2MARKET: CUSTOMER PURCHASING BEHAVIOUR
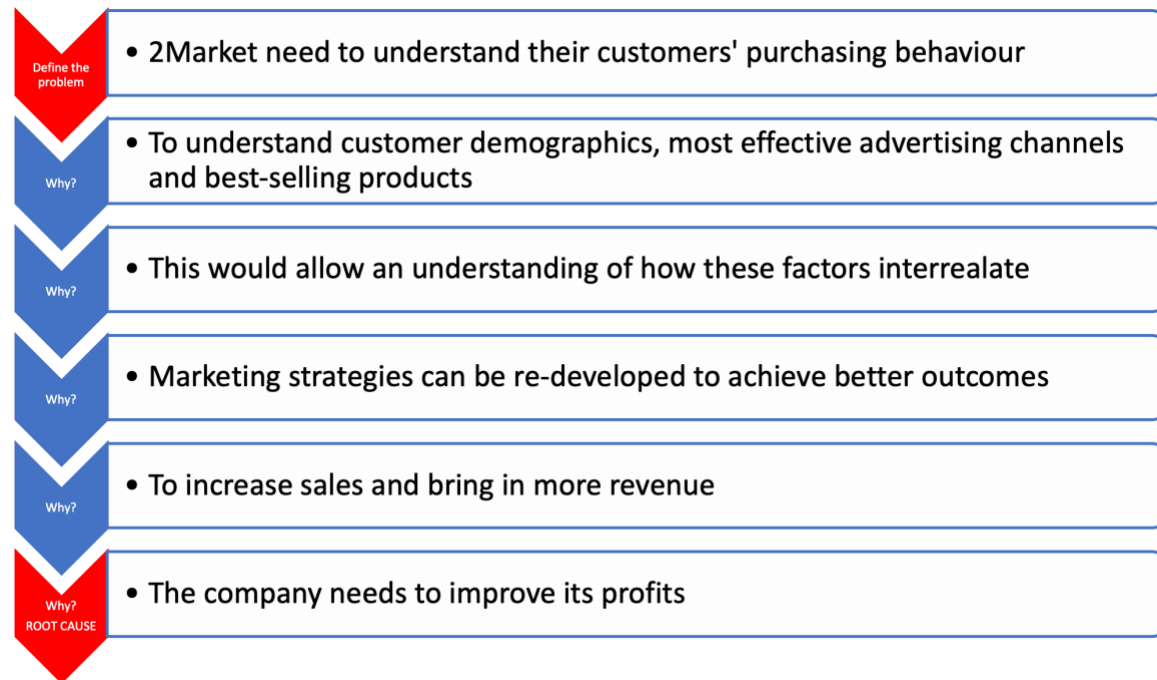
Analysis Report

For the Attention of the Marketing Director

Claire Lawrence

Data Analyst

Introduction

The Business Problem

2Market is a global supermarket operating online and in-store. The problem is identified as below:

Define the problem	• 2Market need to understand their customers' purchasing behaviour
Why?	• To understand customer demographics, most effective advertising channels and best-selling products
Why?	• This would allow an understanding of how these factors interrealate
Why?	• Marketing strategies can be re-developed to achieve better outcomes
Why?	• To increase sales and bring in more revenue
Why? ROOT CAUSE	• The company needs to improve its profits

Report Purpose and Problem Statement

This report analyses 2Market's marketing and ad datasets to inform decision-making by the Marketing Director in developing advertising campaigns, based upon demographics, advertising and product sales. The end goal is to improve profits.

Questions for the Project Team

- Are there background issues (e.g. profits have declined)?
- Are there constraints that the analysis should take account of?

Questions to Ask of the Data

- What is the typical customer demographic?
- Do certain demographics interact with certain advertising channels?
- Do certain product categories sell best?

Analytical Approach

Using Excel

The marketing dataset was cleaned as below:

Action	Result	Rationale	Relevant Observations
Duplicates/blanks identified using conditional formatting/filters.	None found.	Duplicates can skew analysis and blanks can prevent actions working properly.	
Income field data type changed to number. Dollar sign then removed using Find and Replace. Field changed to currency (US$) and decimal places removed.	Income values in correct format and represented as whole numbers.	Ensures data works correctly/is represented clearly.	
Incorrectly-formatted dates corrected to dd/mm/yyyy using the =textsplit() function and concatenated into right order. Corrected values pasted over original column and changed to Date format (as dd/mm/yyyy).	Dates all in same format.	To ensure reliable calculations/pivot table functions.	
New column created to calculate age of customers based upon year of birth using =2022-*cell reference* and applied to all.	Additional column with ages of customers.	Used for age-related analyses.	
Removed outliers by calculating first and third quartiles, interquartile range and upper/lower limits.	Three outliers identified (customers with ages over 120).	Outliers could skew analysis results.	Lower/upper limits were 17 and 89 respectively.
Marital status categories 'Single' and 'Alone' combined into 'Single'. 'Together'	Data combined/renamed.	Accuracy of analysis/clarity.	Ambiguity of 'Together' – left as separate category.

renamed to 'Cohabiting".			
Spend for each product category combined to give total spend per customer (using SUM function).	Combined column.	Valuable for analysis.	
First/third quartiles identified based on total customer spend.	Bottom 25%/top 75% identified. Data extracted into separate sheets.	To understand profiles/behaviour of those customers.	Bottom 25% spent $69 or less. Top 75% spent $1048 dollars or more.

The data was then analysed to understand demographics using:
- **Formulae:** for calculations (e.g. average age);
- **Pivot tables:** to explore relationships within data (i.e. between age, marital status, income, education, country and age);
- **Charts:** to visualise complex data relationships.

Using SQL

SQL was used to explore relationships between the datasets. Queries are shown in the Appendix; not all explorations resulted in useful findings.

The analysis explored the following according to demographics:
- Individual product spend by country;
- Ad lead conversions;
- Most popular products;
- Take-up of deals offered;
- Differences between lowest/highest spenders.

Using Tableau

Tableau was used to understand seasonal trends and the split between online and in-person shopping.

Dashboard Design and Development

The dashboard is designed to address the stakeholder's interests (demographics, effectiveness of advertising and best-selling products). Chosen interactive filters are based on analytical findings and factors found to have greatest impact on trends.

Users can instantly see most popular products and most successful ad types according to country, income range and whether customers have kids/teens. The 'total spend by country' table gives insight into each country's market share.

Filters are grouped top-left for ease of selection/navigation, whilst the two main bar charts are shown parallel to each other and change simultaneously according to filter preferences. A key is shown to explain product categories. All information required for understanding the charts is seen at the top.

The 'colour blind' palette is used for the bottom chart to aid accessibility. Another palette is used for the top chart for differentiation. Text uses a dark colour against a light background and titles/options within filters have been logically named.

The dashboard and its interactivity primarily allow for directed advertisement target planning.

Patterns, Trends and Insights

Customer Demographics

The average customer is:
- 52 years old. Most are in the range 46 to 62;
- Married/living with partner;
- Earning/spending most in their 50s, though trends shift amongst very high earners;
- Well-educated, with 88% of at least graduate level;
- Have kids and/or teens at home (just over half of all customers);
- Reside in Spain (location of almost half (49%)).
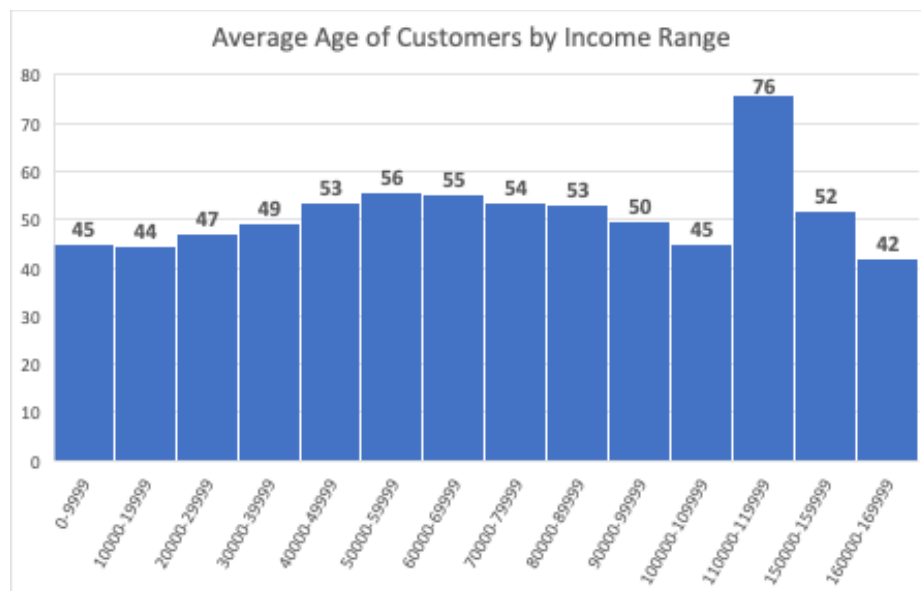
Analysis of the lowest/highest quartiles according to total spend shows the following customer profiles (orange = differences, blue = same trend):

Bottom 25% (Lowest Spenders)	Top 25% (Highest Spenders)
Most have kids and/or teens	*Very few* have kids and/or teens
Earn $20,000 to $40,000	Earn $60,000 to $90,000
Average age = 49	Average age = 53
Majority have taken an offered deal	Majority have taken offered deal
Slightly more purchases in store than online	Slightly more purchases in store than online
Very few have made complaints	Very few have made complaints
Approx. half in Spain	Approx. half in Spain

Therefore, income level and kids/teens are possible predictors for spending level ahead of other trends.

Most Effective Advertising Channels

On average, online channels are of equal success. Brochure advertising is not successful.

Differences are noted relating to the presence of kids/teens:
- Instagram and Facebook conversions were highest where there are *no kids/teens*;
- Bulk email was by far most effective where there *are kids only* at home;
- Twitter was most effective where there *are teens only* at home.

Differences in effectiveness are more noticeable in relation to this indicator than in relation to income or marital status. Additionally, customers with kids are more likely to take up deal offers than those without (possibly linked to general lower spending power).

Best-Selling Products

Alcohol is the best-selling product, followed by meat and then commodities. This is standard across all countries and demographics.

Sales of all products are consistent throughout the year. However, they drop in July and peak in August.

Data Gaps

- There is no detail available regarding product type, online or in-person purchase in relation to deal take-up. Additional data could offer further insight.
- More research needs to be done to understand the relationship between products and successful leads (and why sales drop and peak in July and August).

Recommendations

- Target general ads at the average customer demographic.
- Particularly take income and the presence of kids/teens in the home into account when targeting ads.
- Limit investment in brochure advertising.
- Consider targeting the top 25% spenders, as they have the greatest spending power.

Appendix

```sql
--Joining both tables to make one table:
CREATE TABLE all_data AS
SELECT *
FROM public."marketing_data" md
FULL JOIN public."ad_data" ad USING ("id");

SELECT * FROM public."all_data";

--Total ad conversions by country (joining tables):
SELECT md."country",
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR ad."instagram_ad" = 1 OR
ad."facebook_ad" = 1 OR ad."brochure_ad" = 1
GROUP BY md."country";

--Number of lead conversions where kids/teens are at home:
SELECT "kids_home", "teen_home",
       SUM("bulkmail_ad") AS bulkmail_sum,
       SUM("twitter_ad") AS twitter_sum,
       SUM("instagram_ad") AS instagram_sum,
       SUM("facebook_ad") AS facebook_sum,
       SUM("brochure_ad") AS brochure_sum
FROM public."all_data"
WHERE ("kids_home" >= 1 AND "teen_home" >= 1) AND ("bulkmail_ad" = 1 OR
"twitter_ad" = 1 OR "instagram_ad" = 1 OR "facebook_ad" = 1 OR "brochure_ad" = 1)
GROUP BY "kids_home", "teen_home";

--Number of lead conversions where kids/teens are NOT at home:
SELECT "kids_home", "teen_home",
       SUM("bulkmail_ad") AS bulkmail_sum,
       SUM("twitter_ad") AS twitter_sum,
       SUM("instagram_ad") AS instagram_sum,
       SUM("facebook_ad") AS facebook_sum,
       SUM("brochure_ad") AS brochure_sum
FROM public."all_data"
WHERE ("kids_home" = 0 AND "teen_home" = 0) AND ("bulkmail_ad" = 1 OR "twitter_ad"
= 1 OR "instagram_ad" = 1 OR "facebook_ad" = 1 OR "brochure_ad" = 1)
GROUP BY "kids_home", "teen_home";


--Total product sales where no kids/teens are at home:
SELECT md.kids_home, md.teen_home,
            SUM(amt_liq) total_amt_liq,
            SUM(amt_vege) total_amt_vege,
            SUM(amt_non_veg) total_amt_non_veg,
       SUM(amt_pes) total_amt_pes,
       SUM(amt_chocolates) total_amt_chocolates,
       SUM(amt_comm) total_amt_comm,
       SUM(amt_liq) + SUM(amt_vege) + SUM(amt_non_veg) + SUM(amt_pes) +
SUM(amt_chocolates) + SUM(amt_comm) total_amt
  FROM public.marketing_data md
  WHERE md.kids_home = 0 AND md.teen_home = 0
GROUP BY md.kids_home, md.teen_home


--Total product sales where kids/teens are at home:
SELECT md.kids_home, md.teen_home,
            SUM(amt_liq) total_amt_liq,
            SUM(amt_vege) total_amt_vege,
```

```sql
        SUM(amt_non_veg) total_amt_non_veg,
        SUM(amt_pes) total_amt_pes,
        SUM(amt_chocolates) total_amt_chocolates,
        SUM(amt_comm) total_amt_comm,
        SUM(amt_liq) + SUM(amt_vege) + SUM(amt_non_veg) + SUM(amt_pes) +
SUM(amt_chocolates) + SUM(amt_comm) total_amt
  FROM public.marketing_data md
  WHERE md.kids_home = 1 AND md.teen_home =1
GROUP BY md.kids_home, md.teen_home


--Total product spend by country and marital status:
SELECT md.country, md.marital_status,
            SUM(amt_liq) total_amt_liq,
        SUM(amt_vege) total_amt_vege,
            SUM(amt_non_veg) total_amt_non_veg,
        SUM(amt_pes) total_amt_pes,
        SUM(amt_chocolates) total_amt_chocolates,
        SUM(amt_comm) total_amt_comm,
        SUM(amt_liq) + SUM(amt_vege) + SUM(amt_non_veg) + SUM(amt_pes) +
SUM(amt_chocolates) + SUM(amt_comm) total_amt
  FROM public.marketing_data md
GROUP BY md.country;

--Total number of leads conversions by channel:
SELECT
        SUM("bulkmail_ad") AS bulkmail_sum,
        SUM("twitter_ad") AS twitter_sum,
        SUM("instagram_ad") AS instagram_sum,
        SUM("facebook_ad") AS facebook_sum,
        SUM("brochure_ad") AS brochure_sum
FROM public."all_data"
WHERE "bulkmail_ad" = 1 OR "twitter_ad" = 1 OR "instagram_ad" = 1 OR "facebook_ad"
= 1 OR "brochure_ad" = 1;

--Lead conversions by marital status:
SELECT md."marital_status",
        SUM(ad."bulkmail_ad") AS bulkmail_sum,
        SUM(ad."twitter_ad") AS twitter_sum,
        SUM(ad."instagram_ad") AS instagram_sum,
        SUM(ad."facebook_ad") AS facebook_sum,
        SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR ad."instagram_ad" = 1 OR
ad."facebook_ad" = 1 OR ad."brochure_ad" = 1
GROUP BY md."marital_status";

--Lead conversions without the WHERE clause (same as above but shows YOLO, which
had 0):
SELECT md."marital_status",
        SUM(ad."bulkmail_ad") AS bulkmail_sum,
        SUM(ad."twitter_ad") AS twitter_sum,
        SUM(ad."instagram_ad") AS instagram_sum,
        SUM(ad."facebook_ad") AS facebook_sum,
        SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
GROUP BY md."marital_status";

--table created (for later analysis) showing the total spend per product per
country,
--along with the total lead conversions:
CREATE TABLE country_lead_conversions AS
SELECT md."country",
        'Vege' product,
        SUM(md."amt_vege") AS total_product_spend,
        SUM(ad."bulkmail_ad") AS bulkmail_sum,
```

```
        SUM(ad."twitter_ad") AS twitter_sum,
        SUM(ad."instagram_ad") AS instagram_sum,
        SUM(ad."facebook_ad") AS facebook_sum,
        SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE md."amt_vege" > 0 AND (ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR
ad."instagram_ad" = 1 OR ad."facebook_ad" = 1 OR ad."brochure_ad" = 1)
GROUP BY md."country"
UNION
SELECT md."country",
       'Pes' product,
       SUM(md."amt_pes") AS total_product_spend,
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE md."amt_pes" > 0 AND (ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR
ad."instagram_ad" = 1 OR ad."facebook_ad" = 1 OR ad."brochure_ad" = 1)
GROUP BY md."country"
UNION
SELECT md."country",
       'Non Veg' product,
       SUM(md."amt_non_veg") AS total_product_spend,
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE md."amt_non_veg" > 0 AND (ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR
ad."instagram_ad" = 1 OR ad."facebook_ad" = 1 OR ad."brochure_ad" = 1)
GROUP BY md."country"
UNION
SELECT md."country",
       'Liquor' product,
       SUM(md."amt_liq") AS total_product_spend,
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE md."amt_liq" > 0 AND (ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR
ad."instagram_ad" = 1 OR ad."facebook_ad" = 1 OR ad."brochure_ad" = 1)
GROUP BY md."country"
UNION
SELECT md."country",
       'Comm' product,
       SUM(md."amt_comm") AS total_product_spend,
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE md."amt_comm" > 0 AND (ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR
ad."instagram_ad" = 1 OR ad."facebook_ad" = 1 OR ad."brochure_ad" = 1)
GROUP BY md."country"
UNION
SELECT md."country",
       'Chocolates' product,
```

```sql
        SUM(md."amt_chocolates") AS total_product_spend,
        SUM(ad."bulkmail_ad") AS bulkmail_sum,
        SUM(ad."twitter_ad") AS twitter_sum,
        SUM(ad."instagram_ad") AS instagram_sum,
        SUM(ad."facebook_ad") AS facebook_sum,
        SUM(ad."brochure_ad") AS brochure_sum
FROM public."marketing_data" md
JOIN public."ad_data" ad USING ("id")
WHERE md."amt_chocolates" > 0 AND (ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR
ad."instagram_ad" = 1 OR ad."facebook_ad" = 1 OR ad."brochure_ad" = 1)
GROUP BY md."country";


--Tables created relating to lower and upper quartile spending customers:
CREATE TABLE lower_quartile
(
        ID bigint primary key,
        Year_Birth int,
        Age int,
        Education text,
        Marital_Status text,
        Income varchar(10),
        Kidhome int,
        Teenhome int,
        Dt_Customer varchar(10),
        Recency int,
        AmtLiq int,
        AmtVege int,
        AmtNonVeg int,
        AmtPes int,
        AmtChocolates int,
        AmtComm int,
        Total_Spend int,
        NumDeals int,
        NumWebBuy int,
        NumWalkinPur int,
        NumVisits int,
        Response int,
        Complain int,
        Country varchar(10),
        Count_success int
);

--To check whether created correctly:
SELECT * FROM lower_quartile
LIMIT 10;

CREATE TABLE upper_quartile
(
        ID bigint primary key,
        Year_Birth int,
        Age int,
        Education text,
        Marital_Status text,
        Income varchar(10),
        Kidhome int,
        Teenhome int,
        Dt_Customer varchar(10),
        Recency int,
        AmtLiq int,
        AmtVege int,
        AmtNonVeg int,
        AmtPes int,
        AmtChocolates int,
        AmtComm int,
        Total_Spend int,
        NumDeals int,
        NumWebBuy int,
```

```
        NumWalkinPur int,
        NumVisits int,
        Response int,
        Complain int,
        Country varchar(10),
        Count_success int
);


--Querying lead conversions by marital status (lower and upper quartiles):
SELECT lq."marital_status",
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."lower_quartile" lq
JOIN public."ad_data" ad USING ("id")
WHERE ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR ad."instagram_ad" = 1 OR
ad."facebook_ad" = 1 OR ad."brochure_ad" = 1
GROUP BY lq."marital_status";

SELECT uq."marital_status",
       SUM(ad."bulkmail_ad") AS bulkmail_sum,
       SUM(ad."twitter_ad") AS twitter_sum,
       SUM(ad."instagram_ad") AS instagram_sum,
       SUM(ad."facebook_ad") AS facebook_sum,
       SUM(ad."brochure_ad") AS brochure_sum
FROM public."upper_quartile" uq
JOIN public."ad_data" ad USING ("id")
WHERE ad."bulkmail_ad" = 1 OR ad."twitter_ad" = 1 OR ad."instagram_ad" = 1 OR
ad."facebook_ad" = 1 OR ad."brochure_ad" = 1
GROUP BY uq."marital_status";


--Total product spend (lower and upper quartiles):
SELECT SUM(AmtLiq) total_amt_liq,
       SUM(AmtVege) total_amt_vege,
       SUM(AmtNonVeg) total_amt_non_veg,
         SUM(AmtPes) total_amt_pes,
         SUM(AmtChocolates) total_amt_chocolates,
         SUM(AmtComm) total_amt_comm,
         SUM(AmtLiq) + SUM(AmtVege) + SUM(AmtNonVeg) + SUM(AmtPes) +
SUM(AmtChocolates) + SUM(AmtComm) total_amt
  FROM public.lower_quartile;

SELECT SUM(AmtLiq) total_amt_liq,
       SUM(AmtVege) total_amt_vege,
       SUM(AmtNonVeg) total_amt_non_veg,
         SUM(AmtPes) total_amt_pes,
         SUM(AmtChocolates) total_amt_chocolates,
         SUM(AmtComm) total_amt_comm,
         SUM(AmtLiq) + SUM(AmtVege) + SUM(AmtNonVeg) + SUM(AmtPes) +
SUM(AmtChocolates) + SUM(AmtComm) total_amt
  FROM public.upper_quartile;
```