Background / Context of the Business

2Market is a global supermarket with both online and in-store operations. The company is seeking to understand its customers' purchase behaviour to inform an upcoming marketing campaign by analysing three key areas: customer demographics, product success rates, and advertisement channel effectiveness. Although the brief does not specify the underlying reason for this initiative, I assume—based on industry practice—that 2Market intends to inform the upcoming marketing campaign with the goal of increasing sales and/or reducing costs.

To help frame the problem, I used the 5 Whys technique:

1. Why analyse customer behaviour? → To improve marketing campaign effectiveness
2. Why improve campaigns? → To increase sales efficiency
3. Why increase sales efficiency? → To better target customers with the right products
4. Why better targeting? → To reduce marketing waste and avoid ineffective channels
5. Why reduce waste? → To maximise ROI and gain competitive advantage

- Root Cause: To maximise return on investment (ROI) and maintain a competitive advantage in the market.

In line with this reasoning, this data analytics project aims to extract actionable insights by identifying key demographic trends, understanding which products and ads perform well, and exploring relationships between demographics and product preferences. That will help define relevant KPIs, visualise patterns, and present practical recommendations to support 2Market's marketing and business goals.

Questions I would ask the 2Market Team

- What is the timeline of transactions in the dataset? Are we looking at daily, weekly, or monthly data?
- Who exactly will be the audience for the final presentation? (I am currently assuming it's the Marketing Team, and the presentation time is 8 minutes.)
- How long do I have to present? Will there be time for Q&A?
- Do you have any initial hypotheses about which products or channels perform best?
- What is driving the decision to launch a new marketing campaign at this time?
- Given the tight timeframe, how realistic is it to design a campaign without already knowing your customer demographics?

Questions I'd Like to Ask of the Data

- Are certain demographics consistently purchasing specific products?
- Which advertising channels result in the highest conversion or engagement rates?
- Are there seasonal or time-based trends in customer purchases?
- Are there clusters of customers with similar buying habits?

- Which products are underperforming, and in what customer segments?
- What is the return rate of products or customer churn rate (if available)?
- How do in-store and online purchase behaviours differ across demographics?

Analytical Approach

To begin the analysis, I imported the marketing_data.csv file into Excel and confirmed that the dataset contained 2,216 unique customer records. I first ensured data quality by checking for missing values and duplicates, but none were found. Since the dataset provided birth years rather than age, I created a new column to calculate age using =2024 - YEAR(BirthDate). I noticed a wide age range (27 to 130), with customers over 100 years old flagged as outliers. Many of these hadn't made purchases in over 90 days, suggesting outdated or inaccurate records.

Next, I cleaned the income column by removing currency symbols and converting it to numeric format. I also reformatted the product spending and date columns for clarity and analysis. Using =AVERAGE(range) and the SUBTOTAL function, I calculated the overall average age (53) and the average age by marital status. I filtered income between $90,000 and $100,000 to find this group had an average age of around 52. Most customers had no children or teens at home—an important demographic insight.

I created bar and scatter plots to explore relationships between age, marital status, and income. I also used SQL to validate key findings and followed best practices such as clear formatting, handling nulls with COALESCE, and using LEFT JOIN for data integrity.

Please see the appendix for details, calculations, results, Excel charts, and SQL queries.

Dashboard Design and Development

When designing the dashboard, I kept 2Market's goal front and centre: helping the marketing team better understand their customers and make data-driven decisions for their upcoming campaign. I wanted the dashboard to feel clear, practical, and easy to explore—even for someone who is not a data expert.

I structured the layout in a logical, top-down flow. It starts with a broad demographic overview and gradually dives deeper into purchasing behaviour. Every chart answers a specific question I had during analysis—so the story behind the data naturally unfolds for the user.

I included three main visualisations:

A bar chart to compare the average age across different marital statuses—it shows which segments are older or younger.

A scatter plot to explore the relationship between age and income—perfect for spotting trends and income bands by age.

A horizontal bar chart and filtered table that zoom in on customers earning between $90,000 and $100,000—so the team can quickly analyse this high-value group.

To make it interactive, I added filters for income, marital status, and age. I wanted users to feel free to explore without getting lost or overwhelmed.

In terms of design, I used a neutral background and a simple colour scheme—blue for income, green for age, orange for spending—so everything stayed clean and consistent. I also made sure fonts were readable and colour choices were accessible for everyone, including users with colour vision impairments.

Throughout the process, I focused on keeping things intuitive and visually balanced. I wanted the dashboard to feel approachable, insightful, and helpful—something the 2Market team could use to guide accurate marketing decisions.

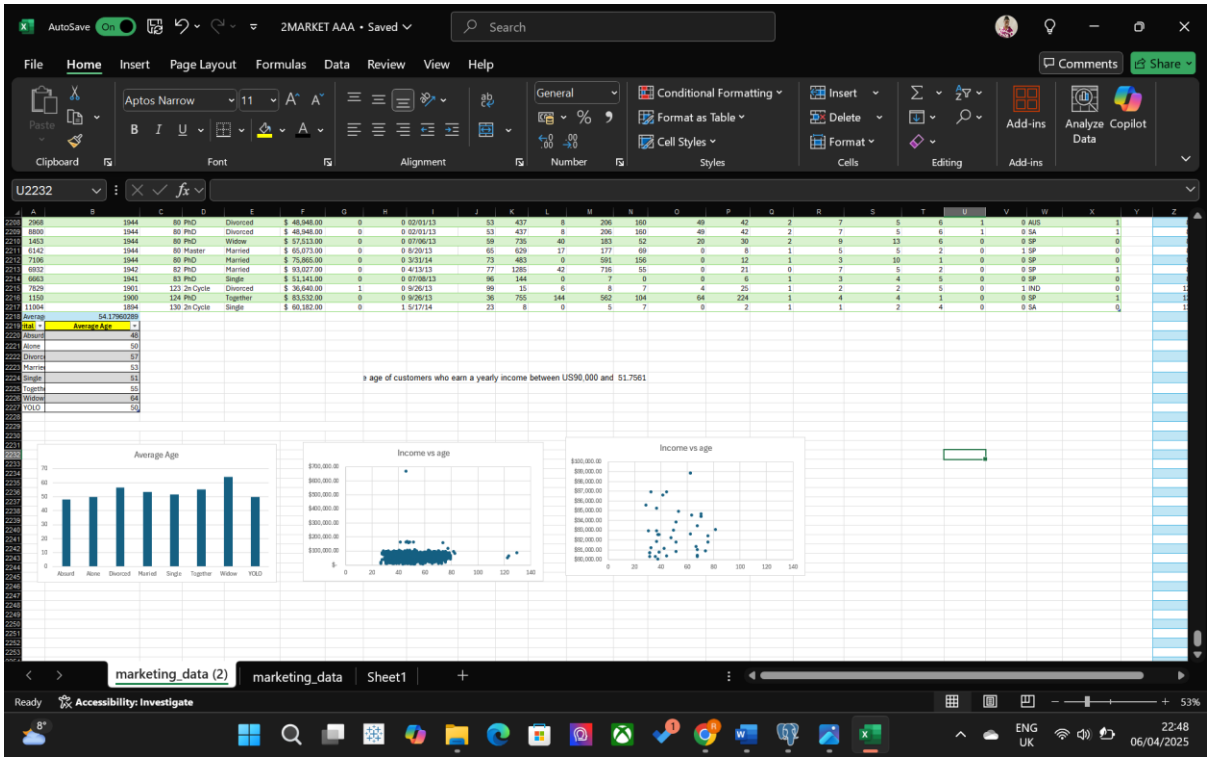
Patterns, Trends, and Insights

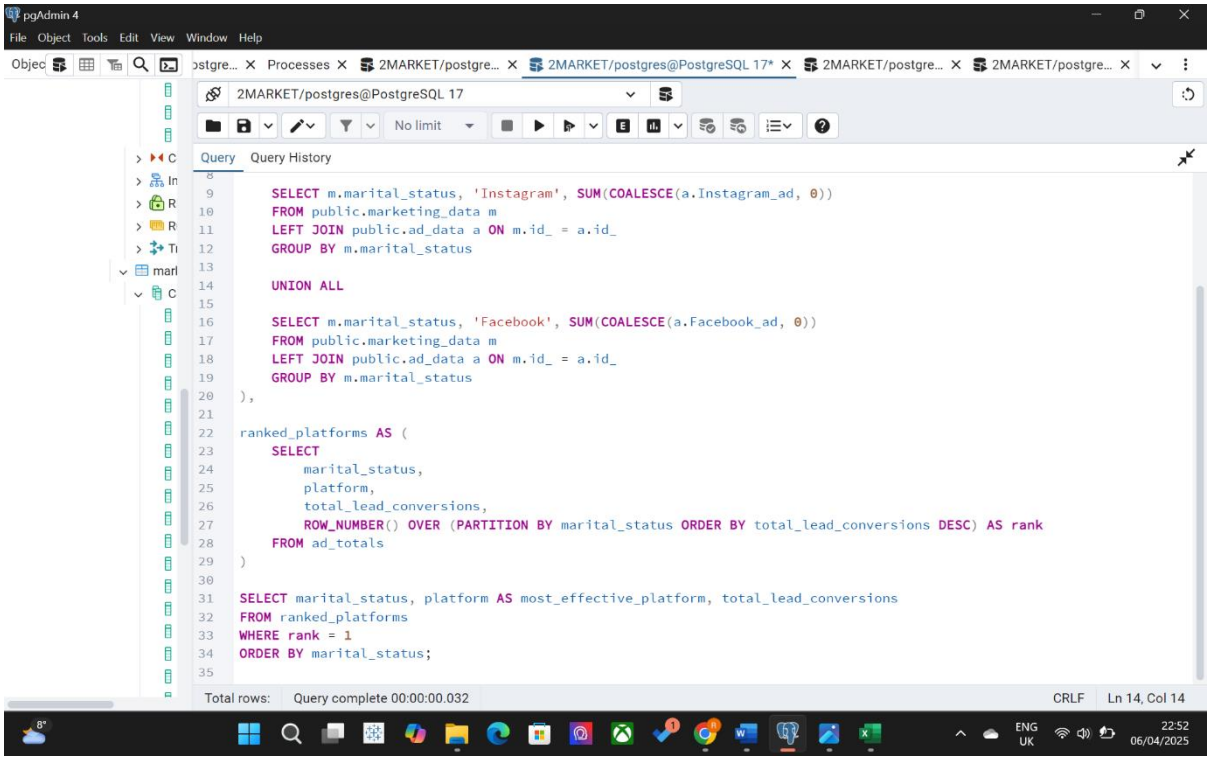
While exploring the data through the dashboard, I noticed several trends that could meaningfully shape 2Market's marketing strategy. One standout insight is the connection between age and marital status—widowed customers are the oldest, while single and divorced customers are generally younger. This suggests that segmenting campaigns by life stage, rather than just income, could improve targeting precision. Income tends to increase with age, but a notable number of younger customers (ages 27–45) are earning high incomes—an emerging segment worth focusing on.

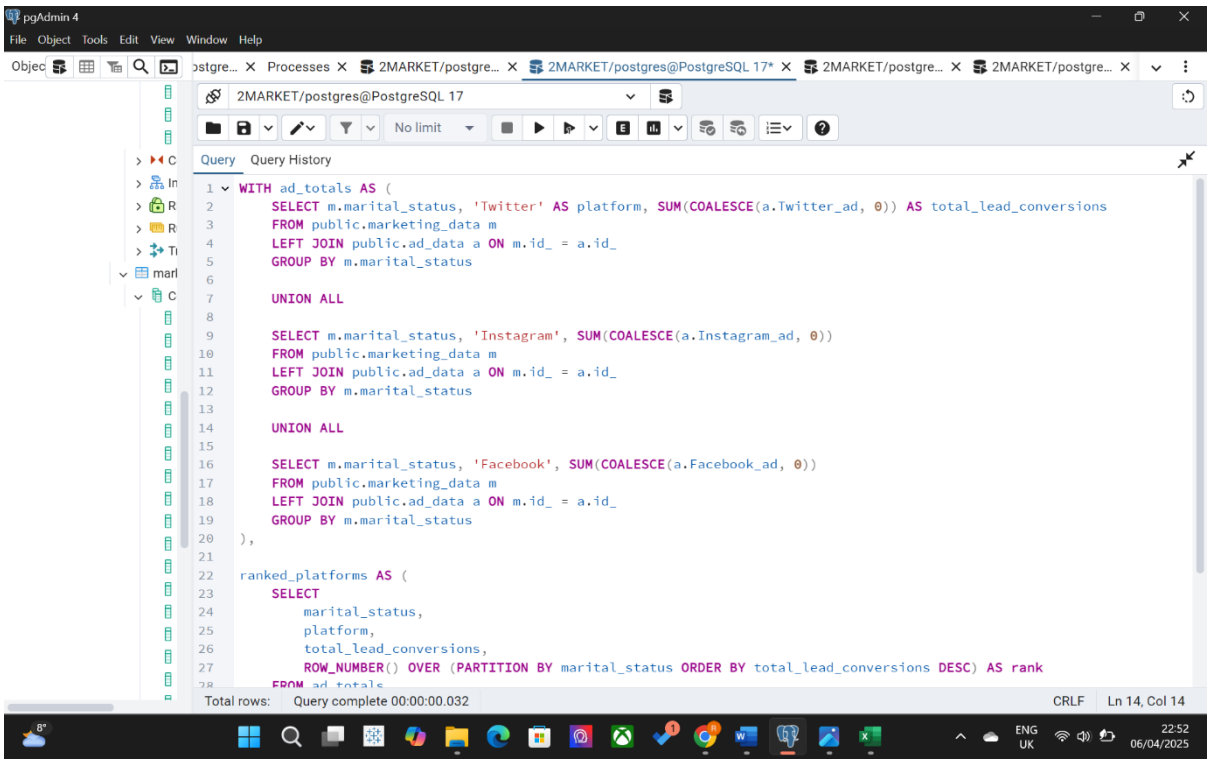
Most customers have no children or teenagers at home, suggesting family-centric messaging might be less effective. I also spotted data anomalies—such as a customer aged 130—that point to outdated records and the need for regular data cleansing. Additionally, the dataset lacks an explicit transaction timestamp, limiting trend analysis.

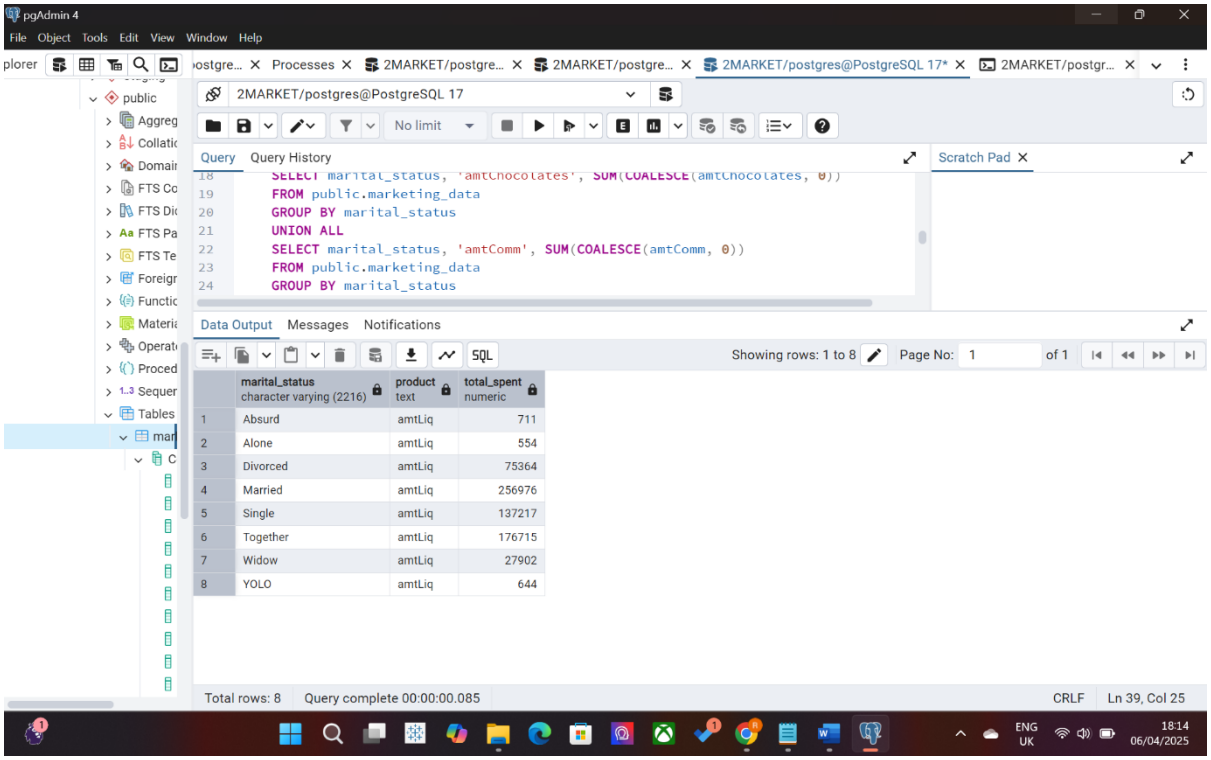
- The top Advertising channels across all demographics are: Twitter, Instagram and Bulk mail
- The most popular products across all demographics are alcoholic beverages.

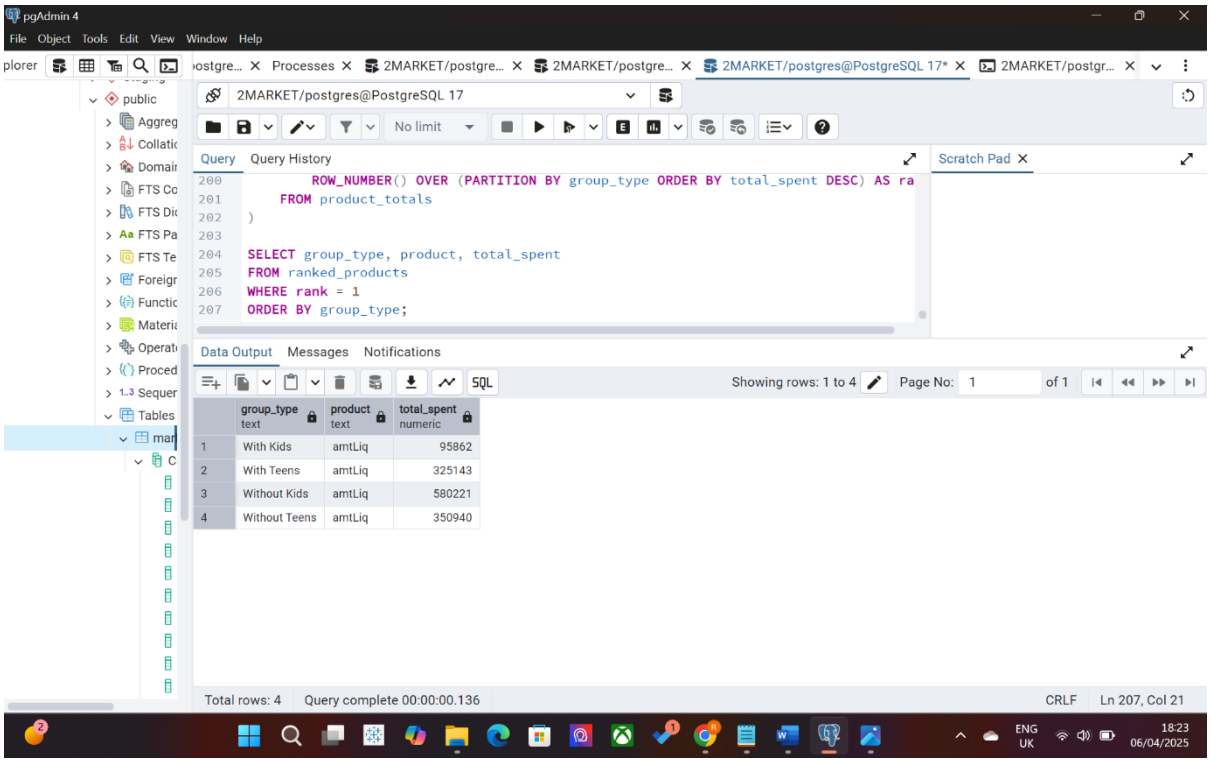
Appendix

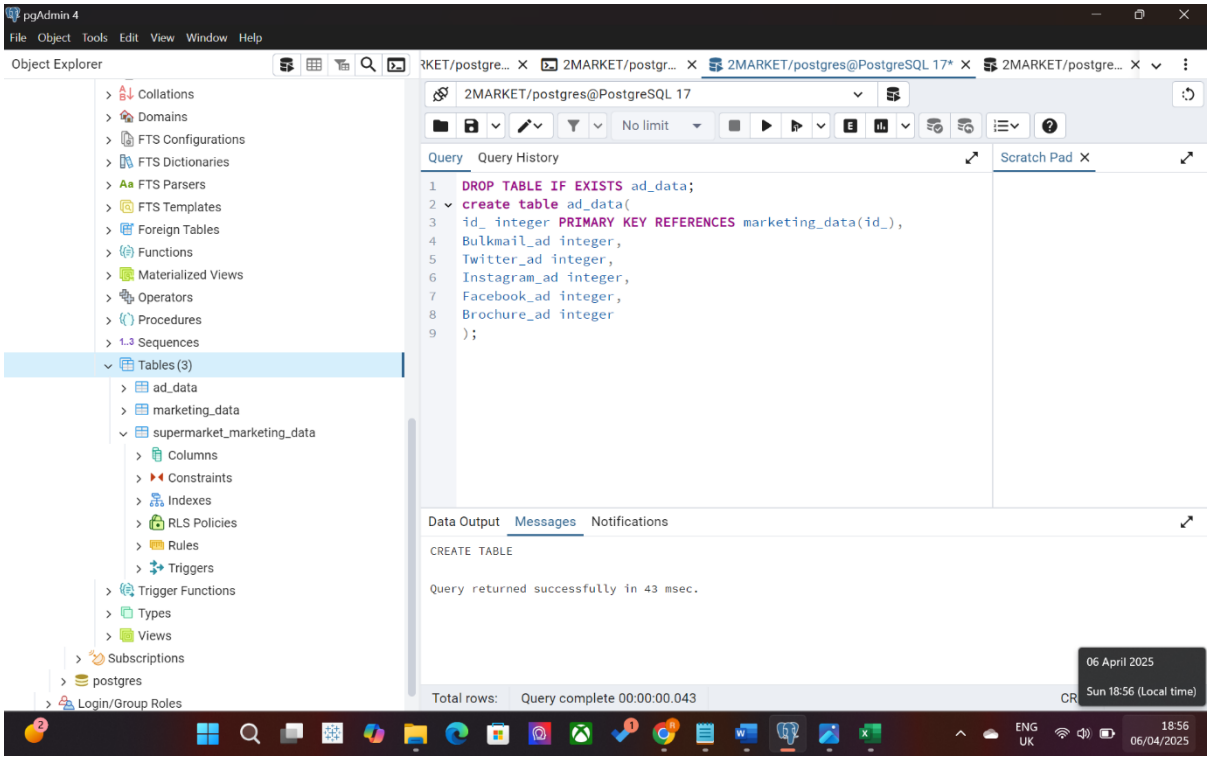


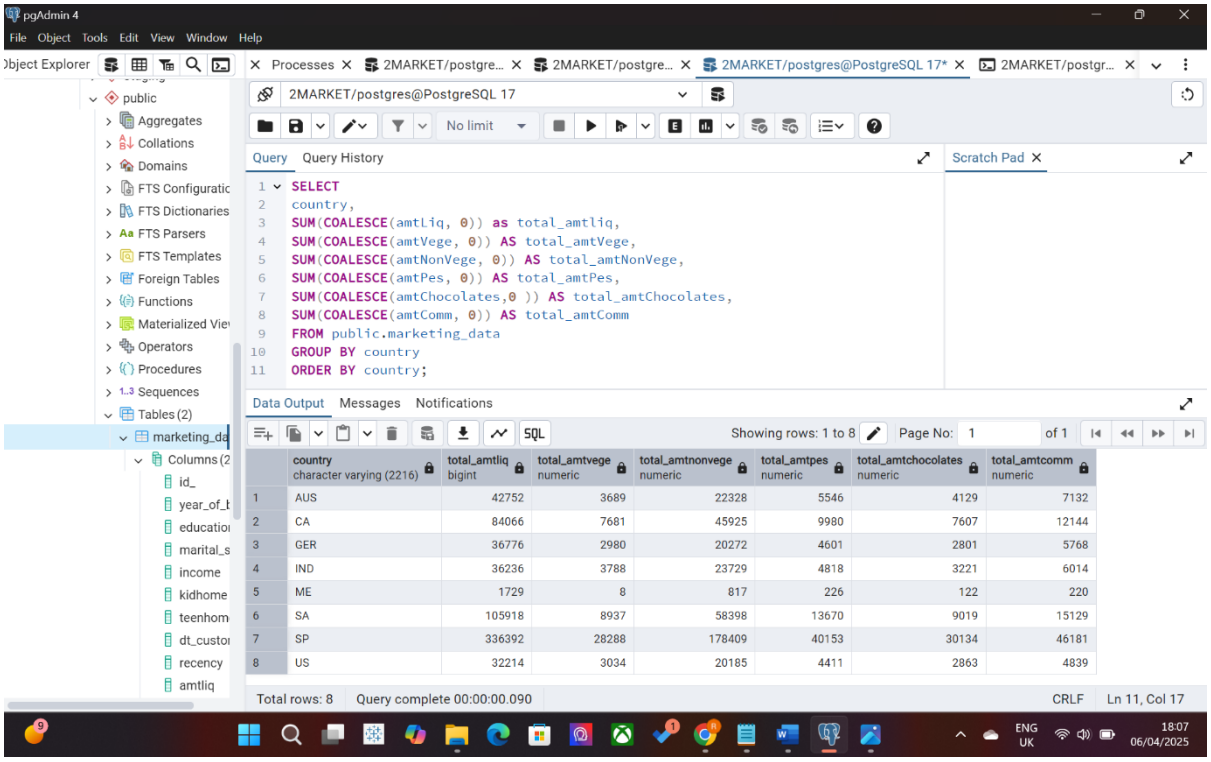

pgAdmin 4
File Object Tools Edit View Window Help
Processes
2MARKET/postgres@PostgreSQL 17*
2MARKET/postgres@PostgreSQL 17
No limit
Query
Query History
SELECT m.marital_status, 'Instagram', SUM(COALESCE(a.Instagram_ad, 0))
FROM public.marketing_data m
LEFT JOIN public.ad_data a ON m.id_ = a.id_
GROUP BY m.marital_status
UNION ALL
SELECT m.marital_status, 'Facebook', SUM(COALESCE(a.Facebook_ad, 0))
FROM public.marketing_data m
LEFT JOIN public.ad_data a ON m.id_ = a.id_
GROUP BY m.marital_status
),
ranked_platforms AS (
SELECT
marital_status,
platform,
total_lead_conversions,
ROW_NUMBER() OVER (PARTITION BY marital_status ORDER BY total_lead_conversions DESC) AS rank
FROM ad_totals
)
SELECT marital_status, platform AS most_effective_platform, total_lead_conversions
FROM ranked_platforms
WHERE rank = 1
ORDER BY marital_status;
Total rows:
Query complete 00:00:00.032
CRLF
Ln 14, Col 14
ENG
UK
22:52
06/04/2025

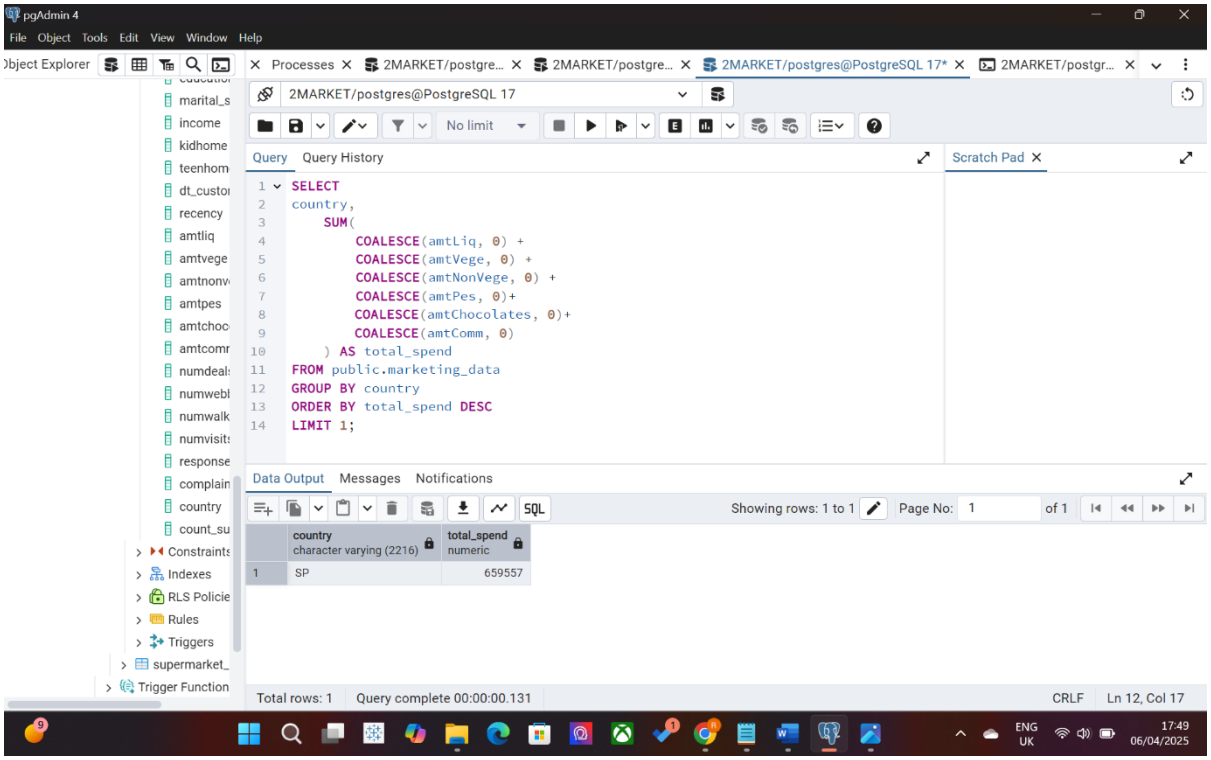

pgAdmin 4
File Object Tools Edit View Window Help
Processes
2MARKET/postgres@PostgreSQL 17*
2MARKET/postgres@PostgreSQL 17
No limit
Query
Query History
WITH ad_totals AS (
SELECT m.marital_status, 'Twitter' AS platform, SUM(COALESCE(a.Twitter_ad, 0)) AS total_lead_conversions
FROM public.marketing_data m
LEFT JOIN public.ad_data a ON m.id_ = a.id_
GROUP BY m.marital_status
UNION ALL
SELECT m.marital_status, 'Instagram', SUM(COALESCE(a.Instagram_ad, 0))
FROM public.marketing_data m
LEFT JOIN public.ad_data a ON m.id_ = a.id_
GROUP BY m.marital_status
UNION ALL
SELECT m.marital_status, 'Facebook', SUM(COALESCE(a.Facebook_ad, 0))
FROM public.marketing_data m
LEFT JOIN public.ad_data a ON m.id_ = a.id_
GROUP BY m.marital_status
),
ranked_platforms AS (
SELECT
marital_status,
platform,
total_lead_conversions,
ROW_NUMBER() OVER (PARTITION BY marital_status ORDER BY total_lead_conversions DESC) AS rank
Total rows:
Query complete 00:00:00.032
CRLF
Ln 14, Col 14
ENG
UK
22:52
06/04/2025


pgAdmin 4
File Object Tools Edit View Window Help
2MARKET/postgres@PostgreSQL 17
Query Query History
Scratch Pad
SELECT marital_status, 'amtChocolates', SUM(COALESCE(amtChocolates, 0))
FROM public.marketing_data
GROUP BY marital_status
UNION ALL
SELECT marital_status, 'amtComm', SUM(COALESCE(amtComm, 0))
FROM public.marketing_data
GROUP BY marital_status
Data Output Messages Notifications
Showing rows: 1 to 8 Page No: 1 of 1
marital_status character varying (2216)
product text
total_spent numeric
1 Absurd amtLiq 711
2 Alone amtLiq 554
3 Divorced amtLiq 75364
4 Married amtLiq 256976
5 Single amtLiq 137217
6 Together amtLiq 176715
7 Widow amtLiq 27902
8 YOLO amtLiq 644
Total rows: 8
Query complete 00:00:00.085
CRLF
Ln 39, Col 25
18:14
06/04/2025


pgAdmin 4
File Object Tools Edit View Window Help
2MARKET/postgres@PostgreSQL 17
Query Query History
Scratch Pad
ROW_NUMBER() OVER (PARTITION BY group_type ORDER BY total_spent DESC) AS ra
FROM product_totals
)
SELECT group_type, product, total_spent
FROM ranked_products
WHERE rank = 1
ORDER BY group_type;
Data Output Messages Notifications
Showing rows: 1 to 4 Page No: 1 of 1
group_type text
product text
total_spent numeric
1 With Kids amtLiq 95862
2 With Teens amtLiq 325143
3 Without Kids amtLiq 580221
4 Without Teens amtLiq 350940
Total rows: 4
Query complete 00:00:00.136
CRLF
Ln 207, Col 21
18:23
06/04/2025


pgAdmin 4
File Object Tools Edit View Window Help
Object Explorer
2MARKET/postgres@PostgreSQL 17
No limit
Query Query History
Scratch Pad
DROP TABLE IF EXISTS ad_data;
create table ad_data(
id_ integer PRIMARY KEY REFERENCES marketing_data(id_),
Bulkmail_ad integer,
Twitter_ad integer,
Instagram_ad integer,
Facebook_ad integer,
Brochure_ad integer
);
Collations
Domains
FTS Configurations
FTS Dictionaries
FTS Parsers
FTS Templates
Foreign Tables
Functions
Materialized Views
Operators
Procedures
Sequences
Tables (3)
ad_data
marketing_data
supermarket_marketing_data
Columns
Constraints
Indexes
RLS Policies
Rules
Triggers
Trigger Functions
Types
Views
Subscriptions
postgres
Login/Group Roles
Data Output Messages Notifications
CREATE TABLE
Query returned successfully in 43 msec.
Total rows:
Query complete 00:00:00.043
06 April 2025
Sun 18:56 (Local time)
ENG
UK
18:56
06/04/2025


pgAdmin 4
File Object Tools Edit View Window Help
2MARKET/postgres@PostgreSQL 17
No limit
Query Query History
Scratch Pad
SELECT
country,
SUM(COALESCE(amtLiq, 0)) as total_amtliq,
SUM(COALESCE(amtVege, 0)) AS total_amtVege,
SUM(COALESCE(amtNonVege, 0)) AS total_amtNonVege,
SUM(COALESCE(amtPes, 0)) AS total_amtPes,
SUM(COALESCE(amtChocolates,0)) AS total_amtChocolates,
SUM(COALESCE(amtComm, 0)) AS total_amtComm
FROM public.marketing_data
GROUP BY country
ORDER BY country;
Data Output Messages Notifications
Showing rows: 1 to 8
Page No: 1 of 1
country character varying (2216)
total_amtliq bigint
total_amtvege numeric
total_amtnonvege numeric
total_amtpes numeric
total_amtchocolates numeric
total_amtcomm numeric
1 AUS 42752 3689 22328 5546 4129 7132
2 CA 84066 7681 45925 9980 7607 12144
3 GER 36776 2980 20272 4601 2801 5768
4 IND 36236 3788 23729 4818 3221 6014
5 ME 1729 8 817 226 122 220
6 SA 105918 8937 58398 13670 9019 15129
7 SP 336392 28288 178409 40153 30134 46181
8 US 32214 3034 20185 4411 2863 4839
Total rows: 8
Query complete 00:00:00.090
CRLF
Ln 11, Col 17
ENG
UK
18:07
06/04/2025


pgAdmin 4
File Object Tools Edit View Window Help
Object Explorer
Processes
2MARKET/postgres@PostgreSQL 17*
2MARKET/postgres@PostgreSQL 17
No limit
Query
Query History
Scratch Pad
SELECT
country,
SUM(
COALESCE(amtLiq, 0) +
COALESCE(amtVege, 0) +
COALESCE(amtNonVege, 0) +
COALESCE(amtPes, 0)+
COALESCE(amtChocolates, 0)+
COALESCE(amtComm, 0)
) AS total_spend
FROM public.marketing_data
GROUP BY country
ORDER BY total_spend DESC
LIMIT 1;
Data Output
Messages
Notifications
Showing rows: 1 to 1
Page No: 1
of 1
country
character varying (2216)
total_spend
numeric
1
SP
659557
Total rows: 1
Query complete 00:00:00.131
CRLF
Ln 12, Col 17
ENG
UK
17:49
06/04/2025