MARKETING CAMPAIGN ANALYSIS AND OPTIMIZATION

INTRODUCTION

- Effective marketing campaigns are crucial for businesses to reach the right customers and maximize return on investment.

- This project analyzes a real-world marketing dataset to uncover customer behavior patterns and predict campaign responses using machine learning.

- Key insights are visualized through an interactive Power BI dashboard to support data-driven marketing decisions.

PROBLEM STATEMENT

Ineffective targeting in marketing campaigns leads to low response rates, highlighting the need for data-driven methods to accurately predict customer behavior and optimize campaign performance.

DATASET OVERVIEW

- The dataset used for this project is the Marketing Campaign Dataset sourced from Kaggle, originally provided by a Portuguese retail company. It contains 2,240 records and 29 features.

Key Feature Categories:

- ***Customer Demographics***:
  - ✓ Age: Derived from Year_Birth
  - ✓ Education: customer's level of education
  - ✓ Marital_Status: customer's marital status
  - ✓ Income: Annual household income

- ***Household Composition:***
  - ✓ Kidhome, Teenhome: Number of children/teenagers at home

- ***Customer Behavior:***
  - ✓ Recency: Days since last purchase
  - ✓ Customer_Tenure_Years: Years since the customer joined
  - ✓ Complain: Whether the customer has complained in the past

- ***Product Spendings:***
  - ✓ MntWines, MntMeatProducts, MntFruits, MntSweetProducts, MntFishProducts, MntGoldProds: Amount spent in each category

- ***Campaign Responses:***
  - ✓ AcceptedCmp1 to AcceptedCmp5: Response to individual campaigns
  - ✓ Response: Overall response to the last campaign

- ***Channel & Purchase Behavior*:**
  - ✓ NumWebPurchases,NumCatalogPurchases, NumStorePurchases: Purchase channels
  - ✓ NumWebVisitsMonth: Website visits in last month
  - ✓ NumDealsPurchases: number of purchases made with discount

- ***Target Variable*:**

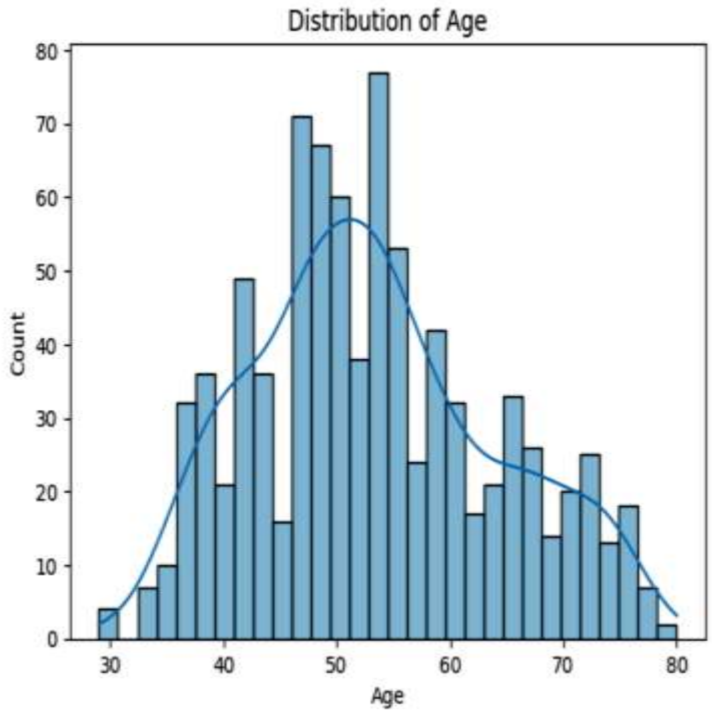
✓ Response: A binary variable (0 or 1) indicating whether the customer responded positively to the latest marketing campaign.

EDA VISUALS

DATA VISUALIZATION

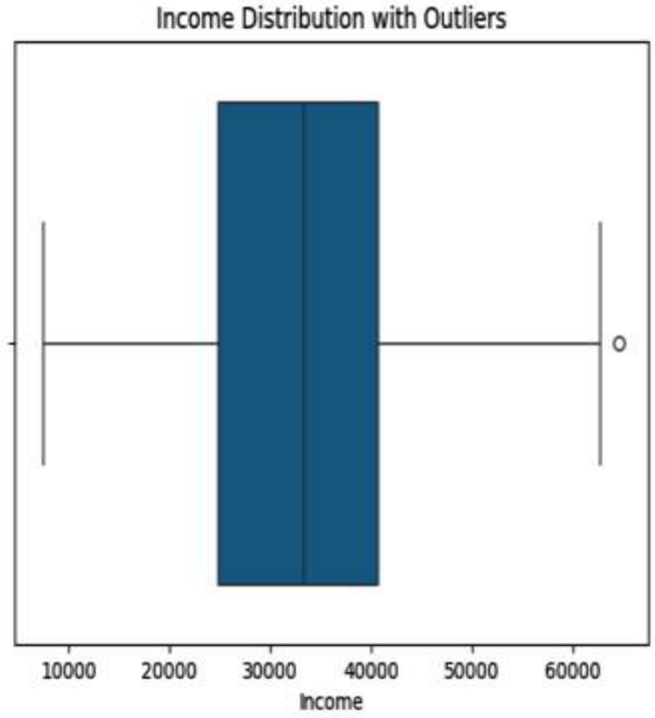
```
[39] # Histogram for Age
     sns.histplot(data['Age'], bins=30, kde=True)
     plt.title("Distribution of Age")
```

```
Text(0.5, 1.0, 'Distribution of Age')
```



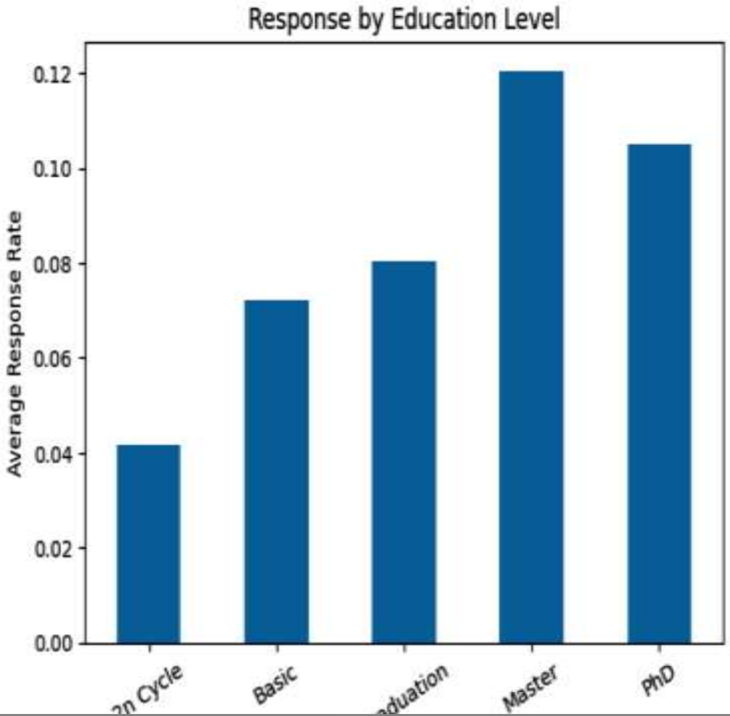
[40]
```
# Boxplot for Income
sns.boxplot(x=data['Income'])
plt.title("Income Distribution with Outliers")
```

```
Text(0.5, 1.0, 'Income Distribution with Outliers')
```



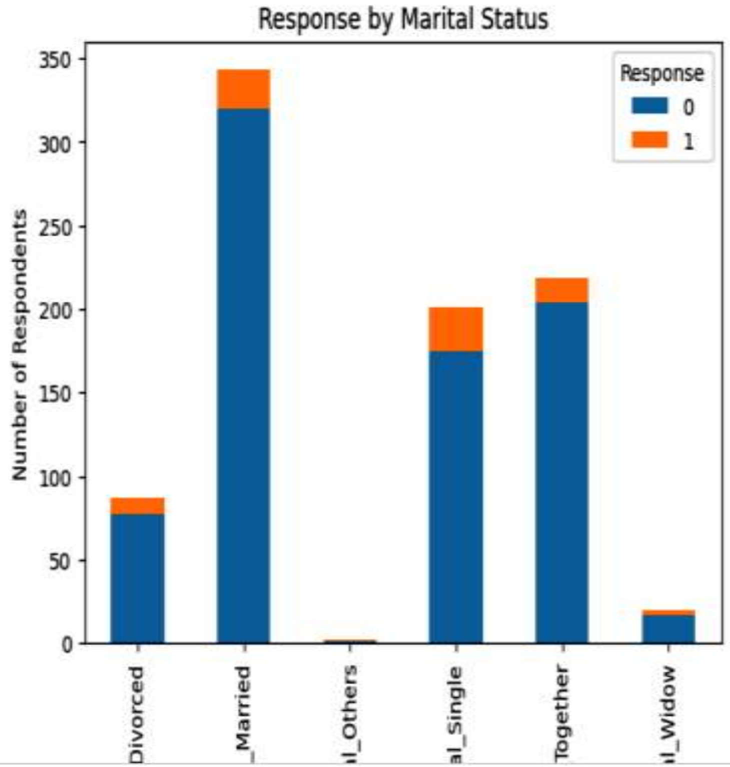
[41]
```
# Count of Education levels
sns.countplot(x='Education', data=data)
```

```python
edu_map = {0: 'Basic', 1: '2n Cycle', 2: 'Graduation', 3: 'Master', 4: 'PhD'}

education_response = data.groupby(data['Education'].map(edu_map))['Response'].mean()

education_response.plot(kind='bar')
plt.title("Response by Education Level")
plt.xlabel("Education")
plt.ylabel("Average Response Rate")
plt.xticks(rotation=30)
plt.show()
```



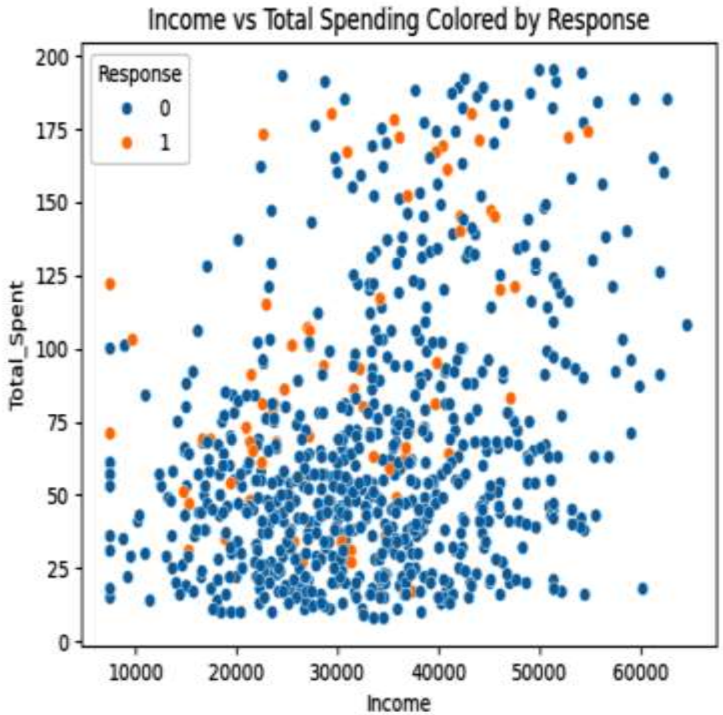
```python
# Response rate by Marital Status
marital_cols = ['Marital_Divorced','Marital_Married','Marital_Others','Marital_Single','Marital_Together','Marital_Widow']
response_by_marital = data.groupby('Response')[marital_cols].sum().T
response_by_marital.plot(kind='bar', stacked=True)
plt.title("Response by Marital Status")
plt.ylabel("Number of Respondents")
```

```
Text(0, 0.5, 'Number of Respondents')
```



```python
# Income vs. Total_Spent
sns.scatterplot(x='Income', y='Total_Spent', hue='Response', data=data)
plt.title("Income vs Total Spending Colored by Response")
```

```
Text(0.5, 1.0, 'Income vs Total Spending Colored by Response')
```




```python
[48] accepted_cols = ['AcceptedCmp1', 'AcceptedCmp2', 'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5']
```

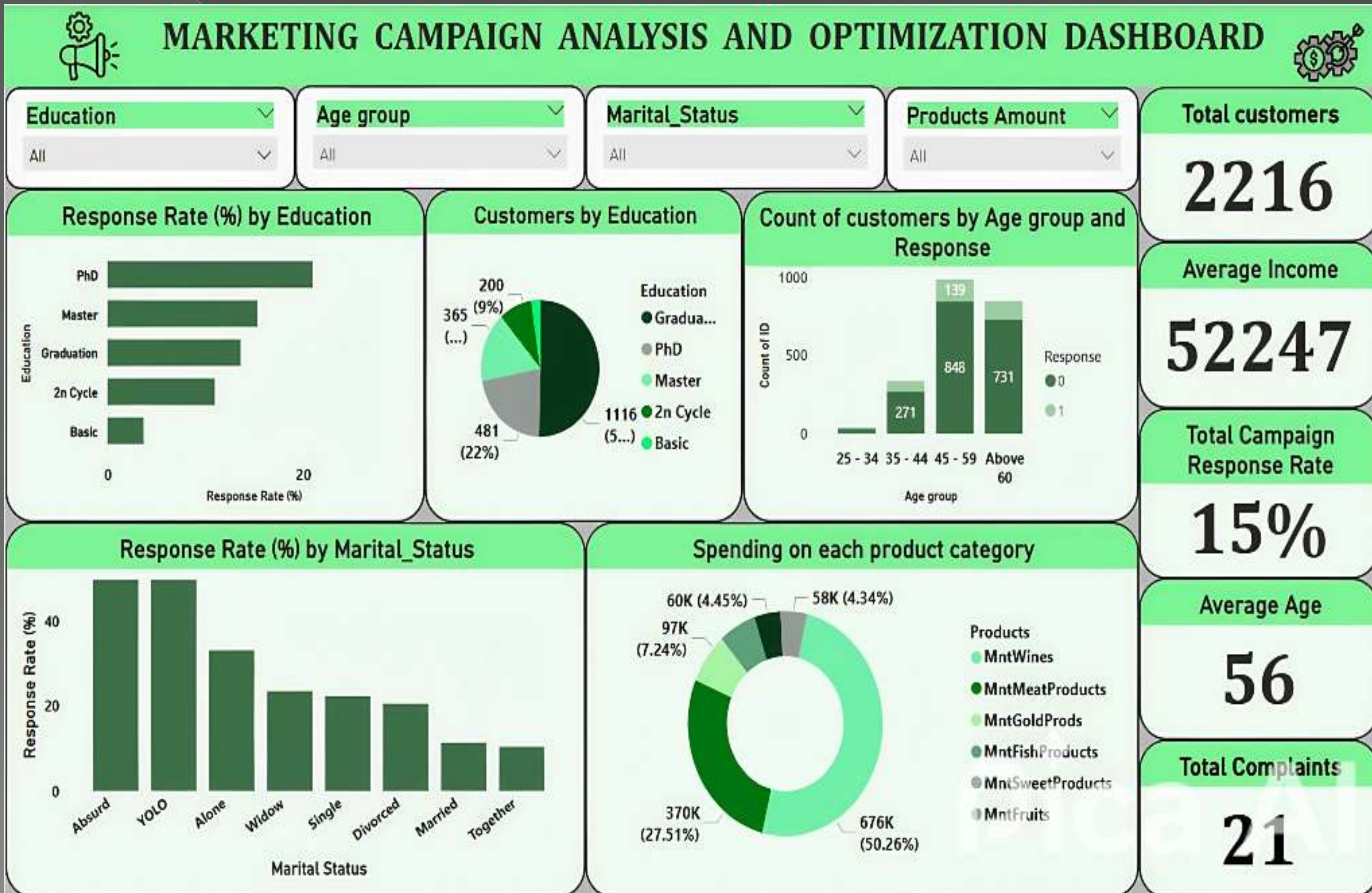
MODELS USED AND RESULTS

Models Evaluated:

- **Logistic Regression**
  - Accuracy (CV=10): **91.39%**
  - Accuracy (CV=50): **91.76%**
  - Precision: **1.00**, Recall: **0.31**, F1 Score: **0.48**
- **Random Forest Classifier**
  - Accuracy (CV=10, n=40): **91.85%**
  - Accuracy (CV=50, n=5): **91.82%**
  - Precision: **0.43**, Recall: **0.21**, F1 Score: **0.29**
- **Support Vector Classifier (SVC)**
  - Accuracy (CV=50): **91.21%**
  - Accuracy (CV=20): **91.16%**

- **K-Nearest Neighbors (KNN)**
  - Accuracy (CV=10): **91.16%**
  - Accuracy (CV=20): **91.16%**
- **Decision Tree Classifier**
  - Accuracy (CV=50): **88.57%**
  - Accuracy (CV=70): **87.88%**
- **Gaussian Naive Bayes**
  - Accuracy (CV=20): **87.15%**
  - Accuracy (CV=70): **87.06%**
- ***Summary:***
- **Best Accuracy**: **Random Forest (CV=10)** with **91.85%**
- **Most Precise Prediction**: **Logistic Regression** with **Precision = 1.00**
- Although Random Forest achieved the highest accuracy, **Logistic Regression** showed better balance in predictive quality based on the F1 Score.

POWER BI DASHBOARD



CONCLUSION

This project combined machine learning and business intelligence to analyze and optimize marketing campaign performance using real-world customer data.

Key Findings:

- Education Impact: Customers with high education showed the highest campaign response rates—over 20%, compared to below 10% for Basic education.

- Channel Preference: Most purchases were made through store visits and web, while catalog purchases had the lowest engagement—indicating a shift toward more modern retail channels.

- Product Spending: Customers spent most on Wines (50.26%), followed by Meat Products (27.5%), highlighting clear product preferences.

- Customer Recency: Customers who made recent purchases (lower recency values) were more likely to respond positively to new campaigns.
- Family Composition: Customers with 0 or 1 child at home responded more to campaigns than larger households, suggesting individual-focused targeting is more effective.
- Demographics Summary:
  - **Average Age**: 56 years
  - **Average Income**: $52,247
  - **Total Customers**: 2,216
  - **Campaign Response Rate**: 15%
- ML Insights:
- Logistic Regression performed best in predicting campaign responses, with a Precision of 1.0, Recall of 0.31, and F1 Score of 0.48, ensuring minimal false positives in campaign targeting.

- Power BI Dashboard:
- The dashboard provided interactive filtering and visualization by education, age, recency, spending behavior, and more, enabling fast and informed decision-making.

- *Recommendations to Improve ROI:*
- Focus campaigns on high-response groups like highly educated customers and those with recent purchases.
- Personalize offers for high-spending segments (e.g., wine and meat buyers).
- Use modern channels (web and store) more than catalog-based outreach.
- Retrain ML models periodically with updated customer data to refine predictions.
- A/B test campaigns before full rollout to optimize messaging for each segment.

THANK YOU……