2Market Customer Research Report

Including Demographics Analysis and Ad Channel Conversion

Table of Contents

Background/Context of the Report

2Market are a global supermarket who sells products both online and instore. They want to gain a data driven understanding of their customer demographics, product performance and advertising channel performance. Their aim is to use data to provide relevant product advertising to each demographic, increasing engagement and sales.

Without accurate customer and advertising information, 2Market cannot target their specific customer markets effectively, potentially leading to loss of revenue.

The goal of this report will be to provide:

- Customer Demographics (age, location, marital status, and income)
- Product performance by demographic.
- Ad channel performance by demographic.
- Revenue generated by each advertising per channel
- Strategic Recommendations

Approach

Data Collection and Handling

Source: Dataset provided by 2Market, consisting of customer details and ad channel conversion data

Cleaning: Removed 47 duplicate records. There were no blank or missing values (See Appendix 1)

Limitations:
The data has no records more recent than 2014, meaning all analysis could be out of date

Outliers identified (e.g., ages >100) were retained (see Appendix 2)

 For insights involving marital status, only legally recognized statuses were considered (see Appendix 3)

Data Analysis Methods

Analysis Tools
- Excel: Initial analysis of age, marital status, and income.
- SQL: In-depth product and ad channel performance analysis.
- Tableau: Interactive dashboard creation.
- Details: See Appendices 4 (Excel) and 5 (SQL).

Dashboard Design and Development

Overview of the Dashboard

The dashboard was designed to allow stakeholders to quickly assess and filter the report data. This allows for a 'drill down' of the data based on customer age and country, and supports product and total spend stats.

Accessibility: The dashboard uses primarily non red-green colours. Where these colours are used, there are other ways to differentiate the data, such as bar size and markers.

All components have alt text descriptions of the chart and the primary info given, in order to support screen readers.

Components of the Dashboards

- Demographics Analysis: A chart showing spend per product category. This is filtered by age and country, for example the spend of 50–60-year-olds in Spain. This component allows us to find out what the most product category is in a particular age and regional segment.
- Total Spend by Age and Country: A symbol map plots total spend against the world map. This is filtered by age group and country. This allows us to quickly see 2Market's largest market regions by spend.
- Ad Channel Conversion: This shows ad performance by channel, country, and age
- Ad Revenue: This show revenue per channel and country
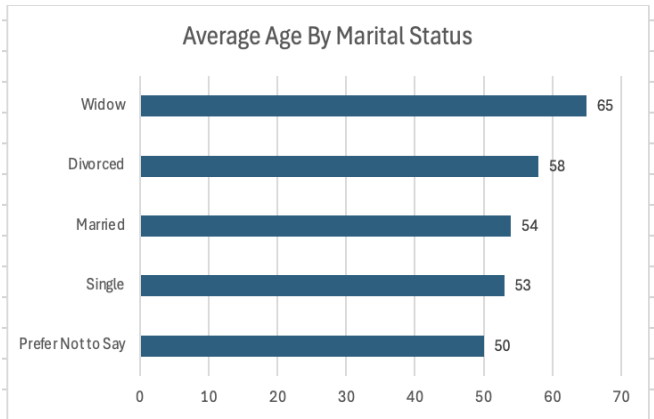
Patterns, Trends, Insights

Key Findings:

The average age of 2Market customers is 55.
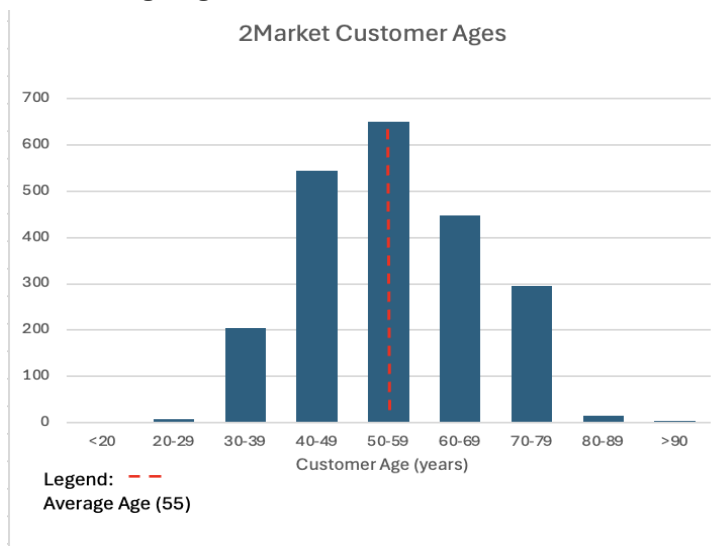
The oldest cohort by marital status is Widowed, while the youngest is 'Prefer not to say' (See Appendix 3)

The largest cohort by marital status is Married customers, who have an average age of 54

The average income is approximately $52,000. This has not been adjusted for current or purchasing power differences. Income peaks at $80k–$90k for ages ~30 and ~80, with minimal age-income correlation.
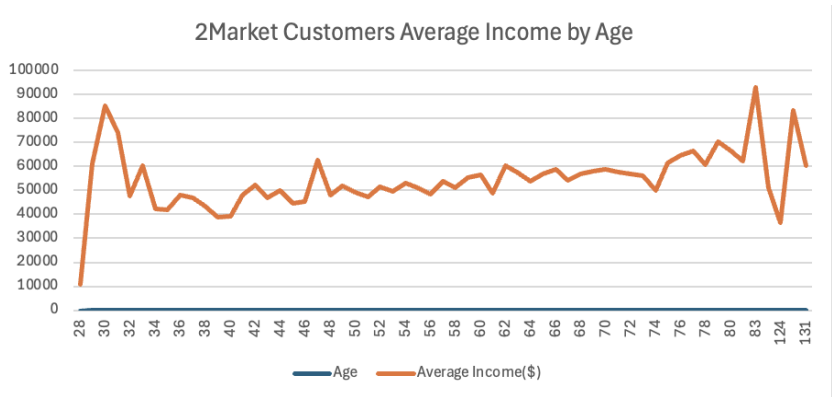
Product Performance and Total Spend Analysis

Top Market: Spain is the highest spending country.

Top Product: Liquor is the best-selling product for all countries, marital statuses and households with teens, kids, or both.

In homes that do not have kids or teens, Liquor is not the best-selling product in only one country - Montenegro. Here the highest category is NonVeg

Liquor is not the bestselling category in the 20-30-year range, in all countries, suggesting the younger generation is drinking less

Ad Channel Conversion

Most effective social media channel per country, based on conversions:

Instagram: Spain, Australia, South Africa
Twitter: Canada,India,USA,Germany
Facebook: None

Highest and Lowest Conversions: Spain and South Africa lead, Montenegro has no conversions across all channels

In the highest performing countries, Instagram primarily drives conversions

Conversions by age:

An unexpected insight is that older customers are more likely to convert from social media channels than younger customers are

Advertising Revenue

In terms of revenue produced by successful conversions, the best performing channels are:

- Twitter
- Facebook
- Instagram
- Bulkmail
- Brochures

Social media channels bring in more revenue than bulk mail - for instance in South Africa, although bulk mail conversions are roughly equal with Facebook, Facebook conversions account for around 150% more revenue

When looking by country, we see that Spain and South Africa generate the most ad revenue. Although Instagram nominally beats Twitter for conversions in these countries, Twitter leads in both by revenue

Brochure ads only account for any revenue in Canada and Spain, showing limited effectiveness

Recap

2Market Customer Persona: Married, 55, earns $52,000, in Spain, largest expenditure is Liquor. Converts via Twitter

Generational Trends: 20–30-year-olds are spending less on Liquor than other age groups

Social Media: Twitter and other social media channels are the most successful at conversions. The overall conversion rate is higher than average, at 15% for social media channels vs 3% industry norm (Forbes, 2018)

Brochure ads have by far the lowest conversion rate.

Recomendations

1. Shift advertising budget from low revenue ad channels to high performing channels (Twitter, Facebook)

2. Use personalised ads to target the customer demographics most likely to convert – married, 50-60 years old, and in Spain or South Africa

3. Research further ways to engage with younger customers on social media (8% conversion rate)

4. Increase offerings of non-alcoholic products to appeal to the 20-30-year-old group, while also keeping a strong Liquor offering for other groups

5. Meat/ nonVeg demand remains strong, avoid over investing in meat alternatives

6. Consider the viability of smaller markets such as Montenegro, which has a total spend of just over $3,000.

Appendices

Appendix 1 – Excel Data Cleaning

The data was firstly cleaned by looking for blank or missing values, duplicate and outliers. Columns were also added where needed for ease of use.

Added columns: Age column had a formula of 2025 - (Birth Year).

Blank or missing values: None were found, and no further actions was taken.

Duplicates: As there were 2000+ rows with multiple columns, I decided on using the CONCAT method along with conditional formatting. In a new column, CONCAT was used on all columns excluding ID. This gave an output of all values in a single cell, with no delimiters. Using conditional formatting, it was then easy to identify identical entries.

Rows where all values were identical were assumed to be identical and removed. In total, 47 rows were removed.

Appendix 2: Outliers

Outliers: 3 values were found that exceed 100 years of age. These were left in the data, as the rest of the data for these entries looked reasonable, and there weren't enough outliers to heavily

skew the data.

There were many (100) values with non-standard responses for 'Marital Status', namely 'Absurd' and 'Yolo'. Again, these were not removed but considered as 'Prefer not to say' in analysis.

Appendix 3: Marital Status

For the purpose of analysis of marital status, only the legal definitions of marital status are included - Married, Divorced, Single and Widowed. 'Together' is counted as Single, as they did not answer Married.

It's assumed the analysis pertains to official Marital status only and not relationship status.

Appendix 4: Excel Data Analysis Methodology

Finding the average age of customers
In order to find the average and visualise it in an easy to understand way, I first created a Pivot table

From this, I grouped the ages in ten year brackets, starting at 20, and found the count for each age bracket

This was put into a simple bar chart, clearly showing the number of customers per age group. To visualise the average, a red dashed line (pre-attentive) was added at the 55-year mark, with a legend to show this is the average age of all customer

Average age of Marital Status
A Pivot table was again used to determine the count of each response type, along with the average age for each group

As mentioned before, only legally defined Marital status' were visualised, along with 'Prefer not to say'

Single, alone and together were all considered as single as they did not match another legally defined status

The Average Age by Marital Status was then plotted on a clustered bar chart

Progression of Income by Customer Age
To find this, a Pivot table was used to determine the average income for each unique age

Eg, there are 3 entries where the age value is 30. The average income of these 3 is $85,449, which is visible on the chart

This was plotted on a line chart, giving a clear view of income progression over age

Customers who earn between $90-100k
First, a helper was used on the main data set, where in a new column this formula was used:

=IF(AND(F2>=90000, F2<=100000), "Yes", "No")

A Pivot table of ID, average age, and Sum of income was filtered by the helper 'In 90k-100k'
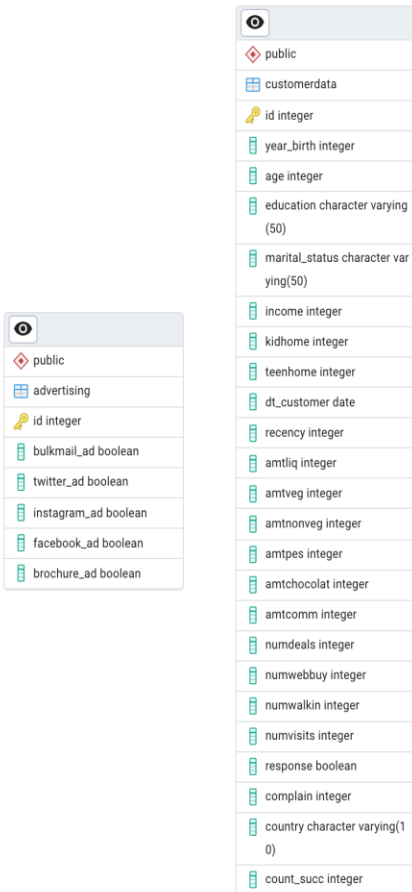
This revealed 39 values with an income in this range

This was shown on a scatter plot, which showed the primary distribution of this income around the 30- 40 year old group, as well as the 70-80 year old group.

Appenndix 5: SQL Data Analysis Methodology

Database: 2Market

Tables:
customerdata
advertising

2Market Database ERD:

Data used in the report was extracted from several queries:

Demograhics:

 Calculating total spend per country

Calculating total spend per product per country

Calculating what product category has the highest spend per country

Calculating what product category has the highest spend per marital status

Calculating what product category has the highest spend depending on if there are children and/or teenagers in the household

Ad Channel Conversion and Revenue:

Calculate ad channel conversions by country

Calculate ad channel conversions by marital status

Calculate ad channel conversions by age

Calculate ad channel conversions and product category spend by country, as well as total ads per country

Calculate revenue generated per channel

All SQL queries used are below with comments:

```
--Calculate total spend per country

SELECT SUM(amtliq + amtveg + amtnonveg + amtpes + amtchocolat + amtcomm) AS
total_spend,"country"

FROM customerdata

GROUP BY "country"

ORDER BY "total_spend" DESC
```

```sql
--Calculate what product is most popular per country
SELECT
    country,
    GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) AS Max_Spend,
    CASE
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Liquor THEN
'Liquor'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Veg THEN 'Veg'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = nonveg THEN
'nonveg'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Pes THEN 'Pes'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Choc THEN
'Choc'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Comm THEN
'Comm'
    END AS Highest_Category
FROM (
    SELECT
        SUM(amtliq) AS Liquor,
        SUM(amtveg) AS Veg,
        SUM(amtnonveg) AS nonveg,
        SUM(amtpes) AS Pes,
        SUM(amtchocolat) AS Choc,
        SUM(amtcomm) AS Comm,
        country
    FROM customerdata
    GROUP BY country
) AS subquery
ORDER BY country;
```

```sql
--Calculate what product is most popular per marital status
SELECT
    marital_status,
    GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) AS Max_Spend,
    CASE
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Liquor THEN 'Liquor'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Veg THEN 'Veg'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = nonveg THEN 'nonveg'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Pes THEN 'Pes'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Choc THEN 'Choc'
        WHEN GREATEST(Liquor, Veg, nonveg, Pes, Choc, Comm) = Comm THEN 'Comm'
    END AS Highest_Category
FROM (
    SELECT
        SUM(amtliq) AS Liquor,
        SUM(amtveg) AS Veg,
        SUM(amtnonveg) AS nonveg,
        SUM(amtpes) AS Pes,
        SUM(amtchocolat) AS Choc,
        SUM(amtcomm) AS Comm,
        marital_status
    FROM customerdata
    GROUP BY marital_status
) AS subquery
ORDER BY marital_status;
```

```sql
--Display ad channel conversions by country
SELECT
        c.country,
    COUNT(CASE WHEN twitter_ad = TRUE THEN 1 END) AS twitter,
    COUNT(CASE WHEN instagram_ad = TRUE THEN 1 END) AS Insta,
    COUNT(CASE WHEN facebook_ad = TRUE THEN 1 END) AS Facebook
FROM advertising a
JOIN customerdata c
ON a.id = c.id
GROUP BY c.country
```

```sql
--Display ad channel conversions by marital status
SELECT
        c.marital_status,
    COUNT(CASE WHEN twitter_ad = TRUE THEN 1 END) AS twitter,
    COUNT(CASE WHEN instagram_ad = TRUE THEN 1 END) AS Insta,
    COUNT(CASE WHEN facebook_ad = TRUE THEN 1 END) AS Facebook
FROM advertising a
JOIN customerdata c
ON a.id = c.id
GROUP BY c.marital_status
```

```sql
--Display ad channel conversions and product category spend by country, as
well as total ads per country

SELECT

    SUM(amtliq) AS Liquor,

    SUM(amtveg) AS Veg,

    SUM(amtnonveg) AS Nonveg,

    SUM(amtpes) AS Pes,

    SUM(amtchocolat) AS Choc,

    SUM(amtcomm) AS Comm,

    COUNT(*) AS total_ads,

    COUNT(CASE WHEN twitter_ad THEN 1 END) AS successful_twitter,

    COUNT(CASE WHEN instagram_ad THEN 1 END) AS successful_insta,

    COUNT(CASE WHEN facebook_ad THEN 1 END) AS successful_facebook,

    COUNT(CASE WHEN response THEN 1 END) AS responded,

    c.country

FROM customerdata c

JOIN advertising a

    ON a.id = c.id

GROUP BY c.country;
```

```sql
--Calculate Ad Revenue per channel

SELECT

  c.id AS customer_id,

  CASE

    WHEN twitter_ad = TRUE THEN 'Twitter'

    WHEN facebook_ad = TRUE THEN 'Facebook'

    WHEN instagram_ad = TRUE THEN 'Instagram'

            WHEN brochure_ad = TRUE then 'Brochure'

            WHEN bulkmail_ad =TRUE then 'BulkMail'

  END AS channel,

  SUM(amtliq + amtveg + amtnonveg + amtpes + amtchocolat + amtcomm) AS total_spend

FROM customerdata c

JOIN advertising a

  ON a.id = c.id

WHERE twitter_ad = TRUE OR facebook_ad = TRUE OR instagram_ad = TRUE OR brochure_ad =
TRUE OR bulkmail_ad =TRUE

GROUP BY

  c.id,

  CASE

    WHEN twitter_ad = TRUE THEN 'Twitter'

    WHEN facebook_ad = TRUE THEN 'Facebook'

    WHEN instagram_ad = TRUE THEN 'Instagram'

            WHEN brochure_ad = TRUE then 'Brochure'

            WHEN bulkmail_ad =TRUE then 'BulkMail'

  END
```