Assignment: Exploratory analysis and presenting insights

Marine Minotte

June 22, 2025

(Word Count: 1028)

1 Introduction - Business context

This analysis is for 2Market, a global supermarket chain operating through physical retail stores and online platforms. Their objective is to gain a deeper understanding of their customers in order to perform data-driven marketing and strategic decision making. Specifically, the analysis evolves around the following questions.

1. What are the demographic characteristics of their customers?
2. Which product groups are the most popular?
3. Which advertising channels are the most successful to engage with their customers?

This will help their objectives as it allows to obtain a clear and concise overview of their customers. The insights gained will help the supermarket to develop more targeted and tailored marketing strategies, therefore improving their business performance.

2 Analytical approach and thought process

The analytical process can be broken down into 4 steps: the data were visualized in Excel, a preliminary draft in Tableau, data cleaning and analysis in SQL, and finally updating the Tableau dashboard based on deeper insights. The data used in the analysis was separated in two csv files, as well as metadata file in text format. The data was on a detailed customer-level and included 2216 IDs. In addition, it showcased multiple variables on demographic characteristics such as age, marital status and income, purchased products and ad channels.

2.1 Visualization in Excel

As mentioned in the previous section, the goal of the approach was to extract insights about customer behavior. The analysis began by using the metadata file to obtain an understanding of all variables and their nature. Filtering and pivoting in Excel provided initial insights such as the average age of the customer (55 years) and a positive correlation between age and income, accompanied by some fluctuations likely caused by some outliers. During this phase, two surprising characters in the marital status, namely YOLO and absurd, were identified, which suggest the possibility of survey selection bias or data quality issues.

2.2 Creating a dashboard in Tableau

During the first draft of the dashboard in Tableau, the goal was to translate key data insights into a visual narrative centering around the company's core questions: the demographic characteristics of their customers, the product preference and the effectiveness of various advertising channels. Thus, the draft focused on:

- Mapping country coverage
- Analyzing income distribution within country by creating different income groups
- Analyzing purchase behavior across income groups and households - Identifying the most effective advertising channel

These preliminary results laid the foundation for further analysis by providing a broad explanatory overview of the customer base.

2.3 SQL and further analysis

Although having already a broad idea of the data, using SQL allowed for clearer understanding, hence a more systematic analysis. Joining the two datasets and filtering the data, while creating an interplay between different variables enabled a deeper exploration of relationships between variables. All output generated in SQL can be found in the Appendix in Tables 1, 2, 3, 4, 5.

However, one important observation emerged. While the income was recorded in US dollars, expenditure figures for product categories lacked defined currencies. Thus, the assumption of the expenditure figures being shown in US dollars had to be drawn, which could be a potential bias and limitation, as it involves customer spending behavior across the world.

2.4 Updated Tableau dashboard

The information gathered during the SQL analysis allowed to update the Tableau dashboard. The aim was to emphasize storytelling centered on customer behavior relative to their life circumstances an d demographic factors. The approach enables 2Market to revise and consider their marketing strategies as they may need to be tailored to local conditions.

Throughout the process outputs were continuously documented. The ideas were altered according to the new results and insights obtained during the different stages of the analysis.

3 Dashboard design and development

The dashboard's purpose was to communicate key findings through a coherent and visual narrative. Thus, the dashboard was updated and altered multiple times throughout the analysis as new insights were uncovered.

First, the goal was to describe the customer base presented in the dataset. To do so, a geographic map was included to display the spread of customers across countries. The size of the circles increases with the number of customers for each country, enabling an assessment of geographic concentration. This step was followed by an income distribution analysis per country. The idea behind this was whether income disparities exist between countries, which would suggest a more country-specific marketing strategy.

Calculating the total expenses of each product group identified product preferences. The figures were visualized using a pie chart. On a more detailed level, expenditures were segmented by household composition and income groups.

The final part of the dashboard focuses on the response rate of various advertising channels. Ranking the channels in a descending order helps identify at first glance which platform yields the biggest marketing return. The results can inform the company which channel to prioritize and engage in more strategic decision making, as it would reduce investments in less effective ones.

4 Results-patterns, trends and insights

The supermarket has customers from 8 different countries spread across the globe, indicating a broad international reach. The largest income group is the Upper-middle income group as shown in Figure 1.

Insights from the excel output showed that the average age of the customers was around 55 years old. A positive correlation between age and income in Table 5 suggested that older individuals have higher purchasing power. The most popular product categories were liquor and meat (Figure

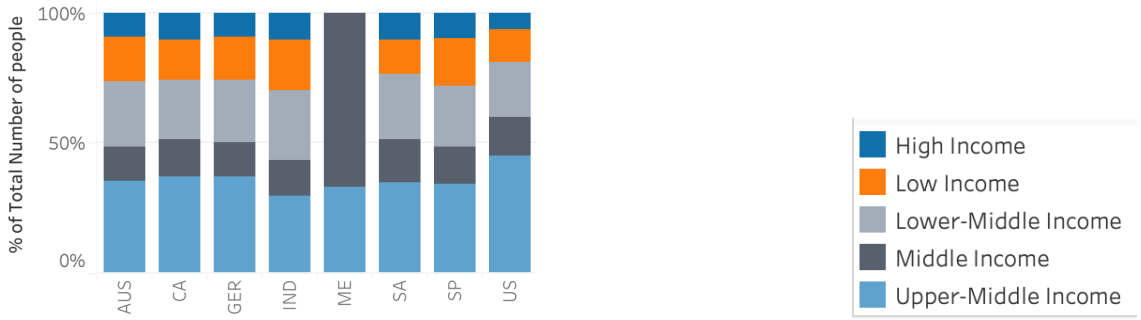
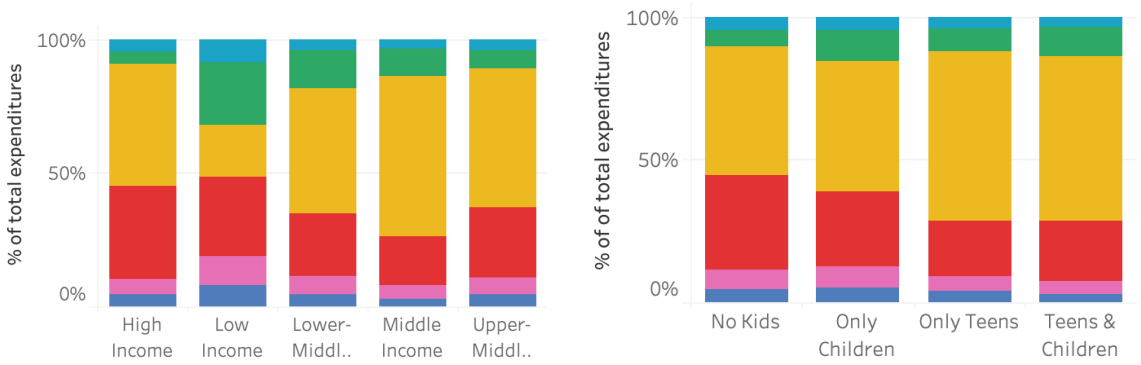


Figure 1: Income group distribution per country in percent

2), remaining consistent when broken down on household or income groups levels as reflected in Figure 3. Twitter emerged as the most successful advertising channel, followed by Bulkmail campaigns (Figure 4). These results align with expectations considering the age bracket of the customers presented in the data.

Figure 2: Purchased product categories

The analysis however pointed to some caveats in the data. The absence of a publication date raises the question of how up to date the data is. Income is reported in US dollars, without any documentation on the currency exchange rate. In addition, the expenditures on the different products are not labeled in any currency, leading to the assumption that it is in US dollars as well. This limitation may introduce bias in cross-country comparisons due to different economic situations and purchasing power parity. The success of the liquor product category is also surprising as the dataset includes a lot of customers from Saudi Arabia, which is a predominantly Muslim country where alcohol consumption is generally restricted.



(a) Income Groups

(b) Household composition

Figure 3: Purchased product categories by household and income groups

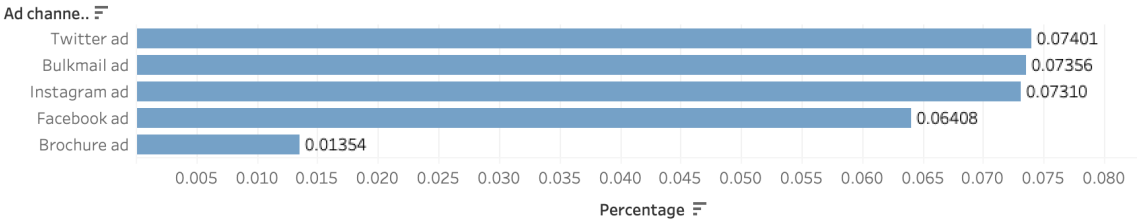


Figure 4: Advertising channels engagement

5 Appendix

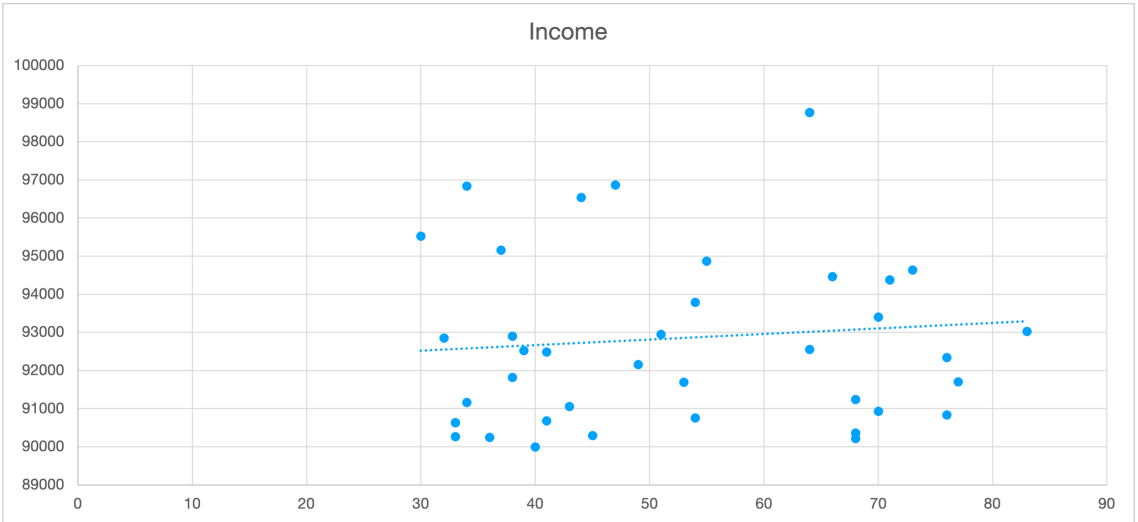


Figure 5: Age and income correlation

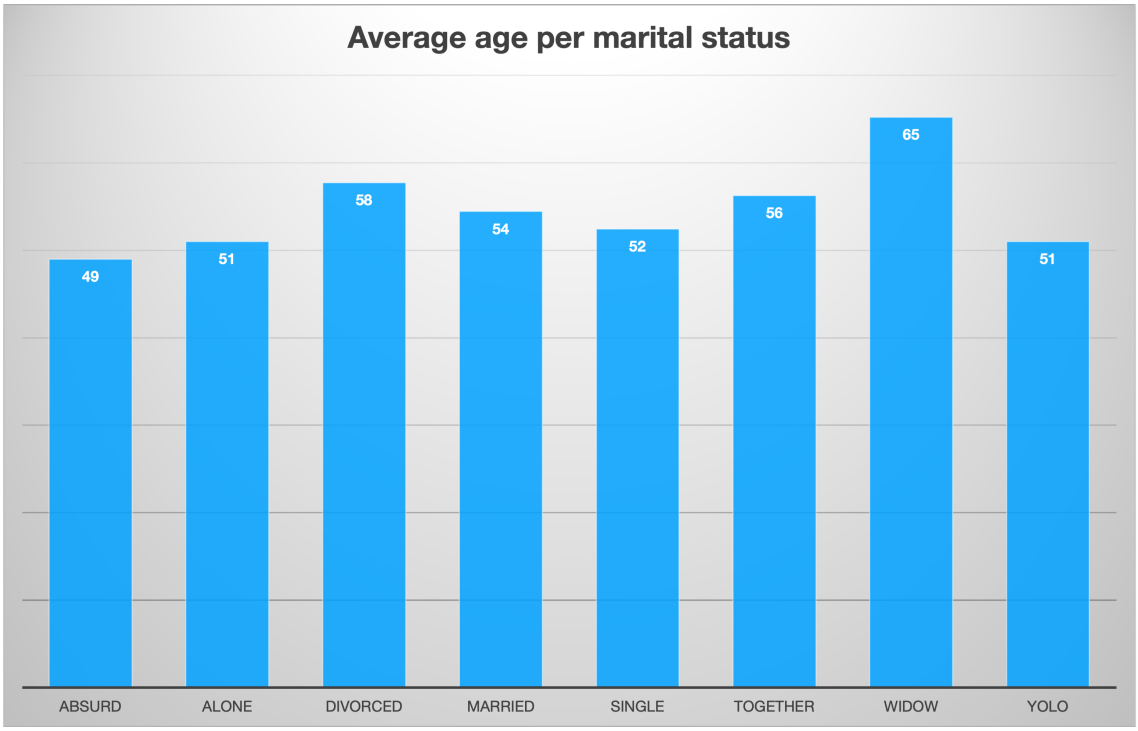


Figure 6: Age per marital status

Country	Total
"SP"	"659557"
"CA"	"167403"
"IND"	"77806"
"AUS"	"85576"
"US"	"67546"
"ME"	"3122"
"SA"	"211071"
"GER"	"73198"

Table 1: Total expenditures per country

Marital status	Liquor	Vegetables	Meat	Fish	Chocolates	Commodities
"Widow"	"27902"	"2422"	"14085"	"3793"	"2878"	"4245"
"YOLO"	"644"	"6"	"100"	"8"	"6"	"84"
"Together"	"176715"	"14612"	"95374"	"22383"	"15031"	"24754"
"Alone"	"554"	"12"	"79"	"23"	"21"	"81"
"Absurd"	"711"	"169"	"625"	"411"	"61"	"408"
"Married"	"256976"	"21981"	"137888"	"30395"	"22926"	"36719"
"Single"	"137217"	"12840"	"87064"	"18262"	"12751"	"20397"
"Divorced"	"75364"	"6363"	"34848"	"8130"	"6222"	"10739"

Table 2: Total expenditures per marital status and product category

Child status	Teen status	Liquor	Vegetables	Meat	Fish	Chocolates	Commodities
"Has children"	"Has teens"	"53872"	"2698"	"19766"	"3723"	"2999"	"9305"
"Has children"	"No teens"	"41990"	"5156"	"25178"	"7487"	"4819"	"11084"
"No children"	"No teens"	"308950"	"33090"	"234758"	"48500"	"33663"	"40661"
"No children"	"Has teens"	"271271"	"17461"	"90361"	"23695"	"18415"	"36377"

Table 3: Total expenditures based on child and teen status

Country	Facebook	Instagram	Twitter	Most effective platform
"SP"	"163"	"182"	"159"	"Instagram"
"CA"	"41"	"47"	"43"	"Instagram"
"IND"	"13"	"12"	"14"	"Twitter"
"AUS"	"15"	"23"	"13"	"Instagram"
"US"	"12"	"10"	"8"	"Facebook"
"ME"	"0"	"0"	"0"	"Facebook"
"SA"	"35"	"42"	"34"	"Instagram"
"GER"	"13"	"14"	"19"	"Twitter"

Table 4: Advertising channel and Country

Marital status	Facebook	Instagram	Twitter	Most effective platform
"Widow"	"9"	"16"	"17"	"Twitter"
"YOLO"	"0"	"0"	"0"	"Facebook"
"Together"	"67"	"89"	"80"	"Instagram"
"Alone"	"0"	"0"	"0"	"Facebook"
"Absurd"	"2"	"2"	"0"	"Facebook"
"Married"	"123"	"129"	"103"	"Instagram"
"Single"	"62"	"63"	"57"	"Instagram"
"Divorced"	"29"	"31"	"33"	"Twitter"

Table 5: Advertising channel and marital status