2Market Technical Report

By: Annora Ng

Introduction

2Market, a global supermarket with online and in-store operations, seeks actionable insights into **customer demographics, advertising channel effectiveness,** and **product preferences.** These insights aim to optimize resource allocation, enhance advertising ROI, and tailor product offerings to customer needs, ultimately driving **revenue growth** and **operational efficiency**. This analysis assumes the dataset reflects data as of 2024, with customer ages calculated using birth year. It also assumes data accuracy, representative advertising campaign data, stable customer purchasing behavior, and equal split of conversions between in-store and online purchases. The report provides evidence-based recommendations to support **strategic decision-making** and **future business growth**.

Analytical Approach

Tools

- **Excel:** Used for data cleaning, initial exploration, and creating new columns to facilitate analysis. Excel's flexibility with formulas and its visual exploration tools allowed for rapid manipulation and validation of the dataset.
- **SQL:** Employed to join datasets, aggregate data, and verify insights. SQL's scalability ensured accuracy in handling large datasets, and its querying capabilities allowed for precise validation of Tableau outputs.
- **Tableau:** Utilized for creating interactive dashboards and visualizations. Tableau's advanced visualization options enabled clear communication of insights tailored to stakeholders.

Data Cleaning in Excel

Key Steps:

- **Age Calculation:** Added an Age column calculated as **2024 - Year_Birth** to categorize customers into age groups for segmentation.
- **Marital Status Grouping:** Consolidated marital status values into logical categories as shown below:

Grouped Marital Status	Marital Status
In a relationship	Married
	Together
Single	Alone
	Divorced
	Single
	Widow
Unknown	YOLO
	Absurd

- **Income Range:** Consolidated Incomes into Income Range column (e.g., <30k, 30k-50k) to simplify high-level trend visualization.

- **Data Standardization:** Used Excel tools to correct blank spaces, standardize column data types, and address typos through the Find & Replace feature, ensuring consistency and reducing potential errors.

Outlier and Duplicate Handling

- Removed three rows where customer ages exceeded 100, as these were logically invalid.
- Retained duplicate rows, as no records were fully identical across all columns, and further investigation was not feasible without stakeholder input.
- Kept an outlier ($666,666) in the Income column to preserve data integrity, avoiding potential bias from arbitrary removal.

Data Preparation and Joining

- Used SQL to perform an inner join on the two datasets by the common **id** column, ensuring only matched records were included in the analysis.
- Verified Tableau visualizations by cross-checking key aggregations in SQL, such as:

```sql
-- Check Total Sales per Country
Select country, SUM(total_spend) as total_sales
From cleanedmarketingdata
Group By country
Order by SUM(total_spend) DESC;
```

Figure 1: Query to check Total Sales Per Country

```sql
-- Check Number of Customers by Educational Level
Select education, COUNT(ID) as number_customers
From cleanedmarketingdata
Group by education
Order by COUNT(ID) DESC;
```

Figure 2: Query to check Number of Customers by Educational Level

```sql
-- Check Total Conversions per Country
Select country, SUM(bulkmail_ad + twitter_ad + instagram_ad + facebook_ad + brochure_ad) as total_conversions
From cleanedmarketingdata INNER JOIN ad_data USING (id)
Group by country
Order by SUM(bulkmail_ad + twitter_ad + instagram_ad + facebook_ad + brochure_ad) DESC;
```

Figure 3: Query to check Total Conversions per Country

Challenges

Limited metadata and stakeholder input constrained the resolution of data anomalies, necessitating cautious assumptions, namely:

- **Data Integrity:** Retaining duplicates and outliers was a calculated decision to preserve data integrity, as the lack of stakeholder clarification prevented confident exclusion or adjustments.
- **Ambiguous Values:** Marital status categories such as “YOLO” and “Absurd” introduced ambiguity. These were carefully grouped into logical segments to ensure meaningful analysis.
- **Missing Product Indicator:** The dataset lacked a direct measure for total product purchases. To address this, the total was estimated by summing the number of web buys and walk-in purchases.

- **Missing Impression Data:** Impression counts were unavailable, leading to the assumption that each customer ID had one impression per advertising channel.

Dashboard Design

The dashboards were created to provide actionable insights for key stakeholders, with a focus on clarity, interactivity, and accessibility.

Customer Demographic Overview

- **Purpose:** Supports **Marketing Managers** and **Customer Insights Teams** in understanding customer profiles to target key segments effectively.
- **Design Choices:**
    - Pie, bar, and map charts simplify comparison across demographics like marital status, income, and geography.
    - Interactive filters (Eg: Age Group, Income Range) facilitate granular analysis to refine segmentation and targeting strategies.

Product Sales and Demographic Variation

- **Purpose:** Designed for **Product Managers** and **Marketing Managers** to assess product preferences and spending patterns.
- **Design Choices:**
    - Bar charts display top-performing product categories by sales, enabling resource allocation toward high-demand items.
    - **Filters** allow Marketing Managers to identify key customer segments driving product sales, enabling targeted campaigns and tailored promotions to boost engagement and revenue.

Advertising Channel Effectiveness and Customer Engagement

- **Purpose:** Provides **Marketing Managers** with a comprehensive view of ad performance and customer responsiveness to optimize channel investments.
- **Design Choices:**
    - Bar and stacked bar charts highlight conversion rates by channel and income range.
    - Bubble charts visualize campaign response rates by country.
    - Filters enable analysis by demographics, helping refine channel strategy.

Consistency Across Dashboards

- Uniform color palettes, standardized age bins, and consistent income ranges ensure readability, reduce cognitive load, and align insights with business objectives.
- The design emphasizes actionable insights to empower stakeholders to drive revenue and refine strategies effectively.

Patterns, Trends and Insights

Customer Demographics

- **65%** of customers are in a relationship.
- **97%** of customers hold a Graduation Level Education or higher.
- Median customer age is **53**.
- **49%** of customers reside in Spain.
- Main Income Groups earns **$30k-70k** annually, followed by **$70k-100k**

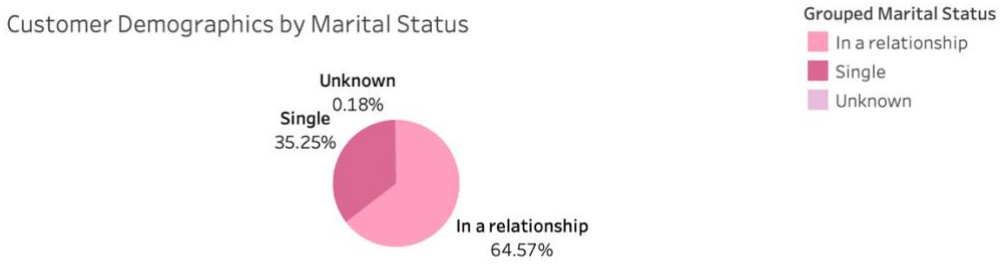

Figure 4: Customer Demographics by Marital Status

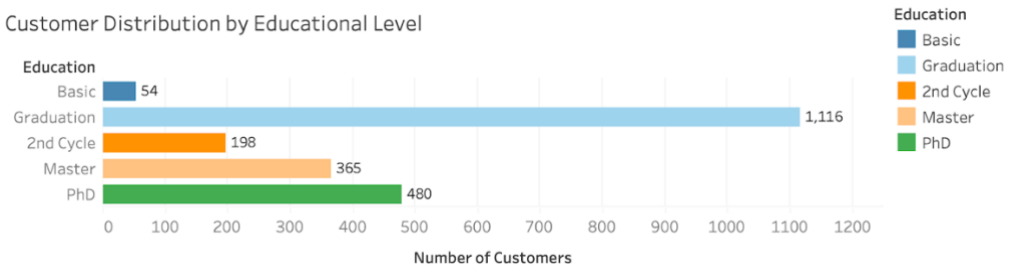

Figure 5: Customer Distribution by Educational Level

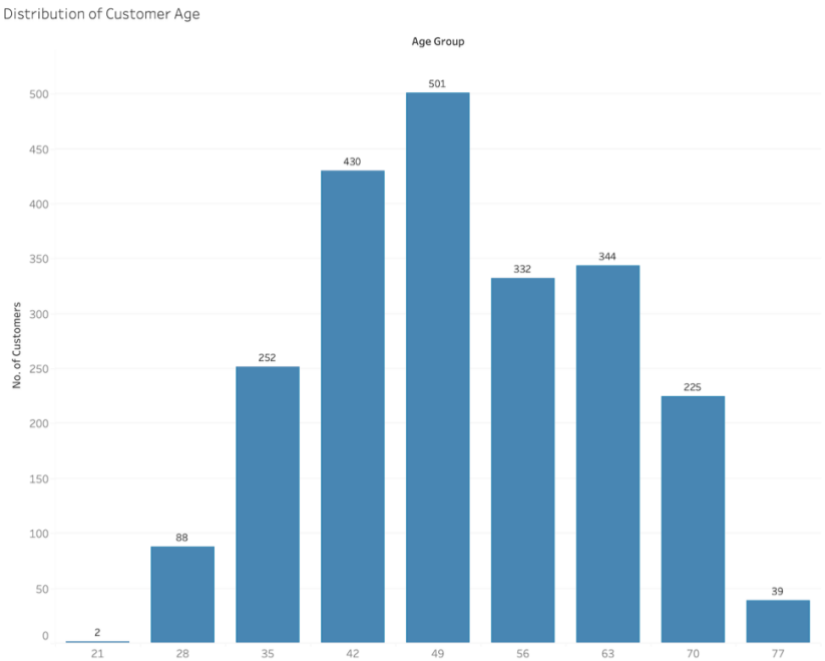

Figure 6: Distribution of Customer Age

Figure 7: Income Range Breakdown of Customers

Figure 8: Geographic Distribution of Customers

Product Performance

- **Alcohol** and **Meat** dominate sales across all demographics.
- **Alcohol** sales nearly double for customers in a relationship compared to singles.

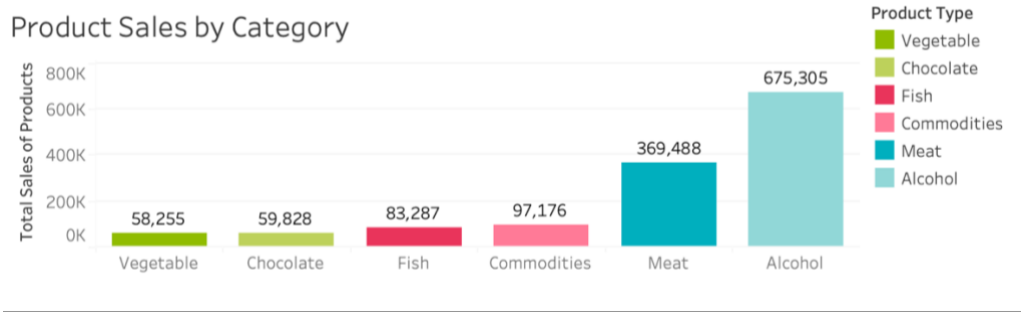

Figure 9: Product Sales by Category

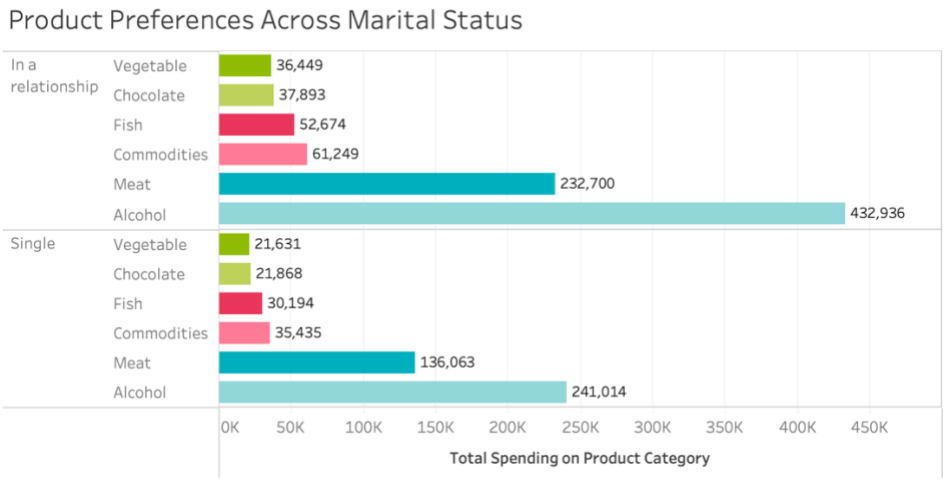

Figure 10: Product Sales Across Marital Status

Relative Spending on Products by Age Group

Age Group	Vegetable	Chocolate	Fish	Commodities	Meat	Alcohol
21 - 27	18	25	18	58	16	3
28 - 34	2,928	2,730	4,215	4,576	21,540	25,692
35 - 41	7,014	7,224	9,367	10,258	44,972	61,771
42 - 48	10,564	10,997	15,604	17,217	59,407	103,785
49 - 55	11,499	11,797	15,803	19,649	73,565	139,787
56 - 62	9,056	8,855	11,784	15,725	53,875	115,585
63 - 69	9,796	10,396	15,025	16,093	63,361	124,417
70 - 76	5,804	6,188	8,635	11,919	43,085	84,485
77 - 83	1,576	1,616	2,836	1,681	9,667	19,780

Figure 11: Relative Spending on Products by Age Group

Customer Spending

- 51% of total sales are generated by customers in the **$70k-100k** income group, indicating a highly profitable segment.

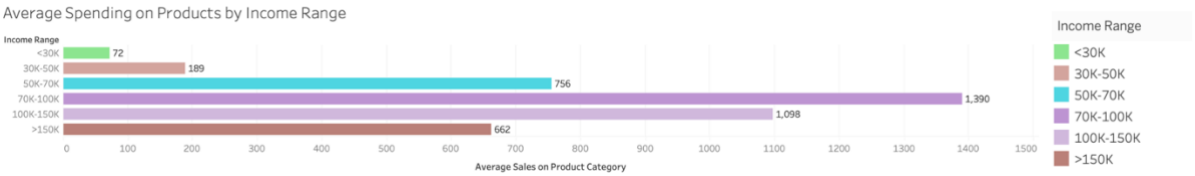

Figure 12: Average Spending on Products by Income Range

Advertising Channel Effectiveness

- **Twitter Ads** outperform other channels with the highest overall conversion rates. However, **Facebook and Instagram Ads** perform the best amongst high-income groups.
- **Montenegro** has the highest campaign response rate, driven solely by Bulkmail Ads, but the low sample size (three customers) limits reliability.

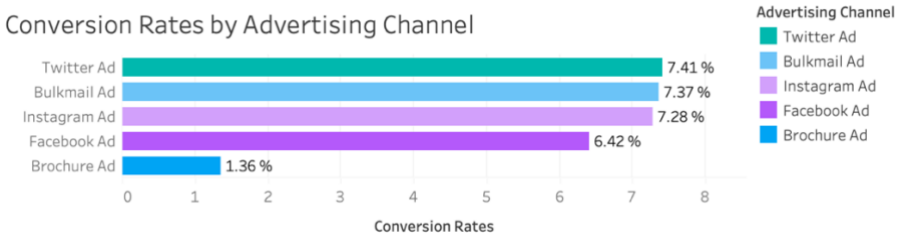

Figure 13: Conversion Rates by Advertising Channel

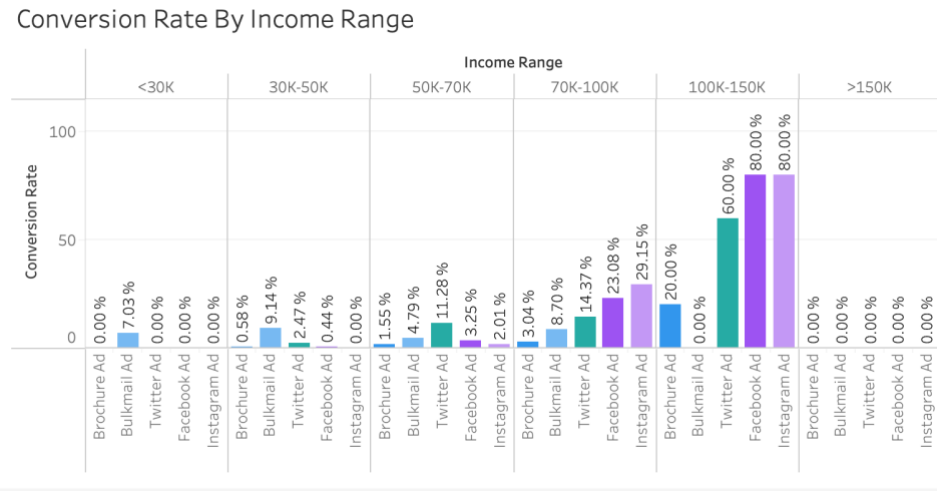

Figure 14: Conversion Rates by Income Range

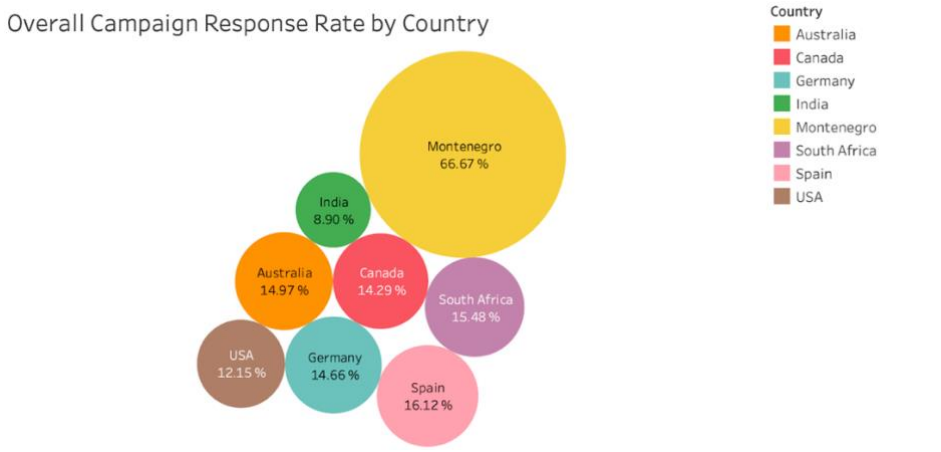

Figure 15: Overall Campaign Response Rate by Country

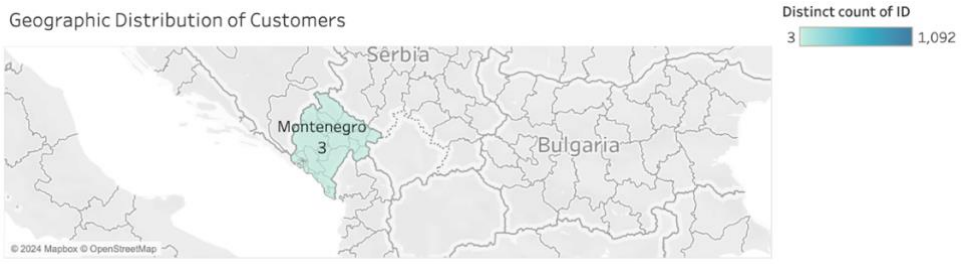

Figure 16: Geographic Distribution of Customers: Montenegro has 3 records

Country-Level Insights

- Alcohol remains the top-performing category across all countries.

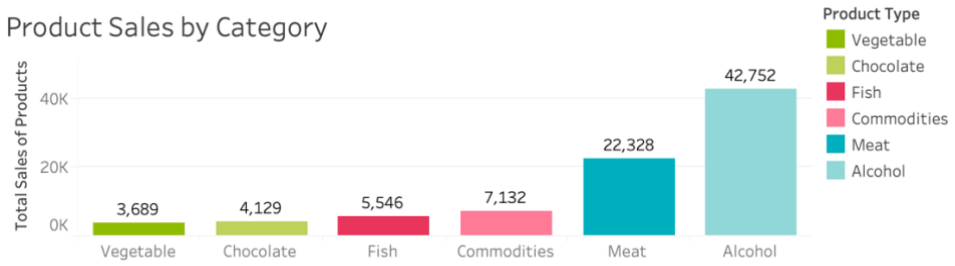

Figure 17: Example 1 - Product Sales by Category in Australia

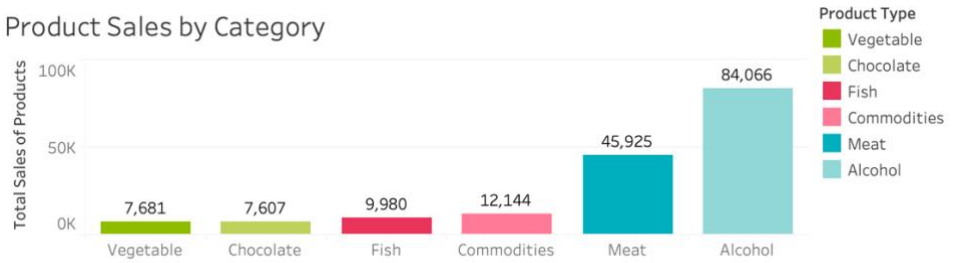

Figure 18: Example 2 - Product Sales by Category in Canada

In-Store vs Online Purchases

- In-store sales consistently outperform online purchases across income ranges and countries, except in Montenegro, likely due to its small sample size.

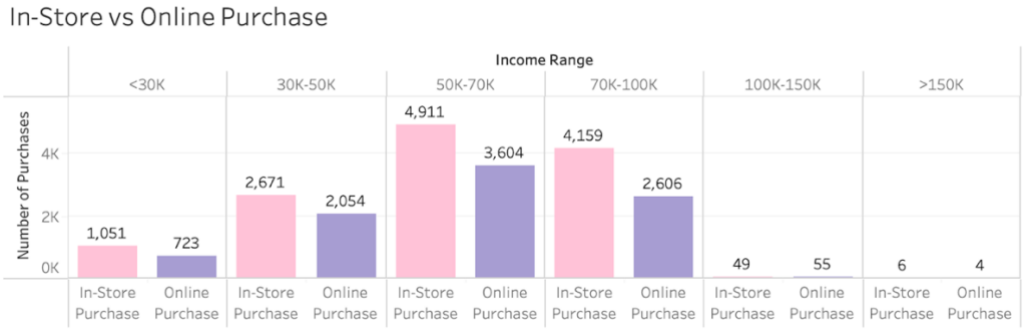

Figure 19: In-Store vs Online Purchase across all countries

Figure 20: In-Store vs Online Purchase for Montenegro

Conclusions and Recommendations

This analysis highlights key insights to inform 2Market's strategy.

1. The most profitable customer segment is the group of customers who are in a relationship, is within the $70k-100k Income Group, and is highly educated.
2. Whilst Twitter Ads perform well overall, Instagram Ads and Facebook Ads perform better for higher income groups.
3. Alcohol and Meat are the best-selling product categories, and In-Store sales perform better than Online Sales.

Recommendations

Channel and Marketing Strategy:

Target customers in a relationship with High Income and Education Levels

- Key Marketing Messaging: Highlight product quality, premium value etc
- Implement 2 target campaigns per quarter focused on this group

<u>Reallocation of Advertising Channel budget for efficient spend:</u>

Proposed New Budget Allocation		
Instagram Ads	35%	Leverage the highest conversion rates
Facebook Ads	25%	Strong performance in the target group
Twitter Ads	20%	Overall strong conversion rates
Bulkmail Ads	10%	Perform well for specific demographics
Reserved Budget to test new Digital Campaigns	10%	Explore other high-potential advertising opportunities

Figure 21: Reallocating Marketing Budget on Advertising Channels

Product Strategy:

- Allocate **50% of inventory resources** to Alcohol and **30% to Meat**, the top-performing categories.
- Launch **in-store bundled promotions** for Alcohol and Meat to boost cross-selling, especially for higher-income customers.

Data Accuracy and Collection:

- Address data inconsistencies, particularly in countries like Montenegro, by ensuring a **minimum of 100 customer data points per country** for analysis.
- Improve data collection processes to **link conversions to purchases** and allocate 10% of the data analytics budget to **refining attribution models.**
- Collect **product-level cost data** and **customer satisfaction surveys** (e.g., NPS) for future campaigns. Allocate 5% of the overall budget to gather these insights quarterly.

- Include a **clear indicator for total product purchases** in future datasets to reduce reliance on inferred metrics like summing web buys and walk-in purchases.
- Establish a **standardized method to record total impressions** per channel, ensuring accurate campaign reach tracking.

APPENDIX

Excel Exploration

Customer Age Distribution

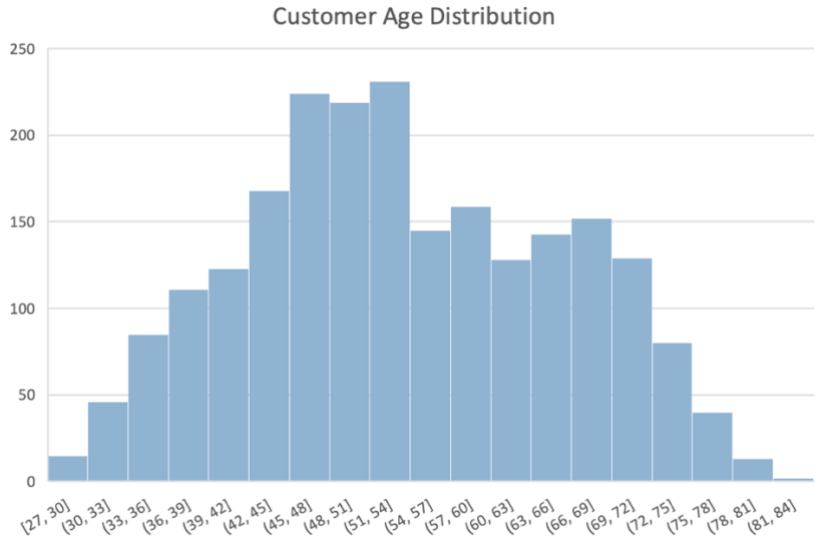

Figure A1: Customer Age Distribution

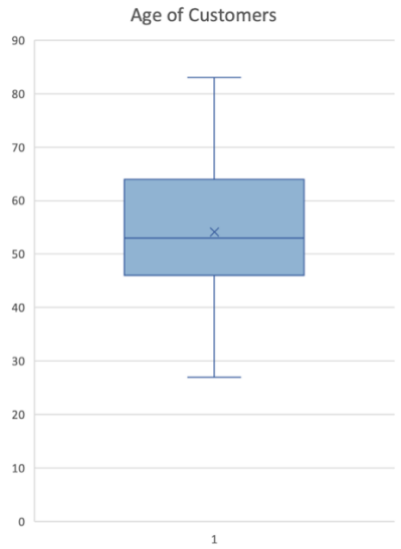

Figure A2: Customer Age Box Plot

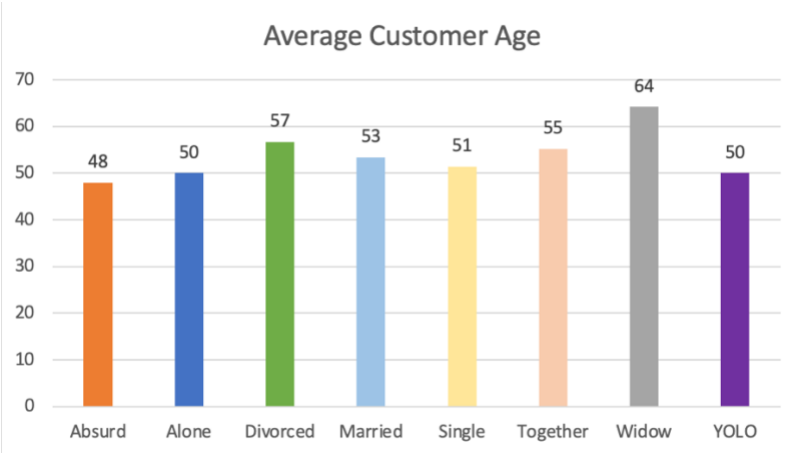

Figure A3: Average Customer Age across Marital Status before grouping.

Note: This chart is included in the appendix to highlight the initial lack of meaningful insights due to illogical categorization of Marital Statuses. To address this limitation, the subsequent chart was created to provide a clearer analysis of Grouped Marital Status.

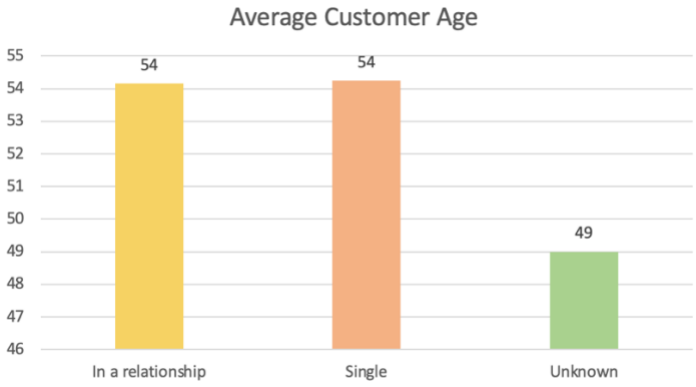

Figure A4: Average Customer Age across Grouped Marital Status.

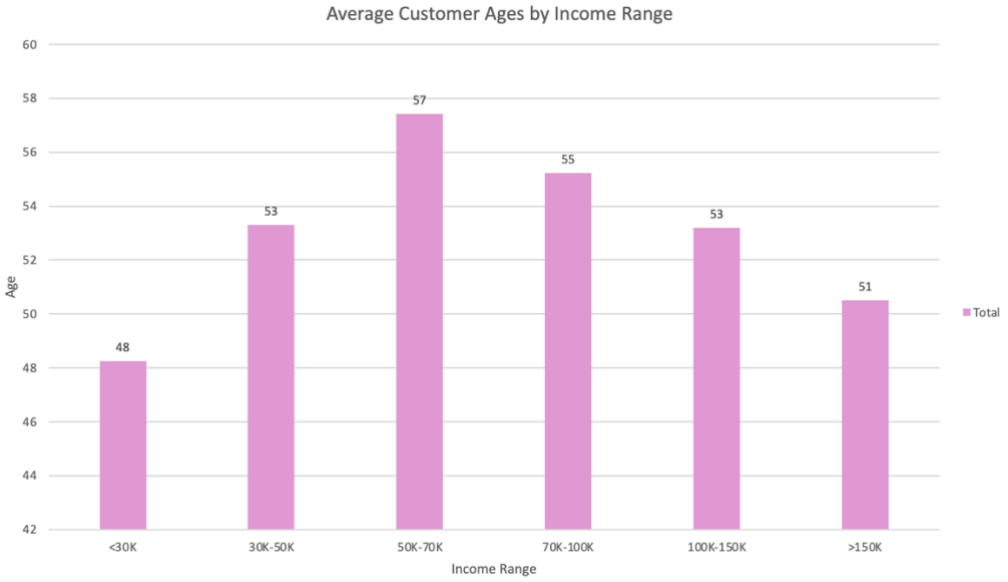

Figure A5: Average Customer Ages by Income Range

Income Exploration

Figure A6: Income Distribution (Histogram)

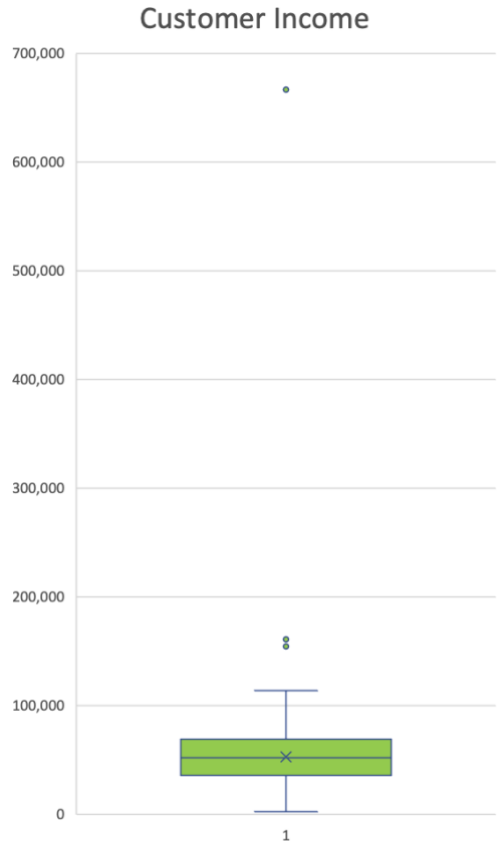

Figure A7: Income Distribution (Boxplot)

Figure A8: Income vs Total Spend

Education Exploration

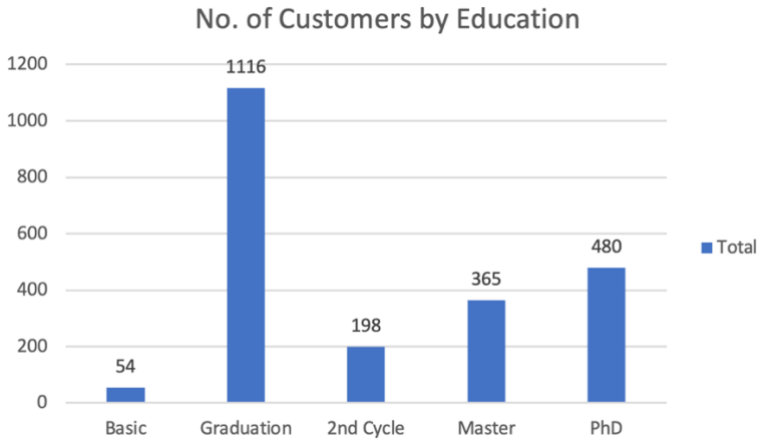

Figure A9: No. of Customers by Education

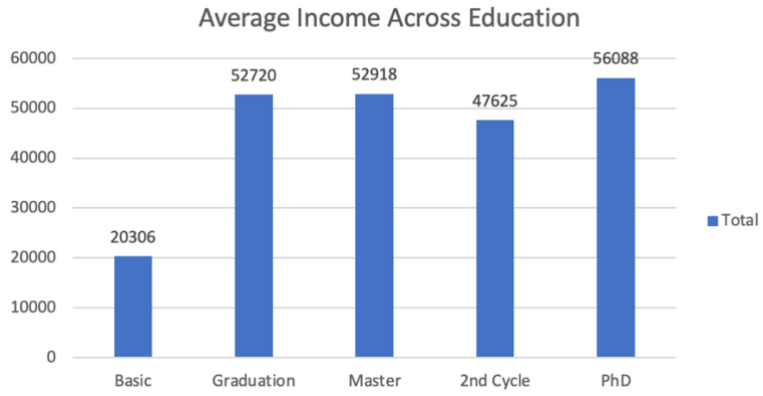

Figure A10: Average Income Across Education Levels

Geographic Exploration

Figure A11: No. of Customers by Country

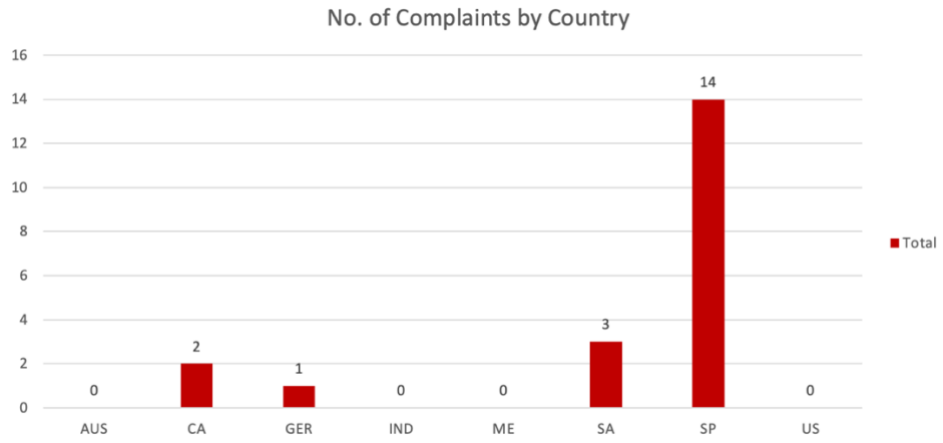

Figure A12: No. of Complaints by Country

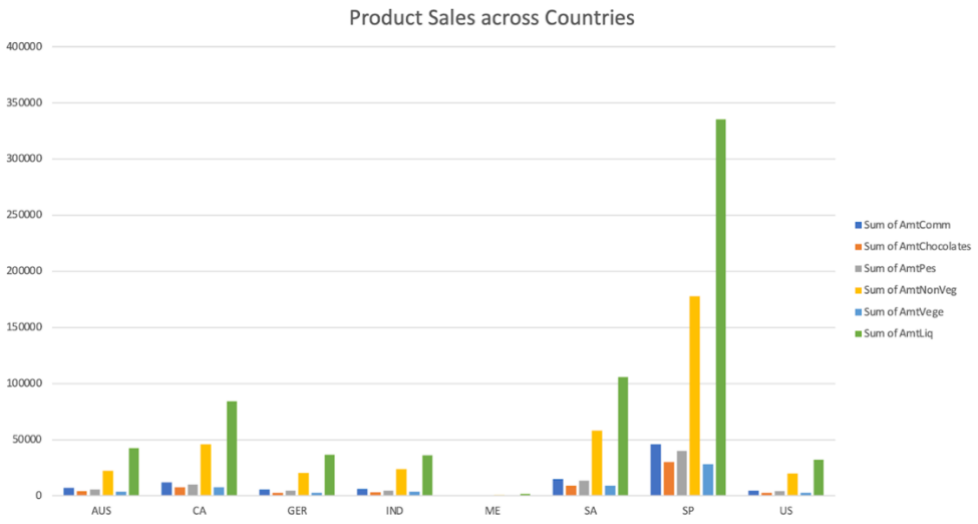

Figure A13: Product Sales Across Countries

Total Spend Distribution

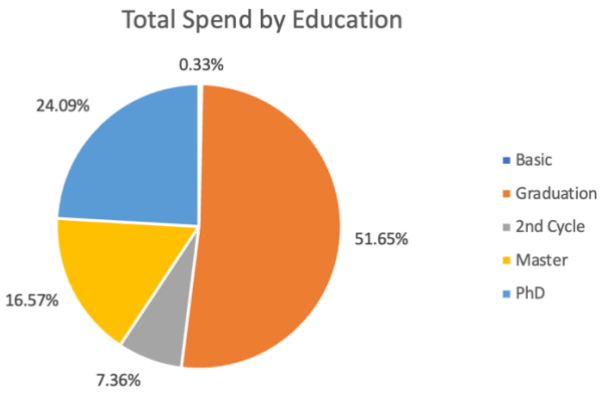

Figure A14: Total Spend by Education

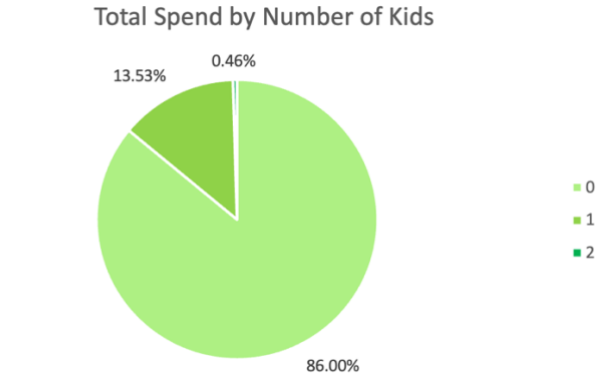

Figure A15: Total Spend by Number of Kids

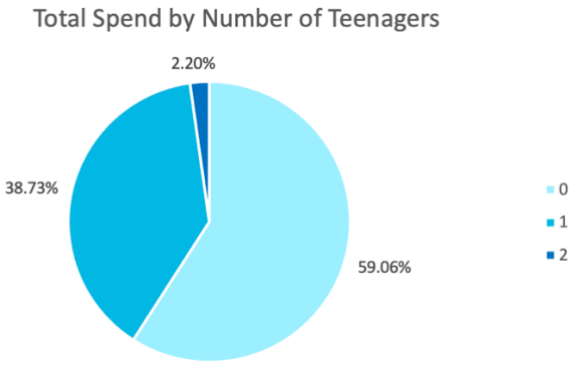

Figure A16: Total Spend by Number of Teenagers

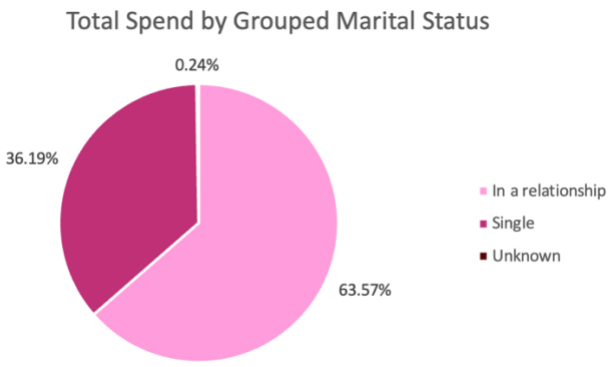

Figure A17: Total Spend by Grouped Marital Status

Tableau Dashboard

Calculated Fields

Total_Sales

```
[Amt Chocolates]+[Amt Comm]+[Amt Liq]+[Amt Non Veg]+[Amt Pes]+[Amt Vege]
```

Figure B1: Calculated Field to obtain Total Sales of Products

Campaign Response Rate

```
(SUM([Response])/COUNT([Response])) *100
```

Figure B2: Calculated Field to obtain Campaign Response Rate

Brochure Ad Conversion Rate

```
SUM([Brochure ad])/COUNT([Brochure ad]) * 100
```

Figure B3: Calculated Field to obtain Advertising Channel Conversion Rate. Similar calculations for the other Advertising Channels were created.

Percentage

Totals summarize values from Table (across).

```
COUNTD([ID]) / TOTAL(COUNTD([ID]))
```

Figure B4: Calculated Field to obtain Percentage, used in Grouped Marital Status Pie Chart

SQL Queries

```sql
CREATE TABLE cleanedmarketingdata (
    id INT,
    year_birth INT,
    age INT,
    education VARCHAR(50),
    marital_status VARCHAR(50),
    grouped_marital_status VARCHAR(50),
    income VARCHAR(20),
    income_range VARCHAR(50),
    income_range_10k VARCHAR(50),
    kidhome INT,
    teenhome INT,
    dt_customer DATE,
    recency INT,
    amtliq INT,
    amtvege INT,
    amtnonveg INT,
    amtpes INT,
    amtchocolates INT,
    amtcomm INT,
    numdeals INT,
    numwebbuy INT,
    numwalkinpur INT,
    numvisits INT,
    response INT,
    complain INT,
    country VARCHAR(10),
    count_success INT,
    total_spend INT
);
```

Figure C1: Query to create cleanedmarketing data table

```sql
CREATE TABLE ad_data (
    id INT,
    bulkmail_ad INT,
    twitter_ad INT,
    instagram_ad INT,
    facebook_ad INT,
    brochure_ad INT
);
```

Figure C2: Query to create ad_data table

```sql
-- Double Checking Excel Calculation for Total Spend
Select amtliq + amtvege + amtnonveg + amtpes + amtchocolates + amtcomm as check_total, total_spend
From cleanedmarketingdata;
```

Figure C3: Query to check Total Spend Calculation

```sql
-- Check Total Sales per Product Category
Select
country,
SUM(amtliq) as total_liq,
SUM (amtvege) as total_veg,
SUM(amtnonveg) as total_nonveg,
SUM(amtpes) as total_pes,
SUM(amtchocolates) as total_choc,
SUM(amtcomm) as total_comm
FROM cleanedmarketingdata
Group by country
Order by SUM(amtliq) DESC;
```

Figure C4: Query to check Total Sales per Product Category