# 2Market Technical Report

## Harshdeep Kohli

## Contents



### 1. Background context

2Market is a global supermarket which sells products both online and in store. 2Market wants to understand 3 key components of their business; customer demographics, product popularity and effectiveness of advertising channels with the hope of increasing business through informed decision-making.

The business problem was identified by using the Five Whys technique *(Appendix 1)* which allows for a structured approach from understanding the problem to determine the root cause. The problem is focussed around the marketing team struggling to generate adequate new business growth due to uncertainty around customer demographics and advertising.

This report will study the following questions:

- *"Who are 2Markets customers?"*
- *"How do their customers interact with products?"*
- *"How effective have their advertising methods been?"*

### 2. Analytical approach

**Data cleaning**

By addressing 5 key data cleaning qualities: 'accuracy', 'completeness', 'consistency', 'uniqueness', and 'timeliness, I made sure that data was suitable for analysis.

- Accuracy – Outliers in income and age fields were identified but not removed, acknowledging the possibility of extreme but true values. Updated income of ID=9432 from $666,666.00 to median income of dataset of $51,373.00 via median imputation. 3 people over the age of 120, but as they don't skew the overall average age, they were not removed *(Appendix 2a).*
- Changed format of 'income' field by removing $ sign, to integer format to allow for numerical calculations.
- Changed format of 'dt_customer' field to UK date format (DD-MM-YYYY) *(Appendix 2b)*
- Completeness – Checks for blank cells and creation of new columns needed for this project. I created various calculated columns such as 'age', 'total spend' amongst others to help with my analysis. Furthermore, both in-store purchases ('NumWalkingPur') and website purchases ('NumWebBuy') cannot both be 0, in this case, 'NumWalkingPur' was changed to 1. There were 10 customers fitting this criterion (Appendix 2c)

- Consistency – Data structure and formats were made uniform and precision and case sensitivity were checked.
- Uniqueness - Involved merging certain marital status and education categories for more consistent analysis *(Appendix 2d)*
- Timeliness - Identified fields for date relevance ('year_birth' and 'dt_customer'). Aware that the data is not recent ('dt_customer' ranges from 2012-2014) therefore could be problematic if conducting a customer time series analysis.

**SQL**

- Reviewed metadata, marketing_data and ad_data files to understand data structure and data types.
- Created marketing_data and ad_data tables within SQL. Assigned 'id' field as primary key and applied correct data types to each column *(Appendix 3a)*
- Successfully imported CSV files for marketing_data and ad_data using pgAdmin4 import data functionality.
- Joined marketing_data and ad_data using a 'inner join' on each tables 'id' field as there were no duplicate or null values in either dataset *(Appendix 3b)*
- Below are the outputs of some questions asked by 2Market:

**Total spend per country?** *(Appendix 4a)*

Data Output | Messages | Graph Visu

| | country character (5) | total_spent numeric |
|---|---|---|
| 1 | SP | 659557.00 |
| 2 | SA | 211071.00 |
| 3 | CA | 167403.00 |
| 4 | AUS | 85576.00 |
| 5 | IND | 77806.00 |
| 6 | GER | 73198.00 |
| 7 | US | 67546.00 |
| 8 | ME | 3122.00 |

**Total spend per product per country?** *(Appendix 4b)*

Data Output | Messages | Graph Visualiser | Notifications

| | country character (5) | total_spend_liquor numeric | total_spend_veg numeric | total_spend_nonveg numeric | total_spend_fish numeric | total_spend_choco numeric | total_spend_comms numeric |
|---|---|---|---|---|---|---|---|
| 1 | US | 32214.00 | 3034.00 | 20185.00 | 4411.00 | 2863.00 | 4839.00 |
| 2 | SP | 336392.00 | 28288.00 | 178409.00 | 40153.00 | 30134.00 | 46181.00 |
| 3 | SA | 105918.00 | 8937.00 | 58398.00 | 13670.00 | 9019.00 | 15129.00 |
| 4 | ME | 1729.00 | 8.00 | 817.00 | 226.00 | 122.00 | 220.00 |
| 5 | IND | 36236.00 | 3788.00 | 23729.00 | 4818.00 | 3221.00 | 6014.00 |
| 6 | GER | 36776.00 | 2980.00 | 20272.00 | 4601.00 | 2801.00 | 5768.00 |
| 7 | CA | 84066.00 | 7681.00 | 45925.00 | 9980.00 | 7607.00 | 12144.00 |
| 8 | AUS | 42752.00 | 3689.00 | 22328.00 | 5546.00 | 4129.00 | 7132.00 |

**Which products are most popular in each country?** *(Appendix 4c)*

Data Output | Messages | Notifications

| | country character (5) | total_spend_liquor numeric | total_spend_veg numeric | total_spend_nonveg numeric | total_spend_fish numeric | total_spend_choco numeric | total_spend_comms numeric | total_spent numeric |
|---|---|---|---|---|---|---|---|---|
| 1 | SP | 336392.00 | 28288.00 | 178409.00 | 40153.00 | 30134.00 | 46181.00 | 659557.00 |
| 2 | SA | 105918.00 | 8937.00 | 58398.00 | 13670.00 | 9019.00 | 15129.00 | 211071.00 |
| 3 | CA | 84066.00 | 7681.00 | 45925.00 | 9980.00 | 7607.00 | 12144.00 | 167403.00 |
| 4 | AUS | 42752.00 | 3689.00 | 22328.00 | 5546.00 | 4129.00 | 7132.00 | 85576.00 |
| 5 | IND | 36236.00 | 3788.00 | 23729.00 | 4818.00 | 3221.00 | 6014.00 | 77806.00 |
| 6 | GER | 36776.00 | 2980.00 | 20272.00 | 4601.00 | 2801.00 | 5768.00 | 73198.00 |
| 7 | US | 32214.00 | 3034.00 | 20185.00 | 4411.00 | 2863.00 | 4839.00 | 67546.00 |
| 8 | ME | 1729.00 | 8.00 | 817.00 | 226.00 | 122.00 | 220.00 | 3122.00 |

**Which products are the most popular based on marital status?** *(Appendix 4d)*

Data Output | Messages | Notifications

| | marital_status character varying (255) | total_spend_liquor numeric | total_spend_veg numeric | total_spend_nonveg numeric | total_spend_fish numeric | total_spend_choco numeric | total_spend_comms numeric | total_spent numeric |
|---|---|---|---|---|---|---|---|---|
| 1 | Married | 256976.00 | 21981.00 | 137888.00 | 30395.00 | 22926.00 | 36719.00 | 506885.00 |
| 2 | In a relationship | 176715.00 | 14612.00 | 95374.00 | 22383.00 | 15031.00 | 24754.00 | 348869.00 |
| 3 | Single | 139126.00 | 13027.00 | 87868.00 | 18704.00 | 12839.00 | 20970.00 | 292534.00 |
| 4 | Divorced | 75364.00 | 6363.00 | 34848.00 | 8130.00 | 6222.00 | 10739.00 | 141666.00 |
| 5 | Widowed | 27902.00 | 2422.00 | 14085.00 | 3793.00 | 2878.00 | 4245.00 | 55325.00 |

**Which products are the most popular based on whether or not there are children or teens in the home?** *(Appendix 4e)*

***Kids at home only***

Data Output | Messages | Notifications

| | kids text | total_spend_liquor numeric | total_spend_veg numeric | total_spend_nonveg numeric | total_spend_fish numeric | total_spend_choco numeric | total_spend_comms numeric | total_spent numeric |
|---|---|---|---|---|---|---|---|---|
| 1 | no kids | 580221.00 | 50551.00 | 325119.00 | 72195.00 | 52078.00 | 77038.00 | 1157202.00 |
| 2 | 1 kid | 92550.00 | 7576.00 | 43560.00 | 10894.00 | 7639.00 | 19613.00 | 181832.00 |
| 3 | 2 kids | 3312.00 | 278.00 | 1384.00 | 316.00 | 179.00 | 776.00 | 6245.00 |

**Which products are the most popular based on whether or not there are children or teens in the home?** *(Appendix 4f)*

***teens at home only***

Data Output Messages Notifications

| | teens text | total_spend_liquor numeric | total_spend_veg numeric | total_spend_nonveg numeric | total_spend_fish numeric | total_spend_choco numeric | total_spend_comms numeric | total_spent numeric |
|---|---|---|---|---|---|---|---|---|
| 1 | no teens | 350940.00 | 38246.00 | 259936.00 | 55987.00 | 38482.00 | 51745.00 | 795336.00 |
| 2 | 1 teen | 307060.00 | 19268.00 | 103811.00 | 26227.00 | 20628.00 | 43365.00 | 520359.00 |
| 3 | 2 teens | 18083.00 | 891.00 | 6316.00 | 1191.00 | 786.00 | 2317.00 | 29584.00 |

**Which social media platform (Twitter, Instagram or Facebook) is the most effective method of advertising in each country?** *(Appendix 4g)*

Data Output Messages Notifications

| | country character (5) | twitter bigint | instagram bigint | facebook bigint |
|---|---|---|---|---|
| 1 | SP | 87 | 89 | 76 |
| 2 | CA | 24 | 21 | 18 |
| 3 | IND | 10 | 6 | 7 |
| 4 | AUS | 6 | 12 | 7 |
| 5 | US | 6 | 5 | 7 |
| 6 | ME | 0 | 0 | 0 |
| 7 | SA | 20 | 21 | 20 |
| 8 | GER | 11 | 8 | 7 |

**Which social media platform is the most effective method of advertising based on marital status?** *(Appendix 4h)*

Data Output | Messages | Notifications

| | marital_status character varying (255) | twitter bigint | instagram bigint | facebook bigint |
|---|---|---|---|---|
| 1 | In a relationship | 42 | 44 | 32 |
| 2 | Widowed | 10 | 7 | 5 |
| 3 | Married | 62 | 66 | 62 |
| 4 | Single | 32 | 32 | 31 |
| 5 | Divorced | 18 | 13 | 12 |

**Which social media platform(s) seem(s) to be the most effective per country? (In this case, assume that purchases were in some way influenced by lead conversions from a campaign)** *(Appendix 4i)*

| country | twitter | instagram | facebook | total_conversions | total_spend_liquor | total_spend_veg | total_spend_nonveg | total_spend_fish | total_spend_choco | total_spend_comms | total_spent |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SP | 87 | 89 | 76 | 252 | 336392 | 28288 | 178409 | 40153 | 30134 | 46181 | 659557 |
| CA | 24 | 21 | 18 | 63 | 84066 | 7681 | 45925 | 9980 | 7607 | 12144 | 167403 |
| SA | 20 | 21 | 20 | 61 | 105918 | 8937 | 58398 | 13670 | 9019 | 15129 | 211071 |
| GER | 11 | 8 | 7 | 26 | 36776 | 2980 | 20272 | 4601 | 2801 | 5768 | 73198 |
| AUS | 6 | 12 | 7 | 25 | 42752 | 3689 | 22328 | 5546 | 4129 | 7132 | 85576 |
| IND | 10 | 6 | 7 | 23 | 36236 | 3788 | 23729 | 4818 | 3221 | 6014 | 77806 |
| US | 6 | 5 | 7 | 18 | 32214 | 3034 | 20185 | 4411 | 2863 | 4839 | 67546 |
| ME | 0 | 0 | 0 | 0 | 1729 | 8 | 817 | 226 | 122 | 220 | 3122 |

**Advertising channel spend by country** *(Appendix 4j)*

Data Output | Messages | Notifications

| | country character (5) | instagram numeric | facebook numeric | twitter numeric | bulkmail numeric | brochure numeric | total_spend numeric | total_ad_spend numeric | percentage numeric |
|---|---|---|---|---|---|---|---|---|---|
| 1 | SP | 147622.00 | 114018.00 | 102271.00 | 55250.00 | 21808.00 | 659557.00 | 440969.00 | 67 |
| 2 | SA | 35738.00 | 31211.00 | 26748.00 | 13287.00 | 4282.00 | 211071.00 | 111266.00 | 53 |
| 3 | CA | 30954.00 | 27650.00 | 25261.00 | 16457.00 | 6397.00 | 167403.00 | 106719.00 | 64 |
| 4 | AUS | 18505.00 | 10480.00 | 7915.00 | 9866.00 | 0 | 85576.00 | 46766.00 | 55 |
| 5 | IND | 9444.00 | 8963.00 | 8129.00 | 9779.00 | 3736.00 | 77806.00 | 40051.00 | 51 |
| 6 | GER | 12551.00 | 9649.00 | 11538.00 | 7949.00 | 3007.00 | 73198.00 | 44694.00 | 61 |
| 7 | US | 7048.00 | 8806.00 | 4698.00 | 3986.00 | 0 | 67546.00 | 24538.00 | 36 |
| 8 | ME | 0 | 0 | 0 | 874.00 | 0 | 3122.00 | 874.00 | 28 |

## 3. Dashboard design & development

This analysis is designed for 2Market's **marketing team** to provide a complete overview featuring key demographic, product and advertising metrics with a focus of providing data driven recommendations. This narrative allows for concise storytelling and keeps the user focused on the problem and possible solutions in a structured manner.

The dashboard was designed strategically to reduce cognitive overload and make key points accessible as possible following web content accessibility guidelines (WCAG). By including relevant metrics which could be filtered for any additional context would provide the user with a top-down customizable view. Finally, the style and design were inspired from tips gained from the book '*Storytelling with data*' by *Cole Nussbaumer Knaflic.* Some of the features I included are:

- Two contrasting colours for majority of visuals
- Colour-blind accessible palette
- Data labels on majority of graphs to aid users with vision impairment
- Dashboard shading in colour 'grey' and graph backgrounds in colour 'white' to accentuate visuals
- Enabled interactivity through filters and highlighting to allow for a personalised view
- Clear focus order by chronologically numbered visuals and 'Z' patterned design layout for easier readability

## 4. Patterns, trends & insights

**Demographic insights**

- Focus on the segment of customers aged between 15-29 and 75-119 These customers have the highest spend per customer, appreciating that they represent a smaller population compared to other age groups *(Appendix 5a)*
- Focus on the segment of customers with income between $50K-$99K as they represent the highest spend per customer *(Appendix 5b)*

- Focus on customers without children, as their spend per customer is significantly higher (c.$700) than those with children *(Appendix 5c)*

**Product insights**

- Liquor and non veg products are the top selling products across all demographics, accounting for 50% and 28% of total spend respectively
- Spain has the highest total spend on liquor, but, from a percentage share of spending on all products, 2Market should focus on expanding their presence in Montenegro, where liquor spending as a proportion of their total basket is the highest *(Appendix 5d)*
- Customers in the age ranges from 45-74 spend the most on liquor as a proportion of their total basket. Focus on promotions on 45-74 age group given their high spending proportion on liquor *(Appendix 5e)*
- Percentage spend on liquor compared to other products is highest in the $50K-$74K income bands which indicates a strong preference for liquor in the middle-income brackets. Focus on promotions for liquor on the $50K-$74K income band *(Appendix 5f)*

**Advertising insights**

- Twitter & Instagram lead the way as the most effective social media platforms by total conversions. But appreciate the effectiveness of Facebook in the US through more US Facebook marketing initiatives. *(Appendix 5g)*
- Bulk mail is the second most effective advertising channel after Twitter in terms of total conversions. Focus on Bulk mail particularly in US and India where it performs the best *(Appendix 5h)*
- Brochures are extremely ineffective. Consider dropping and diverting resources to more effective channels like Twitter and Bulk mail *(Appendix 5i)*

# Appendix 1

**Problem – 2Market's marketing team is unable to generate adequate new business growth.**

1. Why aren't 2Market generating new business growth?
- Marketing isn't focussed at targeting potential customers.

2. Why isn't marketing focussed at targeting potential customers?
- Because their marketing strategies aren't personalised to their customer segments.

3. Why aren't marketing strategies personalised to their customer segments?
- Because they don't understand their different customer segments.

4. Why don't they understand their different customer segments?
- Because their systems are outdated and they haven't invested into data teams to enable them to analyse customer behaviour.

5. Why haven't they invested into systems and data teams?
- Because there is little awareness about potential gains of customer segmentation and systems available to allow for effective marketing campaigns.

**Root cause – The root cause of 2Market not being able to generate new business growth is due to lack of awareness about potential gains of customer segmentation which can be done with the help of skilled data teams and modern systems to allow for more data driven decision making and therefore more targeted marketing campaigns which would help increase business growth.**

**Extra notes**

Potential impacts of not understanding the problem include loss of sales, customers, market share and reputation within the retail sector. By gaining insights into customer demographics, product sales and advertising they can create tailored product offerings to align with their customer segments and boost customer engagement through advertising which would allow them to increase ROI and improve sales performance. Potential solutions are: identifying challenges, and the desired end result the company wants to achieve (i.e. increasing new business growth by 10%), establishing which metrics matter most to the company so that they can be measured effectively. Also using customer segmentation techniques with CRM tools and regularly reviewing performance by adopting a data driven culture.

# Appendix 2

## 2a.

**Income outliers**

- ID 9432 has an 'income' of $666,666.00. This is an extremely high income and may be an outlier if compared to other income values in the dataset. It's possible it could be correct and the person is a high-income earner. For the purpose of accuracy in my analysis and for accurate metrics for age bands, I used **'median imputation'** to update the income of this person from $666,666.00 to $51,373.00, which is the median income of the entire dataset and is a best practice when dealing with outliers in income fields.

**Age outliers**

- ID 11004 - Year of birth = 1894, would mean this person is 130 in the current year (2024)
- ID 1150 - Year of birth = 1900, would mean this person is 124 in the current year (2024)
- ID 7829 - Year of birth = 1901, would mean this person is 123 in the current year (2024)

These people are exceptionally old and it's possible that they are this old, but the reason for this is because we are using the current year (2024) to calculate their age, meaning they would appear older. Why are we using the current year to calculate their age and not looking at this analysis at an earlier point in time, given that we have customer joined dates ranging from 2012-2014? My justification for this is that 2Market is looking at historical data in the current year for an unspecified business problem. Furthermore, the above 3 age outliers don't affect the average age of the entire dataset, therefore I have decided not to remove them from the dataset and kept all their data as it is.

## 2b.

- The 'dt_customer' field was reformatted using the text to columns wizard in Excel to correctly format all dates into UK date format (DD-MM-YYYY)
- The 'Income' field was reformatted to integer type, as this would prove to be a more efficient way at loading the data into SQL/Tableau and allow for easier numerical calculations, and any currency formatting could be applied easily on those tools. Use of find and replace within Excel to achieve this.

## 2c.

- Both 'in-store' purchases ('NumWalkingPur') and 'website' purchases ('NumWebBuy') cannot be 0, in this case, in-store purchase was changed to 1. 10 customers with 'in-store' and 'website purchase' equal to 0. The below screenshot shows the 10 customers who had 'in-store' and 'website' purchases equal to 0, which were subsequently updated to 'in-store'=1. The reason for in-store purchases being updated to 1, rather than website purchases was because in-store purchases heavily outweighs website purchases and increasing either purchase by 10 doesn't change the overall picture.

| | A | Q | R |
|---|---|---|---|
| | ID | NumWebBuy | NumWalkinPur |
| 25 | 4931 | 0 | 0 |
| 54 | 3955 | 0 | 0 |
| 59 | 9931 | 0 | 0 |
| 56 | 5376 | 0 | 0 |
| 82 | 11110 | 0 | 0 |
| 06 | 10749 | 0 | 0 |
| 56 | 6862 | 0 | 0 |
| 12 | 5555 | 0 | 0 |
| 08 | 11181 | 0 | 0 |
| 82 | 8475 | 0 | 0 |

## 2d.

Merged certain marital status and education categories to correctly spelled, and contextualized names using Find and Replace on Microsoft Excel.

- Changed ‘Alone’ & ‘YOLO’ & ‘Absurd’ to ‘Single’
- Changed ‘together’ to ‘In a relationship’
- Changed ‘widow’ to ‘Widowed’

- Changed ‘2n Cycle’ to ‘Second-cycle’
- Changed ‘Graduation’ to ‘Undergraduate’
- Changed ‘Master’ to ‘Masters’

**Extra notes**

Excel, SQL and Tableau were used to conduct this project because they each provide strengths in conducting exploratory analysis to easily identify outliers, trends and insights. Excel allowed me to view the raw data effortlessly as it was only 2,216 rows and conduct descriptive analysis on distributions, correlations and help identify anomalies. SQL allowed me to cross analyse effectively with use of simple queries to join tables and manipulate the data efficiently (i.e. sales by product and country). Tableau allowed me to create informative and interactive visuals to highlight key points and create an effective narrative to help promote growth at 2Market. By adopting simple graphs, mainly bar/column/scatter graphs I made sure that the information was easy to interpret and accessible for the intended audience and force action accordingly.

# Appendix 3

### 3a. Creating the 'marketing_data' table on SQL

```
CREATE TABLE IF NOT EXISTS public.marketing_data
(
    id integer NOT NULL,
    year_birth integer,
    education character varying(255),
    marital_status character varying(255),
    income numeric(10,2),
    kidhome integer,
    teenhome integer,
    dt_customer date,
    recency integer,
    amtliq numeric(10,2),
    amtvege numeric(10,2),
    amtnonveg numeric(10,2),
    amtpes numeric(10,2),
    amtchocolates numeric(10,2),
    amtcomm numeric(10,2),
    numdeals integer,
    numwebbuy integer,
    numwalkinpur integer,
    numvisits integer,
    response integer,
    complain integer,
    country character(5),
    count_success integer,
    CONSTRAINT marketing_data_pkey PRIMARY KEY (id)
)
```

### 3a. Creating the 'ad_data' table on SQL

```
CREATE TABLE IF NOT EXISTS public.ad_data
(
    id integer NOT NULL,
    bulkmail_ad integer,
    twitter_ad integer,
    instagram_ad integer,
    facebook_ad integer,
    brochure_ad integer,
    CONSTRAINT ad_data_pkey PRIMARY KEY (id)
)
```

### 3b. Checking for null values (example)

```
SELECT id,
country,
income,
year_birth
FROM marketing_data
WHERE country IS NULL;
```

### 3b. Checking for duplicate values (example)

```
SELECT id,
COUNT (*) as row_count
FROM marketing_data
GROUP BY id
ORDER BY row_count desc;
```

# Appendix 4

### 4a. Total spend per country?

**SQL Syntax**

```
SELECT
country,
SUM(amtliq+amtvege+amtnonveg+amtpes+amtchocolates+amtcomm) as total_spent
FROM marketing_data
GROUP BY country
ORDER BY total_spent desc;
```

### 4b. Total spend per product per country?

**SQL Syntax**

```
SELECT
country,
SUM(amtliq) as total_spend_liquor,
SUM(amtvege) as total_spend_veg,
SUM(amtnonveg) as total_spend_nonveg,
SUM(amtpes) as total_spend_fish,
SUM(amtchocolates) as total_spend_choco,
SUM(amtcomm) as total_spend_comms,
FROM marketing_data
GROUP BY country
ORDER BY total_spent desc;
```

### 4c. Which products are the most popular in each country?

**SQL Syntax**

```
SELECT
country,
SUM(amtliq) as total_spend_liquor,
SUM(amtvege) as total_spend_veg,
SUM(amtnonveg) as total_spend_nonveg,
SUM(amtpes) as total_spend_fish,
SUM(amtchocolates) as total_spend_choco,
SUM(amtcomm) as total_spend_comms,
SUM(amtliq+amtvege+amtnonveg+amtpes+amtchocolates+amtcomm) as total_spent
FROM marketing_data
GROUP BY country
ORDER BY total_spent desc;
```

### 4d. Which products are the most popular based on marital status?

**SQL Syntax**

```
SELECT
marital_status,
SUM(amtliq) as total_spend_liquor,
SUM(amtvege) as total_spend_veg,
SUM(amtnonveg) as total_spend_nonveg,
SUM(amtpes) as total_spend_fish,
SUM(amtchocolates) as total_spend_choco,
SUM(amtcomm) as total_spend_comms
FROM marketing_data
GROUP BY marital_status;
```

### 4e. Which products are the most popular based on whether or not there are children or teens in the home? *(Kids at home only)*

**SQL Syntax**

```
SELECT
CASE
WHEN kidhome=0 THEN 'no kids'
WHEN kidhome=1 THEN '1 kid'
ELSE '2 kids'
END AS kids,
SUM(amtliq) as total_spend_liquor,
SUM(amtvege) as total_spend_veg,
SUM(amtnonveg) as total_spend_nonveg,
SUM(amtpes) as total_spend_fish,
SUM(amtchocolates) as total_spend_choco,
SUM(amtcomm) as total_spend_comms,
SUM(amtliq+amtvege+amtnonveg+amtpes+amtchocolates+amtcomm) as total_spent
FROM marketing_data
GROUP BY kids
ORDER BY total_spent desc;
```

### 4f. Which products are the most popular based on whether or not there are children or teens in the home? *(Teens at home only)*

**SQL Syntax**

```
SELECT
CASE
WHEN teenhome=0 THEN 'no teens'
WHEN teenhome=1 THEN '1 teen'
ELSE '2 teens'
END AS teens,
SUM(amtliq) as total_spend_liquor,
SUM(amtvege) as total_spend_veg,
SUM(amtnonveg) as total_spend_nonveg,
SUM(amtpes) as total_spend_fish,
SUM(amtchocolates) as total_spend_choco,
```

```
SUM(amtcomm) as total_spend_comms,
SUM(amtliq+amtvege+amtnonveg+amtpes+amtchocolates+amtcomm) as total_spent
FROM marketing_data
GROUP BY teens
ORDER BY total_spent desc;
```

### 4g. Which social media platform (twitter, Instagram or facebook) is the most effective method of advertising in each country?

**SQL Syntax**

```
SELECT
m.country as country,
SUM(a.twitter_ad) as twitter,
SUM(a.instagram_ad) as instagram,
SUM(a.facebook_ad) as facebook
FROM marketing_data m
JOIN ad_data a
ON m.id=a.id
GROUP BY country;
```

### 4h. Which social media platform (twitter, Instagram or facebook) is the most effective method of advertising based on marital status?

**SQL Syntax**

```
SELECT
m.marital_status as marital_status,
SUM(a.twitter_ad) as twitter,
SUM(a.instagram_ad) as instagram,
SUM(a.facebook_ad) as facebook
FROM marketing_data m
JOIN ad_data a
ON m.id=a.id
GROUP BY marital_status;
```

### 4i. Which social media platform(s) seem(s) to be the most effective per country? (In this case, assume that purchases were in some way influenced by lead conversions from a campaign)

**SQL Syntax**

```
SELECT
m.country as country,
SUM(a.twitter_ad) as twitter,
SUM(a.instagram_ad) as instagram,
SUM(a.facebook_ad) as facebook,
SUM(a.twitter_ad+a.instagram_ad+a.facebook_ad) as total_conversions,
SUM(m.amtliq) as total_spend_liquor,
SUM(m.amtvege) as total_spend_veg,
SUM(m.amtnonveg) as total_spend_nonveg,
SUM(m.amtpes) as total_spend_fish,
SUM(m.amtchocolates) as total_spend_choco,
```

```
SUM(m.amtcomm) as total_spend_comms,
SUM(amtliq+amtvege+amtnonveg+amtpes+amtchocolates+amtcomm) as total_spent
FROM marketing_data m
JOIN ad_data a
ON m.id=a.id
GROUP BY country
ORDER BY total_conversions desc;
```

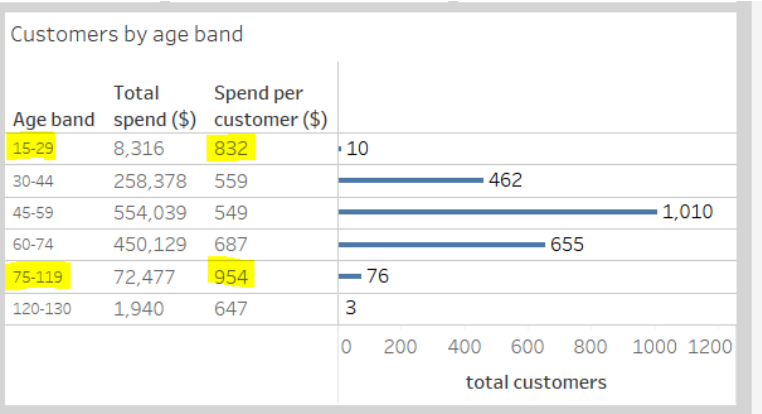
### 4J. Advertising spend per country and channel

**SQL Syntax**

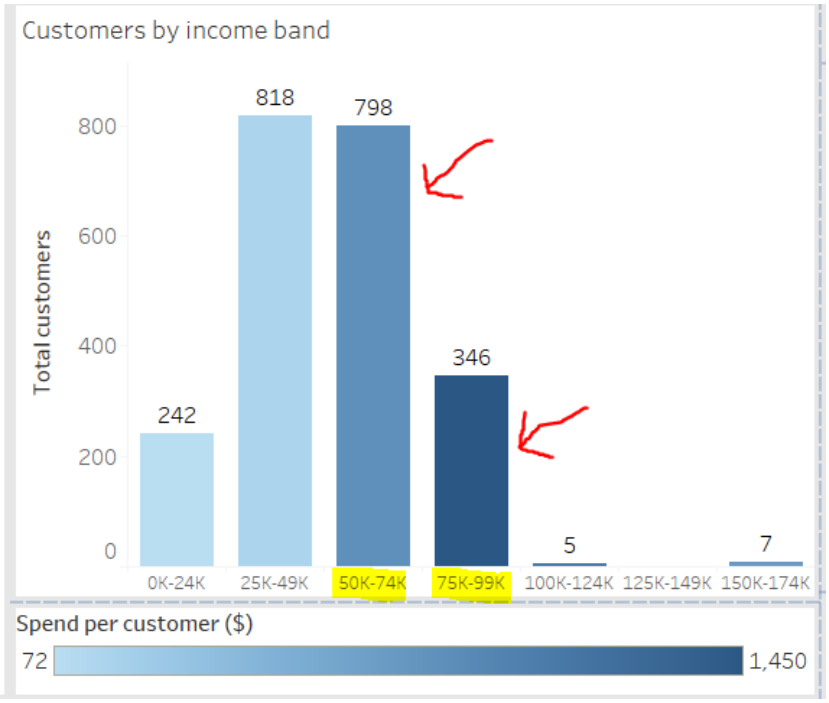
```
WITH t2 as(
SELECT
m.country as country,
SUM(CASE WHEN a.instagram_ad>0 THEN
m.amtliq+m.amtvege+m.amtnONveg+m.amtpes+m.amtchocolates+m.amtcomm ELSE 0 END) as
instagram,
SUM(CASE WHEN a.facebook_ad>0 THEN
m.amtliq+m.amtvege+m.amtnONveg+m.amtpes+m.amtchocolates+m.amtcomm ELSE 0 END) as
facebook,
SUM(CASE WHEN a.twitter_ad>0 THEN
m.amtliq+m.amtvege+m.amtnONveg+m.amtpes+m.amtchocolates+m.amtcomm ELSE 0 END) as
twitter,
SUM(CASE WHEN a.bulkmail_ad>0 THEN
m.amtliq+m.amtvege+m.amtnONveg+m.amtpes+m.amtchocolates+m.amtcomm ELSE 0 END) as
bulkmail,
SUM(CASE WHEN a.brochure_ad>0 THEN
m.amtliq+m.amtvege+m.amtnONveg+m.amtpes+m.amtchocolates+m.amtcomm ELSE 0 END) as
brochure,
SUM(m.amtliq+m.amtvege+m.amtnONveg+m.amtpes+m.amtchocolates+m.amtcomm) as
total_spend
FROM ad_data a
JOIN marketing_data m
ON a.id=m.id
GROUP BY country)

SELECT t2.country,t2.instagram,t2.facebook,t2.twitter,t2.bulkmail,t2.brochure,t2.total_spend,
SUM(t2.instagram+t2.facebook+t2.twitter+t2.bulkmail+t2.brochure) as total_ad_spend,
round((SUM(t2.instagram+t2.facebook+t2.twitter+t2.bulkmail+t2.brochure)/t2.total_spend)*100,0)
as percentage
FROM t2
GROUP BY t2.country,t2.instagram,t2.facebook,t2.twitter,t2.bulkmail,t2.brochure,t2.total_spend
ORDER BY total_spend desc;
```

# Appendix 5

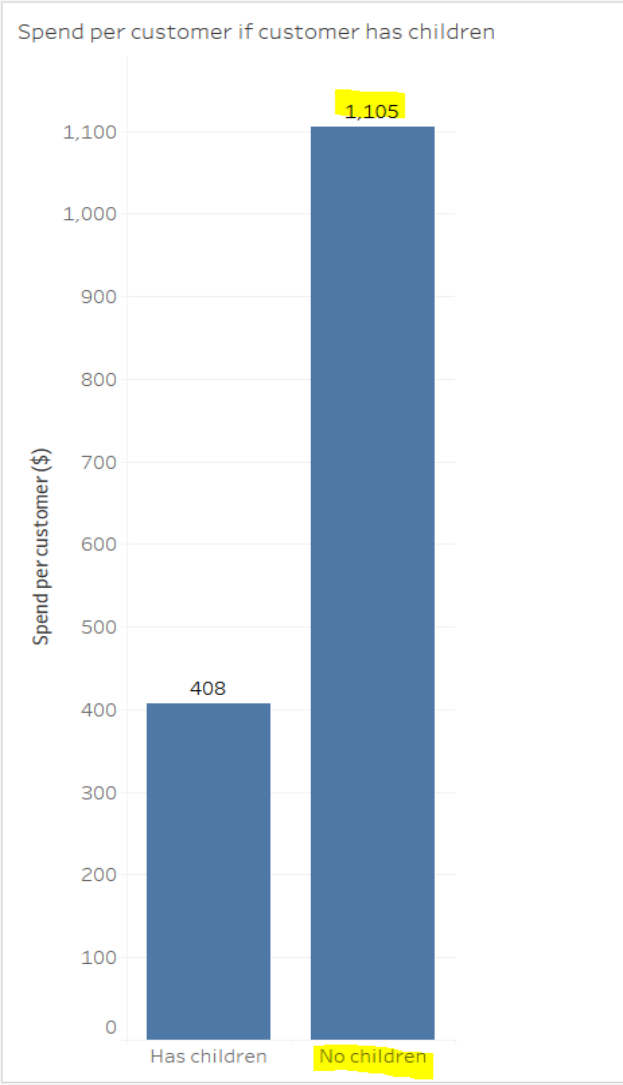
## 5a. Spend per customer by age band




## 5b. Spend per customer by income band



## 5c. Spend per customer if customer has children or not




## 5d. Share of spend per product by country

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | |
| 11 | | AU | CA | GE | IN | MN | SA | SP | US |
| 12 | Liquor | 50% | 50% | 50% | 47% | 55% | 50% | 51% | 48% |
| 13 | Non veg | 26% | 27% | 28% | 30% | 26% | 28% | 27% | 30% |
| 14 | Commodities | 8% | 7% | 8% | 8% | 7% | 7% | 7% | 7% |
| 15 | Fish | 6% | 6% | 6% | 6% | 7% | 6% | 6% | 7% |
| 16 | Chocolates | 5% | 5% | 4% | 4% | 4% | 4% | 5% | 4% |
| 17 | Veg | 4% | 5% | 4% | 5% | 0% | 4% | 4% | 4% |
| 18 | total | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| 19 | | | | | | | | | |

**5e.** **Share of spend per product by age band**

| | 15-29 | 30-44 | 45-59 | 60-74 | 75-119 | 120-130 |
|---|---|---|---|---|---|---|
| Liquor | 41% | 43% | 51% | 53% | 51% | 40% |
| Non veg | 41% | 32% | 27% | 26% | 29% | 30% |
| Commodities | 4% | 8% | 7% | 7% | 5% | 13% |
| Fish | 7% | 7% | 6% | 6% | 6% | 6% |
| Chocolates | 4% | 5% | 5% | 4% | 4% | 4% |
| Veg | 3% | 5% | 4% | 4% | 4% | 8% |

**5f.** **Share of spend per product by income band**

| | 0K-24K | 25K-49K | 50K-74K | 75K-99K | 100K-124K | 124K-149K | 150K-174K |
|---|---|---|---|---|---|---|---|
| Liquor | 15% | 49% | 54% | 47% | 49% | 0% | 4% |
| Non veg | 30% | 22% | 24% | 33% | 14% | 0% | 95% |
| Commodities | 26% | 13% | 7% | 5% | 11% | 0% | 0% |
| Fish | 11% | 7% | 6% | 7% | 6% | 0% | 0% |
| Chocolates | 9% | 4% | 4% | 5% | 13% | 0% | 0% |
| Veg | 8% | 4% | 4% | 4% | 7% | 0% | 0% |

**5g.** **Percentage of conversions by advertising channel and country, showing effectiveness of Facebook in US.**

| country | twitter | instagram | facebook | brochure | bulkmail | total_conversions |
|---|---|---|---|---|---|---|
| SP | 25% | 25% | 22% | 5% | 24% | 100% |
| CA | 28% | 24% | 21% | 7% | 21% | 100% |
| SA | 23% | 24% | 23% | 5% | 24% | 100% |
| GER | 29% | 21% | 18% | 5% | 26% | 100% |
| IND | 26% | 16% | 18% | 5% | 34% | 100% |
| AUS | 18% | 35% | 21% | 0% | 26% | 100% |
| US | 23% | 19% | 27% | 0% | 31% | 100% |
| ME | 0% | 0% | 0% | 0% | 100% | 100% |

**5h.** **Total conversions by advertising channel and country showing effectiveness Bulk mail in India and US.**

| country | twitter | instagram | facebook | brochure | bulkmail | total_conversions |
|---|---|---|---|---|---|---|
| SP | 25% | 25% | 22% | 5% | 24% | 100% |
| CA | 28% | 24% | 21% | 7% | 21% | 100% |
| SA | 23% | 24% | 23% | 5% | 24% | 100% |
| GER | 29% | 21% | 18% | 5% | 26% | 100% |
| IND | 26% | 16% | 18% | 5% | 34% | 100% |
| AUS | 18% | 35% | 21% | 0% | 26% | 100% |
| US | 23% | 19% | 27% | 0% | 31% | 100% |
| ME | 0% | 0% | 0% | 0% | 100% | 100% |

**5i.** **Total conversions by advertising channel and country showing Brochure ineffectiveness**

| country | twitter | instagram | facebook | brochure | bulkmail | total_conversions |
|---|---|---|---|---|---|---|
| SP | 25% | 25% | 22% | 5% | 24% | 100% |
| CA | 28% | 24% | 21% | 7% | 21% | 100% |
| SA | 23% | 24% | 23% | 5% | 24% | 100% |
| GER | 29% | 21% | 18% | 5% | 26% | 100% |
| IND | 26% | 16% | 18% | 5% | 34% | 100% |
| AUS | 18% | 35% | 21% | 0% | 26% | 100% |
| US | 23% | 19% | 27% | 0% | 31% | 100% |
| ME | 0% | 0% | 0% | 0% | 100% | 100% |

# Appendix 6

**Limitations of data provided**

1. Absence of cost/profitability data
2. Outdated information i.e. customer join dates are over 10 years old from current year (2024)
3. GDPR issues i.e. no information available on data collection methods
4. Relatively small sample size of 2,216 customers

**Alternative approaches**

1. Conduct an RFM analysis as this is a commonly used metric in the retail industry and we have data on recency (R), frequency (F), monetary value (M) to conduct this analysis to identify the best to the worst customers.
2. Advanced analytics using predictive and prescriptive modelling using ML tools like Python.
3. A/B testing on products and advertising to understand customer behavior better.
4. Qualitative analysis with the use of surveys/customer feedback.
5. Competitor analysis to help with strategic decisions

---

*References:*

*World Wide Web Consortium (W3C). (2018). Web Content Accessibility Guidelines (WCAG) 2.1. Retrieved from https://www.w3.org/TR/WCAG21/*

*Knaflic, Cole Nussbaumer. Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley, 2015.*