2Market

Marketing Data Analysis

Descriptions and overview of approach, analysis and insights

Toby Draper
06/09/2025

Contents

Background

2Market are a global chain of supermarkets and online stores. This project was undertaken to inform their marketing strategy. Root cause analysis using the 5 Whys framework (Appendix 1) identified a failure to leverage marketing data historically.

Problem Statement:

Previous marketing campaigns generated engagement from a low proportion of registered 2Market customers. This has negatively impacted how soon customers return to make purchases.

Objective:

Attract customers to return and make purchases sooner by leveraging data-driven marketing campaigns that offer compelling discounts.

Analytical Approach

Data Cleaning

Cleaned data files marketing_data.csv (Appendix 2) and ad_data.csv (Appendix 3) using MS Excel, creating versions to import into Tableau (Appendix 4).

Surface Level Insights

Created Dashboard "Surface Level Insights" to introduce business problem, comparing;

- Total number of registered customers in each country compared to those with Count_success >= 1
- Recency, aggregated as median to reduce influence of outliers

High Level Analysis

Analysis in MS Excel (Appendix 5) identifies:

- Relationship between age and marital status
- Relationship between age and income
- Age of customers earning $90k - US$100k ($90k group)

Reproduced visualisations in a Tableau Dashboard for presentation.

SQL Deep Dive #1

Created Schema "raw", importing data into tables (Appendix 6);

- raw.marketing_data

- raw.ad_data

Ran queries to identify (Appendix 7);

- Total spend per country

- Total spend per product per country

- Most popular product types in each country

- Most popular product types based on marital status

- Most popular product types whether or not there are children/teens at home

SQL Deep Dive #2

Consolidated data into one table to simplify queries in new schema "draft" (Appendix 8). Queried (Appendix 9);

- Social media platform effectiveness by country

- Social media platform effectiveness by marital status

- Social media influence on revenues to each product type

Results visualised in Dashboard "Social Media Deep Dive".

Repeated with filter on $90k group

Discount and Recency

- Comparisons made between Deals and Recency as basis for recommendations
- Also highlight the "$90k group" for comparison

Dashboard Design & Development

Surface Level Insights

Visualisation types: Map and Highlight Tables

Colour: Normal Background setting (light blue) on map. Selected other shades of blue for Highlight Tables positioned above and below the map forming distinct sections consistent with map colour.

Interactivity: Top Highlight Table functions as a set of buttons. Clicking applies a filter to the map selecting corresponding country and highlights the matching country in the bottom Highlight Table using a highlight rule to prevent layout shifts.

High Level Analysis

Visualisation types: Line chart, bar charts

Colour: Retained blue as the theme throughout presentation for smooth transitions. Royal blue is used for visualisations relating to "all customers" and teal for focus on the "$90k group".

Notes: Decision to recreate work done in MS Excel in Tableau Dashboards to simplify delivery. No interactivity due to the simplicity of data and design. Axis titles removed throughout the presentation to optimise space. Anticipate questions if this causes confusion.

Social Media Deep Dive

Visualisation types: bar charts

Notes: This is the busiest dashboard and presented challenging design choices. It provides a summary of SQL Deep Dive #1 and #2

- Opted to show selected results in bar charts rather than a table to highlight contrasting results
- Opted to show all data in one dashboard, not split into separate dashboards or use filters to display the results separately to limit slide count and allow direct comparison
- Mitigated small size of bars and text using size setting 1600 x 900 and filtering out results from US, ME and IND

Discount and Recency

Visualisation types: bar charts, box and whisker

Notes: Some countries filtered out that return 0 results to optimise space.

Settled on box and whisker chart type to describe median recency and discuss the desired outcome of improving this.

Detailed Tableau Configuration (appendix 10)

Patterns, Trends & Insights

High Level Analysis

Relationship between age and marital status

- Single customers typically younger (average 52), next Married (54)
- Oldest marital status is Widowed (65)
- Significant number identify as "Together"

Relationship between age and income

- Typically, salaries increase with age.
- Exception: average income for 20s exceeds that of 30s.
- 40s to 80s, income rises from approximately $50k to just over $65k

Comparison with customers earning $90k - US$100k

Customers in their 30s are typically lowest earners. Analysis of customers with incomes between $90 - $100k revealed an exception to the general pattern with more in 30's age group.

Continue to compare this group with all customers for other insights, patterns and exceptions that are potentially worthy of analysis.

SQL Deep Dive #1

- Alcoholic beverages followed by meat products generate highest revenue
- True for all countries and all demographics.

Deals on these products could help 2Market gain better traction from social media campaigns.

Additionally, Spain generates highest revenue ($657704), Montenegro lowest ($3122)

SQL Deep Dive #2

Social media effectiveness

- X (Twitter) marginally more effective for all countries combined
- By individual country, X and Instagram lead
- Highlights: Spain - Instagram 88, X 87. Canada - X 24, Instagram 21
- 54% of all social media engagement in Spain

Most effective social media platform based on Marital Status

- Instagram more popular among Married
- X and Facebook equally popular among that group
- Customers in a relationship ("Married " + "Together") favour Instagram then X. Facebook remains widely used

Social media influencing revenues on product types

- Highest revenues for Alcoholic beverages, whether customers engage with social media or not
- Higher revenues are generated without social media ads than are generated with them for all product types
- Social media drives highest revenues for alcoholic beverages, then meat products (all countries)

Social media effectiveness - $90k Group

- Instagram most popular social media platform
- Instagram popular with Married and Single
- Of all 350 social media engaged customers in Spain, 7 contribute 13% of revenue for alcoholic beverages and 19% for meat products

Discount and Recency

- Significant number of purchases with deals without social media engagement, compared to deals with social media for all customers
- The $90k group make more purchases with deals from social media campaigns.

Measured recency of both sets of customers, highlights;

- Median recency without social media is 50.5 days for all customers
- All customers with social media, drops to 48 days
- Indicates social media engagement positively affects recency
- Reinforced by $90k group with social media where median recency drops to 35.5 days

Recommendations/metrics;

- Reduce deals not linked to social media campaigns
- Reward customers with discount on identified product types purchased through social media campaigns
- Monitor deals with social media engagement
- Capture increased granularity in data;
  - Mean number of days between purchases
  - Product sub-types
  - Number of units bought by product type/sub-type
- Survey social media engaged customers until granularity issues in marketing data addressed

Appendix

Appendix 1

5 Whys – formulate the business problem

A significant proportion of registered customers have typically not made a purchase in three weeks or more (Recency). In many cases, the time since their last purchase is significantly higher than three weeks

Why?

> **Because customers are unaware of 2Market's full range of products and available discounts (Deals)**

Why?

> **Because there is low lead conversion (Count_success), indicating that customers have not responded to social media campaigns in significant numbers**

Why?

> **Because 2Market's previous campaigns were ineffective**

Why?

> **Because they were not informed by a data driven strategy**

Why?

> **Root cause: 2Market failed to leverage data (marketing_data, ad_data) captured previously in order to build successful campaigns**

Assumptions:

- Initially, I interpreted *Recency* as a rough measure of the time between visits. But as the project progressed, I began to understand it more literally *as the time since the last visit*. While the gap between visits is implied in this study, I have flagged this assumption and later recommend that time between visits should be captured directly
- I don't have a comparison of historical sales figures in the revenue data provided. Assume that sales are stagnating and reflected by high recency figures
- Because the business problem was formulated using values in Recency column, assume this project was conducted over 2-3 days and that the recency data has not become inflated by time elapsed since I got the data a number of weeks ago.

Appendix 2

Data Cleaning file marketing_data.csv in Excel (refer to file metadata_2Market.txt for data dictionary)

Column	Header	Excel Data Type (Format)	Actions
A	ID	Number, General	• Conditional format on duplicates. None found. • Go to (Special) on Blanks. None found • Sorted Smallest to Largest
B	Year_Birth	Number, General	• Used Min and Max function to identify outliers. • Found 3 stale or anomalous records (YOB 1894, 1900, 1901) • Deleted records for those 3 customers
C	Education	Text, General	• Created pivot table to review for blanks and bad data • No issues found, deleted sheet containing pivot • Trim function used on text fields to remove spaces • 198 customers list education as "2n Cycle. Assume this should read "2nd Cycle" (did not amend text)

D	Marital_Status	Text, General	• Created pivot table to review for blanks and bad data
			• Trim function used on text fields to remove spaces
			• After inspecting pivot made following changes in data
			   o 2 customers entered "Absurd", changed to NA (Find & Replace)
			   o 3 customers entered "Alone". Changed to Single (Find & Replace)

			- ○ 2 customers entered YOLO. Changed to NA (Find & Replace) • No blanks found
E	Income	Number, General	• Currency symbol entered as text against all values. Removed with Find & Replace. • Formatted as Number with no decimal places as all values are whole • Used AVERAGE function on all rows to indicate any bad data in column. Result returned, none found
F	Kidhome	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned • Min and Max functions to find outliers • No outliers found
G	TeenHome	Number, General	• AVERAGE on column to check for bad data. • None found – result returned • Min and Max functions to find outliers No outliers found
H	Dt_Customer	Number, Date	• The cell format not uniform as Date in all values in column • All values captured as mm/dd/yyyy • Copied Column H to new workbook and save as dates.txt • Opened dates.txt, running Import Wizard to split date mm/dd/yyyy into separate columns using / as delimiter • Some years in yy format. Used date function to correctly create dates in

			2000s: **=DATE(IF(C2<100, 2000+C2, C2), A2, B2)** • Dates now in dd/mm/yyyy • Copied and pasted (paste special as values) into new column in data set. Named Column Header as Dt_Customer and deleted original column
I	Recency	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
J	AmtLiq	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
K	AmtVege	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
L	AmtNonVeg	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
M	AmtPes	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
N	AmtChocolates	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
O	AmtComm	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
P	NumDeals	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
Q	NumWebBuy	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned

R	NumWalkinPur	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
S	NumVisits	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
T	Response	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
U	Complain	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned
V	Country	Text, General	• Inserted pivot table • Checked for blanks • Checked for values not defined in metadata • Trim function used on text fields to remove spaces • No problems found
W	Count_Success	Number, General	• AVERAGE on column to check for errors in data. • None found – result returned

Appendix 3

Data cleaning steps ad_data.csv;

- Conditional formatting to highlight duplicate values in ID column. None found.
- Excel SUM on values in each column to identify empty values or anomalies. None found

Appendix 4

Tableau Data Source and join (left join in ID)

Appendix 5

High Level Analysis

	A	B
	What is the average age of 2Market's customers?	55

Method
Create a new column named year. Enter 2025 in every row of the data set. Calculate year minus birth year to get approximate age. AVERAGE function on approx age column
Note above - all calculations are APPROXIMATE AGE. Cannot know actual age with only birth year available in data.

3			
4	What is the average age of customers by marital stat	Single	52
5		Married	54
5		Together	56
7		Divorced	58
3		Widowed	65
9			

Method
Changed entries so that Single, Married, Divorced, Widow, Together and NA were only values. Changed 3 marked "alone" to single, changed all "absurd" and "yolo" to NA. Retained status "Together" since a significant number of customers registered as this. Filtered by each status in column B, using AGGREGATE(1,7,range) function for each filtered selection to ignore errors and hidden rows
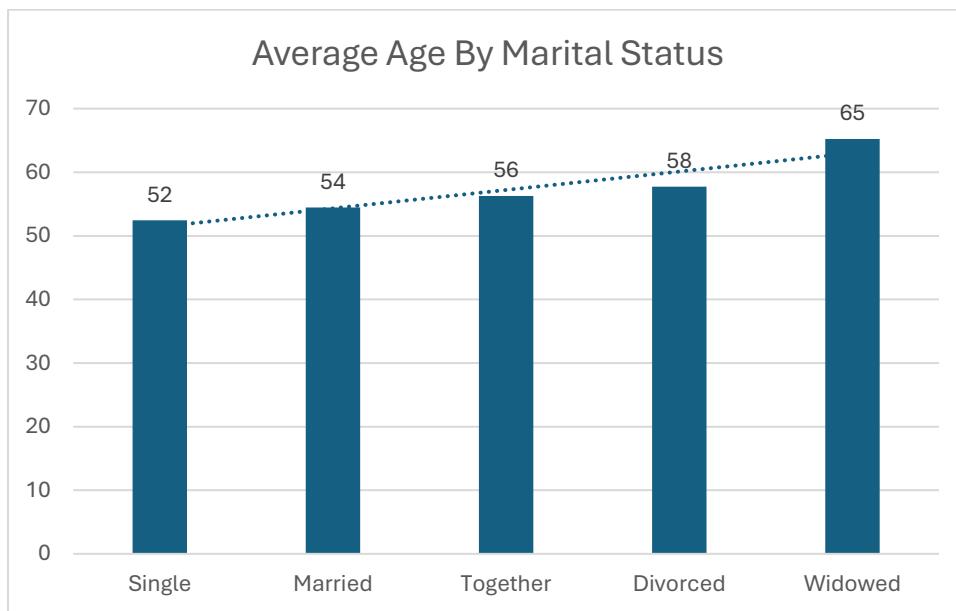
What is the average age of customers who earn a yearly income between US$90,000 and US$100,000?		53

Method
The income field is now formatted as number to use the filter option "between"
Filtered by range >= 90000 <=100000 and used AGGREGATE (1,7,range) function to calculate average age of the results selected by filter

What is the trend of the average age by Marital Status?

Average Age By Marital Status
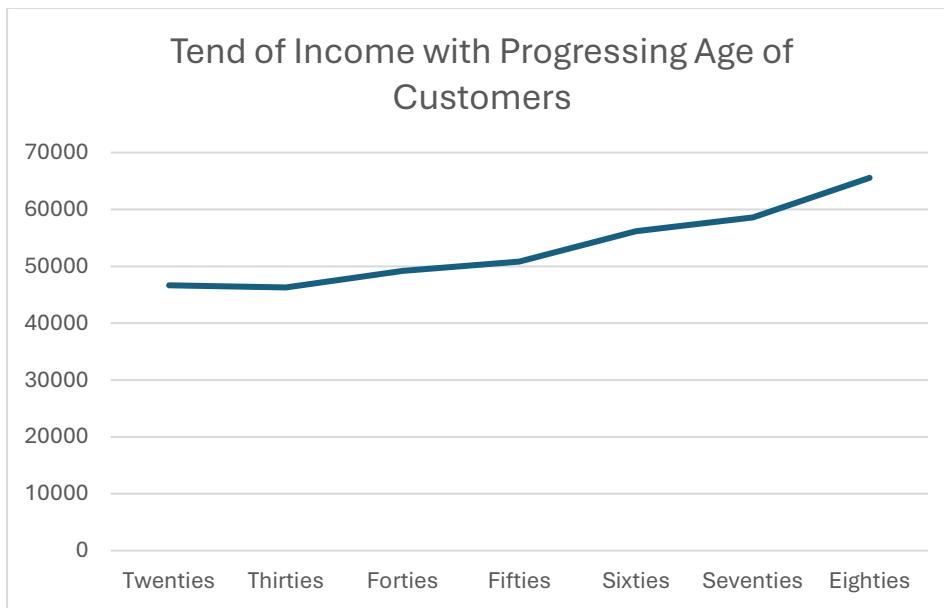
What is the trend of customer income with the progressing age of the customers?

Method

used IFS tp categorise ages into decades
filtered by each decade and captured mean income for each using
AGREGATE(1,7,range) function
copied and pasted values of cells with aggregate function to show just the value
I have removed three records of customers >100 years olf

Age Bracket (Decade)	Average Income
Twenties	46658
Thirties	46283.03
Forties	49224.88
Fifties	50812.91
Sixties	56200.83
Seventies	58611.13
Eighties	65563.53

Tend of Income with Progressing Age of Customers

What is the relationship between income and the age of the customers who earn between US$90,000 and US$100,000?
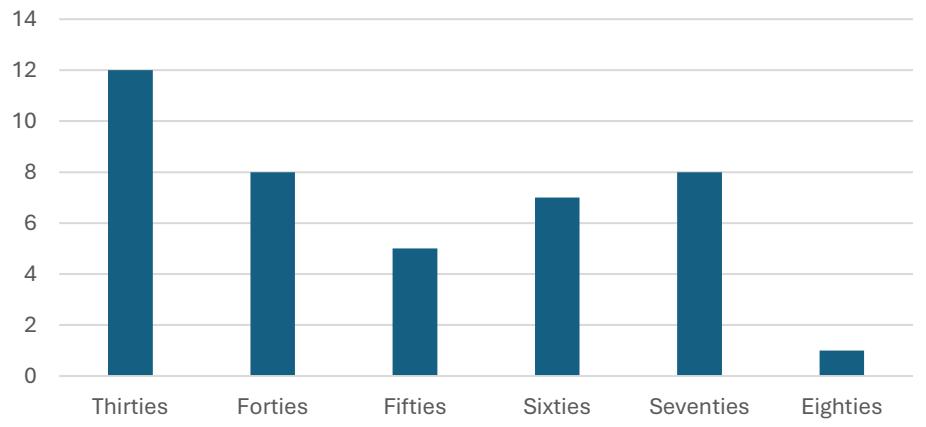
Method
 filtered by salary between 90000 and 100000. copied and pasted into new sheet. Filtered by decade and used aggregate(1,7,range) to calculate mean for each then plotted in a line chart
 Then used AGGREGATE(2,7,range) to obtain a count because I could see there were fewer with each passing decade)
 Conclusion - customers earning $90k - $100k were, younger, with most of them in their thirties, contrary to the general trend of customer's salary increasing proportionately with their age.

Age Bracket (Decade)	Count	Average 90k - 100k
Thirties	12	92552.58
Forties	8	92516.13
Fifties	5	92816.2
Sixties	7	93775.14
Seventies	8	92830.63
Eighties	1	93027

Count of Customers by Age Bracket (Decades), for customers earning $90k - $100k

Appendix 6 – SQL Deep Dive 1

Create Table Queries;

CREATE TABLE raw.marketing_data (

  id   INTEGER,

  year_birth   INTEGER,

  education   VARCHAR,

  marital_status VARCHAR,

  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,

  count_success INTEGER

);

```
CREATE TABLE raw.ad_data (

  id          INTEGER,

  bulkmail_ad   INTEGER,

  twitter_ad    INTEGER,

  instagram_ad  INTEGER,

  facebook_ad   INTEGER,

  brochure_ad   INTEGER

);
```

Appendix 7

SQL Deep Dive #1 Queries and Results

Write SQL queries to format, transform, validate, and analyse the data to determine which products sell the best and whether that varies based on demographic factors, by determining:

1. **Total spend per country**

Answers highlighted in results (below)

QUERY

```
SELECT

 marketing_data.country,

 SUM(marketing_data.amtliq + marketing_data.amtvege + marketing_data.amtnonveg +

    marketing_data.amtpes + marketing_data.amtchocolates +
marketing_data.amtcomm) AS country_total

FROM raw.marketing_data

GROUP BY marketing_data.country;
```

RESULT

country	country_total
SP	657704
CA	167403
IND	77741
AUS	85576
US	67546
ME	3122
SA	211049
GER	73198

2. Total spend per product per country

QUERY

SELECT

 marketing_data.country,

 SUM(marketing_data.amtliq) AS total_liq,

 SUM(marketing_data.amtvege) AS total_vege,

 SUM(marketing_data.amtnonveg) AS total_nonveg,

 SUM(marketing_data.amtpes) AS total_pes,

 SUM(marketing_data.amtchocolates) AS total_choc,

 SUM(marketing_data.amtcomm) AS total_comm,

 SUM(marketing_data.amtliq + marketing_data.amtvege + marketing_data.amtnonveg +

   marketing_data.amtpes + marketing_data.amtchocolates +
marketing_data.amtcomm) AS country_total

FROM raw.marketing_data

GROUP BY marketing_data.country;

RESULT

country	total_liq	total_vege	total_nonveg	total_pes	total_choc	total_comm	country_total
SP	335637	28144	177847	40049	30070	45957	657704
CA	84066	7681	45925	9980	7607	12144	167403
IND	36221	3782	23721	4811	3217	5989	77741
AUS	42752	3689	22328	5546	4129	7132	85576
US	32214	3034	20185	4411	2863	4839	67546
ME	1729	8	817	226	122	220	3122
SA	105910	8937	58393	13663	9019	15127	211049
GER	36776	2980	20272	4601	2801	5768	73198

3. Most popular product types

Answer: Alcoholic Beverages (result matches values in total_liq column in previous question)

QUERY

```
SELECT
 country,
 GREATEST(
  SUM(amtliq),
  SUM(amtvege),
  SUM(amtnonveg),
  SUM(amtpes),
  SUM(amtchocolates),
  SUM(amtcomm)
 ) AS most_pop
 FROM raw.marketing_data
 GROUP BY country;
```

Result

country	most_pop
SP	335637
CA	84066
IND	36221
AUS	42752
US	32214
ME	1729
SA	105910
GER	36776

4. Most popular product types based on marital status

Answer: Alcoholic Beverages

QUERY

```
SELECT marital_status,
 SUM(marketing_data.amtliq) AS total_liq,
  SUM(marketing_data.amtvege) AS total_vege,
  SUM(marketing_data.amtnonveg) AS total_nonveg,
  SUM(marketing_data.amtpes) AS total_pes,
  SUM(marketing_data.amtchocolates) AS total_choc,
  SUM(marketing_data.amtcomm) AS total_comm
 FROM raw.marketing_data
 GROUP BY marital_status;
```

Result

marital_status	total_liq	total_vege	total_nonveg	total_pes	total_choc	total_comm
Together	175960	14468	94812	22279	14967	24530
NA	1355	175	725	419	67	492
Married	256976	21981	137888	30395	22926	36719
Divorced	75349	6357	34840	8123	6218	10714
Widow	27902	2422	14085	3793	2878	4245
Single	137763	12852	87138	18278	12772	20476

5. Most popular product types whether or not there are children/teens at home

Answer(with kids/teens at home): Alcoholic Beverages

QUERY

SELECT

  COUNT(id) AS is_parent,

  SUM(marketing_data.amtliq) AS total_liq,

  SUM(marketing_data.amtvege) AS total_vege,

  SUM(marketing_data.amtnonveg) AS total_nonveg,

  SUM(marketing_data.amtpes) AS total_pes,

  SUM(marketing_data.amtchocolates) AS total_choc,

  SUM(marketing_data.amtcomm) AS total_comm

FROM raw.marketing_data

WHERE (kidhome + teenhome) > 0;

Result

is_parent	total_liq	total_vege	total_nonveg	total_pes	total_choc	total_comm
1581	367110	25309	135292	34891	26229	56739

Answer (without kids/teens at home): Alcoholic Beverages

QUERY

SELECT

  COUNT(id) AS is_not_parent,

  SUM(marketing_data.amtliq) AS total_liq,

  SUM(marketing_data.amtvege) AS total_vege,

  SUM(marketing_data.amtnonveg) AS total_nonveg,

  SUM(marketing_data.amtpes) AS total_pes,

  SUM(marketing_data.amtchocolates) AS total_choc,

  SUM(marketing_data.amtcomm) AS total_comm

FROM raw.marketing_data

WHERE (kidhome + teenhome) = 0;

Result

is_not_parent	total_liq	total_vege	total_nonveg	total_pes	total_choc	total_comm
632	308195	32946	234196	48396	33599	40437

Assumption: the questions ask which products are the "most popular". We can only assume that product types generating highest revenues are the most popular without details of numbers of units sold or unit cost (return to this in Recommendations when suggesting increased granularity).

I have endeavoured to describe the results as "generating highest revenues" rather than "most popular" throughout the report and the presentation.

All figures are in US Dollars and assumed to be gross revenue before taxation and other deductions

Appendix 8

Create new table for further investigation in PostgreSQL

Note: Data was imported to SQL using data types documented in metadata_2market.txt. Result was ad_data containing Boolean values (true/false), making aggregation challenging. Solved by deleting/ recreating tables using data type Integer.

CREATE TABLE draft.marketing_full AS

SELECT

    COALESCE(m.id, a.id) AS id,

    m.year_birth,

    m.education,

    m.marital_status,

    m.income,

    m.kidhome,

    m.teenhome,

    m.dt_customer,

    m.recency,

    m.amtliq,

    m.amtvege,

    m.amtnonveg,

    m.amtpes,

```sql
        m.amtchocolates,

        m.amtcomm,

        m.numdeals,

        m.numwebbuy,

        m.numwalkinpur,

        m.numvisits,

        m.response,

        m.complain,

        m.country,

        m.count_success,

        a.bulkmail_ad,

        a.twitter_ad,

        a.instagram_ad,

        a.facebook_ad,

        a.brochure_ad
FROM raw.marketing_data m
LEFT JOIN raw.ad_data a
    ON m.id = a.id;
```

Appendix 9

SQL Deep Dive #2 Questions, answers, queries and results

- **Social media platform effectiveness by country**

In general, Instagram and X (Twitter) are the most effective methods of advertising in most countries. Where either is first, the other is often a close second. The exception is USA where Facebook is marginally more effective but numbers are low for all methods.

QUERY

SELECT

country,

SUM (twitter_ad) AS twitter,

SUM (instagram_ad) AS insta,

SUM (facebook_ad) AS facebook

FROM draft.marketing_full

WHERE count_success >= 1

GROUP BY country;

RESULTS

Cnty	twitter	insta	facebook
"SP"	87	88	76
"CA"	24	21	18
"IND"	10	6	7
"AUS"	6	12	7
"US"	6	5	7
"ME"	0	0	0
"SA"	20	21	20
"GER"	11	8	7
Total	164	161	142

Note: Total calculated later on in MS Excel after the results by country had been queried in PosgreSQL to assist with framing findings in report

- **Social media platform effectiveness by marital status**

For Together and Married Instagram is most effective. For all others it is X. Insta generates the most hits across all methods due to the higher number of responses from Marrieds and Togethers.

QUERY

SELECT

marital_status,

SUM (twitter_ad) AS twitter,

SUM (instagram_ad) AS insta,

SUM (facebook_ad) AS facebook

FROM draft.marketing_full

GROUP BY marital_status;

RESULTS

Status	twitter	insta	facebook
"Together"	42	43	32
"NA"	0	1	1
"Married"	62	66	62
"Divorced"	18	13	12
"Widow"	10	7	5
"Single"	32	31	30

- **Social media influence on revenues to each product type**

Social media channels are driving sales of alcoholic beverages

More products of all types are sold to customers who have not engaged with social media campaigns

QUERY 1 (without social media)

SELECT

country,

SUM (amtliq) AS booze,

SUM (amtvege)AS vege,

SUM (amtnonveg) AS meat,

SUM (amtpes) AS fish,

SUM (amtchocolates) AS chocs,

SUM (amtcomm) AS stuff

FROM draft.marketing_full

WHERE count_success >=1

GROUP BY country;

RESULT

country	booze	vege	meat	fish	chocs	stuff
SP	188834	18907	107950	27456	19559	31580
CA	48418	5744	30276	7057	5704	8322
IND	23394	3255	16669	3582	2261	4341
AUS	28593	2832	14612	3805	3163	5149
US	21189	2571	16527	3710	2463	3662
ME	1209	0	594	194	73	178
SA	64188	6342	39039	9308	6473	11007
GER	19374	2078	12577	3060	1871	4020

QUERY 2 (with social media)

SELECT

country,

SUM (amtliq) AS booze,

SUM (amtvege)AS vege,

SUM (amtnonveg) AS meat,

SUM (amtpes) AS fish,

SUM (amtchocolates) AS chocs,

SUM (amtcomm) AS stuff,

SUM (twitter_ad) AS twitter,

SUM (instagram_ad) AS insta,

SUM (facebook_ad) AS facebook

FROM draft.marketing_full

WHERE count_success >= 1

GROUP BY country;

RESULT

country	booze	vege	meat	fish	chocs	stuff	twitter	insta	facebook
SP	146803	9237	69897	12593	10511	14377	87	88	76
CA	35648	1937	15649	2923	1903	3822	24	21	18
IND	12827	527	7052	1229	956	1648	10	6	7
AUS	14159	857	7716	1741	966	1983	6	12	7
US	11025	463	3658	701	400	1177	6	5	7
ME	520	8	223	32	49	42	0	0	0
SA	41722	2595	19354	4355	2546	4120	20	21	20
GER	17402	902	7695	1541	930	1748	11	8	7

```
SUM (facebook_ad) AS facebook

FROM draft.marketing_full
WHERE count_success >= 1
AND income BETWEEN 90000 AND 100000
```

Appendix 10

Detailed Tableau Configuration

Calculated Fields

- Create approximate age from Year_Birth Column: YEAR (TODAY()) - [Year Birth]

*Slight inaccuracy due to lack of full date of birth

- "success indicator" Check value of Count_success field for filters: [Count success] >= 1
  - If the value is 1 or more, the result is True.
  - If the value is 0 or null, the result is False.

Surface Level Insights

Dashboard Actions:

Name	Run On	Source	Fields
✐ Highlight1	Select	Surface Level Insights (I	All
▽ Map to Bar Chart Fil	Select	Surface Level Insights	Country

Worksheet 1 – Count Success by Country

Type: Map

Marks:

- Size – SUM(Count success)
- Label – SUM (Count success)
- Detail – Country

Worksheet 2 – Customer IDs by Country

Type: Highlight Table

Marks:

- Colour: CNT (ID)
- Label: CNT (ID)

Columns: Country (discrete),descending

Rows and columns swapped to view results horizontally

Worksheet 3 – Median recency by country

Type: Highlight Table

Marks:

- Colour: MEDIAN(Recency)
- Label: MEDIAN(Recency)

Columns: Country (discrete),descending

Rows and columns swapped to view results horizontally

All Recency averages calculated as median to reduce influence of outliers on results and to reflect typical behaviour more closely

High Level Analysis

Worksheet 4 – Age & Marital Status

Type: Bar

Columns: Marital Status (discrete, ascending)

Rows: AVG (continuous , Age)
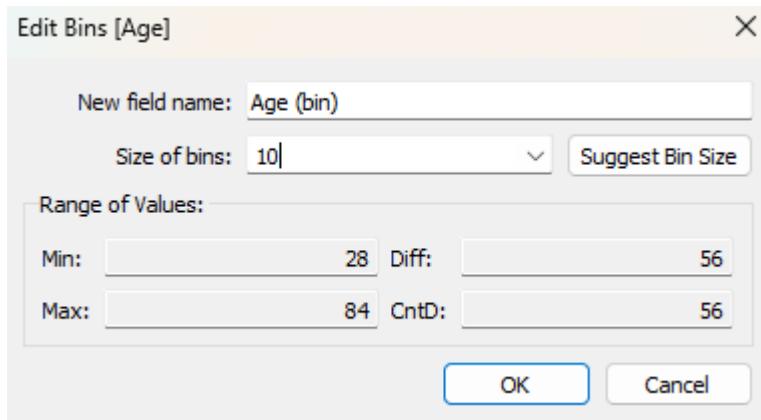
Filters: Marital Status

Marks: Label – AVG(Age)

Worksheet 5 – Age & Income

Type: Line

Columns: Age (discrete, bin)

Edit Bins [Age]

New field name: Age (bin)

Size of bins: 10   Suggest Bin Size

Range of Values:

Min: 28   Diff: 56

Max: 84   CntD: 56

OK   Cancel

Rows: AVG (continuous, Income)

Worksheet 6 – Age of Customers earning $90k - $100k

Type: Bar

Columns: Age (discrete, bin)

Rows: CNT (continuous, Age)

Filters:

- Age (bin)
- Income (bin): 90k selected

Social Media Deep Dive

(all bar charts)

Worksheet 7 – module 4 q1

Columns: Measure Values (continuous);

- SUM (Twitter ad)
- SUM (Instagram ad)
- SUM (Facebook ad)

Rows:

- Country (discrete, ascending)
- Measure Names (discrete, ascending)

Filters:

- Measure Names (social media platform names selected)
- Country (Unselected IND, ME, US)

Marks: Colour – Measure Names

<u>Worksheet 8 – module 4 q2</u>

Columns: Measure Values (continuous);

- SUM (Twitter ad)
- SUM (Instagram ad)
- SUM (Facebook ad)

Rows:

- Marital Status (discrete, ascending)
- Measure Names (discrete)

Filters:

- Measure Names (social media platform names selected)
- Marital Status

Marks: Colour – Measure Names

<u>Worksheet 9 – module 4 q3a</u>

Columns: Measure Values (continuous);

- SUM (Amt Liq)
- SUM (Amt Non Veg)
- SUM (Amt Comm)
- SUM (Amt Pes)
- SUM (Amt Chocolate)
- SUM (Amt Vege)

Rows:

- Country (discrete, ascending)
- Measure Names (discrete, ascending)

Filters:

- Success_indicator: False
- Measure Names (selected to match Measure Values)
- Country (Unselected IND, ME, US)

Marks: Colour – Measure Names

<u>Worksheet 10 – module 4 q3b</u>

Sames as Worksheet 9 except success_indicator set to "True"

Worksheets 11 – 14

(named module 4 q1 (90k). named module 4 q2 (90k), named module 4 q3a (90k), module 4 q3b (90k))

Same as Worksheets 7 – 10 with filter on Income (bin): 90k selected

Discount and Recency

Worksheet 15 – Deals without soc

Type: Bar

Columns: Country (discrete, descending)

Rows: SUM(Num Deals, continuous)

Filters:

- success_indicator: False
- Country (unselect ME)

Worksheet 16 – Deals with soc

Type: Bar

Columns: Country (discrete, descending)

Rows: SUM(Num Deals, continuous)

Filters:

- success_indicator: True
- Country (unselect ME)

Worksheet 17 – recency b&w no soc

Type: Box-and-Whisker

Rows: MEDIAN(Recency, continuous)

Filters:

- success_indicator: False
- Country (unselct ME)

Marks: Detail - Country

Worksheet 18 – recency b&w yes soc

Type: Box-and-Whisker

Rows: MEDIAN(Recency, continuous)

Filters:

- success_indicator: Trues
- Country (unselct ME)

Marks: Detail – Country

Note: Box-and-Whisker charts are only used in the accompanying presentation to describe the median recency. IQR, outliers etc not described. That said the increased richness/detail of this chart type is potentially useful in later cycles of analysis should a deeper understanding of this metric develop.

Worksheets 19 – 22

(named Deals without soc (90k), Deals with soc (90k), recency b&w no soc (90k), recency b&w yes soc (90k))

Same as worksheets 15 – 8 with filter on Income (bin): 90k selected