Background and Context

2Market is a global supermarket that operates both online and in-store. The company aims to better understand its customer purchase behaviour across four key dimensions:

1. Customer demographics
2. Advertising channel effectiveness
3. Product popularity
4. Variation in product sales by demographic group

To address these needs, the following five steps formed the framework for the analysis:

1. Identifying the problem: a lack of insights into customer purchase behaviour.
2. Defining the goal: deriving actionable conclusions from the dataset.
3. Conducting exploratory analysis.
4. Anticipating outcomes and necessary follow-up actions.
5. Reflecting on the process and iterating as needed.

By thoroughly analysing the available data, 2Market seeks to implement an informed strategy that aligns advertising efforts with customer preferences and maximizes product sales.

Analytical Approach

The initial step involved cleaning the raw dataset using Excel. This process focused on formatting inconsistencies, addressing missing values, and identifying outliers. The key steps included:

1. **Calculating Age:** Derived by subtracting the birth year from the current year.
2. **Standardizing Marital Status:** Inconsistent or redundant values were marked as (#N/A) and consolidated (e.g., "Alone" was revised to "Single").
3. **Expanding Country Names:** Abbreviated names were replaced with full names for clarity.
4. **Creating New Fields:** Columns such as Amount Total, Total Children, and Total Ads were introduced to facilitate deeper analysis.
5. **Outlier and Duplicate Detection:** Outliers and duplicates were flagged for removal to ensure data integrity.

Key Insights from Basic Analysis

Using Excel, the following summary metrics were derived:

- **Total Customers:** 2,205
- **Average Age:** 53 years
- **Median Income:** $51,287
- **Average Sale Per Customer:** $607.62

The most prevalent demographics included:

- **Education Levels:** Graduates (1,012), PhDs (424).
- **Marital Status:** Together and Married (1,289 combined).
- **Countries with High Customer Counts:** Spain (997) and South Africa (300).

Visual Insights

Data visualizations were generated to uncover trends and correlations:

1. **Education, Age, and Income:** A strong correlation was observed where higher education levels align with increased age and income.
2. **Marital Status, Age, and Income:** Positive relationships emerged between these three variables, as shown in bar charts.
3. **Income and Sales:** Scatterplots revealed a clear correlation between income and total sales per customer.

SQL Analysis

Further analysis using SQL provided detailed insights by combining sales data with advertising data:

1. **Sales by Country:** Spain consistently recorded the highest total sales, while Montenegro ranked lowest.
2. **Product Preferences:** Liquor emerged as the most popular product category across all demographics.
3. **Advertising Effectiveness:** Instagram was identified as the most impactful advertising channel, followed by Facebook and Twitter. Brochure and bulk mail had the least impact.

Dashboard Design and Development

The dashboard design is based on solving the business problem and presenting this information to the audience with an engaging narrative. In this case, the audience is the Marketing Team, so a detailed knowledge of the business will be assumed. Accessibility considerations were a critical part of the design process to ensure usability for a wide range of users. Key accessibility features include:

- High contrast between text and background.
- Explanatory headers, dimensions, tooltips, and captions on charts.
- User-friendly navigation to facilitate seamless exploration.

At the top of each dashboard page, a summary of key statistics is provided for quick reference, accompanied by a brief explanation of the purpose of each page. The choice of visualisation types was determined by the nature of the data and the insights sought, ensuring clarity and relevance.

Dashboard Features

1. **Interactive Map**: Displays operations by country. Hovering over a country reveals the average age, customer count, and median income. Clicking on a country filters the rest of the dashboard to show data specific to that region.
2. **Histograms**: Visualize the distribution of income, providing a clear view of income ranges across the customer base.
3. **Box-and-Whiskers Plot**: Highlights the average age of customers, showcasing the spread and central tendency of age demographics.
4. **Scatterplots**: Depict relationships between two numerical variables, such as income and average sales, to uncover patterns and correlations.
5. **Bar Charts**: Compare categorical data like marital status, education level, and product categories by a numerical value (e.g., average sales or total sales). These charts can be filtered by demographic attributes such as age group.

Interactivity

Interactivity is a core element of the dashboard, allowing users to drill down into specific subsets of data. Filters enable the exploration of different datasets by product, region, or demographic characteristics. This functionality ensures that the dashboard provides actionable insights tailored to various business needs.

Patterns, Trends, and Insights

The analysis identified several important patterns and trends related to customer behaviour:

1. **Customer Demographics:**
   - Average customer age is 53 years, with a median income of $51,287.
   - Spain has the highest customer count (1,089), while Montenegro has the lowest (3).
   - USA records the highest median income at $55,718, while India has the lowest at $48,432.
2. **Product Popularity:**
   - Liquor is the most lucrative product, generating the highest sales in Spain.
   - Vegetables are the least profitable product category.
3. **Advertising Effectiveness:**
   - Instagram drives the highest revenue per customer ($102.71 per customer).
   - Brochure and bulk mail advertising channels are the least effective.

Recommendations

To Increase Revenue:

1. Focus on high-potential geographies with higher median incomes and customers who are married but have no children or teens at home.

2. Prioritize advertising campaigns for liquor and other high-margin products while reducing focus on low-performing categories like vegetables.
3. Continue investing in Instagram, Facebook, and Twitter for digital advertising campaigns.

To Reduce Costs:

1. Exit underperforming markets with small customer bases.
2. Cease the sale of low-revenue products such as chocolates and vegetables.
3. Minimize expenditure on less effective advertising channels, such as brochures and bulk mail.