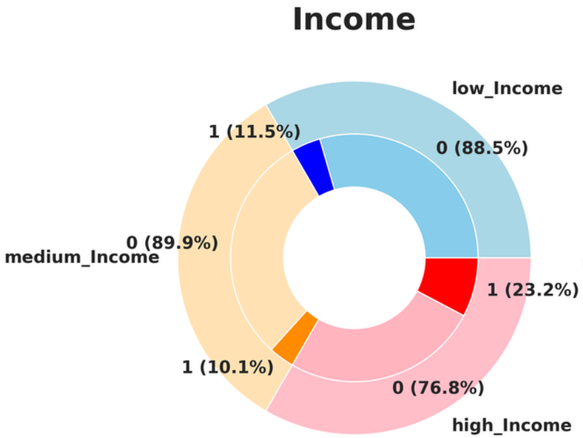


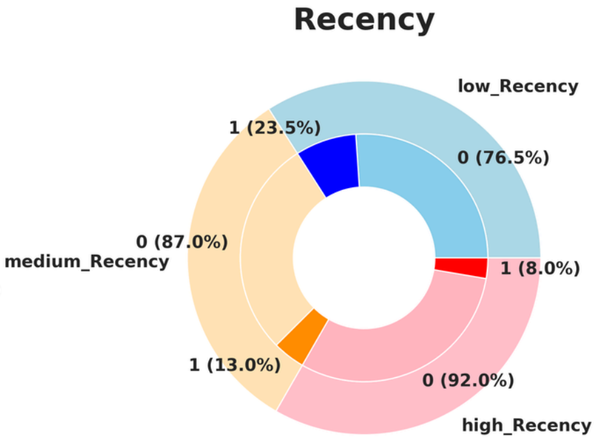


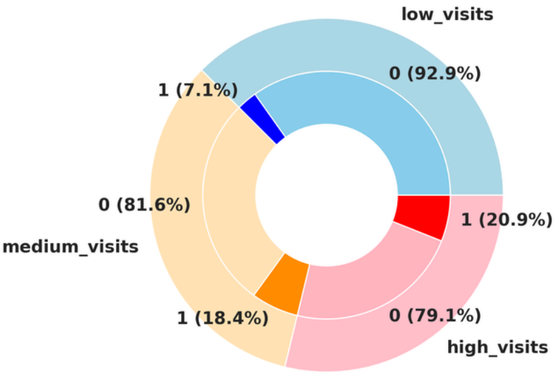
The Power of Visual Charts.

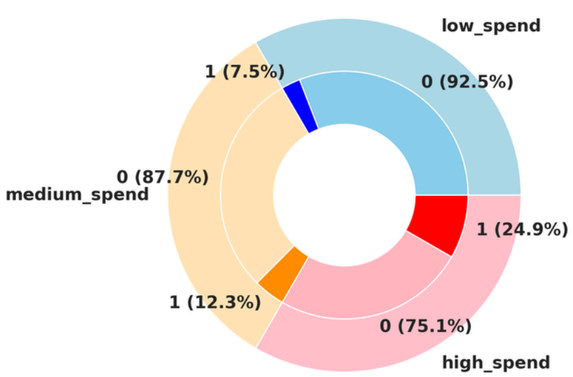
Segmentation Analysis

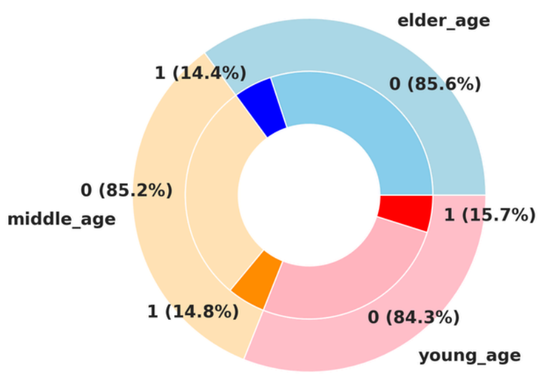












01 *Higher Education*

02 *High Income*

03 *Low Recency*

04 *High Medium Frequency*

05 *High Spend*

The Power of Visual Charts.

4

Model Development and Selection

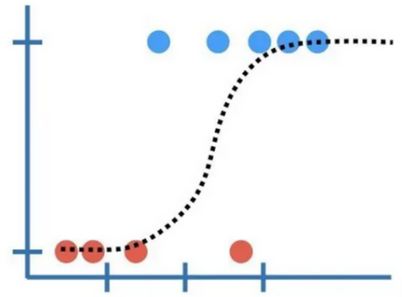
Deliverables:

Trained models (XGBoost, Logistic Regression) and model evaluation summary.

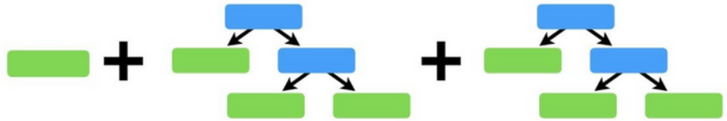
MODELING

Find the magic and fun in presenting with Canva Presentations. Press the following keys while on Present mode! Delete or hide this page before presenting.

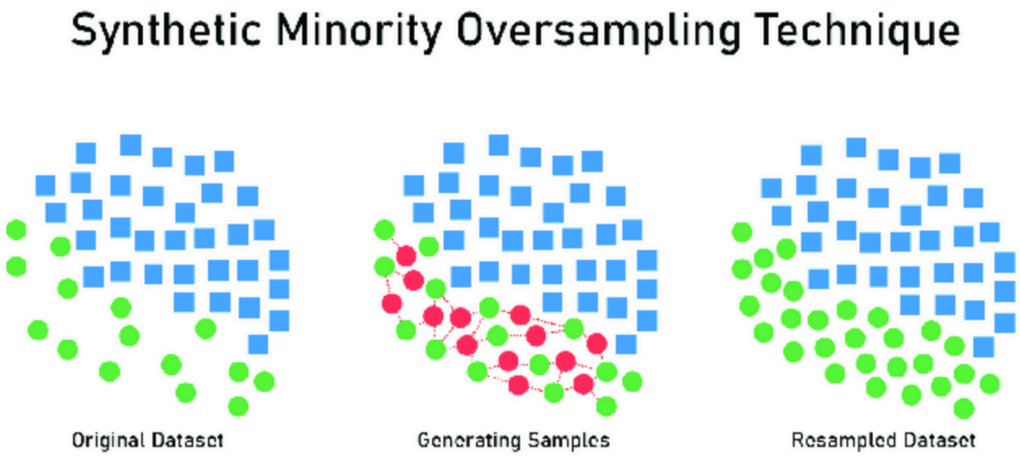
Logistic Regression

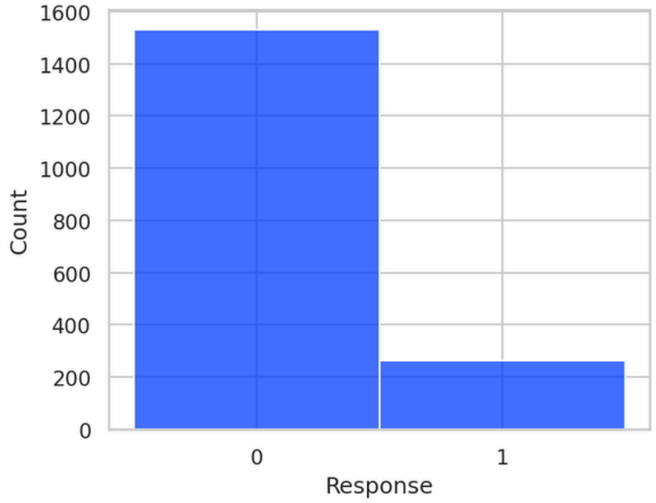


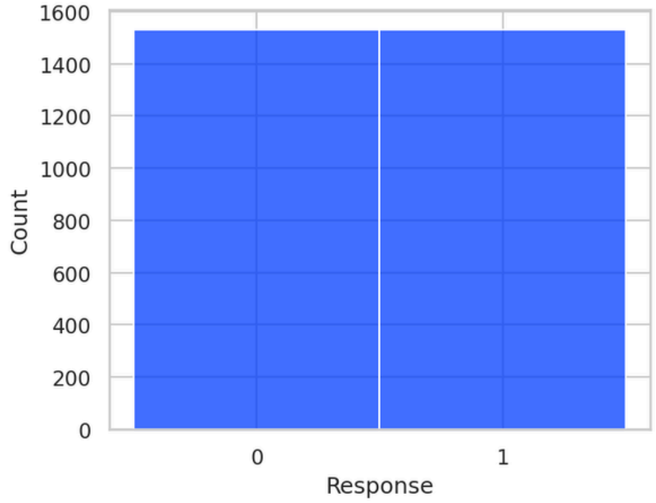
XGBoost



SMOTE-Class Imbalance

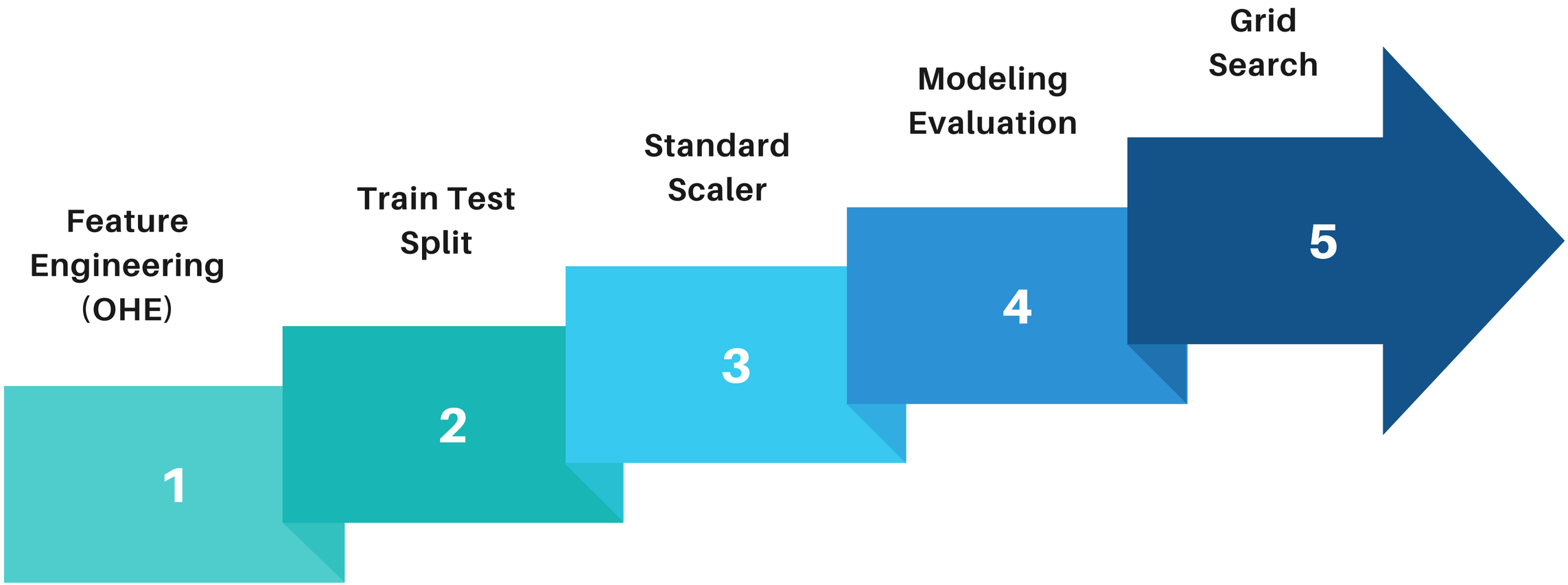




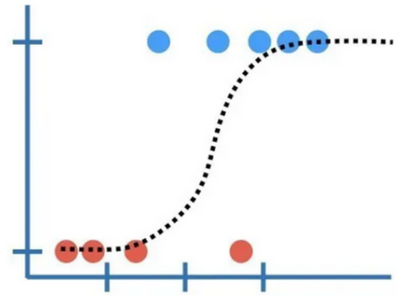


The Power of Visual Charts.

5-STEP MODELING PROCESS



Logistic Regression



01	*Interpretable*
02	Simple Baseline
03	*Feature Weights & p-value*
04	*Lightweight*
05	*Binary Classification*

	Precision	Recall
0	*0.97*	*0.56*
1	*0.28*	*0.90*

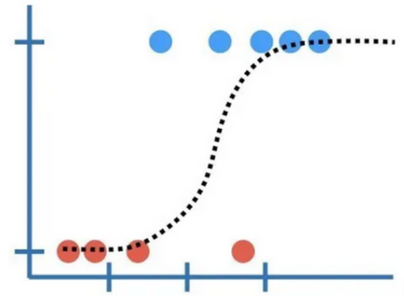
The Power of Visual Charts.

Logistic Regression

Logit Regression Results

Dep. Variable:	Response	**No. Observations:**	3060
Model:	Logit	**Df Residuals:**	3031
Method:	MLE	**Df Model:**	28
Date:	Tue, 23 Apr 2024	**Pseudo R-squ.:**	0.2770
Time:	03:06:12	**Log-Likelihood:**	-1533.6
converged:	False	**LL-Null:**	-2121.0
Covariance Type:	nonrobust	**LLR p-value:**	1.190e-229

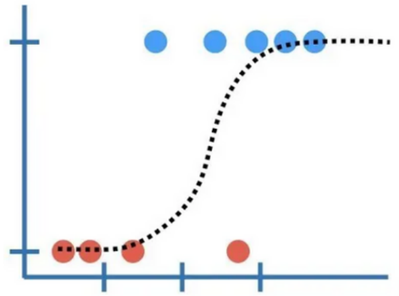
	coef	std err	z	P>|z|	[0.025	0.975]
Year_Birth	0.0345	0.055	0.628	0.530	-0.073	0.142
Income	0.0002	0.096	0.002	0.998	-0.188	0.189
Kidhome	0.0999	0.073	1.369	0.171	-0.043	0.243
Teenhome	-0.4718	0.067	-7.043	0.000	-0.603	-0.341
Recency	-0.6441	0.047	-13.756	0.000	-0.736	-0.552
MntWines	0.1122	nan	nan	nan	nan	nan
MntFruits	-0.0123	nan	nan	nan	nan	nan
MntMeatProducts	0.1537	nan	nan	nan	nan	nan
MntFishProducts	-0.1665	2.79e+05	-5.96e-07	1.000	-5.47e+05	5.47e+05
MntSweetProducts	-0.0431	nan	nan	nan	nan	nan
MntGoldProds	0.1060	1.1e+05	9.67e-07	1.000	-2.15e+05	2.15e+05
NumDealsPurchases	0.2071	nan	nan	nan	nan	nan
NumWebPurchases	0.1893	nan	nan	nan	nan	nan
NumCatalogPurchases	0.3435	nan	nan	nan	nan	nan
NumStorePurchases	-0.6137	nan	nan	nan	nan	nan
NumWebVisitsMonth	0.5688	nan	nan	nan	nan	nan
Complain	0.0154	0.049	0.312	0.755	-0.081	0.112
total_spend_2_years	0.1104	nan	nan	nan	nan	nan
total_visits	0.1812	nan	nan	nan	nan	nan
average_order_items	0.6819	0.198	3.438	0.001	0.293	1.071
Education_Basic	-0.1546	0.063	-2.471	0.013	-0.277	-0.032
Education_Graduation	0.0817	0.089	0.916	0.360	-0.093	0.257
Education_Master	0.1224	0.076	1.620	0.105	-0.026	0.270
Education_PhD	0.3159	0.082	3.861	0.000	0.156	0.476
Marital_Status_Alone	-0.0768	0.065	-1.182	0.237	-0.204	0.051
Marital_Status_Divorced	-0.6668	0.374	-1.782	0.075	-1.400	0.067
Marital_Status_Married	-1.3462	0.572	-2.353	0.019	-2.468	-0.225

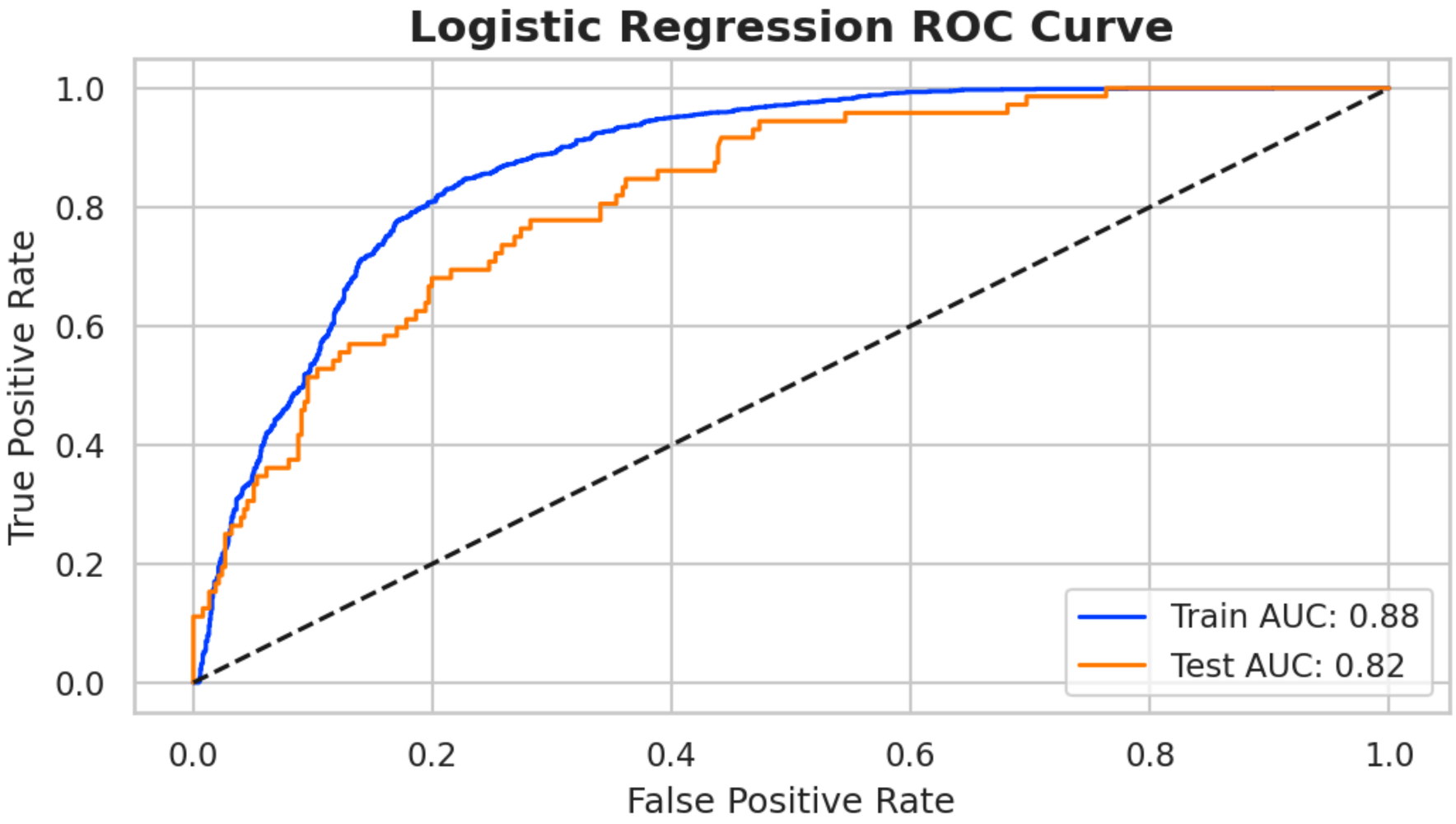


	Precision	Recall
0	0.97	0.56
1	0.28	0.90

The Power of Visual Charts.

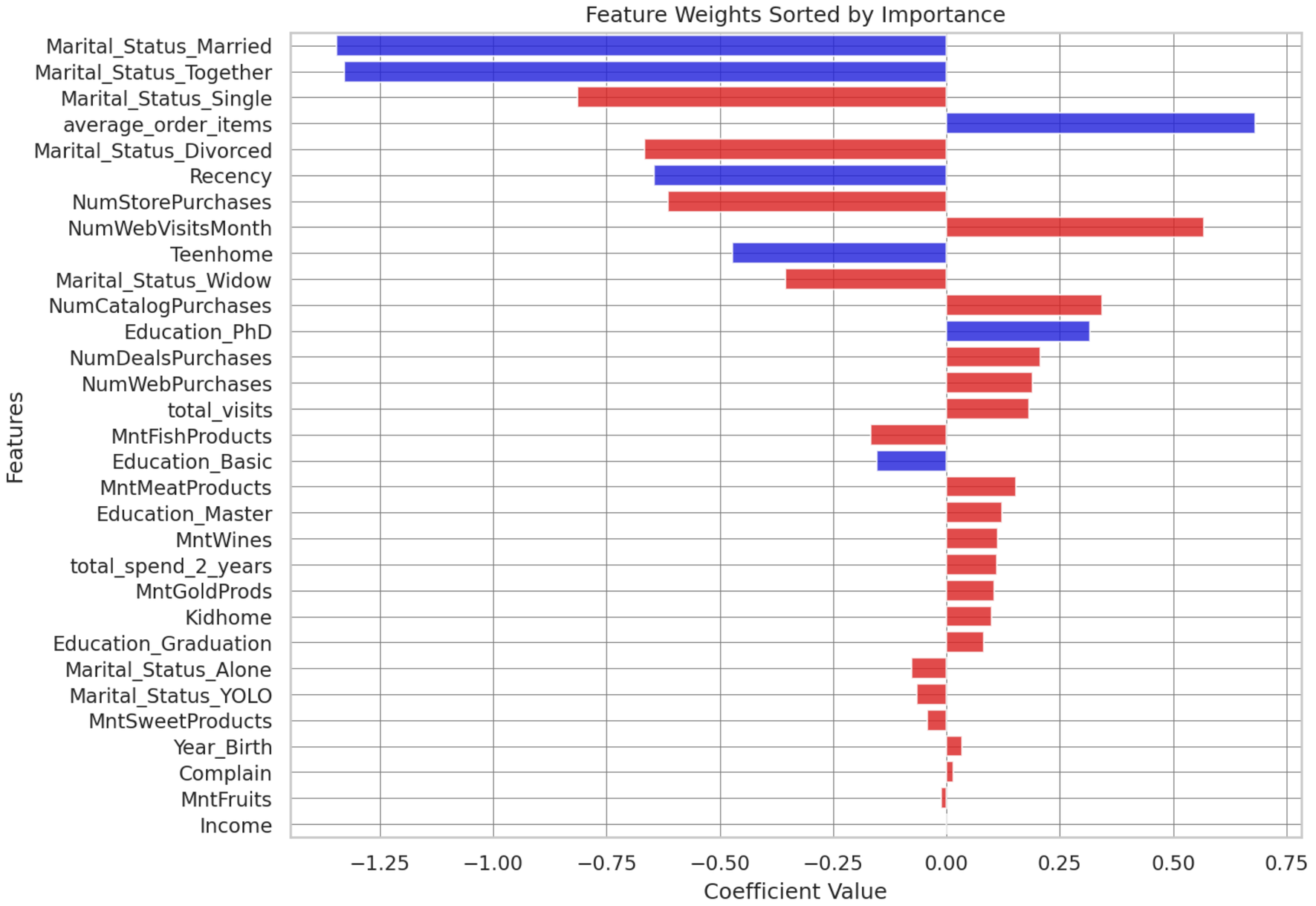
Logistic Regression





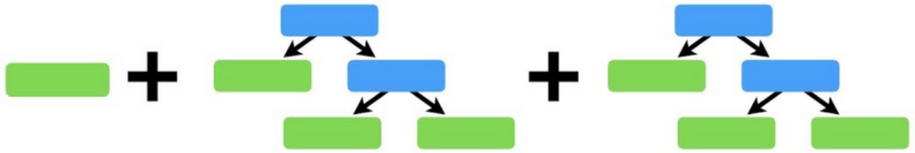
The Power of Visual Charts.

Logistic Regression



The Power of Visual Charts.

XGBoost

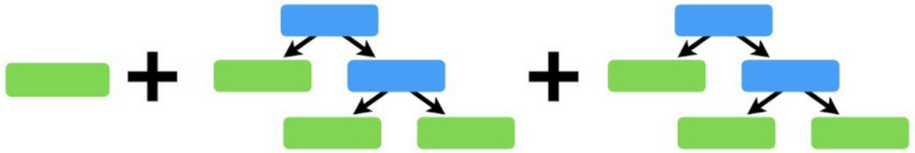


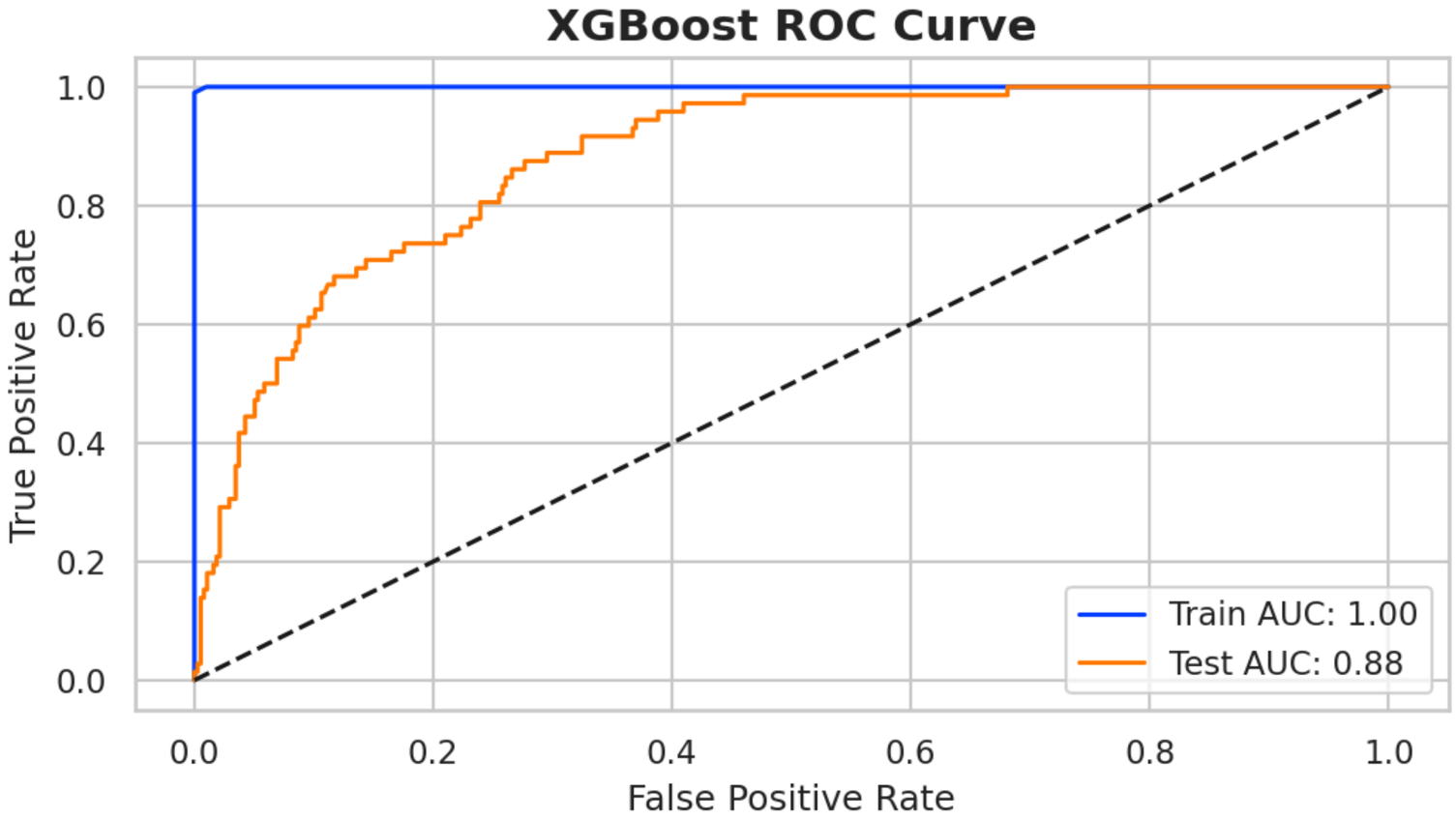
01	*Nonlinearity*
02	Interaction Variables
03	*Parallel Computing*
04	*Emsemble & High Accuracy*
05	*Less Overfitting*

	Precision	Recall
0	*0.91*	*0.92*
1	*0.57*	*0.54*

The Power of Visual Charts.

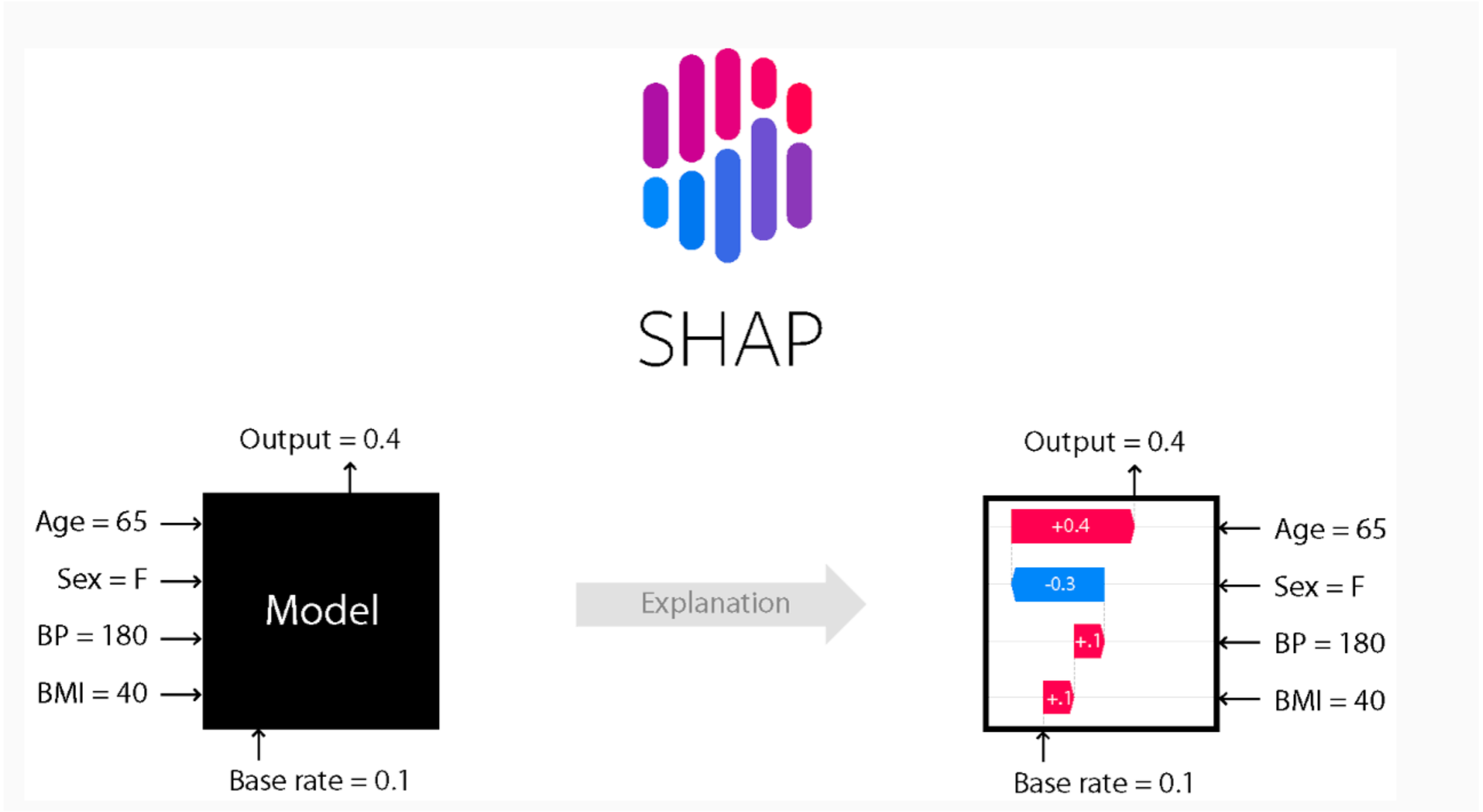
XGBoost





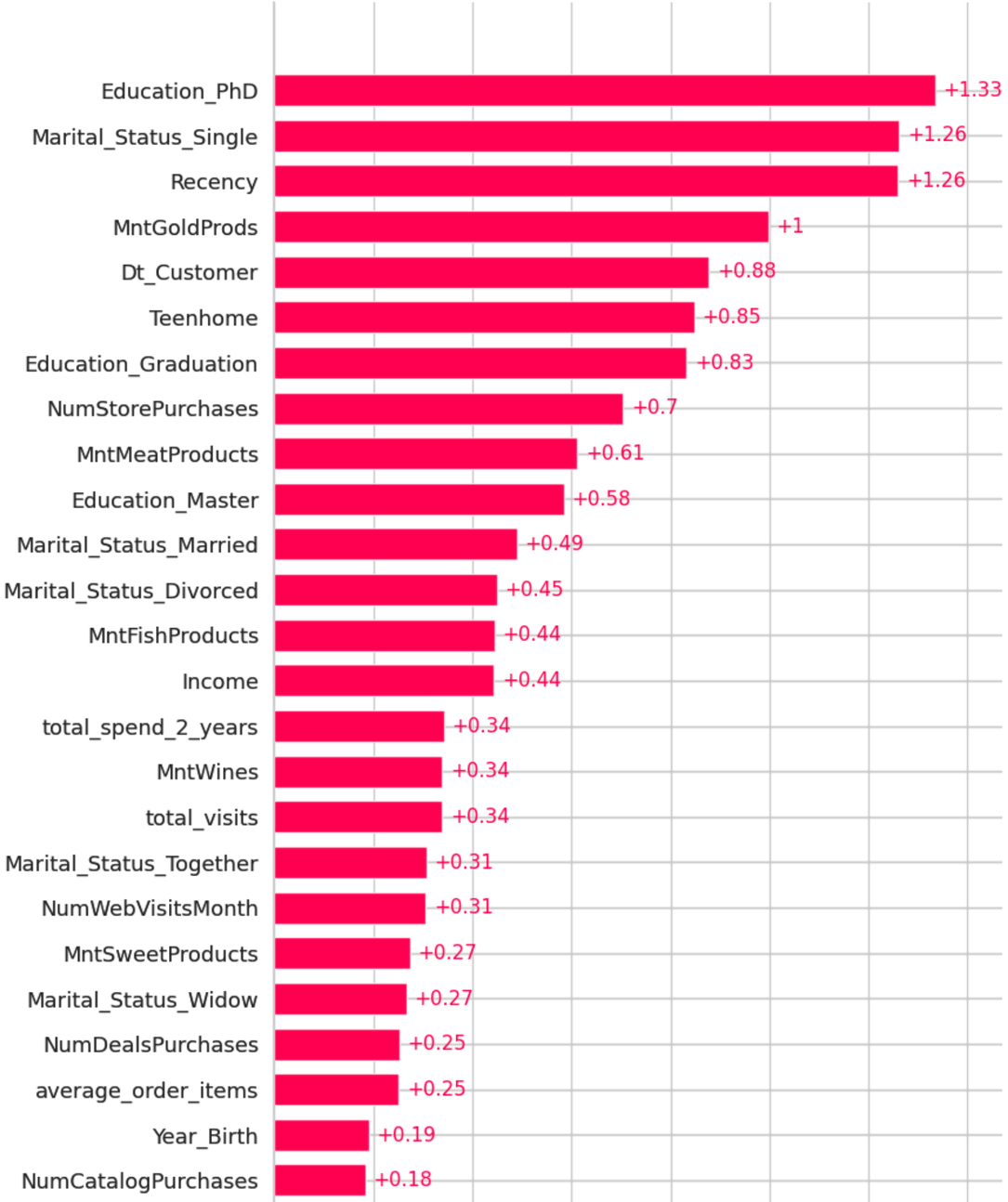
The Power of Visual Charts.

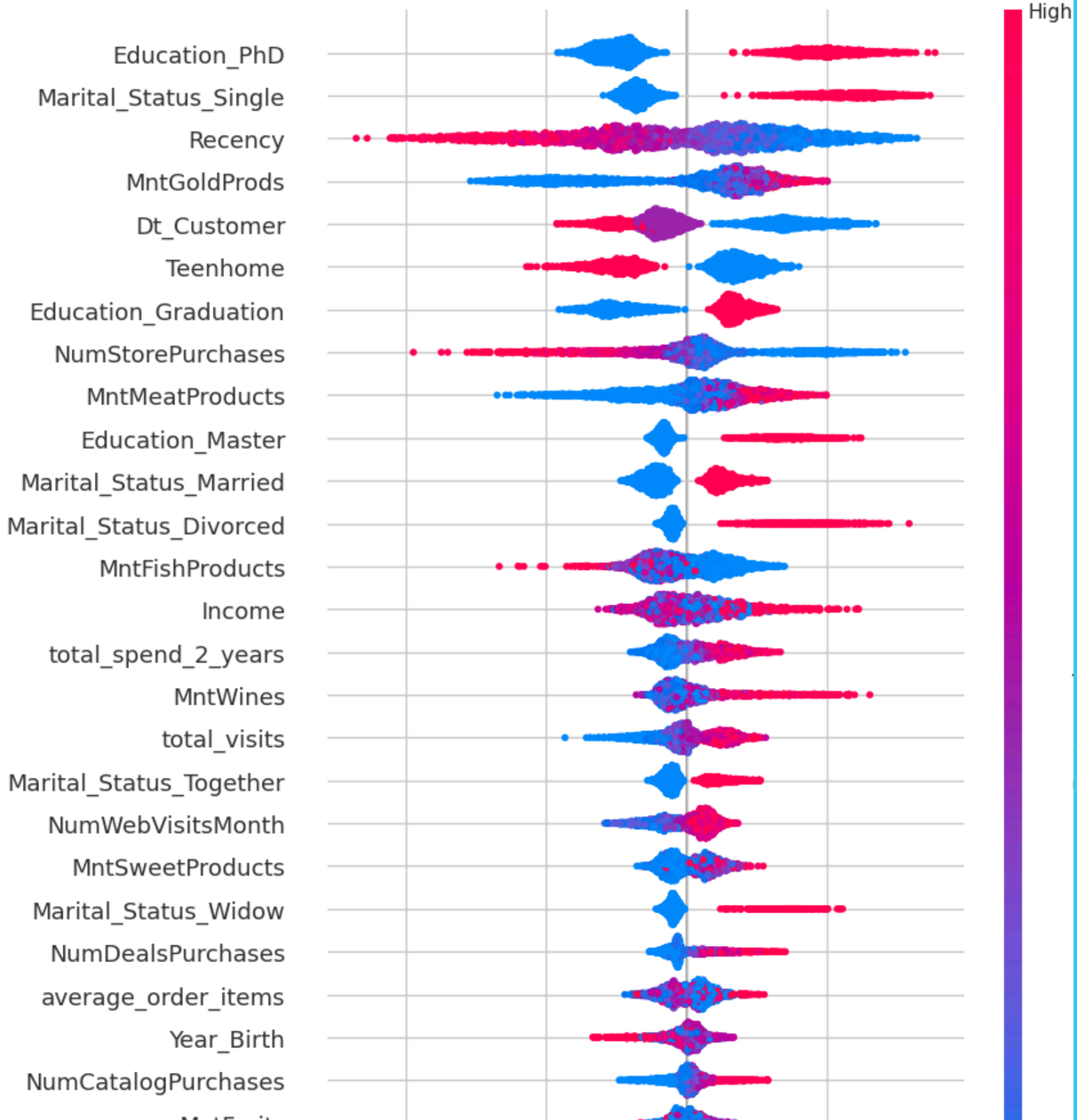
XGBoost Interpret



The Power of Visual Charts.

XGBoost Interpret





The Power of Visual Charts.

5

Resource and Risk Analysis

Deliverables:
Comprehensive resource allocation and risk management plan

RESOURCE REQUIREMENTS

Data Engineer
Design and implement data pipeline.

Data Analyst
Process and explore data to provide insightful analysis for future phases.

Machine Learning Engineers
Machine learning algorithm selection, parameter tuning, and performance optimization.

Risk Manager
Identify potential risks associated with the project and develop strategies to mitigate these risks.

Financial Analyst
Control budgets and forecast ROI

Marketing Specialist
Maintain good customer relationships
Encourage customers to purchase our new Gold memberships

Potential Risk

Risks Around Data: Data Quality and Availability Issues

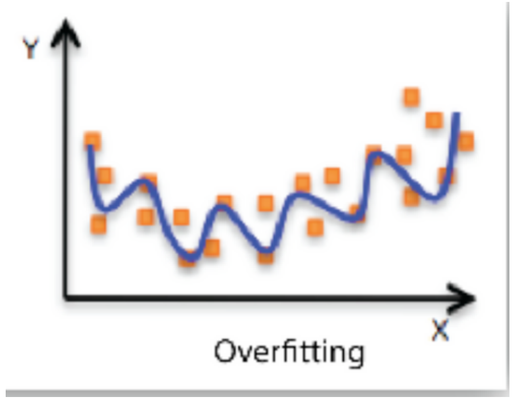
- **Assumption**: well-established data infra
- **Mitigation**: Implement stringent data quality checks before feeding data into the model. This includes handling missing values, correcting anomalies, and standardizing formats.

Risks Around Modeling: Inaccurate Model Predictions

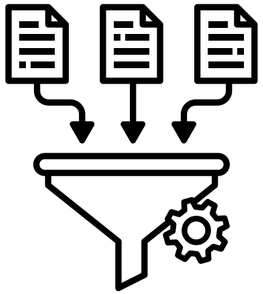
- **Assumption**: customer appeal and pricing
- **Mitigation**: Regularly validate the model against a reserved set of data to ensure accuracy. Employ cross-validation techniques during development to minimize overfitting. Consider incorporating ensemble methods to improve prediction robustness.

Technical Risks: Integration Challenges with Existing CRM Systems

- **Assumption**: well-established data infra
- **Mitigation**: Prior to development, conduct a thorough assessment of the existing IT and CRM infrastructure. Work closely with IT teams to ensure compatibility and smooth integration. Plan for beta testing phases to address any integration issues early.



Overfit - a common modeling problem



Good data infra and processing pipeline saves a lot of time

Potential Risk

Non-technical Risks: Customer Perception and Privacy Concerns

- **Assumption**: regulatory compliance
- **Mitigation**: Ensure all marketing practices comply with privacy laws and regulations. Transparently communicate the benefits and privacy terms of the Gold membership to customers. Offer an easy opt-out option to enhance trust and control.

Organizational Risks: Insufficient Staff Expertise or Resources

- **Assumption**: Staff expertise and training procedure
- **Mitigation**: Evaluate the current capabilities of the data science and marketing teams. If necessary, consider training existing staff or hiring additional experts. Plan resource allocation meticulously to avoid bottlenecks.

Avoid private data leak

Critical Success Factors

- **Accurate and Robust Predictive Modeling:** The success of the project heavily relies on the ability to develop a predictive model that accurately identifies potential buyers of the Gold membership. This involves selecting the right features, using appropriate modeling techniques, and continuously monitoring and refining the model based on performance.

- **Effective Integration with CRM Systems:** Seamless integration of the predictive model with existing CRM systems is crucial for the efficient execution of the campaign. This includes ensuring that the model outputs are effectively used to target potential customers through the CRM.

- **Customer Experience and Satisfaction**: Maintaining or enhancing customer satisfaction and loyalty by making offers that are perceived as relevant and beneficial. This involves understanding customer needs deeply and personalizing communications effectively.

- **Stakeholder Communication**: Regular and clear communication with all stakeholders, including project updates and changes, to ensure alignment and address concerns promptly. Effective communication fosters stakeholder buy-in and supports smoother project execution.

6

Action and Final Report

Deliverables:

Plan for communicating with stakeholders and detailed document summarizing the project.

Communicate project progress to stakeholders

- **Meetings**: Hold regular meetings or teleconferences with key stakeholders to discuss progress, challenges, and strategic decisions. Present key progress indicators, recent findings, upcoming milestones, and any immediate needs or concerns

- **Dashboards**: Develop a real-time dashboard that provides stakeholders with ongoing access to key performance indicators (KPIs), model accuracy metrics, and campaign outcomes. This fosters transparency and allows for quick adjustments.

- **Reports**: Provide detailed reports at each milestone. These will cover what has been achieved, variances from the plan, insights gained, and adjustments to the strategy.

Thank you!