Table of Contents

Context of the business

2Market, a global supermarket, wants to gain a deeper understanding of their customer purchase behaviour and the effectiveness of their marketing efforts.

2Market's key areas of interest include:

- Customers' demographics analysis,
- Advertising channels effectiveness,
- Product sales variation based on demographics

Additional questions to ask:

- What are the overall company goals and objectives?
- Are the data up to date?
- Who has access to the data?
- Have data quality checks or cleaning process has been performed?
- Who are the key stakeholders and what are their information needs?
- What are their concerns?
- How will the analysis result be presented and shared?

Analytical approach

In the process of cleaning and analysing data using Excel and SQL, a systematic and methodical approach was undertaken to ensure the accuracy, completeness, and relevance of the datasets.

Data cleaning and Visualization in Excel

The data cleaning process commenced in Excel, focusing on addressing issues such as missing values, outliers, and formatting inconsistencies. Excel's data filtering and sorting functionalities were leveraged to identify and handle missing data points, ensuring a comprehensive dataset for analysis. The 'Remove Duplicates' feature was employed to identify and eliminate any redundant records, enhancing the dataset's integrity.

Standardizing formats across relevant columns, especially in categorical fields like 'Marital Status' and 'Education.' This involved using Excel functions such as 'PROPER' to capitalize the first letter of each word consistently, ensuring uniformity for ease of analysis. For instance, outliers under the marital status, labelled "YOLO" and "Absurd," were replaced to enhance the meaningfulness of the data. Pivot tables were leveraged for data exploration, and visualizations were created to better understanding of the dataset. (Appendix A)

Data analysis using SQL

Upon completing the initial cleaning, the analysis transitioned to SQL for a more robust and scalable approach. The cleaned datasets were imported into SQL databases, allowing for efficient querying and exploration.

Entity-Relationship Diagram (ERD) is developed to elucidate the data structure and relationships. SQL queries were crafted to perform exploratory data analysis, calculating aggregate metrics, filtering operations, and generating actionable insights. Joins were employed to merge relevant tables and establish relationships between datasets, facilitating a comprehensive understanding of the data landscape. Common Table Expressions (CTEs) were employed to streamline query logic and enhance readability. (Appendix B)

Observations

One key observation was the uneven distribution of customer data across countries. With Spain representing approximately 49% of the total customer base, there exists a potential source of bias in decision-making processes. This insight underscores the importance of considering regional discrepancies to ensure that any future decisions are not disproportionately influenced by a single demographic.

In conclusion, the combined use of Excel and SQL provided a comprehensive and nuanced approach to data cleaning and analysis. This methodology enabled a

thorough exploration of the dataset, ensuring that potential biases were identified and considered in subsequent analyses and decision-making processes.

Dashboard design and development

In the process of designing and developing our dashboards in Tableau, several key decisions were made to ensure the effectiveness, simplicity, and security of the visualizations. Leveraging the CSV files previously analysed via SQL queries, we imported the data into Tableau using the union function to create a unified dataset. The dashboards were structured to provide comprehensive insights into Customers' Demographics, Ad Channels, and Spending Patterns.

A fundamental consideration in our dashboard design was the protection of customers' personal data, aligning with GDPR guidelines. De-identifying techniques, such as data binning, were employed to group age information while maintaining privacy. This approach ensures that sensitive details are safeguarded while still allowing for meaningful analyses.

To cater to the diverse areas of interest for our stakeholders, three distinct dashboards were developed. (Appendix C) The Customers' Demographic dashboard focuses on understanding the demographics of the customer base. The Ad Channels dashboard evaluates the effectiveness of various advertising channels. The Customers' Spending Pattern dashboard examines product sales and explores variations based on demographic factors.

In terms of visualization elements, we opted for bar charts, maps, and filters to present data in a clear and interactive manner. The colour scheme was kept simple and straightforward to enhance visual appeal without overwhelming the audience. We adhered to a principle of simplicity in layout, incorporating no more than four worksheets in each dashboard to maintain clarity and ease of interpretation.

To ensure real-time updates and accessibility for stakeholders, we utilized Live connections. This feature enables any changes in the data source to be promptly

reflected in Tableau, providing stakeholders with the latest information conveniently. The rationale behind these decisions was rooted in a user-centric approach, prioritizing clarity, security, and ease of use in our dashboard design and development process.

Patterns, trends, insights, and recommendations

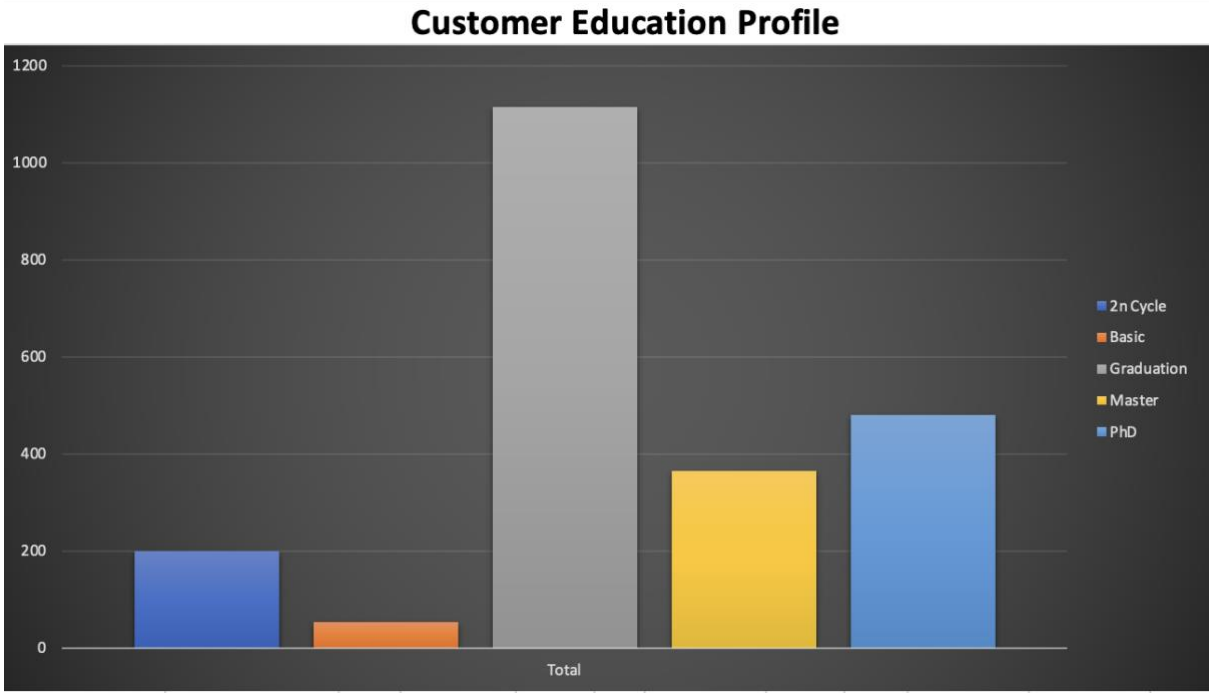
Our analysis of 2 Market's customer data highlights critical patterns and insights. Spain dominates the customer base at 49.34%, urging caution due to potential bias. The 40-59 age group constitutes 55%, emphasizing the need for targeted strategies. Intriguingly, a negative correlation exists between website visits and purchases, suggesting a barrier to online transactions. No clear link between customer income and spending in various product categories implies nuanced consumer behaviour.

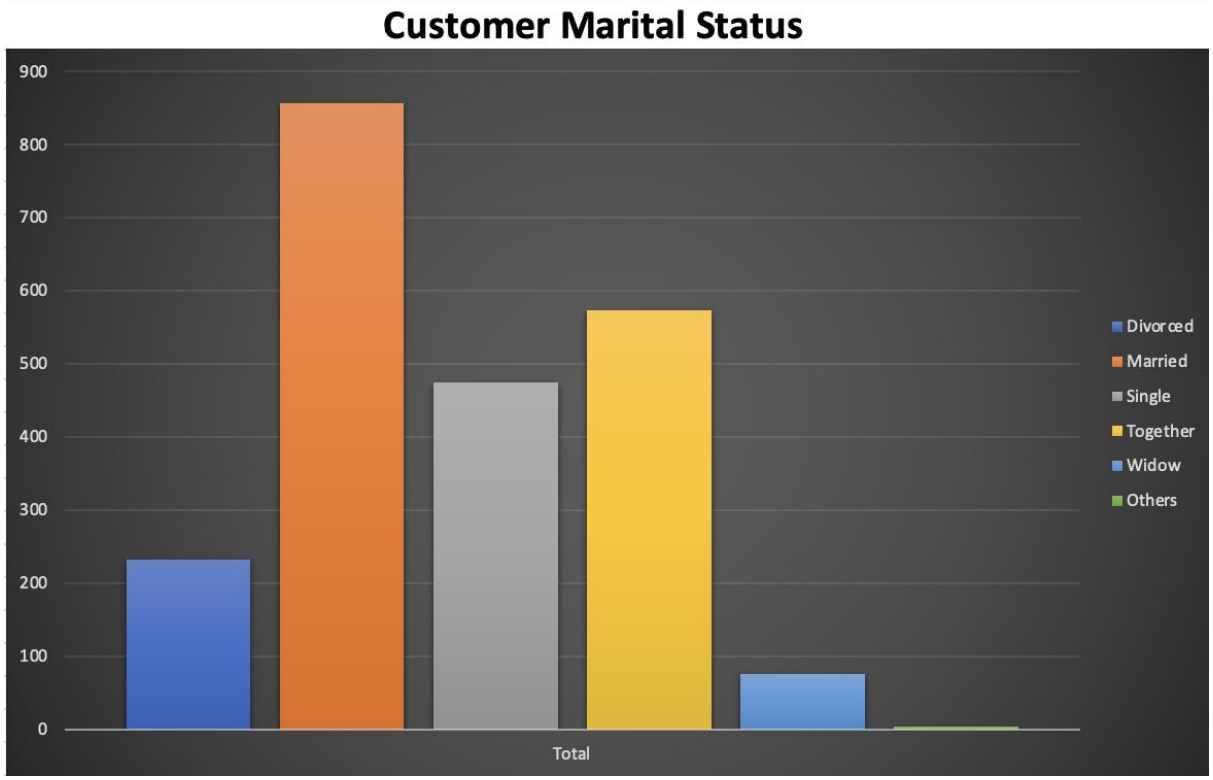
A positive note is the 100% success rate in lead conversions, with alcoholic beverages standing out as the top-selling product globally (50.25%). Twitter, Bulk mail, and Instagram are the most effective advertising channels, each contributing around 24.70%, while brochures lag.

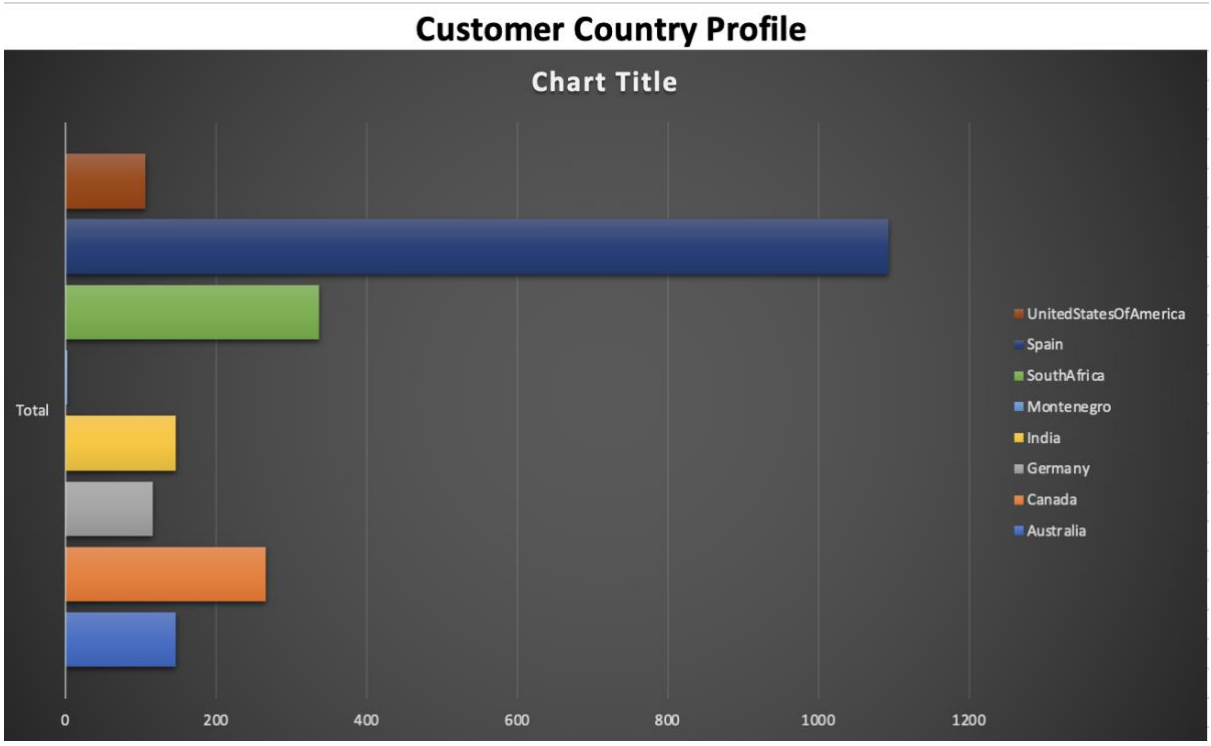
Globally, walk-in purchases dominate at 48%, followed by website purchases at 33% and deals purchases at 13%. Specific product preferences vary regionally, offering strategic insights. Montenegro shows high purchases of alcoholic beverages and fish products, while India leans towards vegetables. Australia customers prefer chocolates and commodities.

Recommendations for further exploration include dissecting average revenue by country and advertising channel, assessing customer retention rates, and understanding the lower web purchase rates compared to walk-ins. These insights equip 2 Market to refine strategies, tapping into nuanced customer behaviours and enhancing overall revenue streams.

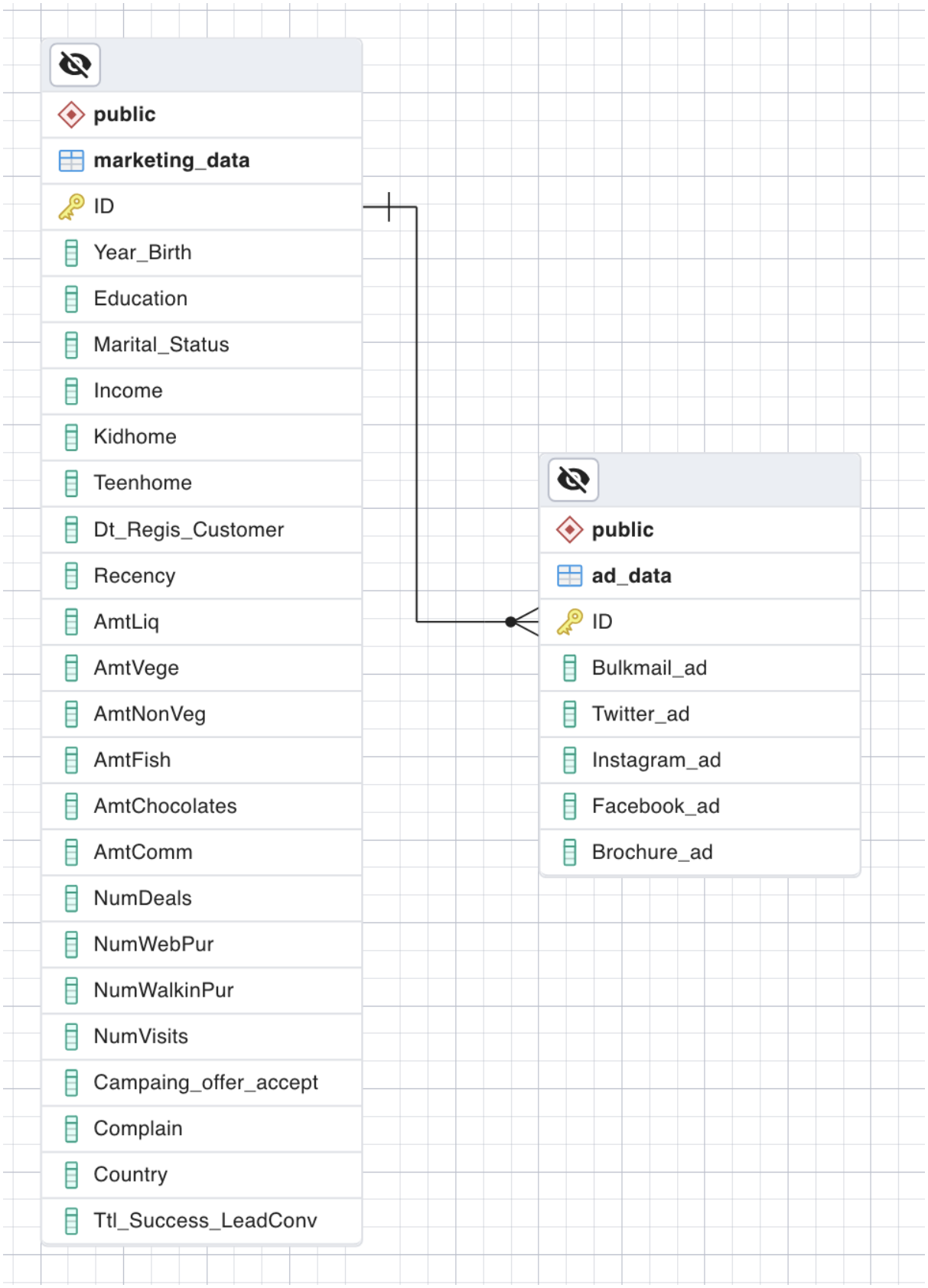
Appendix A. Output from Excel







Appendix B. Output from PgAdmin and SQL Query


public
marketing_data
ID
Year_Birth
Education
Marital_Status
Income
Kidhome
Teenhome
Dt_Regis_Customer
Recency
AmtLiq
AmtVege
AmtNonVeg
AmtFish
AmtChocolates
AmtComm
NumDeals
NumWebPur
NumWalkinPur
NumVisits
Campaing_offer_accept
Complain
Country
Ttl_Success_LeadConv
public
ad_data
ID
Bulkmail_ad
Twitter_ad
Instagram_ad
Facebook_ad
Brochure_ad

percent_customer_country_profile_

Country	customer_count	percentage
SP	1093	49.32
SA	337	15.21
CA	266	12
IND	147	6.63
AUS	147	6.63
GER	116	5.23
US	107	4.83
ME	3	0.14

percent_most_effective_ad_channel_

channel	totalsuccessleads	successpercentage
Twitter_ad	164	24.81
Bulkmail_ad	163	24.66
Instagram_ad	162	24.51
Facebook_ad	142	21.48
Brochure_ad	30	4.54

web_vistVSpurchase_correlation

web_visit_vs_purchase_made_correlation
-0.051226263

percent_purchase_pattern_by_country

Country	liq_percentage	vege_percentage	nonveg_percentage	fish_percentage	chocolates_percentage	comm_percentage
			Minimum %	Maximum %		
SP	51	4.29	27.05	6.09	4.57	7
CA	50.22	4.59	27.43	5.96	4.54	7.25
IND	46.57	4.87	30.5	6.19	4.14	7.73
AUS	49.96	4.31	26.09	6.48	4.82	8.33
US	47.69	4.49	29.88	6.53	4.24	7.16
ME	55.38	0.26	26.17	7.24	3.91	7.05
SA	50.18	4.23	27.67	6.48	4.27	7.17
GER	50.24	4.07	27.69	6.29	3.83	7.88
Minimum	46.57	0.26	26.09	5.96	3.83	7
Maximum	55.38	4.87	30.5	7.24	4.82	8.33
Highest purchase	Montenegro	India	India	Montenegro	Australia	Australia
Lowest purchase	India	Montenegro	Australia	Canada	Germany	Spain

AUS Australia
CA Canada
GER Germany
IND India
ME Montenegro
SA South Africa
SP Spain
US United States of America

Categorise_customer_age_group

age_group	customer_count	percentage
40-59	1228	55
60 or older	690	31
20-39	298	13

income_vs_cusSpendingHabit

age_vs_liquor_correlation	income_vs_vegetable_correlation	income_vs_non_veg_correlation	Min income_vs_fish_correlation	Max income_vs_chocolates_correlation	income_vs_comm_correlation
0.57864975	0.430841681	0.584633357	0.438871359	0.440743792	0.325916446

percent_customer_complain_by_country

Country	total_customers	total_complaints	complaint_percentage
SP	1093	14	1.28
CA	266	2	0.75
IND	147	1	0.68
AUS	147	0	0
US	107	0	0
ME	3	0	0
SA	337	3	0.89
GER	116	1	0.86
Total Number of customers	**2216**		

overall_sucess_rate_of_lead_conversatiion

success_rate
100

percent_best_selling_products

category	spending_percentage
AmtLiq	50.26
AmtNonVeg	27.51
AmtComm	7.24
AmtFish	6.2
AmtChocolates	4.45
AmtVege	4.34

most_effective_ad_channel_by_country

Country	total_customers	total_success_lead_conv_by_country	total_bulkmail_reach_by_country	total_twitter_reach_by_country	total_instagram_reach_by_country	total_facebook_rach_by_country	total_brochure_reach_by_country
AUS	147	34	9	6	12	7	0
CA	266	87	18	24	21	18	6
GER	116	38	10	11	8	7	2
IND	147	38	13	10	6	7	2
ME	3	1	1	0	0	0	0
SA	337	86	21	20	21	20	4
SP	1093	351	83	87	89	76	16
US	107	26	8	6	5	7	0
			1	3	2		

Bulkmal, twitter and IG are most effective channels amoung the countries. Brochure is the least effective one.

customer_marital_status_count

Marital_Status	status_count
Together	573
Others	4
Married	857
Widow	76
Single	474
Divorced	232

customer_purchase_behaviour

Deals	Walkin	Website
5149	9053	12855
19	48	33

2Market_supermarket/postgres@PostgreSQL 13

No limit

Query Query History

```
CREATE TABLE public.ad_data
(
    "ID" INTEGER NOT NULL, "Bulkmail_ad" INTEGER,
    "Twitter_ad" INTEGER, "Instagram_ad" INTEGER,
    "Facebook_ad" INTEGER, "Brochure_ad" INTEGER, CONSTRAINT "PM_ad_data" PRIMARY KEY ("ID")
);
-- Create marketing_data table
CREATE TABLE public.marketing_data
(
    "ID" INTEGER NOT NULL,
    "Year_Birth" INTEGER,
    "Education" VARCHAR(50),
    "Marital_Status" VARCHAR(50),
    "Income" INTEGER,
    "Kidhome" INTEGER,
    "Teenhome" INTEGER,
    "Dt_Regis_Customer" DATE,
    "Recency" INTEGER,
    "AmtLiq" INTEGER,
    "AmtVege" INTEGER,
    "AmtNonVeg" INTEGER,
    "AmtFish" INTEGER,
    "AmtChocolates" INTEGER,
    "AmtComm" INTEGER,
    "NumDeals" INTEGER,
    "NumWebPur" INTEGER,
    "NumWalkinPur" INTEGER,
    "NumVisits" INTEGER,
    "Campaing_offer_accept" INTEGER,
    "Complain" INTEGER,
    "Country" VARCHAR(50),
    "Ttl_Success_LeadConv" INTEGER,
    CONSTRAINT "PM_marketing_data" PRIMARY KEY ("ID")
);
```

2Market_supermarket/postgres@PostgreSQL 13

No limit

Query | Query History

```sql
-- Checking if there is any ID duplication in the table
SELECT
    "ID",
    COUNT(*) AS count_of_duplicates
FROM
    public.marketing_data
GROUP BY
    "ID"
HAVING
    COUNT(*) > 1;

-- Customers' marital status
SELECT
    "Marital_Status",
    COUNT(*) AS status_count
FROM
    public.marketing_data
GROUP BY
    "Marital_Status";


-- Minimum and Maximun year birth
SELECT
    MIN("Year_Birth"), MAX("Year_Birth")
FROM
    public.marketing_data;
```

2Market_supermarket/postgres@PostgreSQL 13

No limit

Query Query History

```sql

-- Income vs Customer spending habits
/*The correlation coefficient ranges from -1 to 1, where -1 indicates a perfect
negative correlation, 1 indicates a perfect positive correlation, and 0 indicates
no correlation. The closer the value is to 1 or -1, the stronger the correlation.*/
SELECT
    corr("Income", "AmtLiq") AS age_vs_liquor_correlation,
    corr("Income", "AmtVege") AS income_vs_vegetable_correlation,
    corr("Income", "AmtNonVeg") AS income_vs_non_veg_correlation,
    corr("Income", "AmtFish") AS income_vs_fish_correlation,
    corr("Income", "AmtChocolates") AS income_vs_chocolates_correlation,
    corr("Income", "AmtComm") AS income_vs_comm_correlation
FROM
    public.marketing_data;

-- Income vs Total spending correlation
SELECT
    corr("Income", "AmtLiq" + "AmtVege" + "AmtNonVeg" + "AmtFish" + "AmtChocolates" + "AmtComm") AS
    income_vs_total_spending_correlation
FROM
    public.marketing_data;


-- What is the relationship between the number of website visits and the number of purchases made?
SELECT
    corr("NumVisits", "NumWebPur") AS web_visit_vs_purchase_made_correlation
FROM
    public.marketing_data;

-- Ans: -0.051226263075050335

```

```sql
-- What is the overall success rate of lead conversions (Count_success)?
/*This query counts the total number of successful lead conversions and divides it by
the total number of rows in the "marketing_data" table.*/
SELECT
    COUNT("Ttl_Success_LeadConv") * 100.0 / COUNT(*) AS success_rate
FROM
    public.marketing_data;
-- Ans: 100%

-- Which advertising channels seem to be the most effective?
WITH ChannelSuccess AS (
    SELECT
        Channel,
        SUM(TotalSuccess) AS TotalSuccessLeads
    FROM (
        SELECT
            'Bulkmail_ad' AS Channel,
            SUM("Bulkmail_ad") AS TotalSuccess
        FROM
            public.ad_data
        UNION ALL
        SELECT
            'Twitter_ad' AS Channel,
            SUM("Twitter_ad") AS TotalSuccess
        FROM
            public.ad_data
        UNION ALL
        SELECT
            'Instagram_ad' AS Channel,
            SUM("Instagram_ad") AS TotalSuccess
        FROM
            public.ad_data
        UNION ALL
        SELECT
```

Query Query History

```sql
            'Facebook_ad' AS Channel,
            SUM("Facebook_ad") AS TotalSuccess
        FROM
            public.ad_data
        UNION ALL
        SELECT
            'Brochure_ad' AS Channel,
            SUM("Brochure_ad") AS TotalSuccess
        FROM
            public.ad_data
    ) AS channels
    GROUP BY
        Channel
)

SELECT
    Channel,
    TotalSuccessLeads,
    ROUND(TotalSuccessLeads * 100.0 / SUM(TotalSuccessLeads) OVER (), 2) AS SuccessPercentage
FROM
    ChannelSuccess
ORDER BY
    SuccessPercentage DESC;

-- Which products seem to sell the best?
SELECT
    category,
    ROUND(total_spending * 100.0 / SUM(total_spending) OVER (), 2) AS spending_percentage
FROM (
    SELECT
        'AmtLiq' AS category,
        SUM("AmtLiq") AS total_spending
    FROM
        public.marketing_data
```

Query Query History

```sql
    UNION ALL
    SELECT
        'AmtVege' AS category,
        SUM("AmtVege") AS total_spending
    FROM
        public.marketing_data
    UNION ALL
    SELECT
        'AmtNonVeg' AS category,
        SUM("AmtNonVeg") AS total_spending
    FROM
        public.marketing_data
    UNION ALL
    SELECT
        'AmtFish' AS category,
        SUM("AmtFish") AS total_spending
    FROM
        public.marketing_data
    UNION ALL
    SELECT
        'AmtChocolates' AS category,
        SUM("AmtChocolates") AS total_spending
    FROM
        public.marketing_data
    UNION ALL
    SELECT
        'AmtComm' AS category,
        SUM("AmtComm") AS total_spending
    FROM
        public.marketing_data
) AS spending_categories
ORDER BY
    spending_percentage DESC;
-- Ans:Liquore with 50.26%
```

```sql
-- Percentage of purchasing pattern by country
WITH product_spending AS (
    SELECT
        "Country",
        SUM("AmtLiq") AS total_liq,
        SUM("AmtVege") AS total_vege,
        SUM("AmtNonVeg") AS total_nonveg,
        SUM("AmtFish") AS total_fish,
        SUM("AmtChocolates") AS total_chocolates,
        SUM("AmtComm") AS total_comm
    FROM
        public.marketing_data
    GROUP BY
        "Country"
)

SELECT
    "Country",
    ROUND(total_liq * 100.0 / (total_liq + total_vege + total_nonveg + total_fish + total_chocolates +
    total_comm), 2) AS liq_percentage,
    ROUND(total_vege * 100.0 / (total_liq + total_vege + total_nonveg + total_fish + total_chocolates +
    total_comm), 2) AS vege_percentage,
    ROUND(total_nonveg * 100.0 / (total_liq + total_vege + total_nonveg + total_fish + total_chocolates +
    total_comm), 2) AS nonveg_percentage,
    ROUND(total_fish * 100.0 / (total_liq + total_vege + total_nonveg + total_fish + total_chocolates +
    total_comm), 2) AS fish_percentage,
    ROUND(total_chocolates * 100.0 / (total_liq + total_vege + total_nonveg + total_fish + total_chocolates +
    total_comm), 2) AS chocolates_percentage,
    ROUND(total_comm * 100.0 / (total_liq + total_vege + total_nonveg + total_fish + total_chocolates +
    total_comm), 2) AS comm_percentage
FROM
    product_spending;
```

```sql
-- Main Query
/* ROUND(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER (), 0) AS percentage: It calculates
the percentage of customers in each age group, rounded to the nearest whole number. */
SELECT
    age_group,
    COUNT(*) AS customer_count,
    ROUND(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER (),0) AS percentage
FROM
    age_groups
GROUP BY
    age_group
ORDER BY
    percentage DESC;

--Ans: 55% 40- 59, 31% 60>, 13% 20-39

-- Percentage of customer country profile
WITH customer_counts AS (
    SELECT
        "Country",
        COUNT(*) AS customer_count
    FROM
        public.marketing_data
    GROUP BY
        "Country"
)

SELECT
    "Country",
    customer_count,
    ROUND(customer_count * 100.0 / SUM(customer_count) OVER (), 2) AS percentage
FROM
    customer_counts
ORDER BY
    customer_count DESC;
```

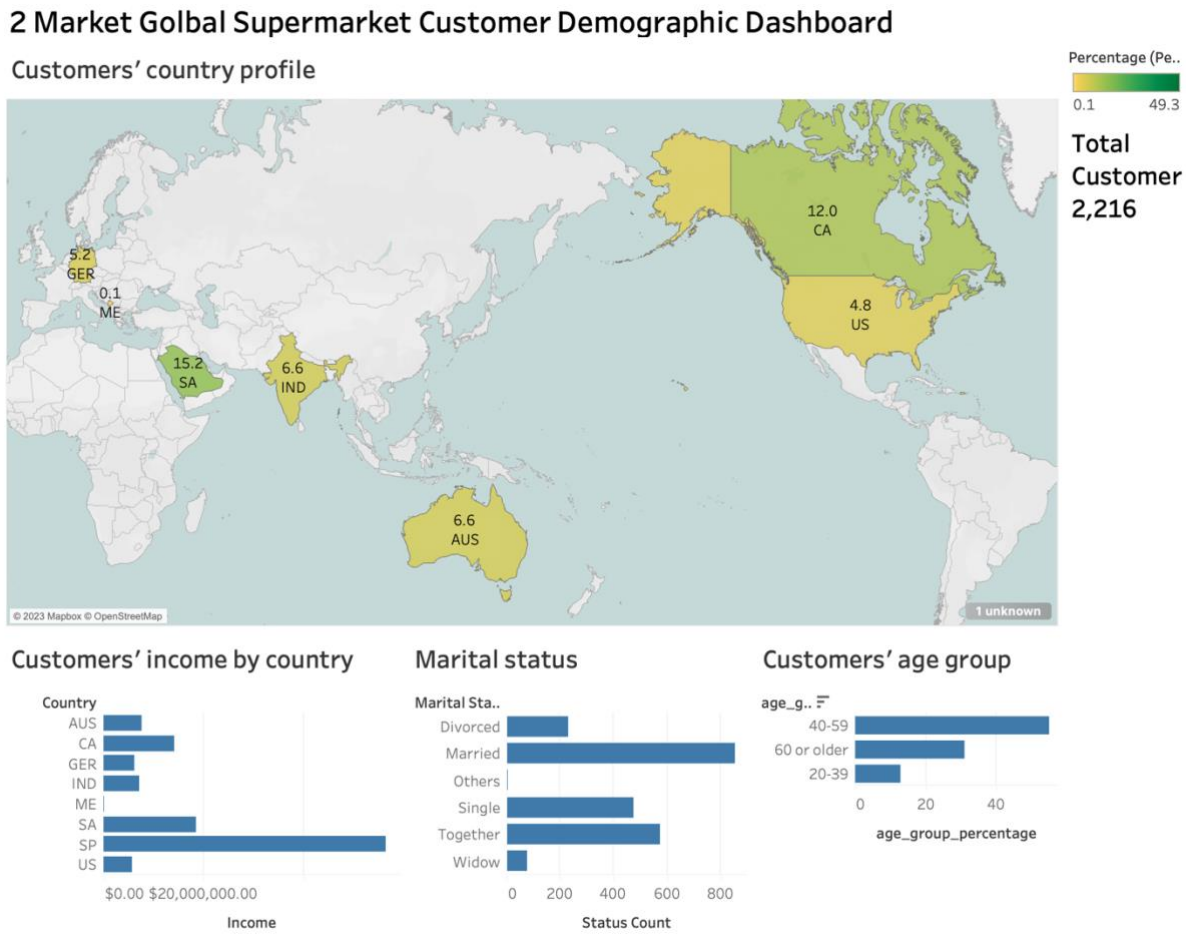
Query Query History

```sql
-- Categorise customer into age group
-- Define Common Table Expression CTE
/* CTE named age_groups is defined. It categorizes
each customer into an age group based on their birth year.*/
WITH age_groups AS (
    SELECT
        CASE
            WHEN "Year_Birth" <= 1963 THEN '60 or older'
            WHEN "Year_Birth" BETWEEN 1964 AND 1983 THEN '40-59'
            WHEN "Year_Birth" BETWEEN 1984 AND 2003 THEN '20-39'
            WHEN "Year_Birth" BETWEEN 2004 AND 2023 THEN '19 or younger'
            ELSE 'Unknown'
        END AS age_group
    FROM
        public.marketing_data
)
```

```sql
-- Percenetage of complain made by customers from different country
WITH customer_complaints AS (
    SELECT
        "Country",
        COUNT("ID") AS total_customers,
        SUM("Complain") AS total_complaints
    FROM
        public.marketing_data
    GROUP BY
        "Country"
)

SELECT
    cc."Country",
    cc.total_customers,
    cc.total_complaints,
    ROUND(cc.total_complaints * 100.0 / cc.total_customers, 2) AS complaint_percentage
FROM
    customer_complaints cc;
```

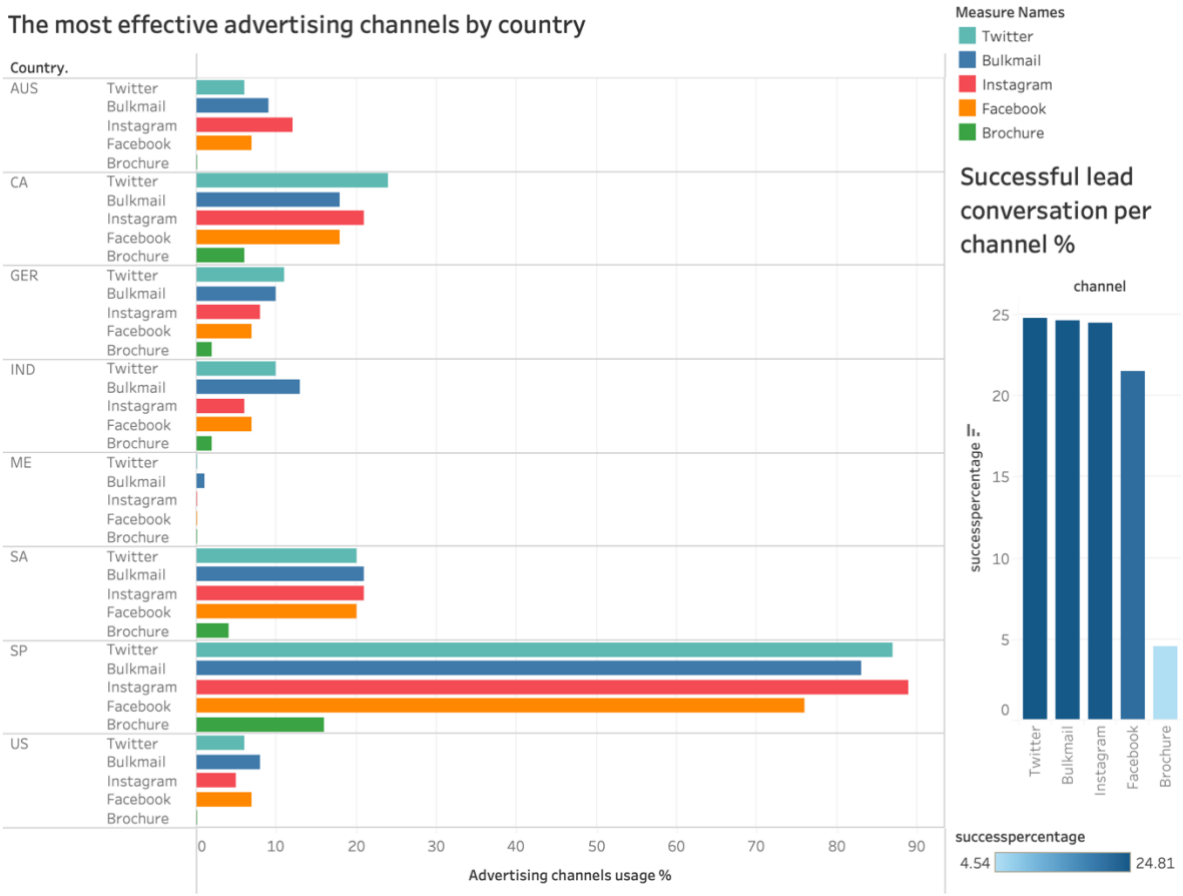
```sql
-- Which ad channel has the highest success lead conversation by country specific?
WITH LeadConversionChannels AS (
    SELECT
        m."ID",
        m."Ttl_Success_LeadConv",
        a."Bulkmail_ad",
        a."Twitter_ad",
        a."Instagram_ad",
        a."Facebook_ad",
        a."Brochure_ad",
        m."Country"
    FROM
        public.marketing_data m
    JOIN
        public.ad_data a ON m."ID" = a."ID"
)

SELECT
    l."Country",
    COUNT(DISTINCT l."ID") AS total_customers,
    SUM(l."Ttl_Success_LeadConv") AS total_success_lead_conv_by_country,
    SUM(l."Bulkmail_ad") AS total_bulkmail_reach_by_country,
    SUM(l."Twitter_ad") AS total_twitter_reach_by_country,
    SUM(l."Instagram_ad") AS total_instagram_reach_by_country,
    SUM(l."Facebook_ad") AS total_facebook_reach_by_country,
    SUM(l."Brochure_ad") AS total_brochure_reach_by_country
FROM
    LeadConversionChannels l
GROUP BY
    l."Country";
```

Appendix C. Tableau Dashboards



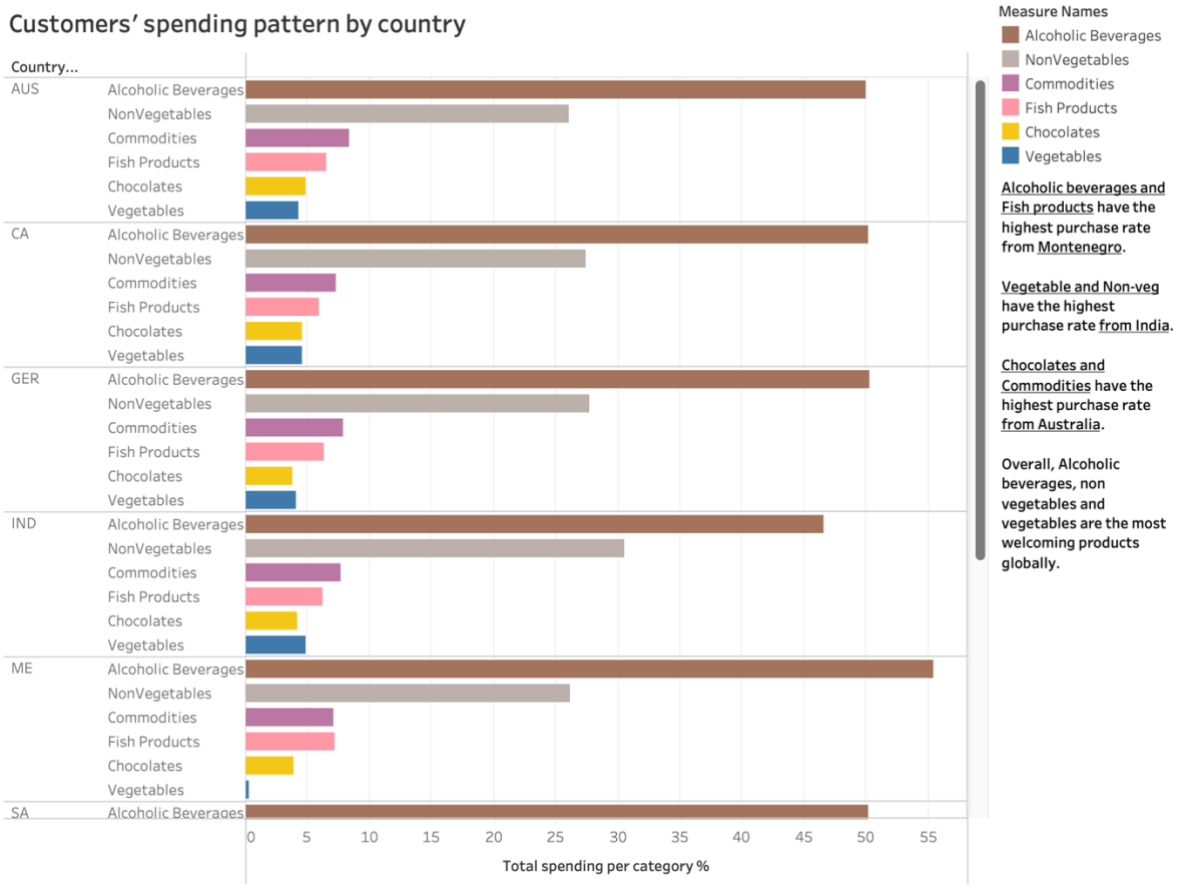
2 Market Global Supermarket Advertising Channel Dashboard

The most effective advertising channels by country



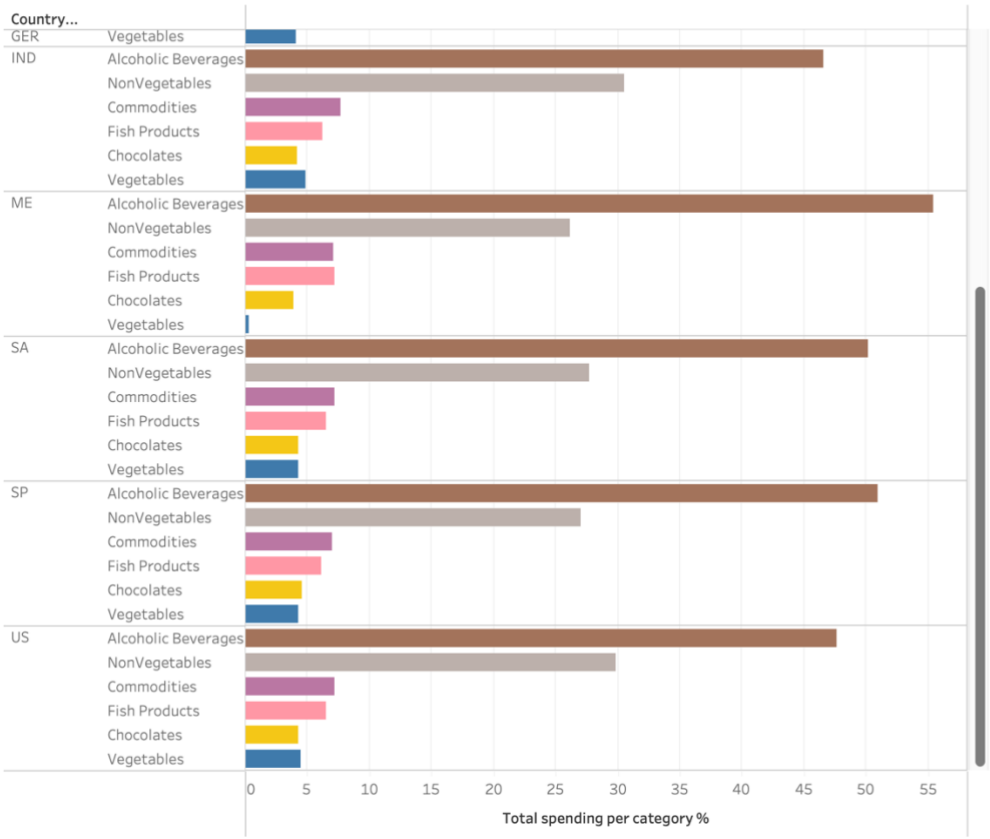
2 Market Global Supermarket Product's Purchasing Power Dashboard

Customers' spending pattern by country



2 Market Global Supermarket Product's Purchasing Power Dashboard

Customers' spending pattern by country



Measure Names
- Alcoholic Beverages
- NonVegetables
- Commodities
- Fish Products
- Chocolates
- Vegetables

Alcoholic beverages and Fish products have the highest purchase rate from Montenegro.

Vegetable and Non-veg have the highest purchase rate from India.

Chocolates and Commodities have the highest purchase rate from Australia.

Overall, Alcoholic beverages, non vegetables and vegetables are the most welcoming products globally.