IITK Course 2 - Marketing Campaign Analysis

Problem Statement

This report documents the complete solution for the marketing campaign data analysis problem.

The steps covered are:
1. Data import and validation
2. Missing value imputation for income
3. Feature engineering
4. Distribution visualization and outlier treatment
5. Categorical encoding
6. Correlation heatmap
7. Hypothesis testing with visualizations
8. Business insights and final conclusions.

Step 1-2: Data Import, Cleaning & Income Imputation

```
# Load dataset and initial cleaning
df = pd.read_csv('marketing_data.csv', encoding='utf-8-sig')
df.columns = df.columns.str.strip()

# Convert Dt_Customer to datetime
df['Dt_Customer'] = pd.to_datetime(df['Dt_Customer'], errors='coerce')

# Clean Income and convert to float
df['Income'] = df['Income'].replace('[\$,]', '', regex=True).astype(float)

# Clean Education and Marital_Status
df['Education'] = df['Education'].str.strip().str.title()
df['Marital_Status'] = df['Marital_Status'].str.strip().str.title()

# Impute missing Income by median grouped by Education and Marital_Status
df['Income'] = df.groupby(['Education', 'Marital_Status'])['Income'].transform(lambda x:
x.fillna(x.median()))
df.dropna(subset=['Income'], inplace=True)
```

The income column was missing values, so we filled missing entries using the median income for groups defined by Education and Marital Status. This maintains consistency in imputation based on similar customer profiles.

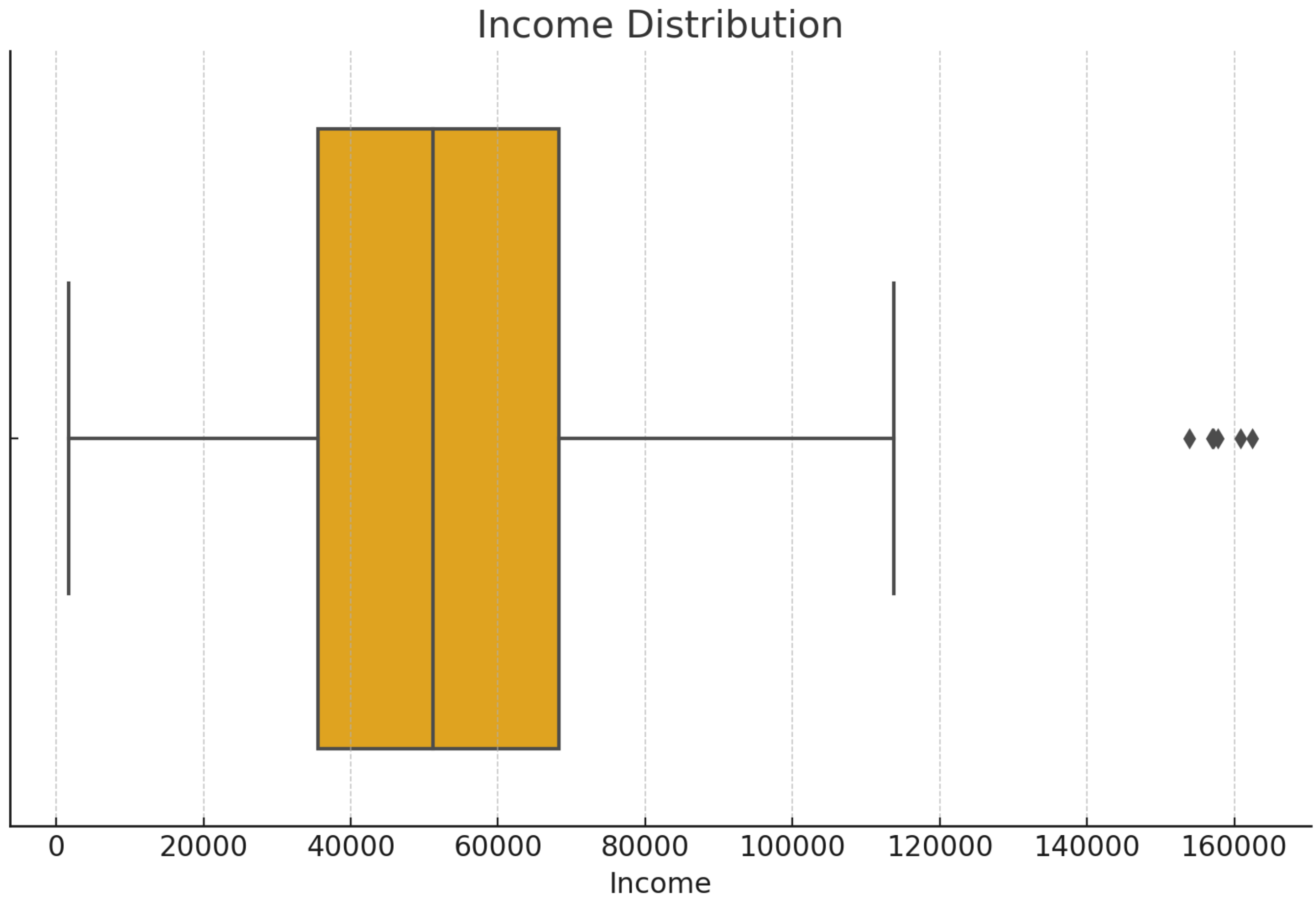
Step 3: Feature Engineering

```
# Create Age, Children, Total Spending, Total Purchases
df['Age'] = dt.datetime.now().year - df['Year_Birth']
df['Children'] = df['Kidhome'] + df['Teenhome']
```

IITK Course 2 - Marketing Campaign Analysis

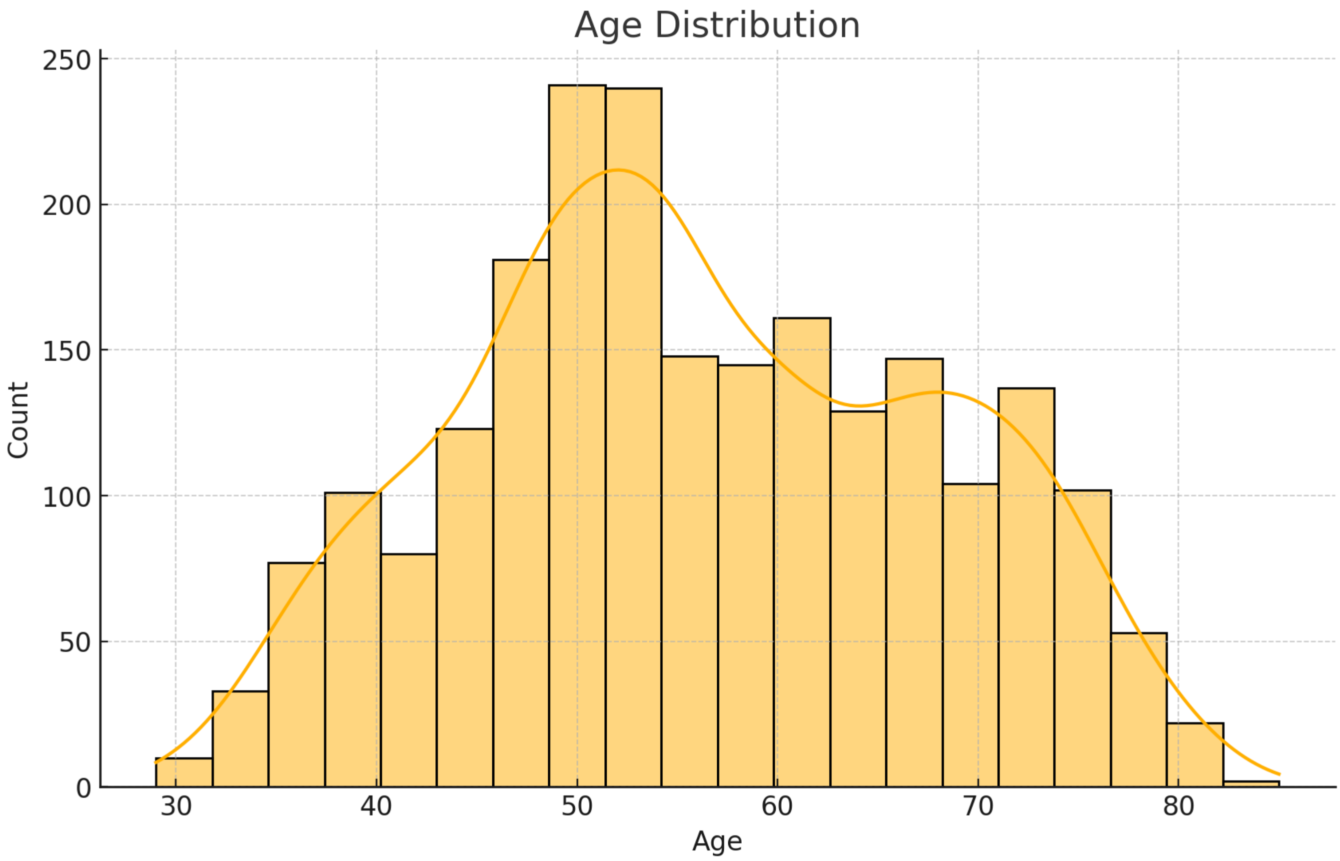
```
df['Total_Spending'] = df[['MntWines', 'MntFruits', 'MntMeatProducts', 'MntFishProducts',
'MntSweetProducts', 'MntGoldProds']].sum(axis=1)
df['Total_Purchases'] = df[['NumWebPurchases', 'NumCatalogPurchases',
'NumStorePurchases']].sum(axis=1)
```

New variables were created to capture the customer's age, total children, total spending on products, and total purchase transactions across channels.



Income Distribution - Boxplot

IITK Course 2 - Marketing Campaign Analysis



Age Distribution - Histogram

Visualizations show the distribution of Income and Age. Outliers in Income (>200,000) and Age (>100) were removed for cleaner analysis.

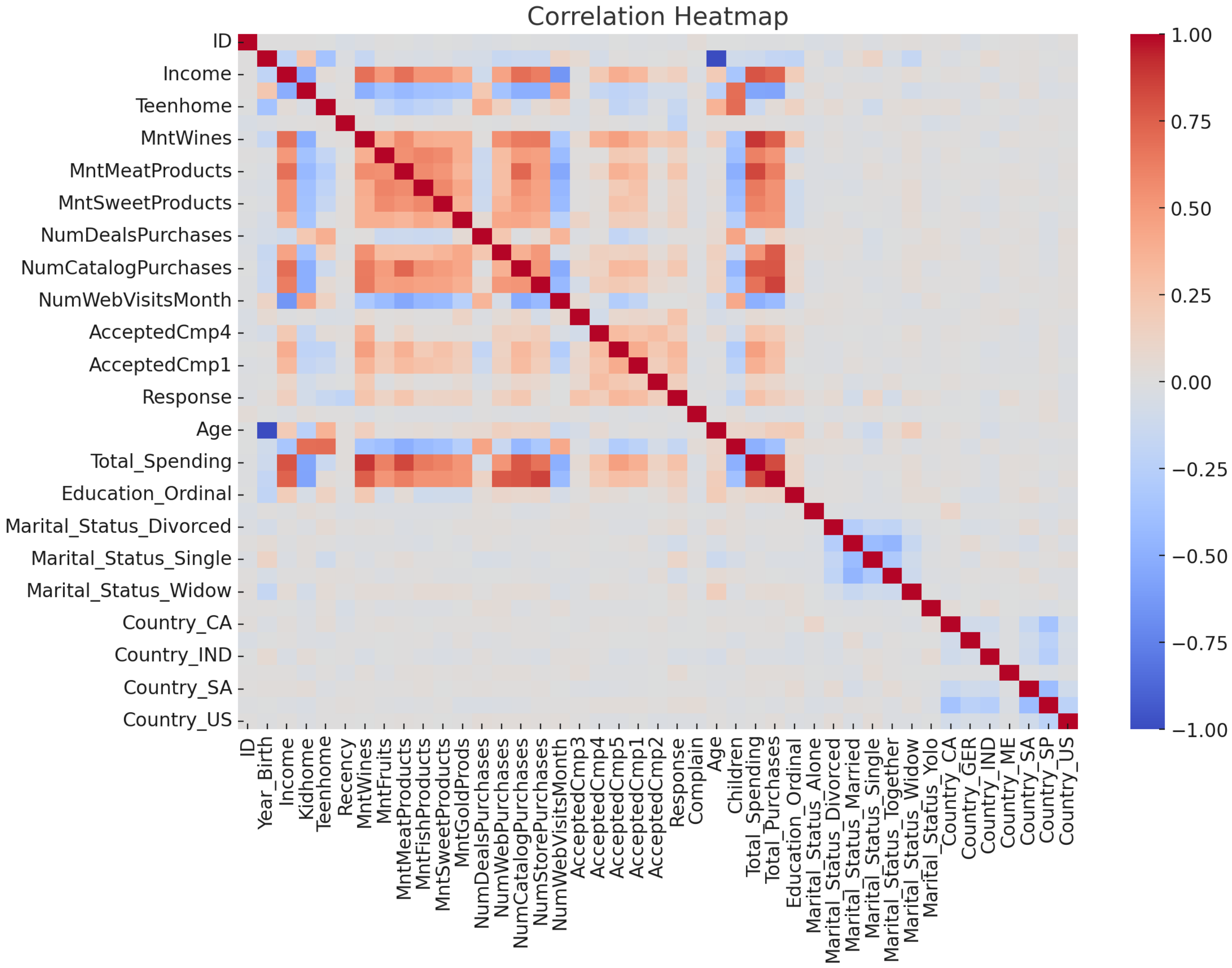
Step 5: Encoding Categorical Variables

```
# Ordinal encoding for Education
ordinal_map = {'Basic':1, '2N Cycle':2, 'Graduation':3, 'Master':4, 'Phd':5}
df['Education_Ordinal'] = df['Education'].map(ordinal_map)

# One-hot encoding for Marital_Status and Country
df_encoded = pd.get_dummies(df, columns=['Marital_Status', 'Country'], drop_first=True)
```

Education was ordinally encoded while Marital Status and Country were one-hot encoded to prepare for correlation and modeling.

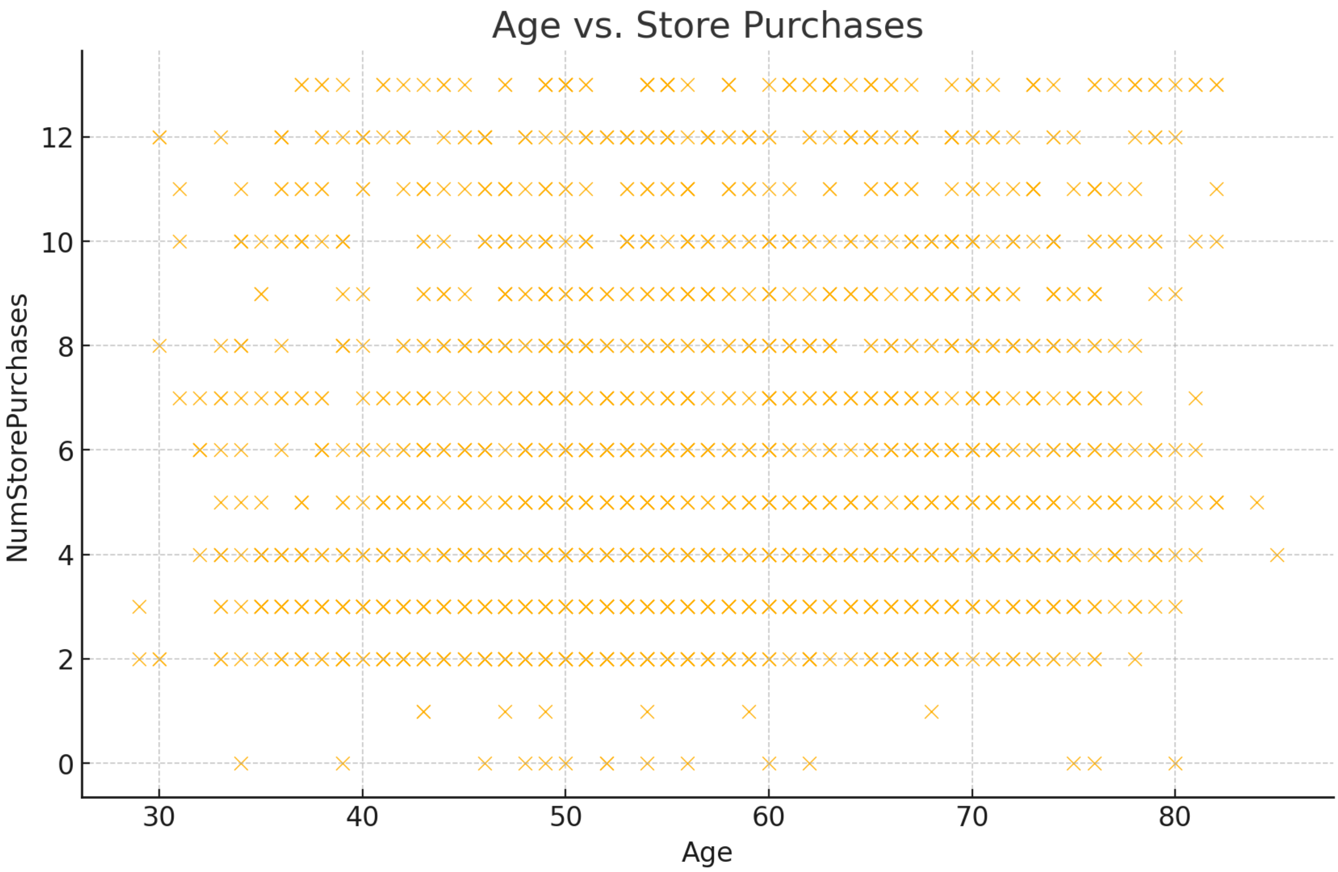
IITK Course 2 - Marketing Campaign Analysis



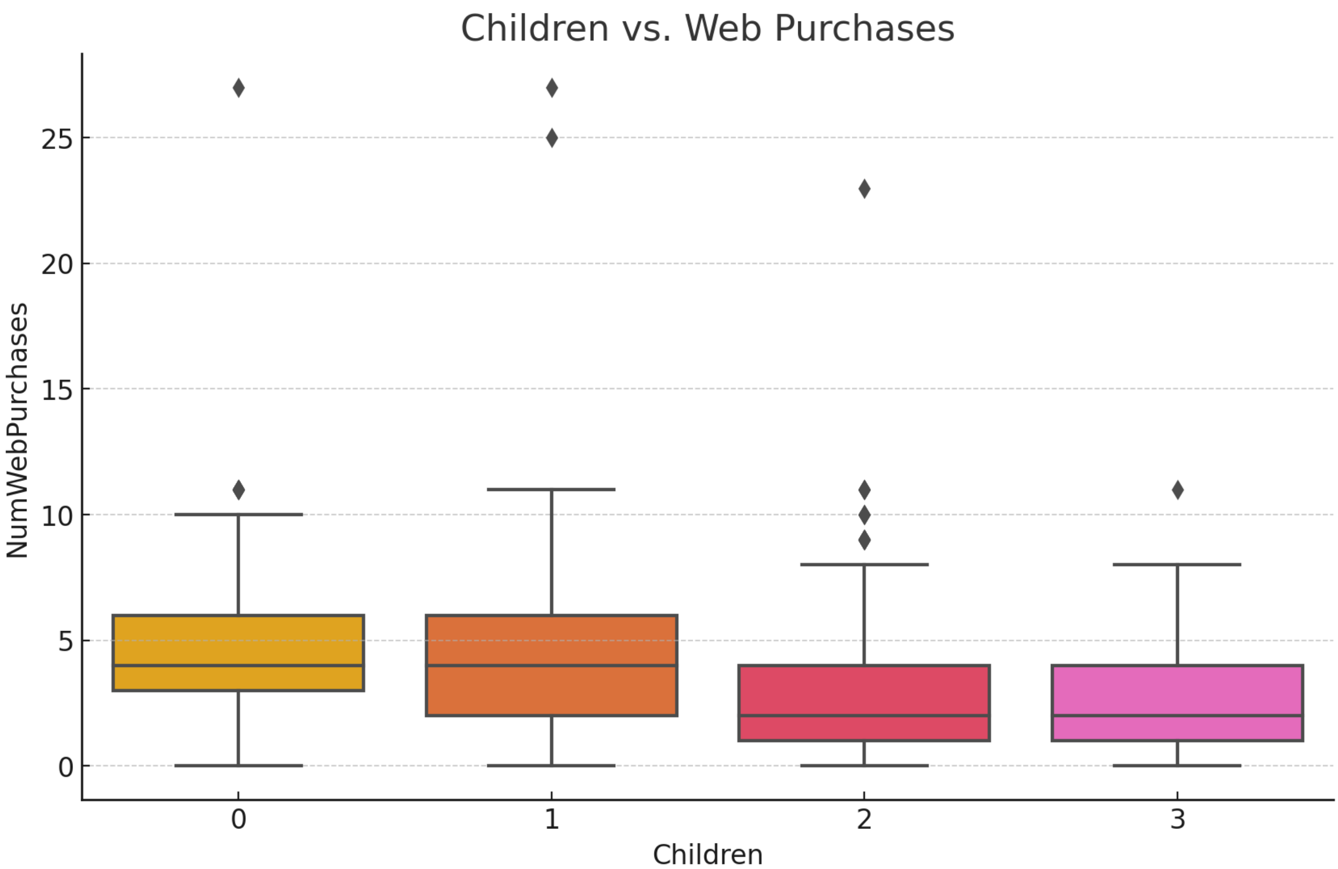
Correlation Heatmap

Step 7: Hypothesis Testing via Visualizations

IITK Course 2 - Marketing Campaign Analysis



Hypothesis: Older customers prefer in-store shopping

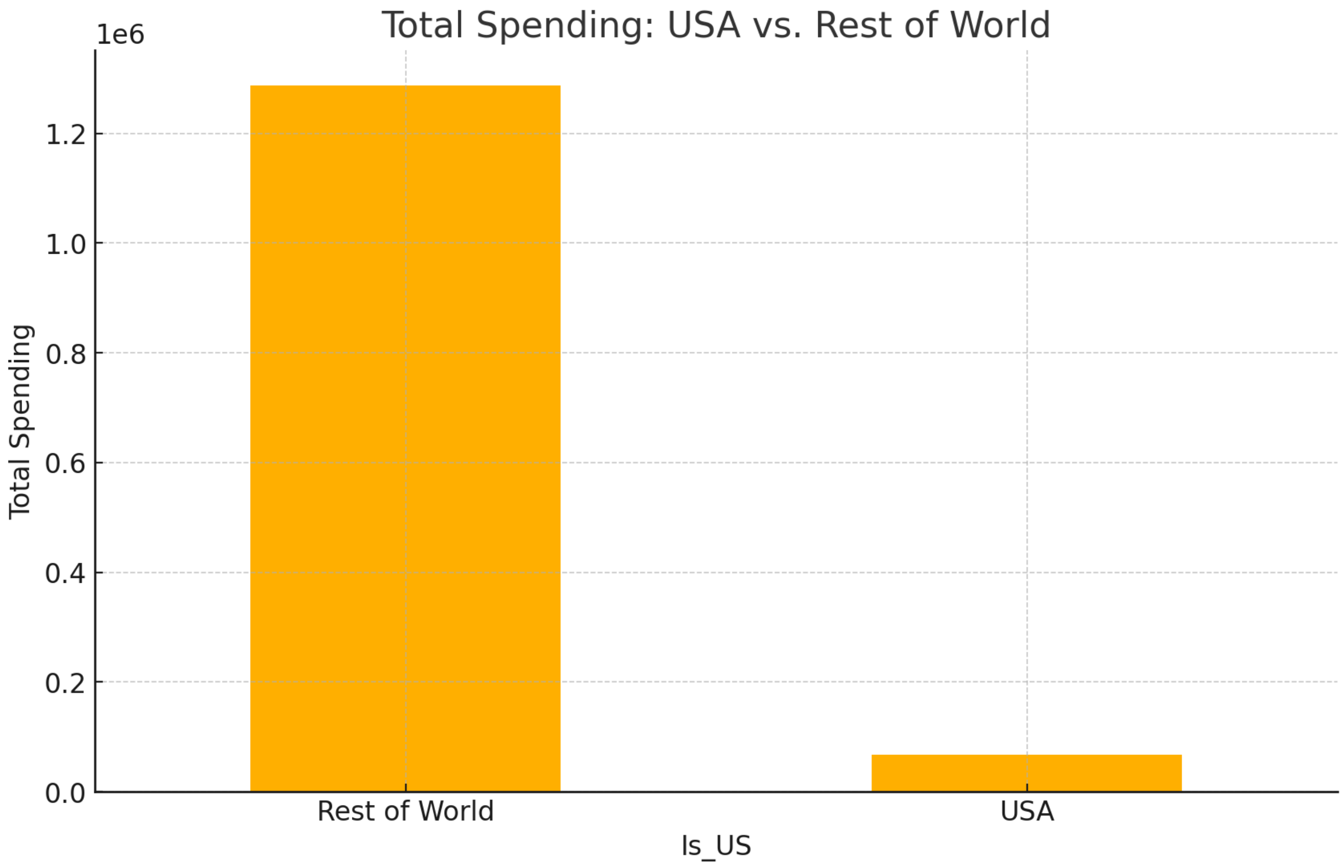


IITK Course 2 - Marketing Campaign Analysis

Hypothesis: Customers with children prefer online shopping

Hypothesis: Online purchases may cannibalize store sales

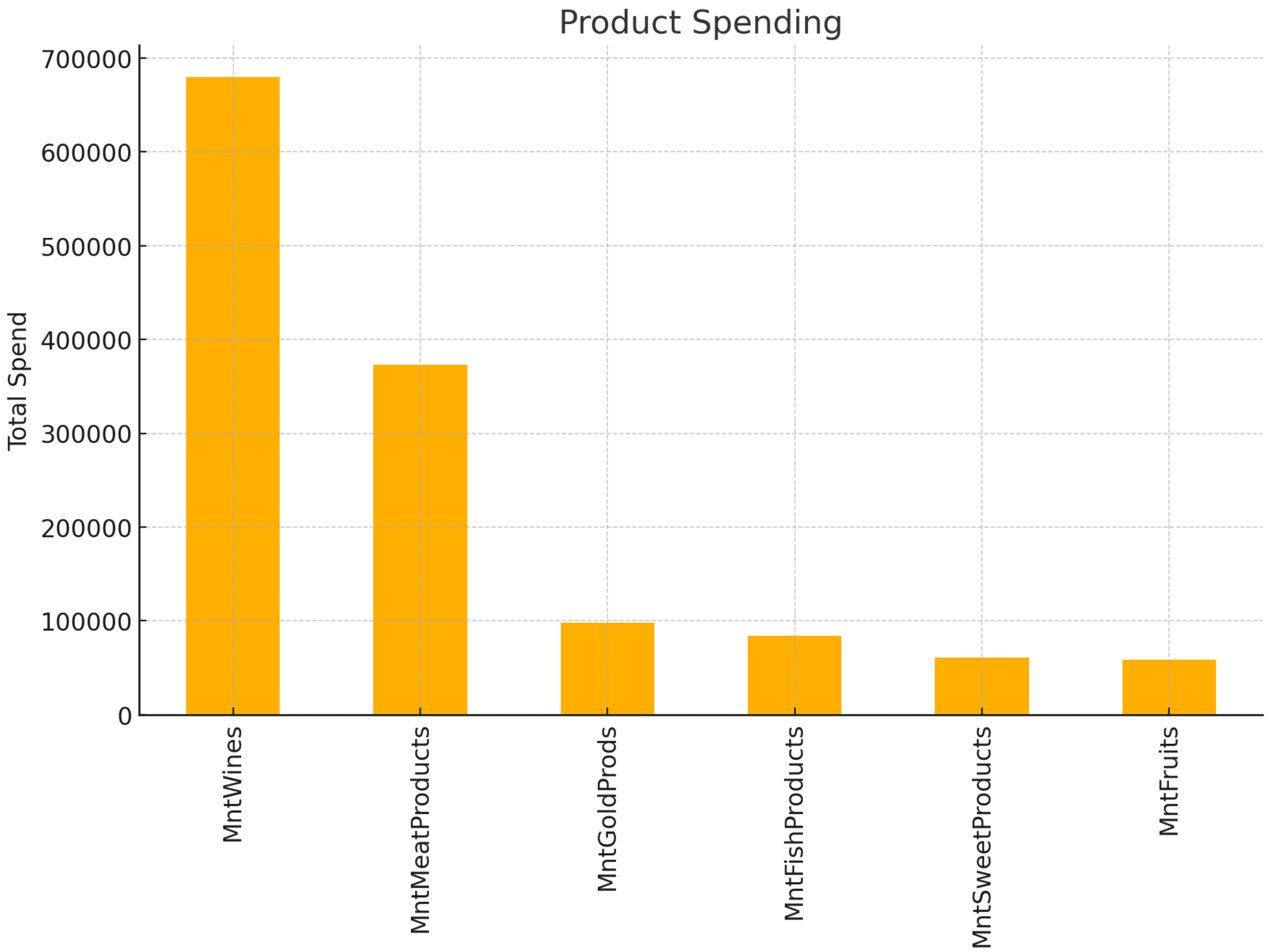
IITK Course 2 - Marketing Campaign Analysis



Hypothesis: USA outperforms rest of world in purchase volumes

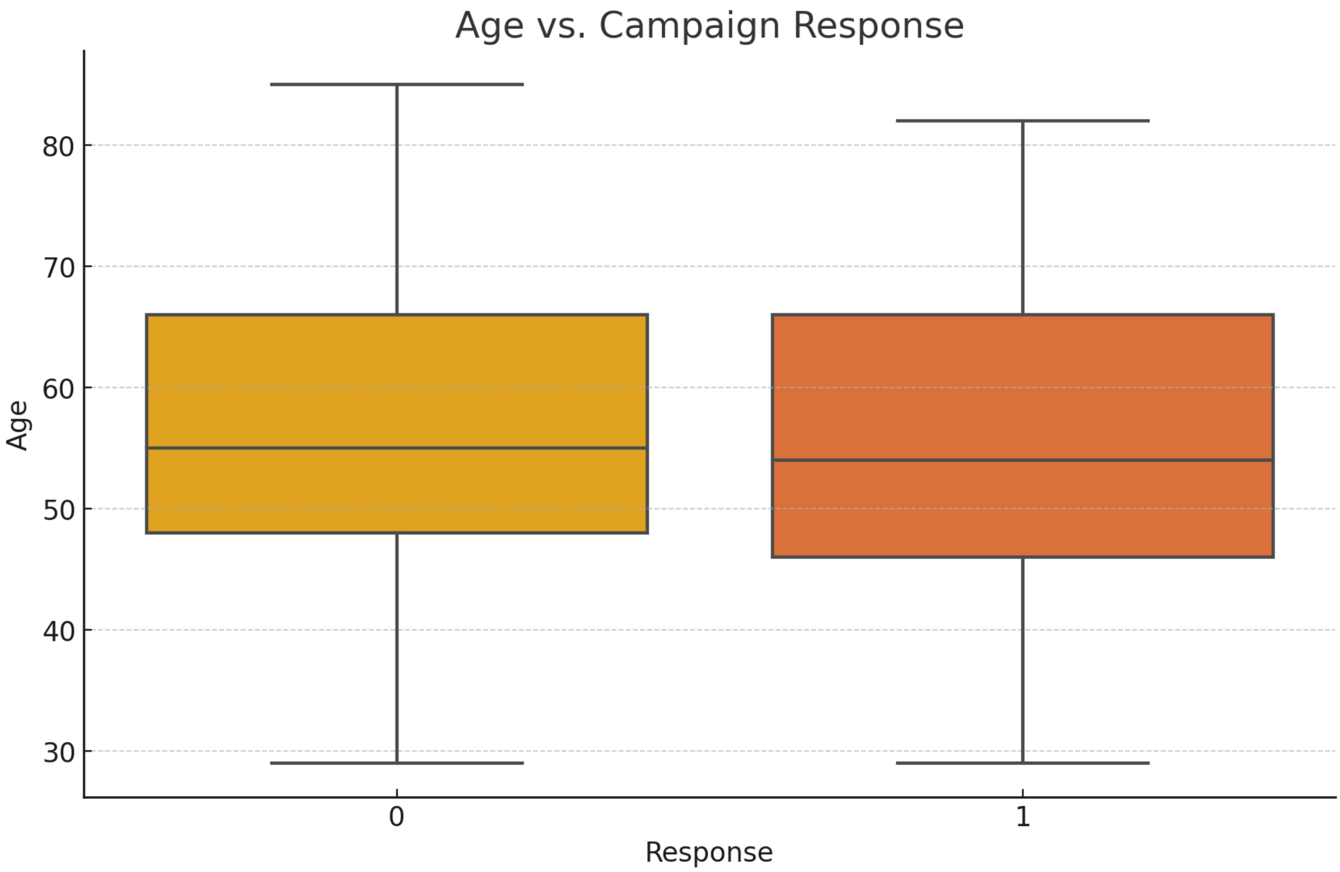
Step 8: Business Insights

IITK Course 2 - Marketing Campaign Analysis

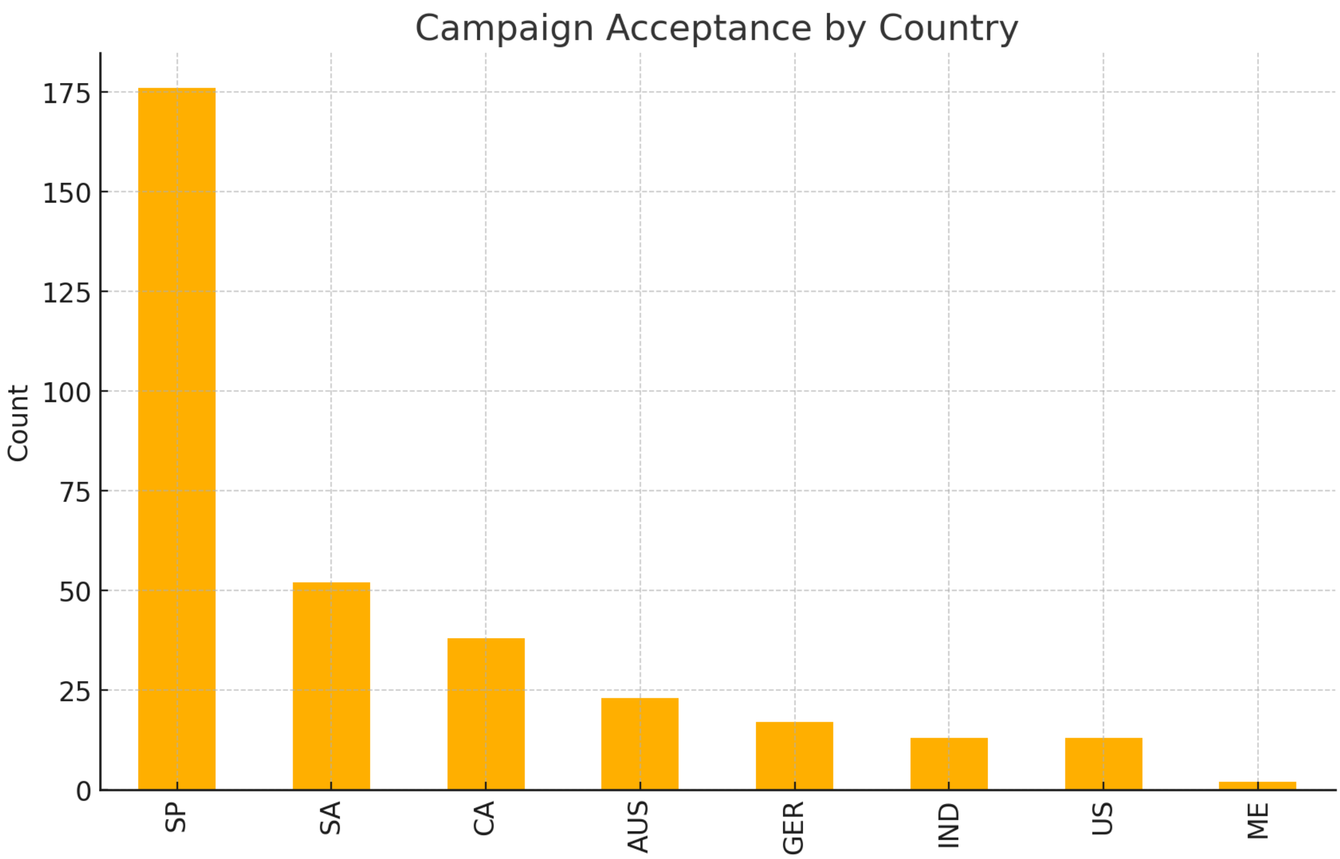


Top-performing and Lowest-performing Products

IITK Course 2 - Marketing Campaign Analysis

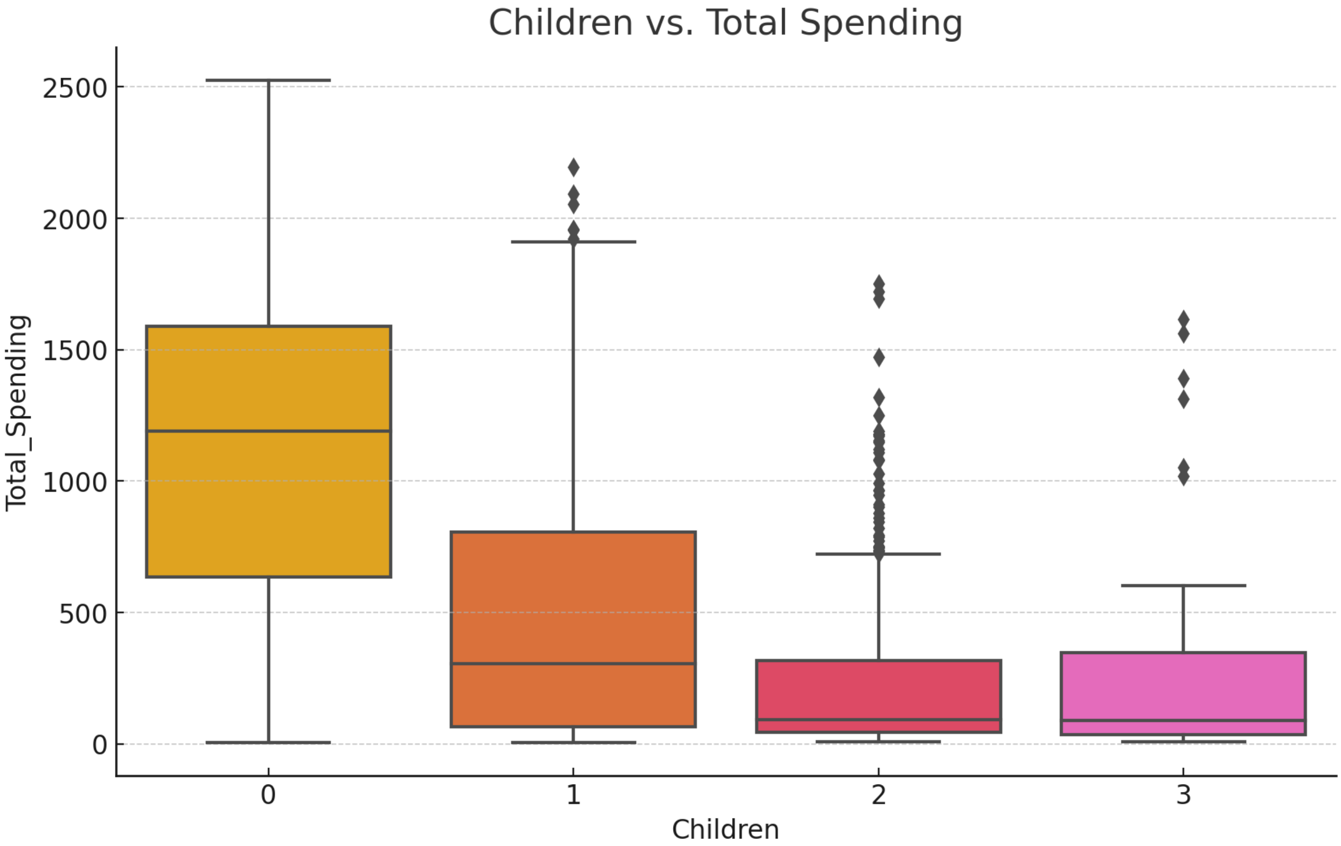


Correlation between Age and Campaign Acceptance



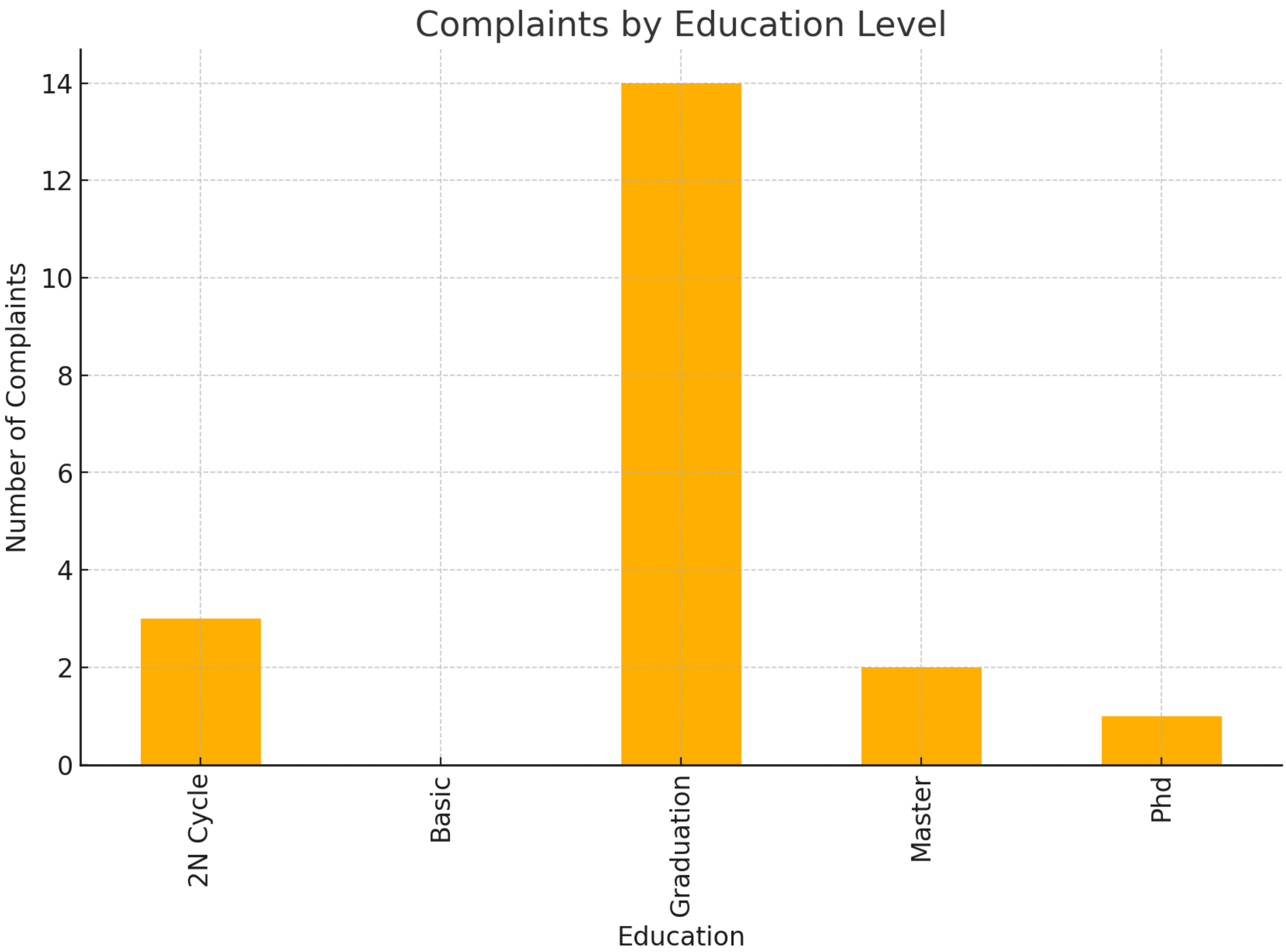
IITK Course 2 - Marketing Campaign Analysis

Country with Highest Campaign Acceptance



Pattern of Children vs Total Expenditure

IITK Course 2 - Marketing Campaign Analysis



Educational Background vs Complaints