2M Market

Technical Report:
Demographics, Sales, and Marketing Channel Analysis

Andrea Rossi

Table of Contents

1. Business Overview and Objectives

2Market is a worldwide retail chain that offers products online and in physical stores. With a commitment to data-driven decision-making, 2Market seeks to deepen its understanding of customer behaviours and preferences to inform the new marketing campaign strategy.

This report aims to address key knowledge gaps, focusing on three main areas:

1. Customer demographics
2. The effectiveness of advertisement channels in driving sales
3. Understanding product categories' performance across demographics and geographic segments.

This information is intended for the marketing department and will empower the business in several ways. Specifically, it will enable recommendations to optimise preliminary marketing strategies, allocate budget efficiently and identify sales growth opportunities.

2. Analytical Approach

1.1. Data Processing, Cleaning and Analysis

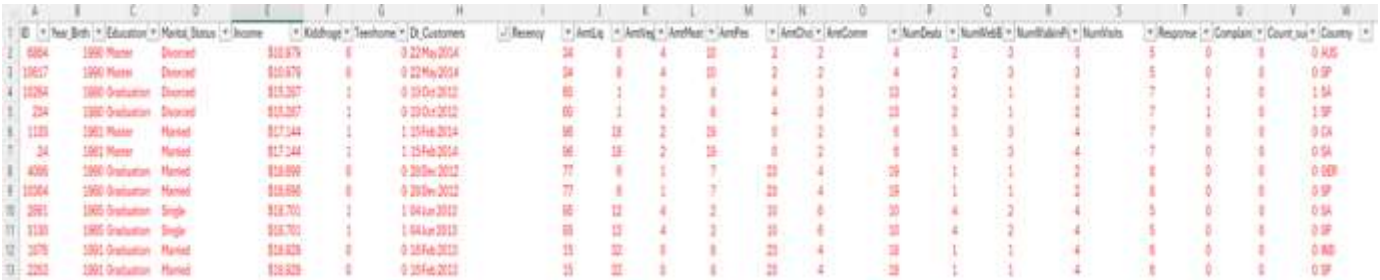
The data preparation and cleaning of marketing_data.csv and ad_data.cvs followed two distinct stages. Stage 1 involved using Excel for bulk data cleaning and initial exploration, suitable for medium to small-sized datasets. In stage 2, the PostgreSQL open-source RDBMS was employed for database creation, manipulation, and more complex descriptive analysis.

It should be noted that due to the discrepancies in the data, some assumptions were made to facilitate analysis. To ensure accuracy, it's essential to verify the data sources with the stakeholders and implement measures to prevent inconsistencies in the future.

1.1.1. Stage 1

The dataset files were checked for data validity, accuracy, completeness, consistency, uniqueness, and timeliness, enhancing overall data quality and integrity.

In particular, Duplicates (N: 203) were identified and removed using the CONCAT function (A2:V2) and Conditional Formatting. Instances with identical attributes except for the Country Code were treated as duplicates based on the unlikelihood of finding these occurrences given the relatively small dataset size. (e.g. ID: 6684-&10617, 1135 & 24; see Fig 1 for examples).



When resolving these duplicates, a simplistic assumption was made: the CustomerID corresponding to the country with the higher occurrences was retained as correct. Therefore, it should be noted that countries with lower frequencies might be underrepresented in this analysis (e.g., Germany, United States, Australia, and India). The same approach was applied to records with identical attributes, but a different boolean value, as boolean-type data, are more susceptible

to data entry errors (e.g., ID: 9671 & 8975, 4491& 873, 535 & 4608). Further data validation led to the removal of IDs 11004, 1159, and 7829, as age inconsistencies were found with values over 100 years of age.

Date formats were standardised as dd-mm-by, and the Income column was converted to a currency format ($). Inconsistencies in the Marital Status categorical column were also handled (See Appendix 1 for segmentation criteria).

1.1.2. Analysis and Key Findings

In Excel, pivot tables and bar charts were leveraged to summarise and visualise common customer demographics, income, age, and marital status:

- The consumer base has a median yearly income of $51,550 and an average age of 52, suggesting that 2Market is currently attracting middle- to high-earners and failing to appeal to the youngest segments.

- Income generally grows with age progression yet beyond the highest income threshold ($ 100.000 and above); this doesn't hold true. The majority of customers are married or in a relationship (64%) and are relatively younger than widowed and divorced, whilst singles (22%) are, on average, the youngest. The customer base also presents high levels of education.

1.2. Stage 2

The 2Market Datasets tables were structured in a database (2M_Analysis, Fig. 1) to facilitate manipulation, data warehousing, and analysis.

At the staging level, the cleaning process primarily involved renaming columns to improve clarity (Appendix 4), segmenting demographic groups (Customer_Age to Customer_Age_Groups, Kiddhome & TeenHome to Household_type), and creating new calculated fields. (see Appendix 2).

To determine how spending behaviours vary across different demographic or geographic segments, SQL queries calculated the average amount spent on products for each category (see Appendix 3).

SELECT and LEFT JOIN clauses were then used to merge the marketing and advertisement tables based on the common PK "Id." With the Marketing_data table cleaned in Excel beforehand, the LEFT JOIN ensured retention of all Marketing_data rows in the result set. This enabled descriptive analysis of ad channels' effectiveness by segment and country (appendix 3)

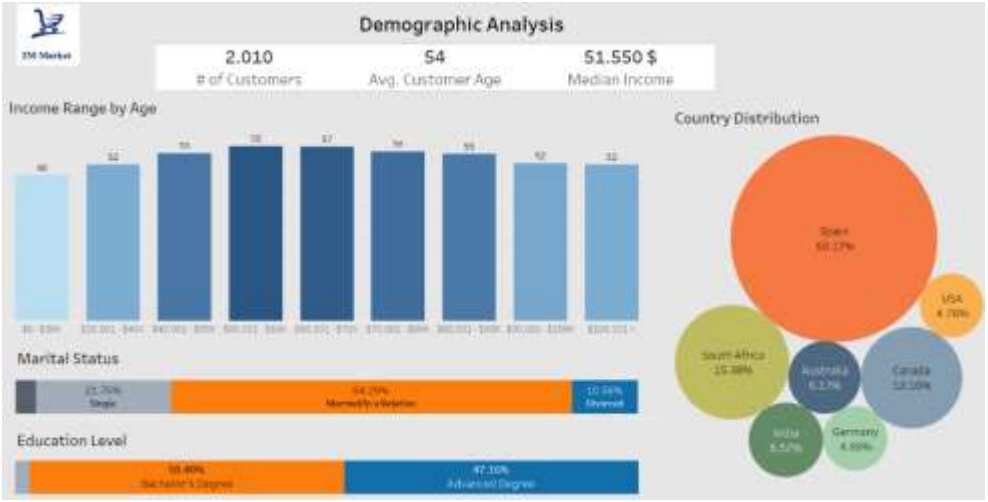
1.2.1. Analysis and Key Findings:

- Spain dominates the market share, but Germany, South Africa, and the USA show higher spending, indicating potential for increased ROI in those regions.
- Liquor products drive sales, with spending variations across demographics and locations, suggesting profit maximization opportunities.
- Digital advertisement outperformed printed promotional materials.
- One-size-all does not work as the platform effectiveness changes across countries and segments.
- High lead conversions via social media corresponded with high sales in Spain and Canada.

3. Dashboard Design and Development

Building on the above data exploration and analysis, Tableau was employed for its advanced visualisation features, allowing for a more comprehensive and intuitive analysis. Through the software, three dashboards were created, sized to ensure compatibility with all devices and tested for functionality. The design prioritises clarity and simplicity, with colours chosen for readability and accessibility.

Demographic Dashboard: offers insight into the customer base composition and will inform the marketing team about the audience characteristics. It establishes a context for the sales variations and will be key to tailoring marketing efforts and strategic planning. (See Fig 1. below)

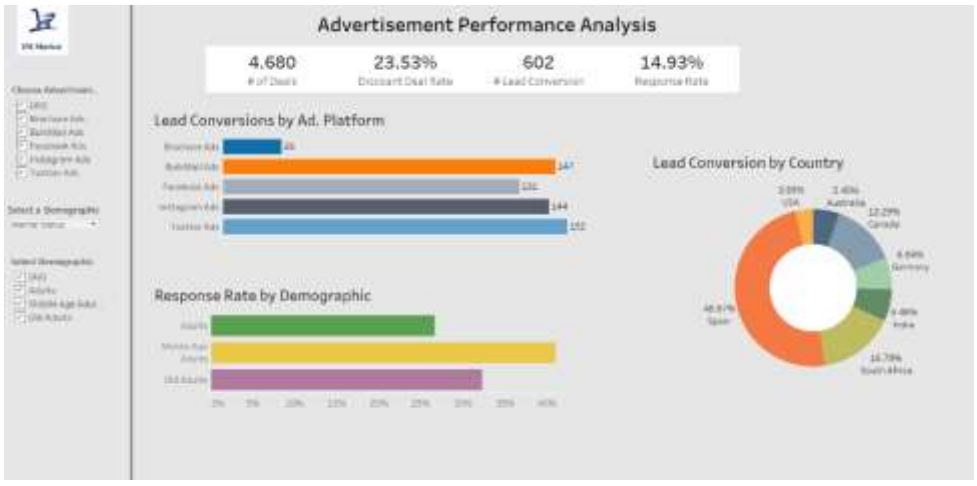


The Key metrics include average consumer age and median income. Dynamic visualisations, such as bar charts, stack charts, and bubble charts, compare the percentage distribution of marital status and education levels as well as the average distribution of age by income range.

Product Performance Dashboard: provides insight into product sales and variation across demographics (bar charts) and geographic locations (doughnut chart). (See Fig. 2 below)

This data aids the 2Market team in visualising valuable information in terms of customer needs, product popularity and purchasing habits. Users can compare them using a Product filter and the "Select a Demographic" parameter. Additionally, the "Avg Amount Spent" filter enables quick comparison of high-spending segments.

Advertisement Performance Dashboard: reports on the effectiveness of the different advertisement platforms in driving sales, and it evaluates, based on past acceptance, the propensity of different segments to respond positively to future marketing campaigns (Response Rate). This data will inform decisions regarding platform selection and budget allocation and highlight demographic groups which are more receptive, aiding in targeted marketing strategies. (Fig 3., below)



4. Recommendations: Marketing Optimisation, Budget allocation and Growth areas

- Alcoholic beverages and meat products are the best sellers, and as such, the marketing strategy should leverage their popularity to drive sales.
- Aggregating data based on average spending allows us to identify low- and high-spending customers. Efforts should be addressed to those segments that present the highest affinity with these products, specifically older, educated Adults with high incomes. Future analysis should aim for more precise correlations using statistical software.
- Spain holds the largest market share, but higher average spending was observed in Germany, South Africa and the US, suggesting potential for reallocating marketing resources and increasing ROI in these countries. Further, this highlights the opportunity to implement strategies to encourage higher spending in Spain.
- Overall, Twitter was the most effective platform in driving sales, yet there are differences across demographic segments and locations and a combination of different digital advertisement platforms is recommended. To target higher spending customers specifically, this should be focused on Instagram, Facebook and Twitter, however for decision making, further analysis of cost-effectiveness by ads is recommended.
- Advertisement efforts were more successful in Spain and South Africa compared to the other countries. It is important to highlight that countries where sales were the highest also observed high success in lead conversions through Social Media, suggesting effective marketing in these countries and opportunities for market expansion in countries with low lead conversion by prioritising these advertisement channels.

5. Appendices

Appendix 1: Segmentation and Grouping Criteria

Demographic fields such as Customer Age, Income, Marital Status, Education, and Household type were further segmented into distinct groups or ranges through SQL queries in PGAdmin and grouping and IF statements in Tableau at convenience.

- **Marital status field:**

  "Yolo" and "Alone" were grouped as "Single".

  "Absurd" was relabelled as "Unknown".

  Other statuses were grouped as being married or in a relationship.

- **Education field:**

  "2nd Cycle", "Masters", and "Ph.D." were grouped as Advanced Degree.

  "Graduation" was renamed as Bachelor's degree.

  "Basic" was renamed as High School Diploma.

- **Income:**

  Income was categorized into "Income Ranges", with each bracket representing a $10,000 increment up to $100,000 and above.

- **Customer Age: SQL query below**

  Ages between 27-44 were categorized as Adults.

  Ages 45-59 were categorized as Middle Age Adults.

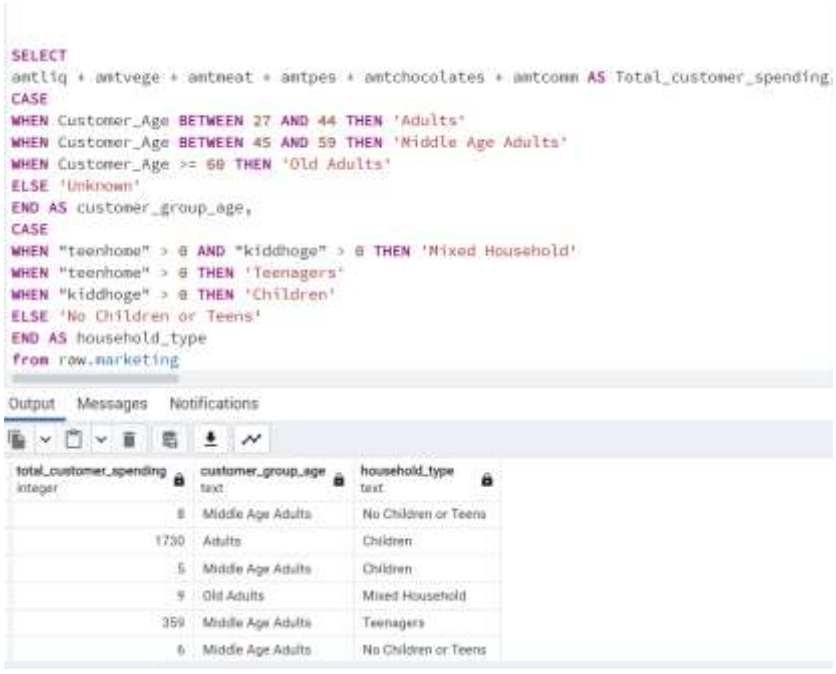
  Ages over 60 were categorized as Old Adults.

- **Household Type** (SQL query below)

  Having children was categorized as "With Children".

  Having teenagers was categorized as "With Teenagers".

  Having both children and teenagers was categorized as "Mixed Household".



```
SELECT
amtliq + amtvege + amtmeat + amtpes + amtchocolates + amtcomm AS Total_customer_spending,
CASE
WHEN Customer_Age BETWEEN 27 AND 44 THEN 'Adults'
WHEN Customer_Age BETWEEN 45 AND 59 THEN 'Middle Age Adults'
WHEN Customer_Age >= 60 THEN 'Old Adults'
ELSE 'Unknown'
END AS customer_group_age,
CASE
WHEN "teenhome" > 0 AND "kiddhoge" > 0 THEN 'Mixed Household'
WHEN "teenhome" > 0 THEN 'Teenagers'
WHEN "kiddhoge" > 0 THEN 'Children'
ELSE 'No Children or Teens'
END AS household_type
from raw.marketing
```

Output | Messages | Notifications

total_customer_spending integer	customer_group_age text	household_type text
8	Middle Age Adults	No Children or Teens
1730	Adults	Children
5	Middle Age Adults	Children
9	Old Adults	Mixed Household
359	Middle Age Adults	Teenagers
6	Middle Age Adults	No Children or Teens

Appendix 2: Column renaming queries:

```
ALTER TABLE staging.stg_marketing
RENAME COLUMN "kiddhoge" TO "Kid_home"

ALTER TABLE staging.stg_marketing
RENAME COLUMN "dt_customers" TO "customer_joindate"

ALTER TABLE staging.stg_marketing
RENAME COLUMN "numwebbuy" TO "num_web_Pur",

ALTER TABLE staging.stg_marketing
RENAME COLUMN "numwalkinpur" TO "num_store_pur"

ALTER TABLE staging.stg_marketing
RENAME COLUMN "numvisits" TO "num_month_webvisits"

ALTER TABLE staging.stg_marketing
RENAME COLUMN count_success TO "total_numleadconv"

ALTER TABLE staging.stg_marketing
RENAME COLUMN amtpes TO "amtfish"
```

Appendix 3: Data Analysis:

- Total spend per country:

```
SELECT "country", SUM("customer_spending") AS total_customer_spending
FROM staging.stg_marketing
GROUP BY "country"
ORDER BY total_customer_spending DESC;
```

Output Messages Notifications

country character	total_customer_spending bigint
SP	616316
SA	196505
CA	143947
AUS	69902
IND	68757
GER	65691
US	59695
ME	758

- Total spend per product per country:

```sql
total spend per product and per country

SELECT "country",
SUM("amtvege") AS vege_spend,
SUM("amtmeat") AS meat_spend,
SUM("amtliq") AS liq_spend,
SUM("amtfish") AS fish_spend,
SUM("amtcomm") AS comm_spend,
SUM ("amtchocolates") AS choc_spend
FROM staging.stg_marketing
GROUP BY "country";
```

Data Output | Messages | Notifications

	country character	vege_spend bigint	meat_spend bigint	liq_spend bigint	fish_spend bigint	comm_spend bigint	choc_spend bigint
1	SP	27206	176306	300748	39615	44494	27947
2	CA	5928	36411	76876	8492	10354	5886
3	IND	2778	18932	35033	3563	5280	3171
4	AUS	2896	18835	36265	4299	4759	2848
5	US	2224	16366	31208	2926	4064	2907
6	ME	32	348	180	76	90	32
7	SA	9909	52248	98242	12595	14335	9176
8	GER	2088	18093	34406	4011	4378	2715

- Most popular product by country

```sql
SELECT "country",
CASE
WHEN SUM("amtvege") >= SUM("amtmeat") AND
SUM("amtvege") >= SUM("amtliq") AND
SUM("amtvege") >= SUM("amtfish") AND
SUM("amtvege") >= SUM("amtchocolates") THEN 'Vegetables'
WHEN SUM("amtmeat") >= SUM("amtvege") AND
SUM("amtmeat") >= SUM("amtliq") AND
SUM("amtmeat") >= SUM("amtfish") AND
SUM("amtmeat") >= SUM("amtchocolates") THEN 'Meat'
WHEN SUM("amtliq") >= SUM("amtvege") AND
SUM("amtliq") >= SUM("amtmeat") AND
SUM("amtliq") >= SUM("amtfish") AND
SUM("amtliq") >= SUM("amtchocolates") THEN 'Liquor'
WHEN SUM("amtfish") >= SUM("amtvege") AND
SUM("amtfish") >= SUM("amtmeat") AND
SUM("amtfish") >= SUM("amtliq") AND
SUM("amtfish") >= SUM("amtchocolates") THEN 'Fish'
WHEN SUM("amtchocolates") >= SUM("amtvege") AND
SUM("amtchocolates") >= SUM("amtmeat") AND
SUM("amtchocolates") >= SUM("amtliq") AND
SUM("amtchocolates") >= SUM("amtfish") THEN 'Chocolates'
ELSE 'Commodities'
END AS most_popular_product
FROM staging.stg_marketing
GROUP BY "country";
```

Data Output | Messages | Notifications

	country character	most_popular_product text
1	SP	Liquor
2	CA	Liquor
3	IND	Liquor
4	AUS	Liquor
5	US	Liquor
6	ME	Meat
7	SA	Liquor
8	GER	Liquor

- Most popular product based on marital status.

```
SELECT "marital_status",
CASE
WHEN SUM("amtvege") >= SUM("amtmeat") AND
SUM("amtvege") >= SUM("amtliq") AND
SUM("amtvege") >= SUM("amtfish") AND
SUM("amtvege") >= SUM("amtchocolates") THEN 'Vegetables'
WHEN SUM("amtmeat") >= SUM("amtvege") AND
SUM("amtmeat") >= SUM("amtliq") AND
SUM("amtmeat") >= SUM("amtfish") AND
SUM("amtmeat") >= SUM("amtchocolates") THEN 'Meat'
WHEN SUM("amtliq") >= SUM("amtvege") AND
SUM("amtliq") >= SUM("amtmeat") AND
SUM("amtliq") >= SUM("amtfish") AND
SUM("amtliq") >= SUM("amtchocolates") THEN 'Liquors'
WHEN SUM("amtfish") >= SUM("amtvege") AND
SUM("amtfish") >= SUM("amtmeat") AND
SUM("amtfish") >= SUM("amtliq") AND
SUM("amtfish") >= SUM("amtchocolates") THEN 'Fish'
WHEN SUM("amtchocolates") >= SUM("amtvege") AND
SUM("amtchocolates") >= SUM("amtmeat") AND
SUM("amtchocolates") >= SUM("amtliq") AND
SUM("amtchocolates") >= SUM("amtfish") THEN 'Chocolates'
ELSE 'Commodities'
END AS most_popular_product
FROM staging.stg_marketing
GROUP BY "marital_status";
```

Data Output | Messages | Notifications

	marital_status character	most_popular_product text
1	Unknown	Liquors
2	Married	Liquors
3	Widow	Liquors
4	Single	Liquors
5	Divorced	Liquors

- Most popular product based on whether or not there are children or teens in the home:

```
SELECT
CASE WHEN "teenhome" > 0 THEN 'Teenagers' WHEN "Kid_home" > 0 THEN 'Children' ELSE 'No Children or Teens'END AS household_type,
CASE WHEN SUM("amtvege") >= SUM("amtmeat") AND
SUM("amtvege") >= SUM("amtliq") AND
SUM("amtvege") >= SUM("amtfish") AND
SUM("amtvege") >= SUM("amtchocolates") THEN 'Vegetables'
WHEN SUM("amtmeat") >= SUM("amtvege") AND
SUM("amtmeat") >= SUM("amtliq") AND
SUM("amtmeat") >= SUM("amtfish") AND
SUM("amtmeat") >= SUM("amtchocolates") THEN 'Meat'
WHEN SUM("amtliq") >= SUM("amtvege") AND
SUM("amtliq") >= SUM("amtmeat") AND
SUM("amtliq") >= SUM("amtfish") AND
SUM("amtliq") >= SUM("amtchocolates") THEN 'Liquors'
WHEN SUM("amtfish") >= SUM("amtvege") AND
SUM("amtfish") >= SUM("amtmeat") AND
SUM("amtfish") >= SUM("amtliq") AND
SUM("amtfish") >= SUM("amtchocolates") THEN 'Fish'
WHEN SUM("amtchocolates") >= SUM("amtvege") AND
SUM("amtchocolates") >= SUM("amtmeat") AND
SUM("amtchocolates") >= SUM("amtliq") AND
SUM("amtchocolates") >= SUM("amtfish") THEN 'Chocolates'
ELSE 'Commodities'
END AS most_popular_product
FROM staging.stg_marketing
GROUP BY CASE WHEN "teenhome" > 0 THEN 'Teenagers'
WHEN "Kid_home" > 0 THEN 'Children'
ELSE 'No Children or Teens' END;
```

Data Output | Messages | Notifications

	household_type text	most_popular_product text
1	Teenagers	Liquors
2	Children	Liquors
3	No Children or Teens	Liquors

- Distribution of Effective Social Media method of advertising by country:

```
SELECT m.country,
       SUM(a.Instagram_ad) AS instagram_success,
       SUM(a.Twitter_ad) AS twitter_success,
       SUM(a.Facebook_ad) AS facebook_success
FROM raw.marketing m
LEFT JOIN raw.advertisement a ON m.id = a.id
GROUP BY m.country;
```

Output | Messages | Notifications

country character	instagram_success bigint	twitter_success bigint	facebook_success bigint
SP	69	71	67
CA	22	17	17
AUS	8	6	9
IND	6	16	8
US	7	3	5
ME	0	6	0
SA	20	28	10
GER	13	11	9

- Most effective social media by country

```
SELECT m.country,
CASE
WHEN SUM(a.twitter_ad) >= SUM(a.facebook_ad) AND
SUM(a.twitter_ad) >= SUM(a.instagram_ad) THEN 'Twitter'
WHEN SUM(a.facebook_ad) >= SUM(a.twitter_ad) AND
SUM(a.facebook_ad) >= SUM(a.instagram_ad) THEN 'Facebook'
ELSE 'Instagram'
END AS most_effective_social
FROM raw.marketing m
LEFT JOIN raw.advertisement a
ON m.id = a.id
GROUP BY m.country;
```

Output Messages Notifications

country character	most_effective_social text
SP	Twitter
CA	Instagram
AUS	Facebook
IND	Twitter
US	Instagram
ME	Twitter
SA	Twitter
GER	Instagram

- Social media most effective method of advertising based on marital status.

```
SELECT m.marital_status,
CASE
WHEN SUM(a.twitter_ad) >= SUM(a.facebook_ad) AND
SUM(a.twitter_ad) >= SUM(a.instagram_ad) THEN 'Twitter'
WHEN SUM(a.facebook_ad) >= SUM(a.twitter_ad) AND
SUM(a.facebook_ad) >= SUM(a.instagram_ad) THEN 'Facebook'
ELSE 'Instagram'
END AS most_effective_social
FROM raw.marketing m
LEFT JOIN raw.advertisement a
ON m.id = a.id
GROUP BY m.marital_status;
```

Output Messages Notifications

marital_status character	most_effective_social text
Unknown	Facebook
Married	Instagram
Widow	Twitter
Single	Twitter
Divorced	Twitter

- Social media platforms' effectiveness and amount spent per country relation:

```
SELECT
    m.country,
    SUM(a.twitter_ad) AS total_twitter_ad,
    SUM(a.facebook_ad) AS total_facebook_ad,
    SUM(a.instagram_ad) AS total_instagram_ad,
    SUM("amtchocolates") + SUM("amtcomm") + SUM("amtpes") + SUM("amtliq") + SUM("amtmeat") + SUM("amtvege") AS total_spending
FROM raw.marketing m
LEFT JOIN  raw.advertisement a ON m.id = a.id
GROUP BY m.country
ORDER BY m.country DESC;
```

Output Messages Notifications

country character	total_twitter_ad bigint	total_facebook_ad bigint	total_instagram_ad bigint	total_spending bigint
US	3	5	7	59695
SP	71	67	69	616316
SA	28	18	20	196505
ME	0	0	0	758
IND	16	8	6	68757
GER	11	9	13	65691
CA	17	17	22	143947
AUS	6	8	8	69902