2Market Technical Report

By Chris Buck

Introduction

2Market is an in-store and online supermarket with a global presence. It commissioned this data analytics project to better understand customer demographics, product popularity, and advertising effectiveness to maximise profitability by increasing sales and optimising its marketing spend. The project supports this objective by identifying target markets for advertising through a Mutually Exclusive Collectively Exhaustive (MECE) framework to categorising customers according to demographic characteristics (awware 2021). The MECE method is effective at identifying groups who outperform others with respect to key metrics such as sales per customer. That said, there is a chance that some groupings seem arbitrary from a sociological perspective (age ranges might not coincide with generations, for example). The project also highlights the most effective advertising approaches measured by sales per customer as well as the product categories that contribute most to total sales. These insights drive two marketing recommendations to increase 2Market's profitability.

Analytical Approach

Here are some steps I took to clean and analyse data in **Excel**. I used Excel because it provides visual cues to identify duplicates and outliers through conditional formatting and box-and-whisker charts. Please refer to Appendix 1 for screenshots.

- Conducted a "Special" search for blanks using the Go To dialog box: none found
- Scanned for duplicate rows using Conditional Formatting to highlight duplicate values in the ID column: none found
- Scanned for duplicate rows using Conditional Formatting to highlight duplicate values in the Income column: many near-duplicate rows detected. Did not remove any without consulting stakeholders about how to proceed. The rows in question are spread throughout the data when sorting in ascending order by each column so they should not affect the results of the analysis significantly.
- Added a column for age, calculated by subtracting birth year from current year. If 2014 (the most recent year during which a customer registered with 2Market) was used, then the youngest customers would be 17 years old. This did not seem plausible because one of these 17-year-olds would have made an alcohol purchase and both had "Married" as their marital status.
- Checked for outliers in numerical fields using pivot tables and box-and-whisker charts: replaced Ages > 120 and Incomes > $175,000 with #N/A

- Checked for anomalies in categorical fields using pivot tables to show count for unique values: in Marital Status field changed “Absurd” and “YOLO” to #N/A and changed “Alone” to “Single” using Find and Replace.
- Used pivot table to group customers into income bands to calculate average age

Here are some steps I took to test and analyse data in **PostgreSQL**. I used PostgreSQL because it can generate tables and conduct sense tests of data quickly. Please refer to Appendix 2 for screenshots.

- Queried total sales for each Product Category by Country
- Joined the tables
- Conducted Date and Numeric Range Tests on joined table: no rows returned
- Queried total Ad Channel Conversions by Country

Here are some steps I took to clean and analyse data in **Tableau**. I used Tableau for analysis due to the ease with which it creates visuals. Please refer to Appendix 3 for screenshots.

- Used group function to combine Masters and 2nd Cycle in Education field
- Created tables analysing sales patterns for nearly every demographic field
- Created a calculated field to group customers who only used one ad channel by ad channel to explore its relationship to sales per customer
- Edited location data so South Africa and Spain appeared properly on the map

Dashboard Design and Development

- Visualisation Type
    - Bar charts used to display total sales and sales per customer with respect to demographics and ad channels
    - Scatter plots used to display relationships between measures such as sales in a particular product category with total sales
    - Map used to display geographic location of customers so stakeholders could see key metrics in the tooltip and use it to filter by country
- Colour
    - Blue gradient consistently used to represent total sales
    - Colours selected from the colour-blind palette for accessibility
- Interactivity
    - Dashboards have filters for the most significant demographic dimensions
    - Where appropriate, visualisations can also be used as filters
    - Customer dashboard has KPIs that change as filters are applied
    - Tooltips used to display essential metrics on visualisations

- Layout
    - Visualisations arranged to be read left-to-right, top-to-bottom
    - Chart headers removed where appropriate to prevent dashboard clutter

Patterns, Trends, and Insights

- Customer Insights
    - *Country:* Spain accounts for approximately half of 2Market's customers and approximately half of its total sales.
    - *Education:* The overwhelming majority of 2Market customers are highly educated and generate higher total sales and sales per customer than customers with a basic education.
    - *Income:* Customers earning $60,000 or more generate more sales and sales per customer than customers earning less than $60,000.
    - *Age:* Sales per customer amongst 26-35 year olds is relatively high and comparable to that of customers in the 56-65 and 66-75 age ranges.

- Advertising Insights
    - In the sample of 322 customers who purchased through only one ad channel, Instagram ads generate the highest sales per customer followed by Facebook and then Twitter. Note that there's no way to determine the specific amount of sales generated by a successful conversion with the data provided.
    - Customers who made purchases through two or more different ad channels generated higher sales per customer than those who did not.

- Product Insights
    - Alcohol sales outnumber sales of all other product categories combined.
    - Meat product sales have the strongest positive correlation with sales across all categories.
    - Meat sales per customer are highest in 26-35 and 76-85 age groups, which are the age groups with the highest total sales per customer.

Marketing Campaign Recommendations

1. 2Market should use targeted ads across multiple social media channels to reach potential customers ages 26-35 who earn $60,000 or more. These potential customers could be found in countries in like Canada, Germany, and the United States where young adults have higher average incomes than their counterparts in Spain, where most of 2Market's customers are located (World Bank 2023).

2. 2Market should launch a promotional campaign to encourage existing customers, especially those ages 36-75, to purchase more meat products because sales of these products have a strong relationship with total sales across all product categories.

Technical Recommendations

1. 2Market should collect and analyse more granular data from individual sales transactions to better understand the relationships between customer demographics, advertising effectiveness, and product popularity.

2. 2Market should review existing customer data for near-duplicate entries. It appears as though some customers are being registered twice but in different countries. If these customers are in fact the same, then associating them with a single customer ID will improve the accuracy of insights discovered through data analytics.

Reference List

awware (2021) MECE framework: general principles and examples in marketing. Available at: https://awware.co/blog/mece-framework/ (Accessed: 15 December 2024).

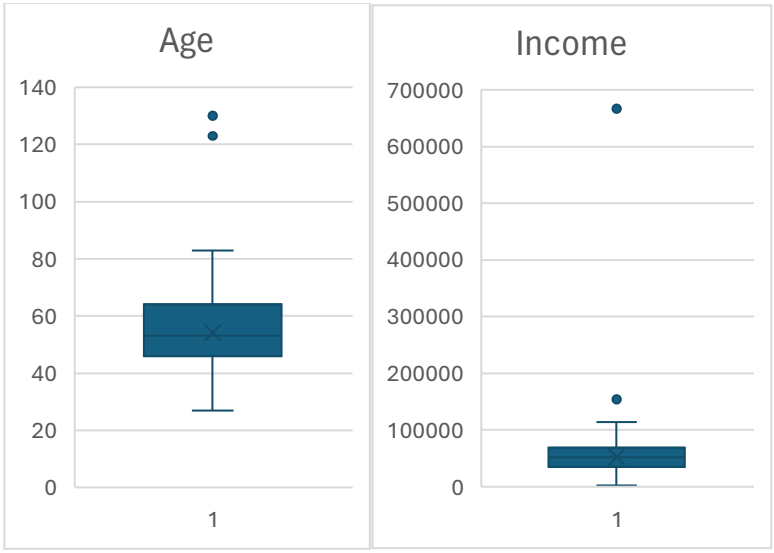
World Bank (2023) OECD income distribution database. Available at: https://prosperitydata360.worldbank.org/en/dataset/OECD+IDD (Accessed: 15 December 2024).

Appendix 1: Excel Screenshots

Example of a Near-Duplicate Row Identified through Conditional Formatting

ID	Year_Birth	Age	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	AmtLiq	AmtVege	AmtNonVeg	AmtPes	AmtChocolates	AmtComm
5184	1996	28	Graduation	Single	34824.00	0	0	3/26/14	65	4	2	11	2	0	4
8315	1996	28	Graduation	Single	34824.00	0	0	3/26/14	65	4	2	11	2	0	4

Box-and-Whiskers Plots to Identify Outliers in Age and Income



Pivot Table to Identify Anomalies in the Marital Status Field by Counting Unique Values

Row Labels	Count of Marital_Status
Absurd	2
Alone	3
Divorced	232
Married	857
Single	471
Together	573
Widow	76
YOLO	2
Grand Total	**2216**

Pivot Table to Group Customers into Income Bands and Calculate Average Age

Row Labels	Average of Age
0-9999	47
10000-19999	46
20000-29999	49
30000-39999	52
40000-49999	55
50000-59999	58
60000-69999	57
70000-79999	56
80000-89999	56
90000-99999	52
100000-109999	47
110000-119999	78
150000-159999	54
160000-169999	44
660000-670000	46

Appendix 2: PostgreSQL Screenshots

Syntax to Create Table for Data imported from the Ad Data .csv File

```
CREATE TABLE raw.ad_data(
customer_id INTEGER,
bulkmail_ad INTEGER,
twitter_ad INTEGER,
instagram_ad INTEGER,
facebook_ad INTEGER,
brochure_ad INTEGER
);

SELECT *
FROM raw.ad_data
```

Syntax to Create Table for Data imported from the Marketing Data .csv File

```
CREATE TABLE raw.marketing_data(
customer_id INTEGER,
year_birth DATE,
age INTEGER,
education VARCHAR(15),
marital_status VARCHAR(15),
income NUMERIC,
kidhome INTEGER,
teenhome INTEGER,
dt_customer DATE,
recency INTEGER,
amtliq NUMERIC,
amtvege NUMERIC,
amtnonveg NUMERIC,
amtpes NUMERIC,
amtchocolates NUMERIC,
amtcomm NUMERIC,
numdeals INTEGER,
numwebbuy INTEGER,
numwalkinpur INTEGER,
numvisits INTEGER,
response INTEGER,
complain INTEGER,
country VARCHAR(10),
count_success INTEGER
);

SELECT *
FROM raw.marketing_data
```

Syntax to Query Total Sales for Product Category by Country

```
-- total spend per product per country

SELECT country,
SUM(amtliq) AS total_alcohol,
SUM(amtvege) AS total_veg,
SUM(amtnonveg) AS total_meat,
SUM(amtpes) AS total_fish,
SUM(amtchocolates) AS total_chocolates,
SUM(amtcomm) AS total_commodities
FROM raw.marketing_data
GROUP BY country;
```

	country character varying (10)	total_alcohol numeric	total_veg numeric	total_meat numeric	total_fish numeric	total_chocolates numeric	total_commodities numeric
1	SP	336392	28288	178409	40153	30134	46181
2	CA	84066	7681	45925	9980	7607	12144
3	AUS	42752	3689	22328	5546	4129	7132
4	IND	36236	3788	23729	4818	3221	6014
5	US	32214	3034	20185	4411	2863	4839
6	ME	1729	8	817	226	122	220
7	SA	105918	8937	58398	13670	9019	15129
8	GER	36776	2980	20272	4601	2801	5768

Syntax to Join the Tables

```
-- join the two tables

CREATE TABLE raw.joined_table AS
SELECT *
FROM raw.marketing_data
JOIN raw.ad_data
USING(customer_id)

SELECT *
FROM raw.joined_table
```

Syntax for Numeric Range Check

```
--Numeric Range Check
SELECT * FROM
raw.joined_table
WHERE
    age < 0 OR
    income < 0 OR
    kidhome < 0 OR
    teenhome < 0 OR
    recency < 0 OR
    amtliq < 0 OR
    amtvege < 0 OR
    amtnonveg < 0 OR
    amtpes < 0 OR
    amtchocolates < 0 OR
    amtcomm < 0 OR
    numdeals < 0 OR
    numwebbuy < 0 OR
    numwalkinpur < 0 OR
    numvisits < 0 OR
    response < 0 OR
    complain < 0 OR
    count_success < 0 OR
    bulkmail_ad < 0 OR
    twitter_ad < 0 OR
    instagram_ad < 0 OR
    facebook_ad < 0 OR
    brochure_ad < 0
```

Syntax for Date Range Check

```
--Date Range Check
SELECT *
FROM raw.joined_table
WHERE dt_customer > CURRENT_DATE
```

Syntax to Query Ad Channel Conversions by Country

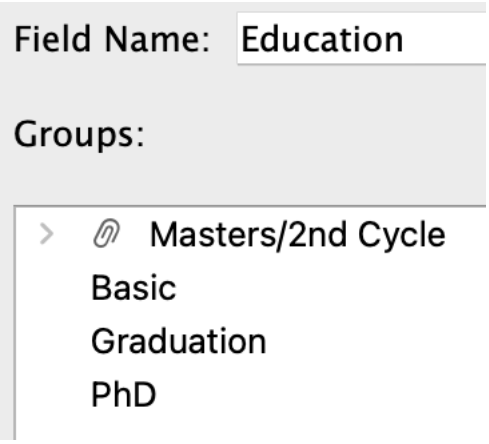
```
-- Which social media platform (Twitter, Instagram, or Facebook) is most effective in each country?

SELECT country, SUM(twitter_ad) AS twitter, SUM(instagram_ad) AS instagram, SUM(facebook_ad) AS facebook
FROM raw.joined_table
GROUP BY country
ORDER BY country ASC
```

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

Appendix 3: Tableau Screenshots

Grouped Masters and 2nd Cycle in Education Field

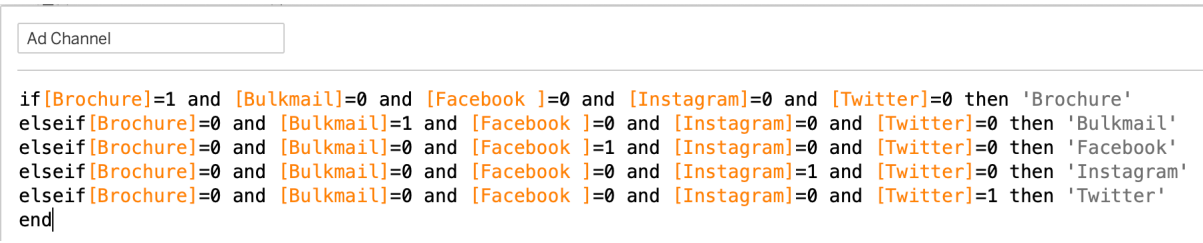


Example of Table Examining Sales by a Demographic Field (e.g. Education)

Education, Purchases, and Sales Table

Education	Total Purchases	Customers	Total Sales	Sales per Customer	Sales per Purchase	Purchases per Customer
Basic	256	54	$4K	$82	17	5
Graduation	11,046	1,110	$688K	$620	62	10
Masters/2nd Cycle	5,478	562	$321K	$572	59	10
PhD	4,999	474	$319K	$672	64	11

Calculated Field to Isolate Customers who Purchased through Only One Ad Channel



Edited Locations so South Africa and Spain Appeared Correctly on Map