Enhancing Marketing Strategies Through **Customer Segmentation** and **Retention Analysis**

Capstone Project #2 | Inggar Gumintang | JCDSOL - 014 - 2

CONTENT PILLARS

Conclusion and Future Work

Summary of Key Findings

Final Strategic Recommendations

Data Analysis

Customer Segmentation Insights

Analysis of Customer Retention

Campaign Insights Unveiled

Data Understanding and Preparation

Description of the Dataset

Data Cleaning and Preparation Processes

Introduction to Customer Segmentation and Retention

Background and Context

Objectives of the Analysis

Scope and Methodology

Introduction to **Customer Segmentation** and **Retention**

Background and Context

- Supermarkets face intense competition and need to **retain customers** to maintain profitability.
- **Customer segmentation and retention analysis** help understand different customer groups and their behaviors, enabling personalized marketing strategies and improved engagement.

Objectives of the Analysis

- **Customer Profiling:** Develop detailed profiles of customer segments.
- **Retention Insights:** Analyze recency data to understand retention patterns.
- **Campaign Insights Unveiled:** eveal valuable insights from marketing campaigns to improve strategic decision-making.

Scope and Methodology

- **Stakeholders:** Business and Marketing Strategy Manager.
- Case study in a **four-season country**.
- Data Understanding and Preparation.
- Customer Segmentation Insights.
- Analysis of Customer Retention.
- Unveiling Campaign Insight.

Data Understanding and Preparation - Description of the Dataset

The data can be accessed at the following link: **Customer Supermarket**, which has 2240 rows and 29 columns.

Content Attributes

People	Products
• **ID:** Customer's unique identifier • **Year_Birth:** Customer's birth year • **Education:** Customer's education level • **Marital_Status:** Customer's marital status • **Income:** Customer's yearly household income • **Kidhome:** Number of children in customer's household • **Teenhome:** Number of teenagers in customer's household • **Dt_Customer:** Date of customer's enrollment with the company • **Recency:** Number of days since customer's last purchase • **Complain:** 1 if the customer complained in the last 2 years, 0 otherwise	• **MntWines:** Amount spent on wine in last 2 years • **MntFruits:** Amount spent on fruits in last 2 years • **MntMeatProducts:** Amount spent on meat in last 2 years • **MntFishProducts:** Amount spent on fish in last 2 years • **MntSweetProducts:** Amount spent on sweets in last 2 years • **MntGoldProds:** Amount spent on gold in last 2 years

Data Understanding and Preparation - Description of the Dataset

Content Attributes	
Promotion	**Place**
● **NumDealsPurchases:** Number of purchases made with a discount ● **AcceptedCmp1:** 1 if the customer accepted the offer in the 1st campaign, 0 otherwise ● **AcceptedCmp2:** 1 if the customer accepted the offer in the 2nd campaign, 0 otherwise ● **AcceptedCmp3:** 1 if the customer accepted the offer in the 3rd campaign, 0 otherwise ● **AcceptedCmp4:** 1 if the customer accepted the offer in the 4th campaign, 0 otherwise ● **AcceptedCmp5:** 1 if the customer accepted the offer in the 5th campaign, 0 otherwise ● **Response:** 1 if the customer accepted the offer in the last campaign, 0 otherwise	● **NumWebPurchases:** Number of purchases made through the company's website ● **NumCatalogPurchases:** Number of purchases made using a catalog ● **NumStorePurchases:** Number of purchases made directly in stores ● **NumWebVisitsMonth:** Number of visits to the company's website in the last month

Data Understanding and Preparation - Data Cleaning and Preparation Processes

1. Data Type Changes
2. Handling Missing Value
3. Checking Duplicate Update
4. Checking Outliers
5. Data Transformation:
   - Change Marital Status Column
   - Change Education Column
   - Making Age Column
   - Making Childern Column
   - Making Monthly Income Column
   - Making Monthly Spending Column
   - Rename Column

Data Understanding and Preparation - Data Cleaning and Preparation Processes

1 Data Type Changes

```
Data columns (total 29 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   ID                   2240 non-null   int64
 1   Year_Birth           2240 non-null   int64
 2   Education            2240 non-null   object
 3   Marital_Status       2240 non-null   object
 4   Income               2216 non-null   float64
 5   Kidhome              2240 non-null   int64
 6   Teenhome             2240 non-null   int64
 7   Dt_Customer          2240 non-null   object
 8   Recency              2240 non-null   int64
 9   MntWines             2240 non-null   int64
 10  MntFruits            2240 non-null   int64
 11  MntMeatProducts      2240 non-null   int64
 12  MntFishProducts      2240 non-null   int64
 13  MntSweetProducts     2240 non-null   int64
 14  MntGoldProds         2240 non-null   int64
 15  NumDealsPurchases    2240 non-null   int64
 16  NumWebPurchases      2240 non-null   int64
 17  NumCatalogPurchases  2240 non-null   int64
 18  NumStorePurchases    2240 non-null   int64
...
 27  Z_Revenue            2240 non-null   int64
 28  Response             2240 non-null   int64
dtypes: float64(1), int64(25), object(3)
memory usage: 507.6+ KB
```

```python
# Change Education and Marital_Status to categorical data types
df['Education'] = df['Education'].astype('category')
df['Marital_Status'] = df['Marital_Status'].astype('category')

# Change Dt_Customer to datetime data type
df['Dt_Customer'] = pd.to_datetime(df['Dt_Customer'], format='%d-%m-%Y')
```

```
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   ID              2240 non-null   int64
 1   Year_Birth      2240 non-null   int64
 2   Education       2240 non-null   category
 3   Marital_Status  2240 non-null   category
 4   Income          2216 non-null   float64
 5   Kidhome         2240 non-null   int64
 6   Teenhome        2240 non-null   int64
 7   Dt_Customer     2240 non-null   datetime64[ns]
 8   Recency         2240 non-null   int64
 9   MntWines        2240 non-null   int64
```

Data Understanding and Preparation - Data Cleaning and Preparation Processes

2 Handling Missing Value

```
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   ID              2240 non-null   int64
 1   Year_Birth      2240 non-null   int64
 2   Education       2240 non-null   category
 3   Marital_Status  2240 non-null   category
 4   Income          2216 non-null   float64
 5   Kidhome         2240 non-null   int64
 6   Teenhome        2240 non-null   int64
```

```
df.isna().sum()/df.shape[0]*100
```

```
Marital_Status    0.000000
Income            1.071429
Kidhome           0.000000
```

Missing Value < 5%

```
df_new = df.copy()
df_new['Income'].fillna(df['Income'].mean(), inplace=True)
df_new[['Income']].isna().sum()
```

```
 3   Marital_Status   2240 non-null   object
 4   Income           2240 non-null   float64
 5   Kidhome          2240 non-null   int64
```

Data Understanding and Preparation - Data Cleaning and Preparation Processes

3 Checking Duplicate Update

```python
# Set the display option for better readability
pd.set_option('display.max_colwidth', None)

# Data in each column
listItem = []
for col in df_new.columns:
    listItem.append([col, df_new[col].nunique(), df_new[col].unique()[:5]])  # Displaying first 5 unique values as a sample

# Create a DataFrame for column uniqueness
tabel1Desc = pd.DataFrame(columns=['Column Name', 'Number of Unique', 'Unique Sample'], data=listItem)

# Check for duplicate rows
duplicate_rows = df_new.duplicated()
num_duplicates = duplicate_rows.sum()

# Display the uniqueness table
print("Uniqueness Table:")
print(tabel1Desc)

# Report on duplicate rows
print("\nNumber of duplicate rows:", num_duplicates)

# Optionally, display the duplicate rows if any
if num_duplicates > 0:
    print("\nDuplicate Rows:")
    print(df_new[duplicate_rows])
else:
    print("\nNo duplicate rows found")
```

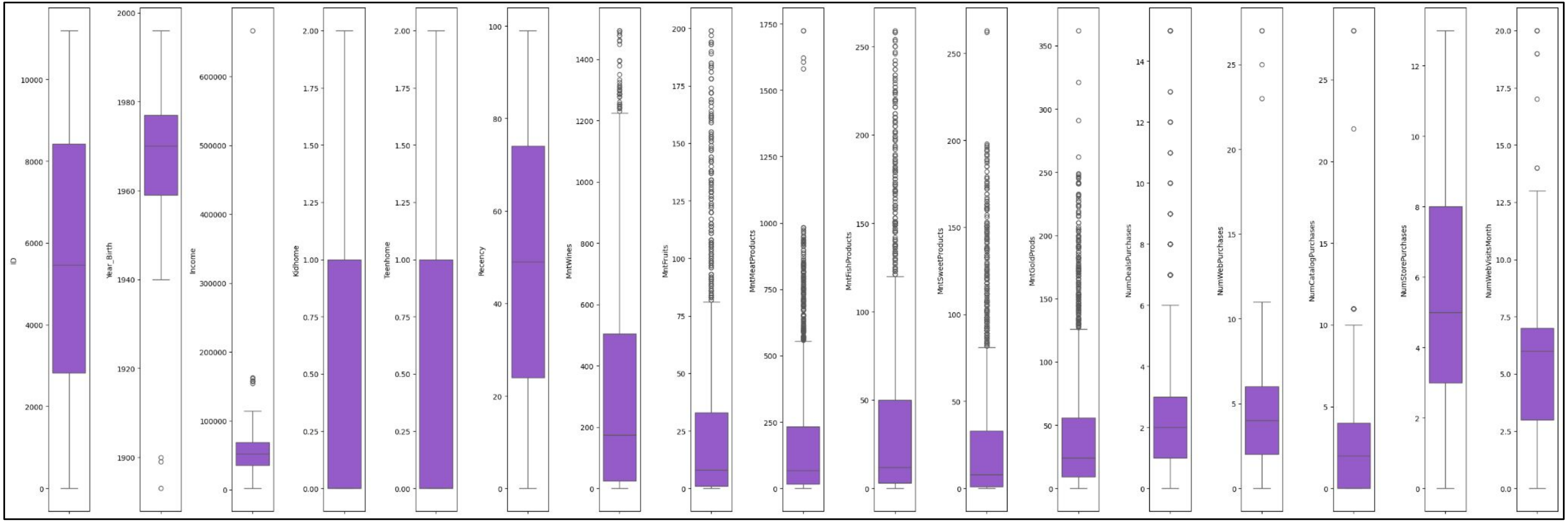
➡

```
Number of duplicate rows: 0

No duplicate rows found
```

Data Understanding and Preparation - Data Cleaning and Preparation Processes

4 Checking Outliers



Data Understanding and Preparation - Data Cleaning and Preparation Processes

5 Data Transformation

```
df_new.Marital_Status = df.Marital_Status.replace({'Together': 'Relationship',
                                                   'Married': 'Relationship',
                                                   'Divorced': 'Single',
                                                   'Widow': 'Single',
                                                   'Alone': 'Single',
                                                   'Absurd': 'Single',
                                                   'YOLO': 'Single'})
```

Change Marital Status

```
df_new["Education"] = df_new["Education"].replace({"Graduation": "Postgraduate",
                                                   "PhD": "Postgraduate",
                                                   "Master": "Postgraduate",
                                                   "2n Cycle": "Undergraduate",
                                                   "Basic": "Undergraduate"})
```

Change Education Column

```
# Calculate the age of each customer
df_new['Age'] = 2014 - df['Year_Birth']
```

Making Age Column

```
# Calculate the number of children
df_new['Children'] = df_new['Kidhome'] + df_new['Teenhome']
```

Making Children Column

```
# Adding the 'MonthlyIncome' column by dividing the 'Income' column by 12
df_new['Monthly Income'] = df_new['Income'] / 12
```

Making Monthly Income Column

Data Understanding and Preparation - Data Cleaning and Preparation Processes

5 Data Transformation

```
# Calculate total spent per customer over two years
df_new['Total Spending'] = df_new['MntWines'] + df_new['MntFruits'] + df_new['MntMeatProducts'] + df_new['MntFishProducts'] + df_new['MntSweetProducts'] + df_new['MntGoldProds']

# Calculate monthly spending by dividing the total spending by 24 (two years)
df_new['Monthly Spending'] = df_new['Total Spending'] / 24
```

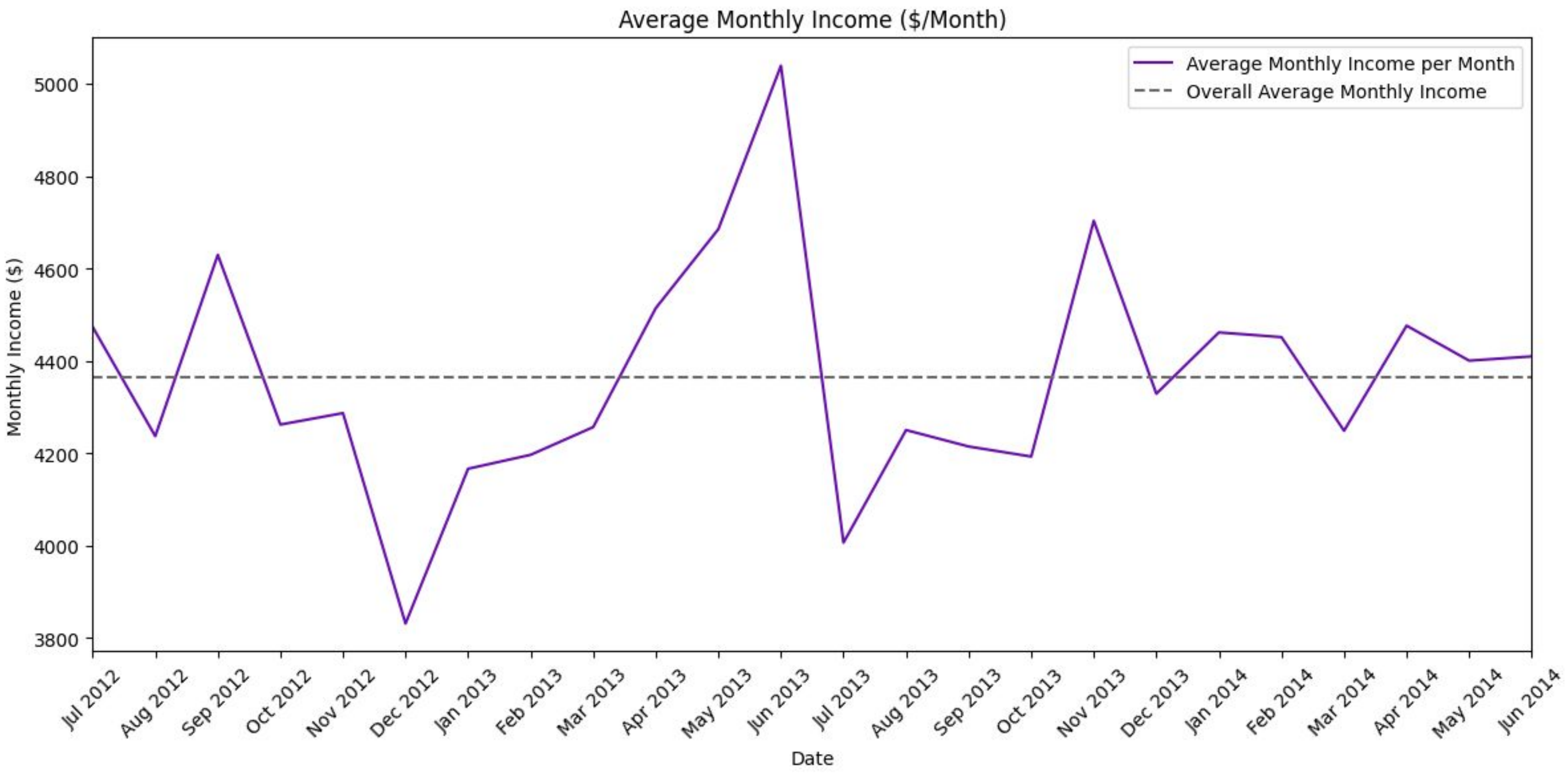
Making Monthly Spending Column

```
df_new.rename(columns = {'MntWines':'Wines',
                          'MntFruits':'Fruits',
                          'MntMeatProducts' : 'Meats',
                          'MntFishProducts' : 'Fish',
                          'MntSweetProducts' : 'Sweets',
                          'MntGoldProds' : 'Gold',
                          'NumDealsPurchases' : 'Deals Purchase',
                          'NumWebPurchases' : 'Web Purchase',
                          'NumCatalogPurchases' : 'Catalog Purchase',
                          'NumStorePurchases' : 'Store Purchase',
                          'NumWebVisitsMonth' : 'Web Visit',
                           'AcceptedCmp3' : 'Campaign 3',
                           'AcceptedCmp4' : 'Campaign 4',
                           'AcceptedCmp5' : 'Campaign 5',
                           'AcceptedCmp1' : 'Campaign 1',
                           'AcceptedCmp2' : 'Campaign 2'}, inplace = True)
```

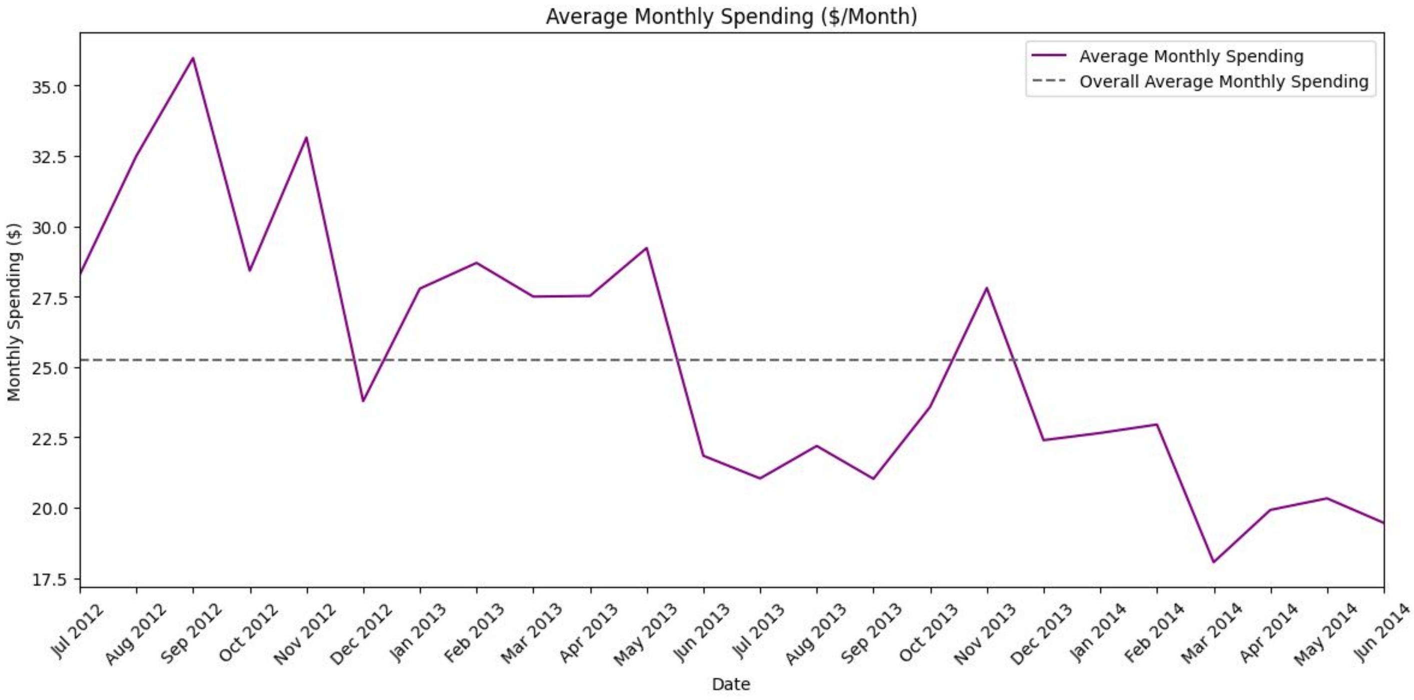
Rename Column

Data Analysis - Customer Segmentation Insights

Comparing Monthly Income and Spending



Data Analysis - **Customer Segmentation Insights**



Highlight: The declining average monthly spending approaching 2014 and beyond, alongside the insight that the **average monthly spending percentage of 0.58%** indicates a potentially low level of consumer expenditure relative to their income, presents an opportunity for us to potentially boost revenue. This opportunity is supported by the stable average monthly income, which shows no significant fluctuations.

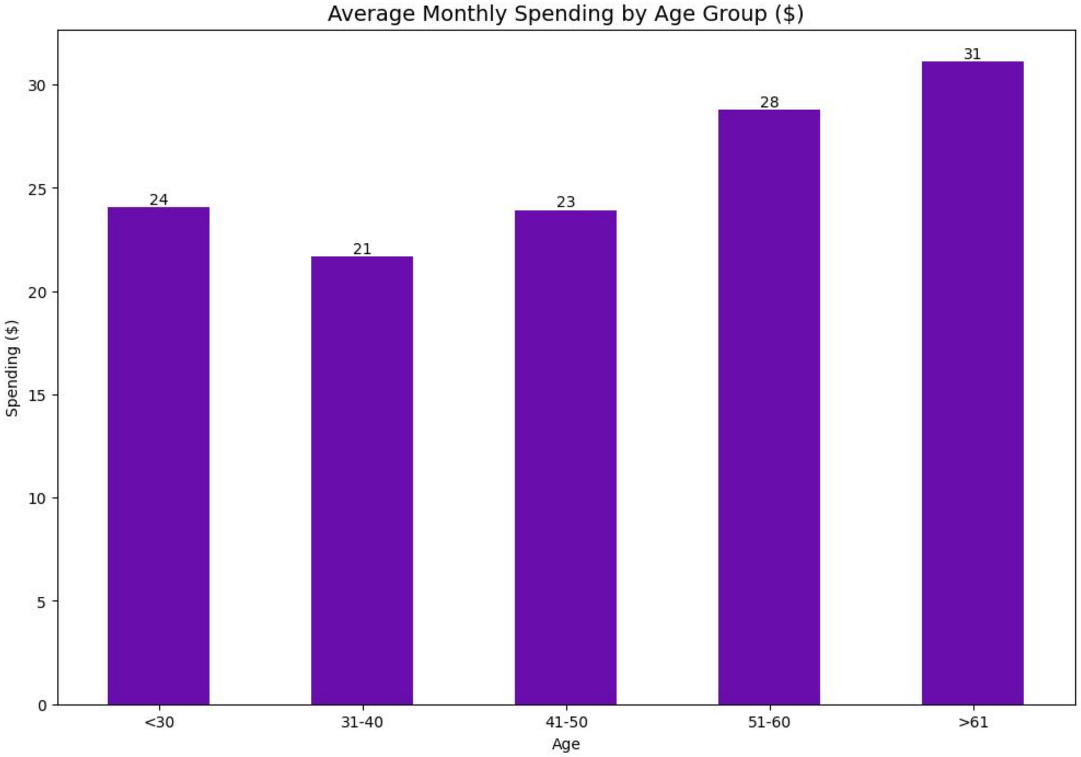
Data Analysis - Customer Segmentation Insights

Action Plan

1. Diversification of Product Offerings
2. Behavioral Economics Approach
3. Segment-Specific Strategies
4. Promotion of High-Margin Products
5. Data-Driven Personalization
6. Cross-Selling and Upselling Initiatives
7. Continuous Monitoring and Optimization
8. Collaboration with Suppliers and Partners

Data Analysis - Customer Segmentation Insights

Analyzing Consumer Spending Patterns Across Different Age Groups



Highlight:

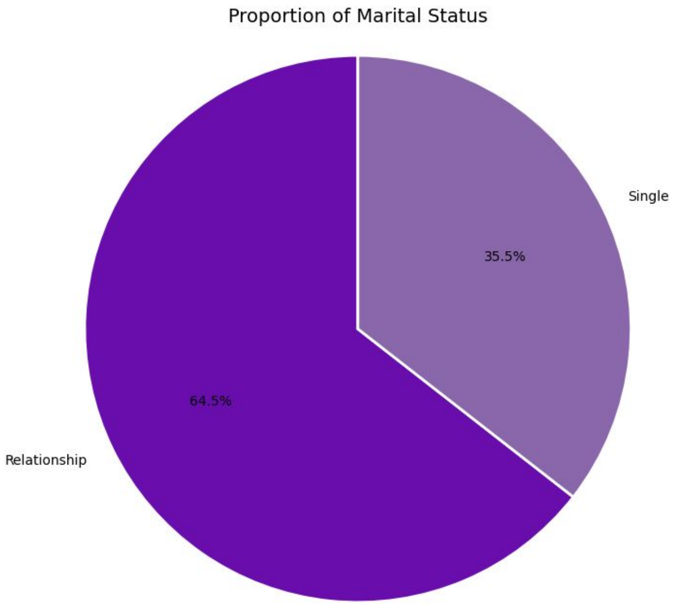
- The average spending increases progressively with age.
- Customers aged greater than 61 (>61) exhibit the highest average spending.

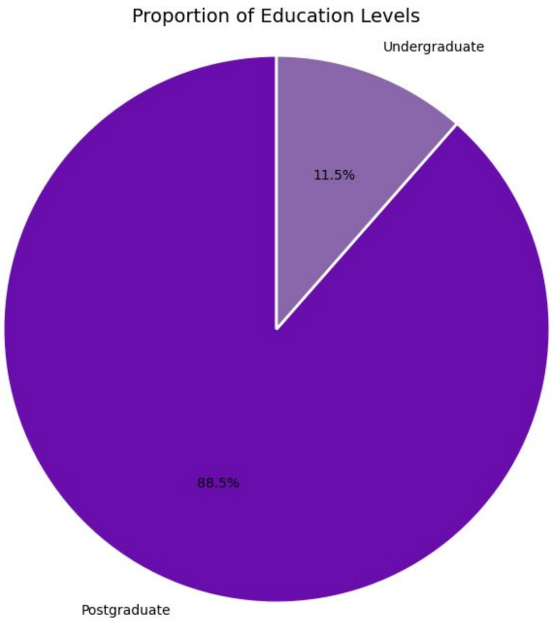
Action Plan:

- Targeted Marketing for Older Age Groups.
- Engagement Strategies for Middle-Aged Groups.
- Attracting Younger Customers.
- Personalized Customer Experiences.

Data Analysis - Customer Segmentation Insights

Understanding Marital Status and Education Levels Composition in Customer Segmentation

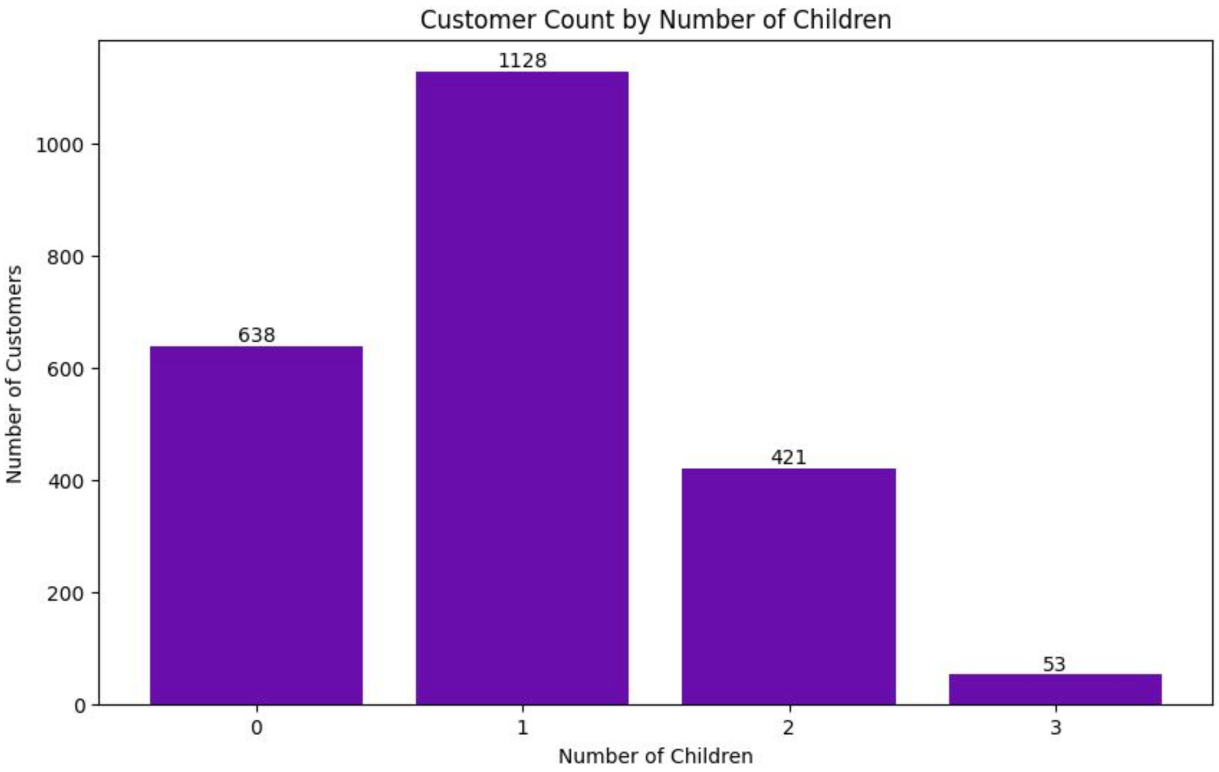




Highlight: The majority of customers are in the 'Relationship' category and have 'Postgraduate' education levels. This suggests that a significant portion of the customer base consists of educated couples or individuals in stable relationships. Marketing strategies should prioritize this dominant group by offering premium products or services that cater to their sophisticated tastes and family-oriented needs.

Data Analysis - Customer Segmentation Insights

Holiday Magic: Winning Over Customers with Strategic Family Packages

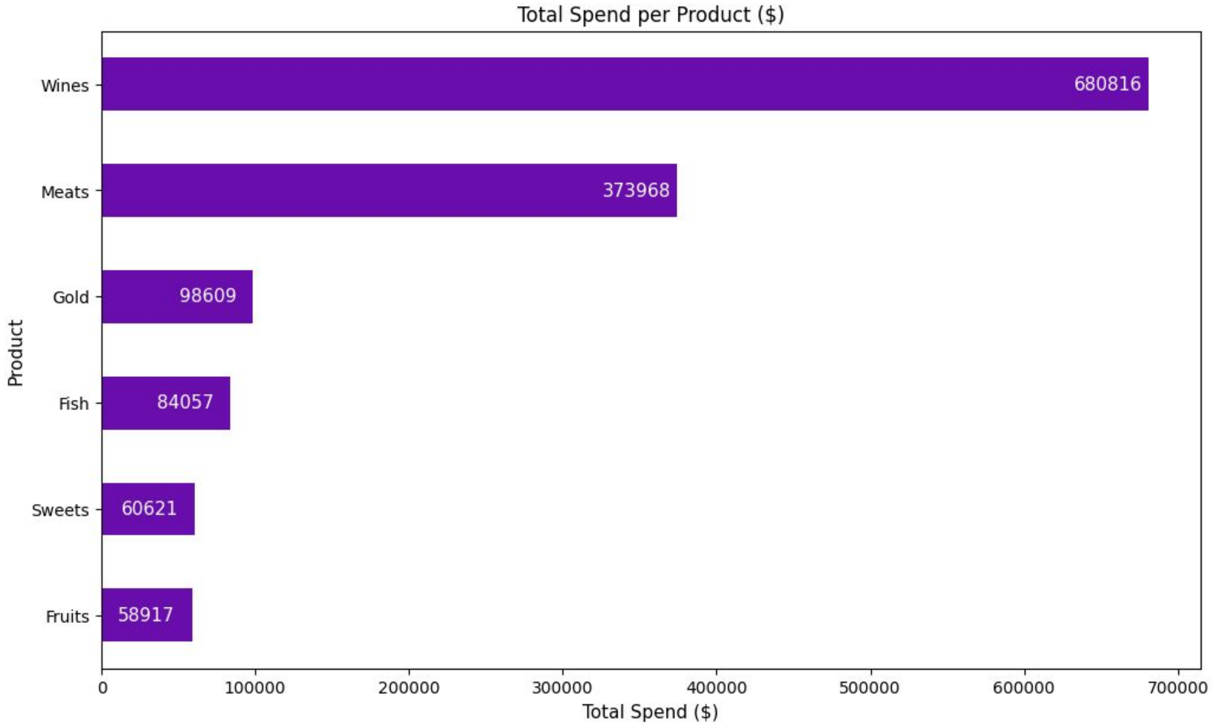


Highlight: The results indicate that customers with one child outnumber those without children, with two children, and with three children. This suggests that the group of customers with one child holds significant influence in purchasing patterns or consumer preferences within the supermarket dataset. Such analysis provides insights for marketing teams to tailor more specific strategies or adjust product offerings that better suit the preferences of this customer segment.

Furthermore, this data highlights the potential for implementing holiday package promotions. Marketing teams could develop targeted promotional campaigns, such as offering holiday packages to popular destinations as part of a purchase incentive, particularly for customers with one child who form a significant segment of the customer base.

Data Analysis - Customer Segmentation Insights

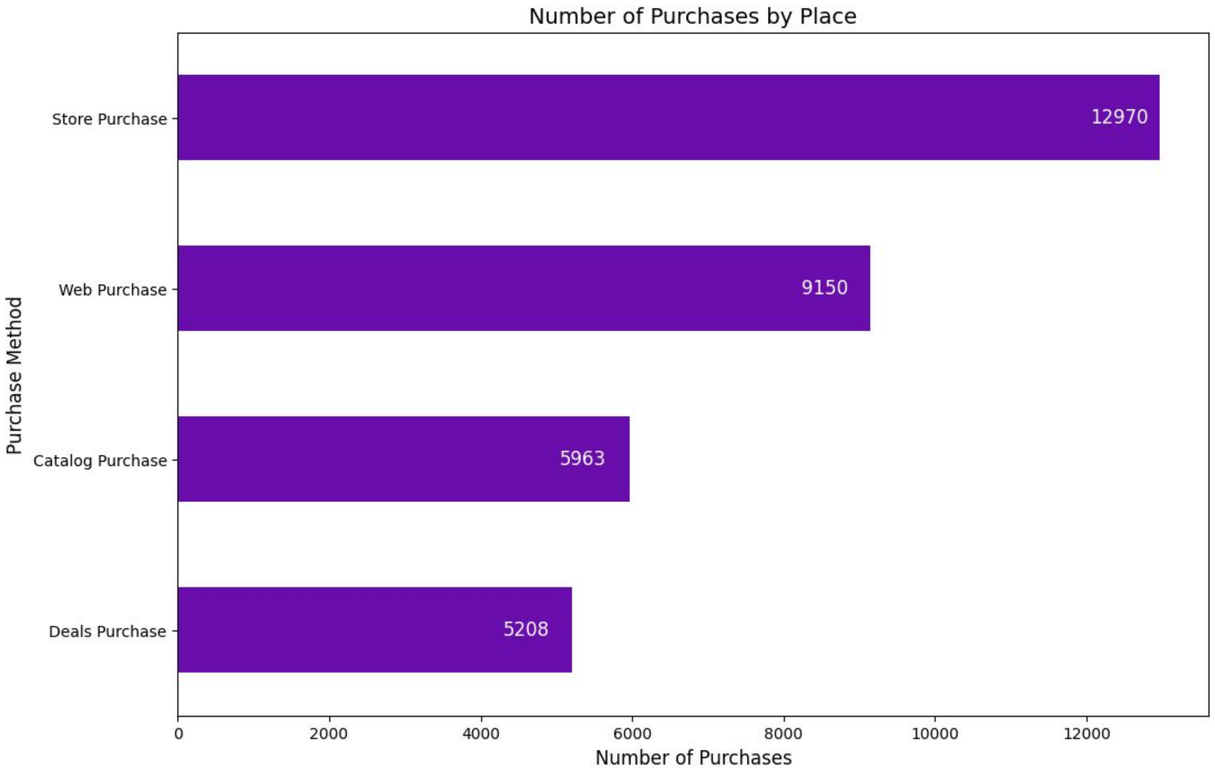
Targeted Treats: Maximizing Profit with Family-Focused Products



Highlight: Given the low spending on family-related products, there's a clear opportunity to develop and promote items that cater to family needs, aligning with the influence of customers with one child to increase spending in these areas.

Data Analysis - Customer Segmentation Insights

Beyond the Checkout: Unveiling Opportunities in Store and Web Purchases



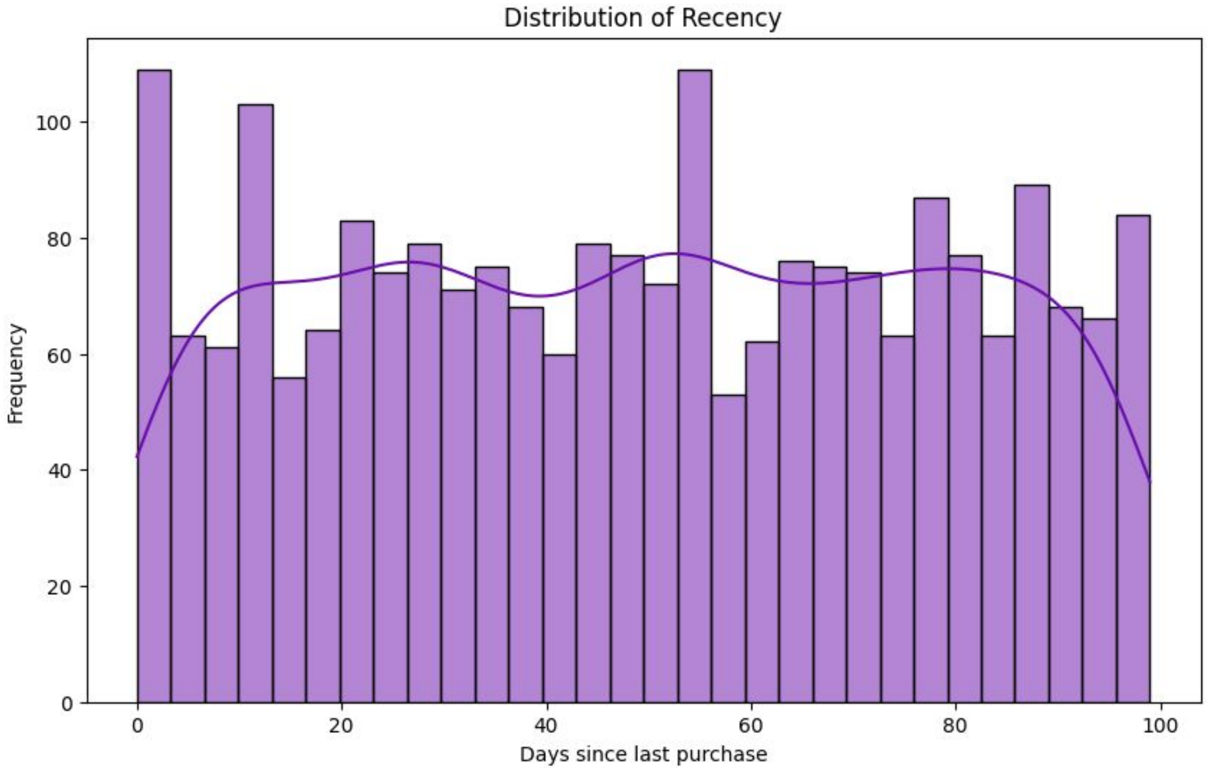
Highlight: Store purchases are the most common, with web purchases growing in importance. Catalog and deals purchases are less popular, suggesting opportunities for targeted marketing.

Action Plan:

- Enhance Online Presence and User Experience.
- Boost Marketing Campaigns.
- Onsite Branding and In-Store Experience.
- Integrate Online and Offline Strategies.
- Customer Engagement and Feedback.

Data Analysis - Analysis of Customer Retention

Recency Radar: Navigating Peaks in Customer Engagement



Highlight: The distribution of Recency in the data shows distinct peaks on specific days (around 0, 10, 30, 50, and 90 days). These high-frequency days indicate increased purchasing activity from customers, although the overall distribution of Recency does not follow a normal pattern.

Data Analysis - Customer Segmentation Insights

Action Plan

1 On-site Branding

- Target Day 0: Customers who have just made a purchase. Enhance in-store branding to encourage impulse purchases.
- Target Day 10: Customers who tend to return within 10 days. Display special promotional banners in-store to capture their attention.

2 Email Marketing Campaigns

- Day 30 and Day 50: Send promotional or reminder emails a few days before customers typically return, incentivizing them to shop again with special offers or discounts.

3 Social Media Promotions

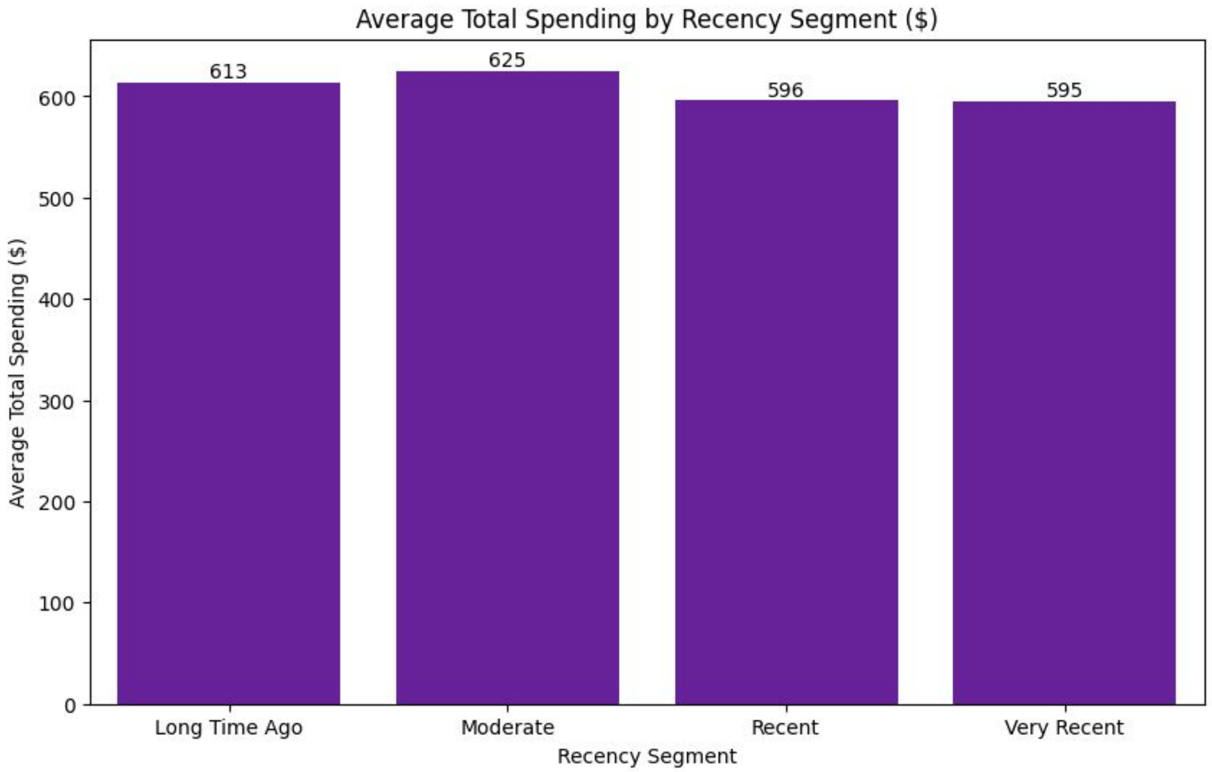
- Day 90: Use social media to remind and promote products to customers who tend to return after a longer period.

4 Store Layout Optimization

- General: Optimize store layout to maximize exposure to products frequently purchased by customers with high Recency on these specific days. With this action plan, you can leverage Recency data to enhance customer retention and drive more in-store purchases.

Data Analysis - **Analysis of Customer Retention**

Segment Spotlight: Unveiling Recency-Based Spending

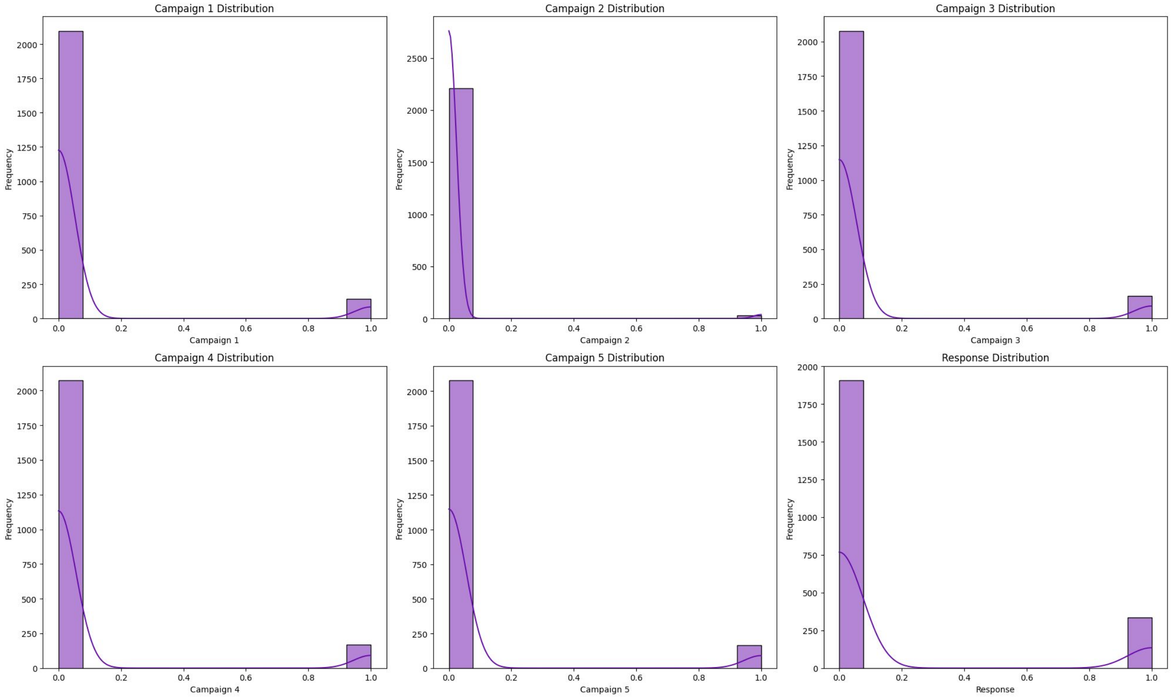


Highlight: While the distribution of customers based on Recency segments is fairly balanced, there is no significant difference observed in the average total spending across segments.

Action Plan: Focus on additional promotional strategies or incentives for customers in the 'Very Recent' and 'Moderate' segments to enhance retention and repeat purchases. Given the proximity in behavior, the impact on the 'Recent' segment from strategies targeting 'Moderate' and 'Very Recent' segments could help optimize promotional budgets effectively. Additionally, consider further personalization for the 'Long Time Ago' segment.

Data Analysis - Campaign Insights Unveiled

Campaign Data Distribution and Response Analysis



Campaign 1 normality test: stat = 1608.7064215330126, p-value = 0.0
Campaign 2 normality test: stat = 3155.124172595476, p-value = 0.0
Campaign 3 normality test: stat = 1485.6762386704095, p-value = 0.0
Campaign 4 normality test: stat = 1461.5140249268186, p-value = 0.0
Campaign 5 normality test: stat = 1485.67623867041, p-value = 0.0
Response normality test: stat = 755.7190160561258, p-value = 7.901339609771405e-165

Highlight: The results of the normality tests indicate that all campaign columns (Campaign 1, Campaign 2, Campaign 3, Campaign 4, Campaign 5) and the response column have very low p-values (close to 0.0), meaning that the data in these columns are not normally distributed. Therefore, non-parametric tests such as the Chi-Square test are more appropriate.

Based on these results, the use of the Chi-Square test is justified because:

- The data are not normally distributed.
- The data are categorical.

All p-values are very low (well below the significance level of 0.05), meaning we can conclude that there is a statistically significant relationship between each campaign and the response. Thus, the initial use of the Chi-Square test was appropriate, and the results indicate a significant relationship between the variables.

Data Analysis - Campaign Insights Unveiled

Statistical Correlation Between Campaigns and Response

```python
# Columns used
campaign_columns = ['Campaign 1', 'Campaign 2', 'Campaign 3', 'Campaign 4', 'Campaign 5', 'Response']

# Function to calculate and display Chi-Square test results
def chi2_test(col1, col2):
    contingency_table = pd.crosstab(df_new[col1], df_new[col2])
    chi2, p, dof, expected = chi2_contingency(contingency_table)
    return chi2, p

# Test correlation between campaigns and response
chi2_results = []
for col in campaign_columns[:-1]:
    chi2, p = chi2_test(col, 'Response')
    chi2_results.append({'Campaign': col, 'Chi2': chi2, 'p-value': p})

# Create DataFrame to display results
chi2_df = pd.DataFrame(chi2_results)

# Display the DataFrame
print(chi2_df)
```

	Campaign	Chi2	p-value
0	Campaign 1	190.241783	2.815498e-43
1	Campaign 2	60.130297	8.878118e-15
2	Campaign 3	142.074884	9.364996e-33
3	Campaign 4	68.312456	1.395368e-16
4	Campaign 5	235.467749	3.828045e-53

Highlight:

- All campaign offers show a statistically significant relationship with the customer's response to the last campaign. This indicates that customers who responded positively to earlier campaigns are more likely to respond positively to future campaigns.
- Campaign 5 has the highest chi-square value, indicating the strongest association with customer response.
- Possible Reasons for the Strong Association:
  - More Attractive Offer.
  - Increased Customer Trust.
  - Better Customer Segmentation.
  - Improved Communication Channels.
  - Optimal Timing.

Action Plan:

- Leverage Past Campaign Data.
- Maximize Web Platform Potential.
- Campaign Optimization.
- Focus on High-Value Customers.

Conclusion and Future Work - **Summary of Key Findings**

1 Customer Spending Behavior

Customers allocate only 0.58% of their monthly income on average to supermarket purchases, indicating potential for increased spending with targeted strategies.

2 Marital Status and Education

Majority of customers are in relationships and have postgraduate education, suggesting a focus on premium products and family-oriented offerings.

3 Product Preferences

High spending is observed on wines, meats, and gold, with lower spending on fish, sweets, and fruits, highlighting opportunities for product diversification.

4 Purchase Methods

Store purchases dominate, but there's growing importance of web purchases, suggesting a need to enhance online marketing and sales strategies.

5 Campaign Effectiveness

All campaigns (Campaigns 1-5) show significant associations with customer response, with Campaign 5 exhibiting the strongest impact.

Conclusion and Future Work - **Final Strategic Recommendations**

1 Diversification of Product Offerings

Introduce new products and promotions tailored to family needs, including holiday package promotions to increase average transaction value.

2 Enhanced Digital Marketing

Optimize online platforms with personalized offers and targeted promotions, particularly for high-spending segments like customers with families.

3 Segment-Specific Campaigns

Develop marketing campaigns that resonate with different demographic segments (e.g., couples, families, singles) based on spending behaviors and preferences.

4 Cross-Channel Integration

Integrate online and offline channels to provide a seamless customer experience, promoting family-oriented products and holiday packages consistently.

Conclusion and Future Work - **Final Strategic Recommendations**

5 Loyalty Programs and Incentives

Implement loyalty programs with rewards for repeat purchases and high-value customer segments, including incentives for purchasing holiday packages and family-oriented products.

Thank **You !**

Capstone Project #2 | Inggar Gumintang | JCDSOL - 014 - 2