

2Market

Customers' Profile, Purchase Behaviours & Marketing Channels' Effectiveness

Data Analysis Report

LSE - Data Analytics – Career Accelerator

Assignment 1

Cristina Alonso Robles

April 2023

1. Methodology

1. Background/context of the business

2. MS Excel Analysis

- - Sorting, cleaning, and organising the datasets.
- - Exploratory Visualisations.
- - Initial Insights.

3. Tableau Desktop

- - Definition and calculation of the metrics.
- - Charts creation.
- - Dashboard Design.

4. PostgreSQL – pgAdmin

- - Querying the data to gather further insights.

2. Background/context of the business

- **Online Research** - Food retail market and grocery and general merchandise retailer industry.
- **Company Name and business definition:** 2Market is a global supermarket which sells products online and in-store.
- **Industry:** Grocery and general merchandise retailer.
- **Sector:** Online and in-store supermarkets (Food retail market).
- **Distribution channels:** Physical Supermarket Store and Online (Website).
- **Five Whys framework approach:**

Q1: Why 2Market doesn't know their customer's profile nor their purchase behaviour?

A1: Because they haven't analysed their data.

Q2: Why 2Market hasn't analysed their customer and advertising data?

A2: Because they were not applying a customer centric approach when making business decisions.

Q3: Why it wasn't their priority?

A3: Because they were just trying to increase their revenue without analysing their customer and advertising data.

Q4: Why 2Market was only focused on their revenue?

A: Because they did not have analytical resources and they thought that analysing data was a waste of time and a big investing in resources, such as data analytics tools and data analysts.

Q5: Why 2Market want to change their business approach now?

A5: Because they noticed they could invest their money more efficiently in different marketing channels depending on their customer's profile and customer's behaviour.

- **2Market's Business Problem**

2Market is not efficient in its business decisions because they are not analysing its customer's profile data nor measuring the effectiveness of its marketing campaigns

- **2Market's Business Objectives:**

- Understand of their customers' demographics
- Customers' purchase behaviour
- Potential relationships between advertising and purchases
- Advertising channels effectiveness
- Best-selling products and if that varies based on demographic

- **Why 2Market should try and solve this problem?**

By analysing and better understanding their customers' profiles and purchase behaviour as well as how their marketing channels perform, 2Market will be able to:

Implement more advanced targeting strategies based on the customers that spend the most so they can reach their audience and implement loyalty programs to maintain their current customers while also making sure they grow their business share in every market.

Make investing decisions more efficiently by assigning the right amount of budget to the right marketing channels. Those that are more effective at reaching their intended audience and that are driving more sales.

- **Stakeholders involved:** 2Market: Global and Local Sales, Finance, Marketing Leads, CEO and CFO.

3. Analytical Approach

A. MS Excel – Exploratory Analysis & Visualisations

- I used the sort and filter tab to make sure the ad_data excel file contains only the inputs (1 and 0) for each of the marketing channels, as specified by the metadata file.
- I filtered and sorted both datasets to clean the data and to ensure they do not contain any duplicates, blanks or format errors. No duplicate values were found. However, after looking at the maximum and minimum values, I found some outliers that I decided to remove from the data to make sure the final insights are accurate and robust. I removed the below customers id's, identified as outliers from the marketing data and ad data datasets. The customer Ids where the customers' age was not in line with the rest of the values and were outside of what we would consider as acceptable:

ID	Year_Birth	Age
11004	1894	129
1150	1900	123
7829	1901	122

- I removed the YOLO marital status, "You Only Live One"? – It's difficult to say the real marital status of these 2 customers so I removed them to make sure the data is consistent and integral.

ID	Year_Birth	Age	Education	Marital_Status
492	1974	49	PhD	YOLO
11133	1974	49	PhD	YOLO

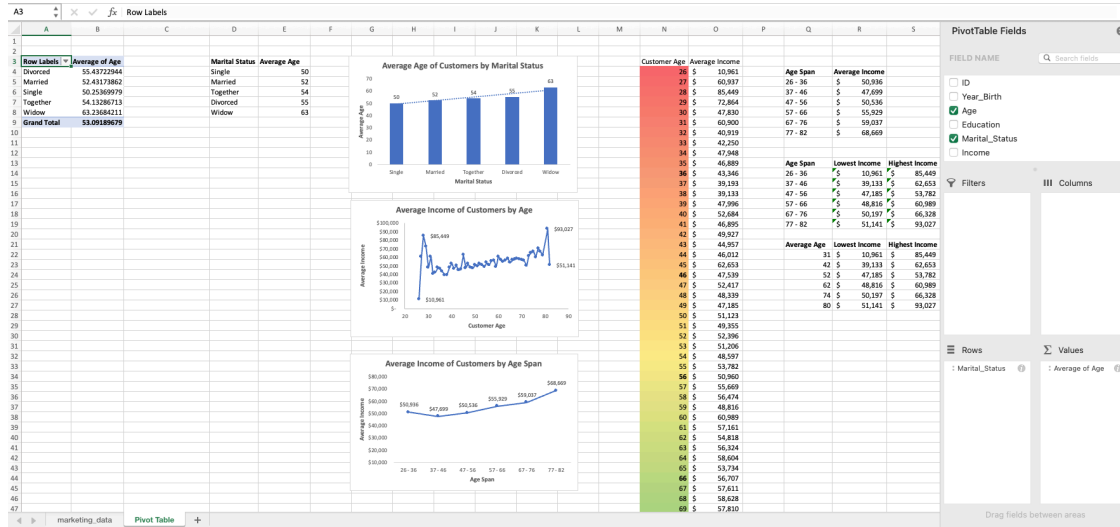
- I also removed the below 2 customers ids under the Absurd marital status from both data sets. After I finished cleaning the data in Excel, I saved a new CSV file for each of the datasets.

ID	Year_Birth	Age	Education	Marital_Status
4369	1958	65	Master	Absurd
7734	1994	29	Graduation	Absurd

- I merged all the records under the 'Alone' and 'Single' marital status into one, by changing the Alone marital status to 'Single'.
- Assuming that the data is up to date, I calculated the customers' age by deducting the year of birth of each customer from 2023.

	ID	Year_Birth	Age	Education
2184	11039	1976	47	Graduation
2185	11051	1957	66	PhD
2186	11056	1975	48	Graduation
2187	11071	1985	38	Graduation
2188	11074	1978	45	Graduation

- Once I considered that the data was cleaned, I created a pivot table from the main marketing data sheet to perform my analysis.



B. SQL – Querying the Data for Further Insights

- Before importing the csv files into SQL, I sense-checked the data in Excel and I made the following changes:
 - I set up the 'Income' attribute to General format. I removed the \$ symbol to make it readable when importing to SQL as INTEGER data type.
 - I changed the date format in Excel from DD-MM-YYYY to YYYY-MM-DD from the marketing_data spreadsheet.
- I saved the cleaned spreadsheets as a new csv cleaned file.
- **Creating marketing_data TABLE in SQL:**

```
1 CREATE TABLE marketing_data(  
2     "ID" BIGSERIAL PRIMARY KEY,  
3     "Year_Birth" INTEGER,  
4     "Age" INTEGER,  
5     "Education" VARCHAR,  
6     "Marital_Status" VARCHAR,  
7     "Income" INTEGER,  
8     "Kidhome" INTEGER,  
9     "Teenhome" INTEGER,  
10    "Dt_Customer" DATE,  
11    "Recency" INTEGER,
```

Data Output Messages Notifications

ID	Year_Birth	Age	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	Amt
[PK] bigint	integer	integer	character varying	character varying	integer	integer	integer	date	integer	intec

Query Query History Scratch Pad x

```
1 SELECT * FROM public.marketing_data  
2 ORDER BY "ID" ASC LIMIT 100  
3
```

Data Output Messages Notifications

ID	Year_Birth	Age	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	A
[PK] bigint	integer	integer	character varying	character varying	integer	integer	integer	date	integer	in
1	0	1986	37	Graduation	Married	70951	0	0	2013-05-04	66
2	1	1962	61	Graduation	Single	57091	0	0	2014-06-15	0
3	9	1976	47	Master	Single	46098	1	1	2012-08-18	86
4	13	1948	75	PhD	Widow	25358	0	1	2013-07-22	57
5	17	1972	51	PhD	Married	60491	0	1	2013-09-06	81
6	20	1966	57	2n Cycle	Married	46891	0	1	2013-09-01	91
7	22	1977	46	Graduation	Divorced	46310	1	0	2012-12-03	99
8	24	1961	62	Master	Together	17144	1	1	2014-02-15	96
9	25	1959	64	Graduation	Married	65148	0	1	2012-11-16	9
10	35	1988	35	Graduation	Married	25545	1	0	2012-09-08	35
11	48	1965	58	Graduation	Together	55761	0	1	2014-04-24	97
12	49	1971	52	Graduation	Single	20587	1	0	2014-05-11	39
13	55	1964	59	Graduation	Together	56253	0	1	2012-12-07	83

Total rows: 100 of 100 Query complete 00:00:00.201 Ln 1, Col 1

- **Creating ad_data TABLE in SQL:**

Scratch Pad X

Query Query History

```
1 SELECT * FROM public.ad_data
2 ORDER BY "ID" ASC
```

Data Output Messages Notifications

	ID [PK] bigint	Bulkmail_ad integer	Twitter_ad integer	Instagram_ad integer	Facebook_ad integer	Brochure_ad integer
1	0	0	0	0	0	0
2	1	0	0	0	0	1
3	9	0	0	0	0	0
4	13	0	0	0	0	0
5	17	0	0	0	0	0
6	20	0	0	0	0	0
7	22	0	0	0	0	0
8	24	0	0	0	0	0
9	25	0	0	0	0	0
10	35	0	0	0	0	0
11	48	0	1	0	0	0
12	49	0	0	0	0	0
13	55	0	0	0	0	0
14	67	0	0	0	0	0
15	73	0	0	0	0	0
16	75	0	0	0	0	0
17	78	0	0	0	0	0
18	87	0	0	0	0	0
19	89	0	0	0	0	0

○ Total Spend by Country

Query

Query History

```

1 SELECT "Country" ,
2 SUM("AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtPes" + "AmtChocolates" + "AmtComm") AS "Total_Spend"
3 FROM public."marketing_data"
4 GROUP BY "Country"
5 ORDER BY "Total_Spend" DESC;

```

Data Output

Messages

Notifications

≡

📄

▼

📄

🗑️

📦

📥

📈

	Country character varying 🔒	Total_Spend bigint 🔒
1	SP	657704
2	SA	211049
3	CA	165810
4	AUS	84360
5	IND	77317
6	GER	73198
7	US	67546
8	ME	3122

○ Total Spend by Marital Status

Query

Query History

```

1 SELECT "Marital_Status" ,
2 SUM("AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtPes" + "AmtChocolates" + "AmtComm") AS "Total_Spend"
3 FROM public."marketing_data"
4 GROUP BY "Marital_Status"
5 ORDER BY "Total_Spend" DESC;

```

Data Output

Messages

Notifications

≡

+

📄

⌵

📋

🗑️

💾

⬇️

🔄

	Marital_Status character varying 🔒	Total_Spend bigint 🔒
1	Married	506885
2	Together	347016
3	Single	289279
4	Divorced	141601
5	Widow	55325

○ Total Spend by Product by Country

Scratch Pad X

Query	Query History
1	SELECT
2	SUM("AmtLiq") AS "Total_AmtLiq",
3	SUM("AmtVege") AS "Total_AmtVege",
4	SUM("AmtNonVege") AS "Total_AmtNonVege",
5	SUM("AmtPes") AS "Total_AmtPes",
6	SUM("AmtChocolates") AS "Total_AmtChocolates",
7	SUM("AmtComm") AS "Total_AmtComm" , "Country"
8	FROM public."marketing_data"
9	GROUP BY "Country"
10	ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;
11	
12	

Data Output Messages Notifications

	Total_AmtLiq bigint	Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Country character varying
1	335637	28144	177847	40049	30070	45957	SP
2	105910	8937	58393	13663	9019	15127	SA
3	83504	7611	45375	9777	7604	11939	CA
4	42281	3587	22203	5334	4068	6887	AUS
5	36776	2980	20272	4601	2801	5768	GER
6	35899	3779	23671	4807	3214	5947	IND
7	32214	3034	20185	4411	2863	4839	US
8	1729	8	817	226	122	220	ME

○ Products Popularity based on Marital Status

Scratch Pad X

Query	Query History
1	SELECT
2	SUM("AmtLiq") AS "Total_AmtLiq",
3	SUM("AmtVege") AS "Total_AmtVege",
4	SUM("AmtNonVege") AS "Total_AmtNonVege",
5	SUM("AmtPes") AS "Total_AmtPes",
6	SUM("AmtChocolates") AS "Total_AmtChocolates",
7	SUM("AmtComm") AS "Total_AmtComm" , "Marital_Status"
8	FROM public."marketing_data"
9	GROUP BY "Marital_Status"
10	ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;
11	
12	

Data Output Messages Notifications

	Total_AmtLiq bigint	Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Marital_Status character varying
1	256976	21981	137888	30395	22926	36719	Married
2	175960	14468	94812	22279	14967	24530	Together
3	137763	12852	87138	18278	12772	20476	Single
4	75349	6357	34840	8123	6218	10714	Divorced
5	27902	2422	14085	3793	2878	4245	Widow

- **Products Popularity based on whether there are children or teens in the home or not.**

A. Kids in the home:

Query

Query History

```

1  SELECT
2    SUM("AmtLiq") AS "Total_AmtLiq",
3    SUM("AmtVege") AS "Total_AmtVege" ,
4    SUM("AmtNonVege") AS "Total_AmtNonVege",
5    SUM("AmtPes") AS "Total_AmtPes",
6    SUM("AmtChocolates") AS "Total_AmtChocolates",
7    SUM("AmtComm") AS "Total_AmtComm" , "Kidhome"
8  FROM public."marketing_data"
9  WHERE "Kidhome" > 0
10 GROUP BY "Kidhome"
11 ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;
```

Data Output

Messages

Notifications

	Total_AmtLiq bigint	Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Kidhome integer
1	92535	7570	43552	10887	7635	19588	1
2	3312	278	1384	316	179	776	2

B. Teens in the home:

Query

Query History

1

SELECT

2

SUM("AmtLiq") AS "Total_AmtLiq",

3

SUM("AmtVege") AS "Total_AmtVege" ,

4

SUM("AmtNonVege") AS "Total_AmtNonVege",

5

SUM("AmtPes") AS "Total_AmtPes",

6

SUM("AmtChocolates") AS "Total_AmtChocolates",

7

SUM("AmtComm") AS "Total_AmtComm" , "Teenhome"

8

FROM public."marketing_data"

9

WHERE "Teenhome" > 0

10

GROUP BY "Teenhome"

11

ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;

Data Output

Messages

Notifications

	Total_AmtLiq bigint	Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Teenhome integer
1	306408	19262	103706	26212	20622	43279	1
2	18083	891	6316	1191	786	2317	2

C. No kids in the home:

Query

Query History

1

SELECT

2

SUM("AmtLiq") AS "Total_AmtLiq",

3

SUM("AmtVege") AS "Total_AmtVege" ,

4

SUM("AmtNonVege") AS "Total_AmtNonVege",

5

SUM("AmtPes") AS "Total_AmtPes",

6

SUM("AmtChocolates") AS "Total_AmtChocolates",

7

SUM("AmtComm") AS "Total_AmtComm" , "Kidhome"

8

FROM public."marketing_data"

9

WHERE "Kidhome" = 0

10

GROUP BY "Kidhome"

11

ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;

Data Output

Messages

Notifications

Total_AmtLiq

Total_AmtVege

Total_AmtNonVege

Total_AmtPes

Total_AmtChocolates

Total_AmtComm

Kidhome

bigint

bigint

bigint

bigint

bigint

bigint

integer

1

578103

50232

323827

71665

51947

76320

0

D. No Teens in the home:

Query

Query History

1

SELECT

2

SUM("AmtLiq") AS "Total_AmtLiq",

3

SUM("AmtVege") AS "Total_AmtVege" ,

4

SUM("AmtNonVege") AS "Total_AmtNonVege",

5

SUM("AmtPes") AS "Total_AmtPes",

6

SUM("AmtChocolates") AS "Total_AmtChocolates",

7

SUM("AmtComm") AS "Total_AmtComm" , "Teenhome"

8

FROM public."marketing_data" |

9

WHERE "Teenhome" = 0

10

GROUP BY "Teenhome"

11

ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;

Data Output

Messages

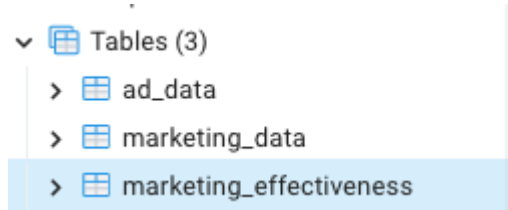
Notifications

	Total_AmtLiq bigint	Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Teenhome integer
1	349459	37927	258741	55465	38353	51088	0

SQL – Joining Tables

Since I did not find any duplicates and/or unmatched values when analysing both datasets in Excel, and both tables contain the exact same number of customers (ID “Unique Customer IDs”), I first used the Inner Join to join the marketing_data and the ad_data tables. Both tables have one column in common (ID) that I will use as the primary key to perform the join.

This means we must join the marketing_data and the ad_data tables. “Unique Customer ID” will be used as the primary key to perform the join between the two tables.



After creating the “marketing_effectiveness” table using the INNER join, there was not data retrieved in the table, so I decided to use the FULL OUTER JOIN.

Query

Query History

1 SELECT * FROM public.marketing_effectiveness

2

Scratch Pad

Data Output

Messages

Notifications

<

○ Social media platforms’ effectiveness by country.

Query rationale: First, I aggregated all social media lead conversions by platforms, using the ad data table and renaming the platforms lead conversions to Total Conversions. Secondly, I used the full outer join command to join the marketing data since we will be looking at each Country performance. Finally, I grouped the table by country and ordered the results by Total Conversions from each platform in descending order.

Scratch Pad X

Query

Query History

```
1 SELECT
2 SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
3 SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
4 SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions", md."Country"
5 FROM public."ad_data" ad
6 FULL OUTER JOIN public."marketing_data" md
7 ON ad."ID" = md."ID"
8 GROUP BY "Country"
9 ORDER BY "Total_Twitter_Conversions" DESC , "Total_Instagram_Conversions" DESC , "Total_Facebook_Conversions" DESC;
```

Data Output

Messages

Notifications

	Total_Twitter_Conversions bigint	Total_Instagram_Conversions bigint	Total_Facebook_Conversions bigint	Country character varying
1	87	88	76	SP
2	24	21	18	CA
3	20	21	20	SA
4	11	8	7	GER
5	10	6	7	IND
6	6	11	6	AUS
7	6	5	7	US
8	0	0	0	ME

○ Social Media Platform Marketing Effectiveness based on Marital Status:

Query rationale: First, I aggregated all social media lead conversions by platforms, using the ad data table and renamed the platforms lead conversions to Total Conversions. Secondly, I used the full outer join command to join the marketing data since we will be looking at each marital status individually. Finally, I grouped the table by marital status and ordered the results by Total Conversions from each platform in descending order.

Query

Query History

```
1 SELECT
2 SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
3 SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
4 SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions", md."Marital_Status"
5 FROM public."ad_data" ad
6 FULL OUTER JOIN public."marketing_data" md
7 ON ad."ID" = md."ID"
8 GROUP BY "Marital_Status"
9 ORDER BY "Total_Twitter_Conversions" DESC , "Total_Instagram_Conversions" DESC , "Total_Facebook_Conversions" DESC;
```

Data Output

Messages

Notifications

	Total_Twitter_Conversions bigint	Total_Instagram_Conversions bigint	Total_Facebook_Conversions bigint	Marital_Status character varying
1	62	66	62	Married
2	42	43	32	Together
3	32	31	30	Single
4	18	13	12	Divorced
5	10	7	5	Widow

I wanted to have a wider view of all the marketing activities together. The query used is the same as above, but including the Bulkmail and Brochure ads by Country as follow:

Scratch Pad X

QueryQuery History

1

SELECT

2

SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",

3

SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",

4

SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions",

5

SUM(ad."Bulkmail_ad") AS "Total_Bulkmail_Conversions",

6

SUM(ad."Brochure_ad") AS "Total_Brochure_Conversions" , md."Country"

7

FROM public."marketing_data" md

8

FULL OUTER JOIN public."ad_data" ad

9

ON md."ID" = ad."ID"

10

GROUP BY "Country"

11

ORDER BY "Total_Twitter_Conversions" DESC ,

12

"Total_Instagram_Conversions" DESC , "Total_Facebook_Conversions" DESC ,

13

"Total_Bulkmail_Conversions" , "Total_Brochure_Conversions" DESC;

Data OutputMessagesNotifications

	Total_Twitter_Conversions bigint	Total_Instagram_Conversions bigint	Total_Facebook_Conversions bigint	Total_Bulkmail_Conversions bigint	Total_Brochure_Conversions bigint	Country character varying
1	87	88	76	83	16	SP
2	24	21	18	18	6	CA
3	20	21	20	21	4	SA
4	11	8	7	10	2	GER
5	10	6	7	13	2	IND
6	6	11	6	9	0	AUS
7	6	5	7	8	0	US
8	0	0	0	1	0	ME

- **Correlation between the amount spent per product in Marital Status and the social media platforms that were used for advertising:**

Query rationale: First, I aggregated all the amounts spent by categories, using the marketing data table and renamed the amounts spent by category to Total amounts. Secondly, I used the full outer join command to join the ad data table since we will be looking at each social platform individually. I used the customer ID as the primary key. Finally, I grouped the table by marital status and ordered the results by Total Conversions from each platform in descending order.

Scratch Pad X

QueryQuery History

```
1 SELECT
2 SUM(md."AmtLiq") AS "Total_AmtLiq",
3 SUM(md."AmtVege") AS "Total_AmtVege",
4 SUM(md."AmtNonVege") AS "Total_AmtNonVege",
5 SUM(md."AmtPes") AS "Total_AmtPes",
6 SUM(md."AmtChocolates") AS "Total_AmtChocolates",
7 SUM(md."AmtComm") AS "Total_AmtComm",
8 SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
9 SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
10 SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions", md."Marital_Status"
11 FROM public."marketing_data" md
12 FULL OUTER JOIN public."ad_data" ad
13 ON md."ID" = ad."ID"
14 GROUP BY "Marital_Status"
15 ORDER BY "Total_Twitter_Conversions" DESC , "Total_Instagram_Conversions" DESC , "Total_Facebook_Conversions" DESC;
16
```

Data OutputMessagesNotifications

	Total_AmtLiq bigint	Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Total_Twitter_Conversions bigint	Total_Instagram_Conversions bigint	Total_Facebook_Conversions bigint	Marital_Status character varying
1	256976	21981	137888	30395	22926	36719	62	66	62	Married
2	175960	14468	94812	22279	14967	24530	42	43	32	Together
3	137763	12852	87138	18278	12772	20476	32	31	30	Single
4	75349	6357	34840	8123	6218	10714	18	13	12	Divorced
5	27902	2422	14085	3793	2878	4245	10	7	5	Widow

- Same query but including Bulkmail_ad and Brochure_ad by Country:

Scratch Pad X

Query Query History

```

1 SELECT
2 SUM(md."AmtLiq") AS "Total_AmtLiq",
3 SUM(md."AmtVege") AS "Total_AmtVege",
4 SUM(md."AmtNonVege") AS "Total_AmtNonVege",
5 SUM(md."AmtPes") AS "Total_AmtPes",
6 SUM(md."AmtChocolates") AS "Total_AmtChocolates",
7 SUM(md."AmtComm") AS "Total_AmtComm",
8 SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
9 SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
10 SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions",
11 SUM(ad."Bulkmail_ad") AS "Total_Bulkmail_Conversions",
12 SUM(ad."Brochure_ad") AS "Total_Brochure_Conversions", md."Country"
13 FROM public."marketing_data" md
14 FULL OUTER JOIN public."ad_data" ad
15 ON md."ID" = ad."ID"
16 GROUP BY "Country"
17 ORDER BY "Total_Twitter_Conversions" DESC,
18 "Total_Instagram_Conversions" DESC, "Total_Facebook_Conversions" DESC,
19 "Total_Bulkmail_Conversions", "Total_Brochure_Conversions" DESC;

```

Data Output Messages Notifications

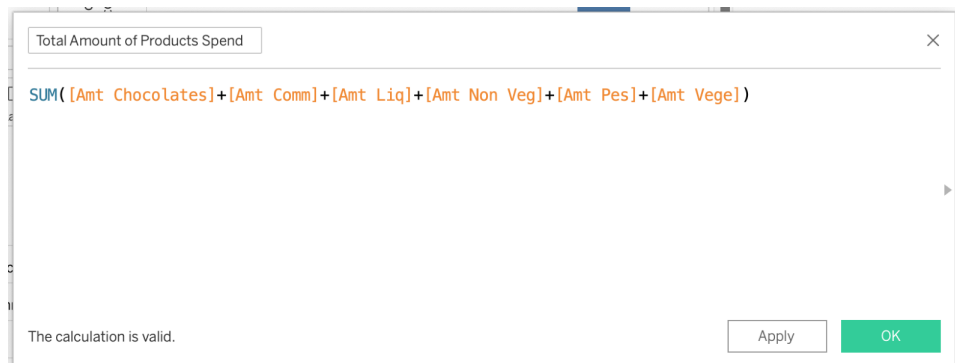
		Total_AmtVege bigint	Total_AmtNonVege bigint	Total_AmtPes bigint	Total_AmtChocolates bigint	Total_AmtComm bigint	Total_Twitter_Conversions bigint	Total_Instagram_Conversions bigint	Total_Facebook_Conversions bigint	Total_Bulkmail_Conversions bigint	Total_Brochure_Conversions bigint	Country character varying
1	7	28144	177847	40049	30070	45957	87	88	76	83	16	SP
2	4	7611	45375	9777	7604	11939	24	21	18	18	6	CA
3	0	8937	58393	13663	9019	15127	20	21	20	21	4	SA
4	6	2980	20272	4601	2801	5768	11	8	7	10	2	GER
5	9	3779	23671	4807	3214	5947	10	6	7	13	2	IND
6	1	3587	22203	5334	4068	6887	6	11	6	9	0	AUS
7	4	3034	20185	4411	2863	4839	6	5	7	8	0	US
8	9	8	817	226	122	220	0	0	0	1	0	ME

3. Dashboard Design & Development. Tableau

- Before I started to build my dashboard in Tableau, I wrote my ideas and designed it on paper. I made a list of the main stakeholders and the main questions they want to have answers for. Secondly, I linked those questions with the metrics and insights gathered from my data analysis to make sure that the selected charts would compile the main insights that I want to convey.
- I imported both data sets and joined the two tables with an inner join. I could have chosen any outer join too since both tables contain the exact number of customers with the same customer's id.
- I created a calculated field called “Total Purchases” as per below to build a chart looking using this metric and filter the view by marital status and by country.



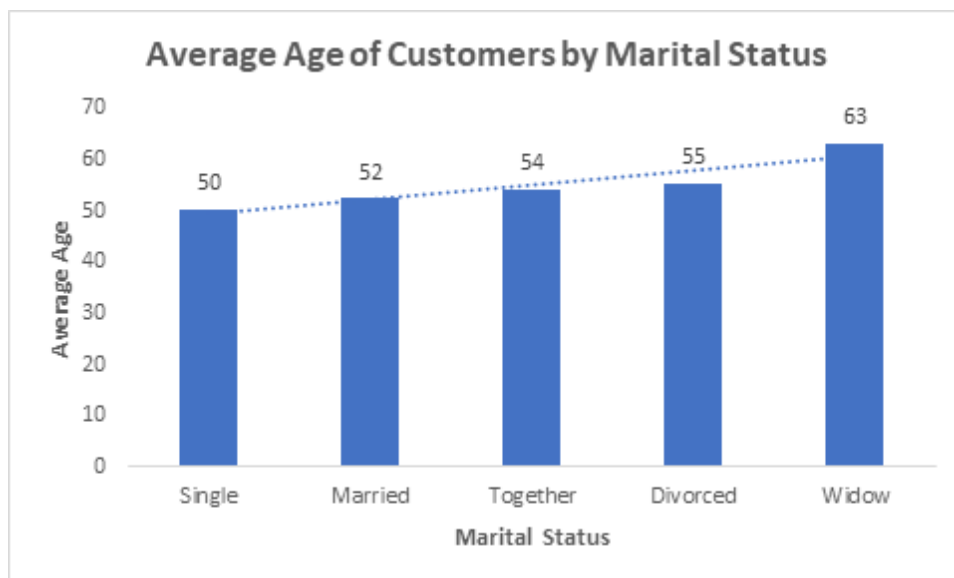
- I also calculated the Total Amount of Product Spend as per below, but I decided to select the chart looking at the amount spend by product category, to convey more insights into the spend by each category.



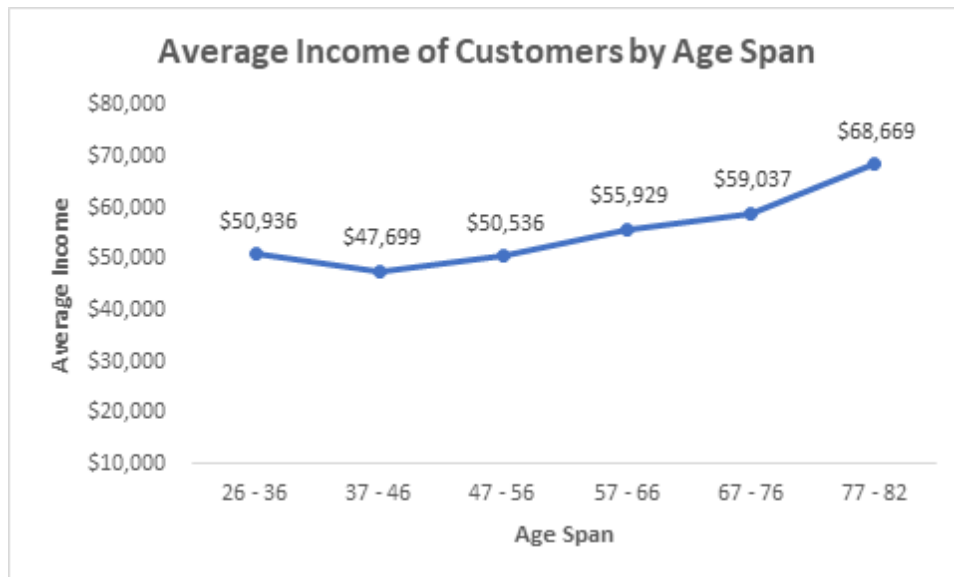
- After working on my charts, I showed them to my Success Manager to have her view and based on the feedback received I amended some of the charts.
- Building my dashboard. I selected the charts that were providing more insights into the business problem, and I made sure that I was following the accessibility requirements as below:
 - I adjusted the font size of the labels, making them bolder.
 - I applied the colour-blind palette when selecting the colours for my charts and I used shape marks encoded with colours to make sure the differences in the numbers are clearly displayed not only based on the colours but their shape (thinner and wider bar charts).
 - I made the tittle's charts dynamics so that when the filters are selected by the audience, the title adjusts accordingly.
 - I used accessible filters.
 - I added a descriptive caption for each chart with a brief description of the main metrics and instructions on how to use the filters.
 - I selected the automatic size so that the dashboard is resized to fit any device and screen it is displayed on.

4. Patterns, Trends & Insights

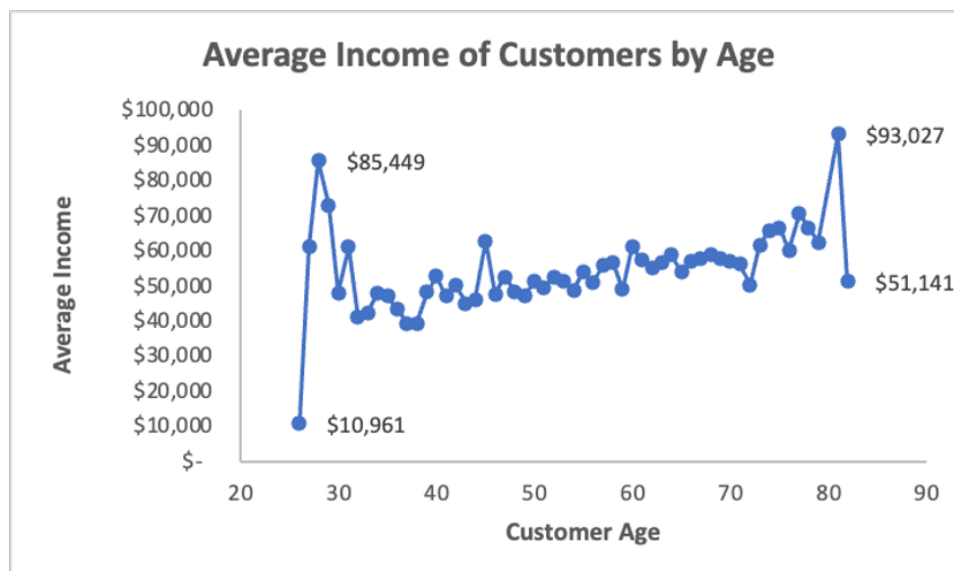
- The average age of 2Market's customers is 53.
- The highest average age by marital status is 63 for Widow Customers.
- The marital status that has the lowest average age is 'Single' (50) whereas the highest average age was reported for "Widow" (63).
- Even though the trend of the average age across the various types of marital status is upward, there is not a big difference between them.



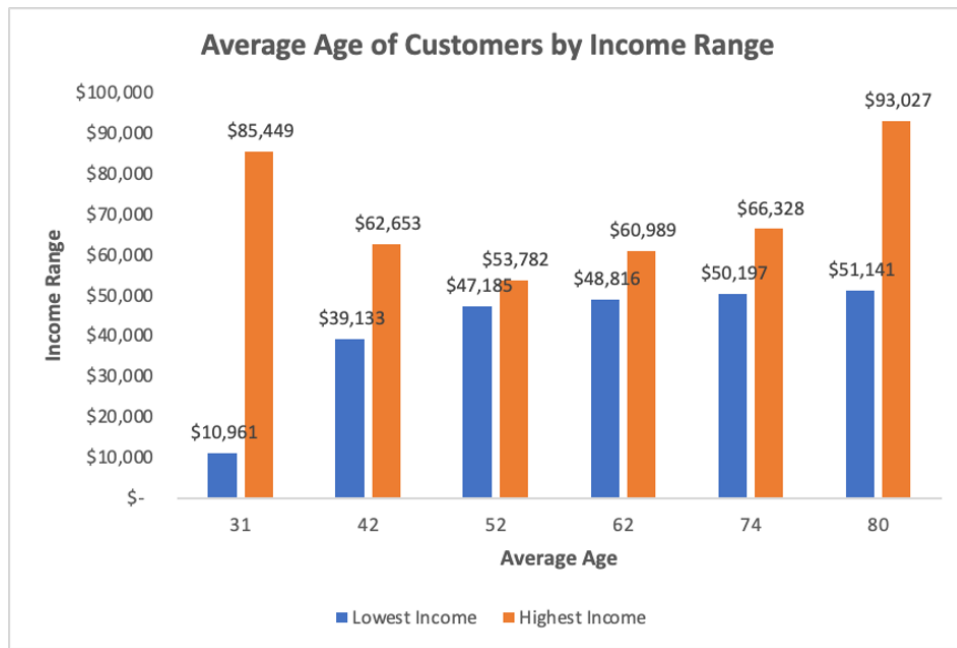
- The average income across all customers is US\$ 52,237.
- The average age of customers who earn a yearly income between US\$90,000 and US\$100,000 is 51.



- In the chart below, we can see the distribution of the average income by customer's age. Even though the lowest income is reported among the youngest customers, we can also see that the second highest average income is also reported among the youngest group.
- The same happens for the lowest and highest average income among the oldest group, indicating that the average income is not necessarily an indication of an older customer.



- There is no correlation between the age of the customers and the income of those who earn between US\$90,000 and US\$100,000.



	Total Amount of Product Spend	Products Popularity
Marital Status	<p>The customers that spend the most by marital status are those who are married, followed by those who are “together” in a relationship and the “single” in the third place.</p> <p>Divorced and widow customers are the ones who spend the least.</p>	<p>Alcoholic beverages is the category most popular across all countries and marital status considering the amount spend by product in each country and by each marital status group, followed by meat items, commodities, fish products and chocolates. Vegetables is the category less consumed by customers.</p>
Country	<p>Spain leads the list of countries by total amount of product spend, followed by South Africa and Canada.</p> <p>Montenegro is at the far away for these, being at the bottom of the list.</p>	

Products Popularity based on whether there are children or teens in the home	Amount spent per product vs. Marketing Channel Effectiveness
<p>When there are kids or teens in the home, customers spend less in products in general. By category, alcoholic beverages remain the most product bought by customers with kids or teens at home, followed by meat items, commodities and fish products and this trend is the same across all households regardless of the presence of kids or teenagers in the home.</p>	<p>Married customers are the ones who buy more products influenced by social media campaigns.</p> <p>Instagram is the primary platform that influences their purchases, followed very closely by Twitter and Facebook. The next customers on the list are those that declare to be ‘together’, followed by the single and divorced customers.</p> <p>Widow’s customers are the least likely to purchase influenced by social media ads. The effectiveness of all the marketing campaigns is greater in Spain, followed by Canada and South Africa. Lead conversions are coming primarily from Instagram ads and Twitter, followed by Facebook.</p> <p>Montenegro remains the country with the least lead conversions.</p>

Conclusions & Recommendations

- 2Market's **Customer-Persona** is a 53-year-old individual who is either married or is in a relationship.
- They earn on average an annual **income** of **USD 52K**. They might or not have kids or teens in the home but regardless of this, **the trend in the purchase behaviour by product category and by marital status across all countries remains static**.
- 2Market should consider applying a **customer-centric approach** when developing their business and marketing strategies and apply **different strategies by Country due to different lifestyles and socio-economic factors**.

Key Insights To Consider In The Future

- Develop and grow your **alcoholic beverages** range of products and **meat items**.
- Send a **survey** to your customers to know more about their preferred brands of alcoholic beverages and to get insights on whether they value **organic meat products**.
- Make sure you reward your most valued customers by sending them **Birthday's cards** and **seasonal deals**.
- When it comes to investing in your marketing campaigns, **Twitter** is the marketing channel where you should invest the most, followed by **Bulkmail** campaigns and advertising on **Instagram** and **Facebook**.

Appendix

Appendix 1

- **Questions I would ask to better understand the project:**
 - When did 2Market was funded?
 - When did 2Market go online with its website/online store?
 - For how long has 2Market been gathering customer data and advertising data?
 - How often do they update the marketing data and ad data databases? Are the marketing data and ad data files collected monthly as the number of web visits suggests?
 - Did they perform data analysis before to inform their business decisions? If so, did they use the insights from their data analysis to inform their business strategies? If so, what did they find more interesting from their analysis in the past? Do they have access to this historical data (customer and advertising data) from previous months/years?
 - Who are the main stakeholders of the project whom we will be presenting at our analysis and insights?
 - The global food and grocery retail market is segmented based on product type, category, distribution channel and geography. Has 2Market tailored its strategies to address their customers' profile, purchase behaviour and considering the location of their stores in different countries?
 - Do they apply different business strategies based on its distribution channels?
 - Do they account for the new consumption models emerging from the social and economic changes, like the COVID-19 pandemic and the financial crisis that directly impact their customers purchase behaviour?
 - How seriously do they take investing in innovation, such as monitoring and innovating with new delivery models, auto check-out tills or facilitating a seamless user experience within their website/online store?
 - Do they account for the consumer polarization in terms of listening to different customer demands such as offering more healthy and sustainable products for those willing to pay more or on the one hand, consumers that are becoming more price sensitive, (mainly the lower-income households) that are trading down and reducing costs by looking more for promotions and cheaper options.
 - How seriously do they take investing in their personnel to make sure they hire the right talent within their stores?

- **Questions I would like to try and answer by analysing the data:**

What is the main revenue source, in-store purchases or online buying?

What demographic profile is the online shopper?

What demographic profile is the in-store shopper?

What products are contributing the most to the 2Market revenue?

- What are their customer profiles? (Based on their age, marital status, income, kids/teens at home and location)
- What's the customer's purchase behaviour? Based on the highest amount spent per product category. Do they buy in-store or online? What's the customer profile of those who purchase the most with a discount? (Based on the number of deals purchased made with a discount)
- What type of products show the biggest spends? Does this differ based on their customer profile and customer's location?
- Number of website purchases vs in-store purchases per customer profile and customer's location.
- Relationship between the number of lead conversions per customer and the type of media or ad that drove the highest conversion rates.
- What are the more effective channels driving more lead conversions by customer demographic and location (country)?

Appendix 2

*Reference: customers' count per Country

AUS: 146

CA: 264

GER: 116

IND: 145

ME: 3

SA: 336

SP: 1092

US: 107

Appendix 3

- **SQL Query to create the marketing_data TABLE in SQL:**

```
CREATE TABLE marketing_data(  
    "ID" BIGSERIAL PRIMARY KEY,  
    "Year_Birth" INTEGER,  
    "Age" INTEGER,  
    "Education" VARCHAR,  
    "Marital_Status" VARCHAR,  
    "Income" INTEGER,  
    "Kidhome" INTEGER,  
    "Teenhome" INTEGER,  
    "Dt_Customer" DATE,  
    "Recency" INTEGER,  
    "AmtLiq" INTEGER,  
    "AmtVege" INTEGER,  
    "AmtNonVege" INTEGER,  
    "AmtPes" INTEGER,  
    "AmtChocolates" INTEGER,  
    "AmtComm" INTEGER,  
    "NumDeals" INTEGER,  
    "NumWebBuy" INTEGER,  
    "NumWalkinPur" INTEGER,  
    "NumVisits" INTEGER,  
    "Response" INTEGER,  
    "Complain" INTEGER,  
    "Country" VARCHAR,
```

```
"Count_success" INTEGER);
```

```
SELECT * FROM public.marketing_data  
ORDER BY "ID" ASC;
```

- **SQL Query to create the ad_data TABLE in SQL:**

```
CREATE TABLE ad_data(  
"ID" BIGSERIAL PRIMARY KEY,  
"Bulkmail_ad" INTEGER,  
"Twitter_ad" INTEGER,  
"Instagram_ad" INTEGER,  
"Facebook_ad" INTEGER,  
"Brochure_ad" INTEGER);
```

```
SELECT * FROM public.ad_data  
ORDER BY ID ASC
```

- **SQL Query to retrieve the Total Spend by Country:**

```
SELECT "Country" ,  
SUM("AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtPes" + "AmtChocolates" +  
"AmtComm") AS "Total_Spend"  
FROM public."marketing_data"  
GROUP BY "Country"  
ORDER BY "Total_Spend" DESC;
```


- **SQL Query to retrieve the Total Spend by Marital Status:**

```
SELECT "Marital_Status" ,  
      SUM("AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtPes" + "AmtChocolates" +  
        "AmtComm") AS "Total_Spend"  
FROM public."marketing_data"  
GROUP BY "Marital_Status"  
ORDER BY "Total_Spend" DESC;
```

- **SQL Query to retrieve the Spend by Product by Country:**

```
SELECT  
      SUM("AmtLiq") AS "Total_AmtLiq",  
      SUM("AmtVege") AS "Total_AmtVege",  
      SUM("AmtNonVege") AS "Total_AmtNonVege",  
      SUM("AmtPes") AS "Total_AmtPes",  
      SUM("AmtChocolates") AS "Total_AmtChocolates",  
      SUM("AmtComm") AS "Total_AmtComm" , "Country"  
FROM public."marketing_data"  
GROUP BY "Country"  
ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC ,  
"Total_AmtNonVege" DESC , "Total_AmtPes" DESC , "Total_AmtChocolates"  
DESC;
```

- **SQL Query to retrieve the Products Popularity based on Marital Status:**

```
SELECT
    SUM("AmtLiq") AS "Total_AmtLiq",
    SUM("AmtVege") AS "Total_AmtVege",
    SUM("AmtNonVege") AS "Total_AmtNonVege",
    SUM("AmtPes") AS "Total_AmtPes",
    SUM("AmtChocolates") AS "Total_AmtChocolates",
    SUM("AmtComm") AS "Total_AmtComm" , "Marital_Status"
FROM public."marketing_data"
GROUP BY "Marital_Status"
ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege"
DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;
```

SQL Query to retrieve the Product's Popularity based on whether there are children or teens in the home or not.

A. When there are Kids in the home:

```
SELECT
    SUM("AmtLiq") AS "Total_AmtLiq",
    SUM("AmtVege") AS "Total_AmtVege" ,
    SUM("AmtNonVege") AS "Total_AmtNonVege",
    SUM("AmtPes") AS "Total_AmtPes",
    SUM("AmtChocolates") AS "Total_AmtChocolates",
    SUM("AmtComm") AS "Total_AmtComm" , "Kidhome"
FROM public."marketing_data"
WHERE "Kidhome" > 0
GROUP BY "Kidhome"
ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege"
DESC , "Total_AmtPes" DESC , "Total_AmtChocolates" DESC;
```

B. When there are Teens in the home:

```
SELECT
SUM("AmtLiq") AS "Total_AmtLiq",
SUM("AmtVege") AS "Total_AmtVege" ,
SUM("AmtNonVege") AS "Total_AmtNonVege",
SUM("AmtPes") AS "Total_AmtPes",
SUM("AmtChocolates") AS "Total_AmtChocolates",
SUM("AmtComm") AS "Total_AmtComm" , "Teenhome"
FROM public."marketing_data"
WHERE "Teenhome" > 0
GROUP BY "Teenhome"
ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC , "Total_AmtNonVege" DESC ,
"Total_AmtPes" DESC , "Total_AmtChocolates" DESC;
```

C. No kids in the home:

```
SELECT
SUM ("AmtLiq") AS "Total_AmtLiq",
SUM ("AmtVege") AS "Total_AmtVege" ,
SUM ("AmtNonVege") AS "Total_AmtNonVege",
SUM ("AmtPes") AS "Total_AmtPes",
SUM("AmtChocolates") AS "Total_ AmtChocolates"
SUM ("AmtComm") AS "Total_AmtComm" , "Teenhome"
FROM public. "marketing_data"
WHERE "Teenhome" = 0
GROUP BY "Teenhome"
ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC, "Total_AmtNonVege" DESC,
"Total_Amtes" DESC, "Total_AmtChocolates" DESC;
```

D. No teens in the home:

```
SELECT
SUM("AmtLiq") AS "Total_AmtLiq",
SUM ("AmtVege") AS "Total_AmtVege",
SUM ("AmtNonVege") AS "Total_AmtNonVege",
SUM("AmtPes") AS "Total_AmtPes",
SUM ("AmtChocolates") AS "Total_AmtChocolates"
SUM ("AmtComm") AS "Total_AmtComm" , "Teenhome"
FROM public. "marketing_data"
WHERE "Teenhome" = 0
GROUP BY "Teenhome"
ORDER BY "Total_AmtLiq" DESC, "Total_AmtVege" DESC, "Total_AmtNonVege" DESC,
"Total_AmtPes" DESC, "Total_AmtChocolates" DESC;
```

Appendix 4

- **Joining Tables in SQL:**

```
CREATE TABLE marketing_effectiveness  
  
AS SELECT ad."ID", "Bulkmail_ad", "Twitter_ad", "Instagram_ad", "Facebook_ad",  
"Brochure_ad",  
  
md."Count_success", "Country", "Marital_Status", "AmtLiq", "AmtVege",  
"AmtNonVege", "AmtPes", "AmtChocolates", "AmtComm"  
  
FROM public."ad_data" ad  
  
FULL OUTER JOIN public."marketing_data" md  
  
ON ad."ID" = md."ID";  
  
SELECT * FROM public.marketing_effectiveness;
```

- **SQL Query to retrieve the Social media platforms' effectiveness by country:**

```
SELECT  
  
SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",  
SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",  
SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions", md."Country"  
  
FROM public."ad_data" ad  
  
FULL OUTER JOIN public."marketing_data" md  
  
ON ad."ID" = md."ID"  
  
GROUP BY "Country"  
  
ORDER BY "Total_Twitter_Conversions" DESC , "Total_Instagram_Conversions" DESC ,  
"Total_Facebook_Conversions" DESC;
```

- **SQL Query to retrieve the Social Media Platform Marketing Effectiveness based on Marital Status:**

```
SELECT
SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions", md."Marital_Status"
FROM public."ad_data" ad
FULL OUTER JOIN public."marketing_data" md
ON ad."ID" = md."ID"
GROUP BY "Marital_Status"
ORDER BY "Total_Twitter_Conversions" DESC , "Total_Instagram_Conversions" DESC ,
"Total_Facebook_Conversions" DESC;
```

- **SQL Query to retrieve all the Marketing Channels Effectiveness by Country:**

```
SELECT
SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions",
SUM(ad."Bulkmail_ad") AS "Total_Bulkmail_Conversions" ,
SUM(ad."Brochure_ad") AS "Total_Brochure_Conversions" , md."Country"
FROM public."marketing_data" md
FULL OUTER JOIN public."ad_data" ad
ON md."ID" = ad."ID"
GROUP BY "Country"
ORDER BY "Total_Twitter_Conversions" DESC ,
"Total_Instagram_Conversions" DESC , "Total_Facebook_Conversions" DESC ,
"Total_Bulkmail_Conversions" , "Total_Brochure_Conversions" DESC;
```

- **SQL to retrieve the correlation between the amount spent per product in Marital Status and the social media platforms that were used for advertising:**

```
SELECT
SUM(md."AmtLiq") AS "Total_AmtLiq",
SUM(md."AmtVege") AS "Total_AmtVege",
SUM(md."AmtNonVege") AS "Total_AmtNonVege",
SUM(md."AmtPes") AS "Total_AmtPes",
SUM(md."AmtChocolates") AS "Total_AmtChocolates",
SUM(md."AmtComm") AS "Total_AmtComm",
SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions", md."Marital_Status"
FROM public."marketing_data" md
FULL OUTER JOIN public."ad_data" ad
ON md."ID" = ad."ID"
GROUP BY "Marital_Status"
ORDER BY "Total_Twitter_Conversions" DESC , "Total_Instagram_Conversions" DESC ,
"Total_Facebook_Conversions" DESC;
```

- **Same query as above but including Bulkmail_ad and Brochure_ad by Country:**

```
SELECT
SUM(md."AmtLiq") AS "Total_AmtLiq",
SUM(md."AmtVege") AS "Total_AmtVege",
SUM(md."AmtNonVege") AS "Total_AmtNonVege",
SUM(md."AmtPes") AS "Total_AmtPes",
SUM(md."AmtChocolates") AS "Total_AmtChocolates",
SUM(md."AmtComm") AS "Total_AmtComm",
SUM(ad."Twitter_ad") AS "Total_Twitter_Conversions",
SUM(ad."Instagram_ad") AS "Total_Instagram_Conversions",
SUM(ad."Facebook_ad") AS "Total_Facebook_Conversions",
SUM(ad."Bulkmail_ad") AS "Total_Bulkmail_Conversions",
SUM(ad."Brochure_ad") AS "Total_Brochure_Conversions" , md."Country"
FROM public."marketing_data" md
FULL OUTER JOIN public."ad_data" ad
ON md."ID" = ad."ID"
GROUP BY "Country"
ORDER BY "Total_Twitter_Conversions" DESC ,
"Total_Instagram_Conversions" DESC , "Total_Facebook_Conversions" DESC ,
"Total_Bulkmail_Conversions" , "Total_Brochure_Conversions" DESC;
```


- **Some errors spotted when working with SQL and how I actioned them:**

At some point in the middle of my analysis in SQL, when I was reviewing the written queries to sense-check the data, I noticed that when creating the marketing_data table in SQL, I slightly change the “AmtNonVeg” as it is written in the marketing_data csv file and the metadata file to “AmtNonVege”. I was wondering how SQL did not give me any error when creating the table and I thought that it was worth double checking that the data for this attribute was correct, so I performed a few queries in SQL and cross-checked the result values with the ones in Excel as follow:

Query

Query History

1

SELECT "AmtNonVege"

2

FROM public."marketing_data"









3


WHERE "ID" = 0;

Data Output

Messages

Notifications



	AmtNonVege 
1	554

When performing the query in SQL and filtering the data in Excel using the same parameters in both tools, I saw that the same output was retrieved from SQL and Excel, so I decided to continue working on the analysis without having to make any changes retrospectively.

Query Query History		L1			
1 SELECT SUM("AmtNonVege") AS "Total_AmtNonVege" , "ID" 2 FROM public."marketing_data" 3 GROUP BY "ID" 4 ORDER BY "Total_AmtNonVege" DESC ;					
Data Output Messages Notifications					
	Total_AmtNonVege bigint		ID [PK] bigint		
1	1725		4931		
2	1725		5376		
3	1622		1501		
4	1582		8475		
5	984		1065		
6	981		1619		
7	974		4947		
8	968		3104		
9	961		9220		
10	951		2109		
11	951		11071		
12	946		2147		
13	940		3179		
14	936		7396		
15	935		1172		
16	932		2002		
17	929		5848		
18	925		10955		
19	925		10489		

Appendix 5

- **Recommendations on the customer's sample by country:**

There are massive differences in the number of customers by Country collected in the datasets. Montenegro had only 3 customers in the list and Spain had 1092. This is a clear indicator of the differences in the product spend by Country, making Spain the Country with the greatest spend of products across all categories and where most of the lead's conversions take place.

It is important to note that the insights gather from this analysis might not be comparable between countries and taken only as a reference of the customers behaviours and marketing channels efficiency in those specific markets.

It is highly advisable that for the next analysis, the number of customers by Country should be approximately the same, or at least establish a minimum sample of customers per Country.

The Beginning of Investing More Efficiently

2Market