Appendix

SQL SCRIPTS

2.1 Import to Tableau and PostgreSQL

The cleaned data was imported into Tableau and left joined with the Ad data provided.

The marketing_data and ad_data were imported into 2 tables created in PostgreSQL by the following syntax:

```sql
CREATE TABLE marketing_data(
    "ID" INT PRIMARY KEY,
    "Year_Birth" INT,
    "Age" INT,
    "Education" VARCHAR(50),
    "Marital_Status" VARCHAR(50)
    "Income" INT,
    "Kidhome" INT,
    "Teenhome" INT,
    "Dt_Customer" DATE,
    "Recency" INT,
    "AmtLiq" INT,
    "AmtVege" INT,
    "AmtNonVeg" INT,
    "AmtPes" INT,
    "AmtChocolates" INT,
    "AmtComm" INT,
    "Total_Sales" INT,
    "NumDeals" INT,
    "NumWebBuy" INT,
    "NumWalkinPur" INT,
    "NumVisits" INT,
    "Response" BOOL,
    "Complain" BOOL,
    "Country" VARCHAR(50),
    "Count_success" INT
);
```

```sql
CREATE TABLE ad_data(
    "ID" INT PRIMARY KEY,
    "Bulkmail_ad" BOOL,
    "Twitter_ad" BOOL,
    "Instagram_ad" BOOL,
    "Facebook_ad" BOOL,
    "Brochure_ad BOOL"
);
```

2.2 Children / No Children

- Aggregated into those with and without children.

```sql
CREATE TABLE kid_data AS
SELECT "ID",
        'No children' child_type
    FROM public.marketing_data
    WHERE "Kidhome" = '0'
    AND "Teenhome" = '0'
UNION ALL
    SELECT "ID",
    'Children' child_type
    FROM public.marketing_data
    WHERE "Kidhome" != '0'
    OR "Teenhome" != '0'
```

2.4

- Aggregated into Products: PostgreSQL to create new table enabling each product line to be incorporated into records.

```
CREATE TABLE product_analysis AS
SELECT *
    FROM (
    SELECT "ID",
        'Alcohol' product_type,
        "AmtLiq" amount
    FROM public.marketing_data
UNION ALL
        SELECT "ID",
        'Vegetables' product_type,
        "AmtVege" amount
    FROM public.marketing_data
UNION ALL
        SELECT "ID",
        'Meat' product_type,
        "AmtNonVeg" amount
    FROM public.marketing_data
UNION ALL
        SELECT "ID",
        'Fish' product_type,
        "AmtPes" amount
    FROM public.marketing_data
UNION ALL
        SELECT "ID",
        'Chocolate' product_type,
        "AmtChocolates" amount
    FROM public.marketing_data
UNION ALL
        SELECT "ID",
        'Commodities' product_type,
        "AmtComm" amount
    FROM public.marketing_data
);
```

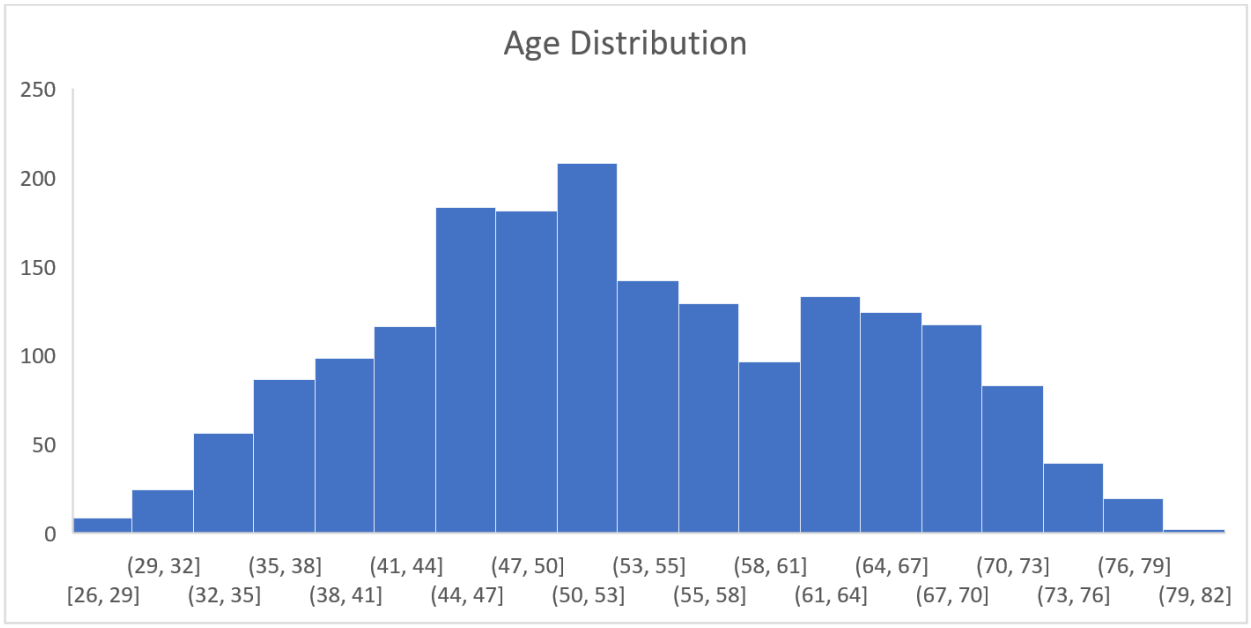
- The new table was joined with relevant columns from marketing data in SQL.

Excel Exploratory Analysis.

SUPPLEMENTS

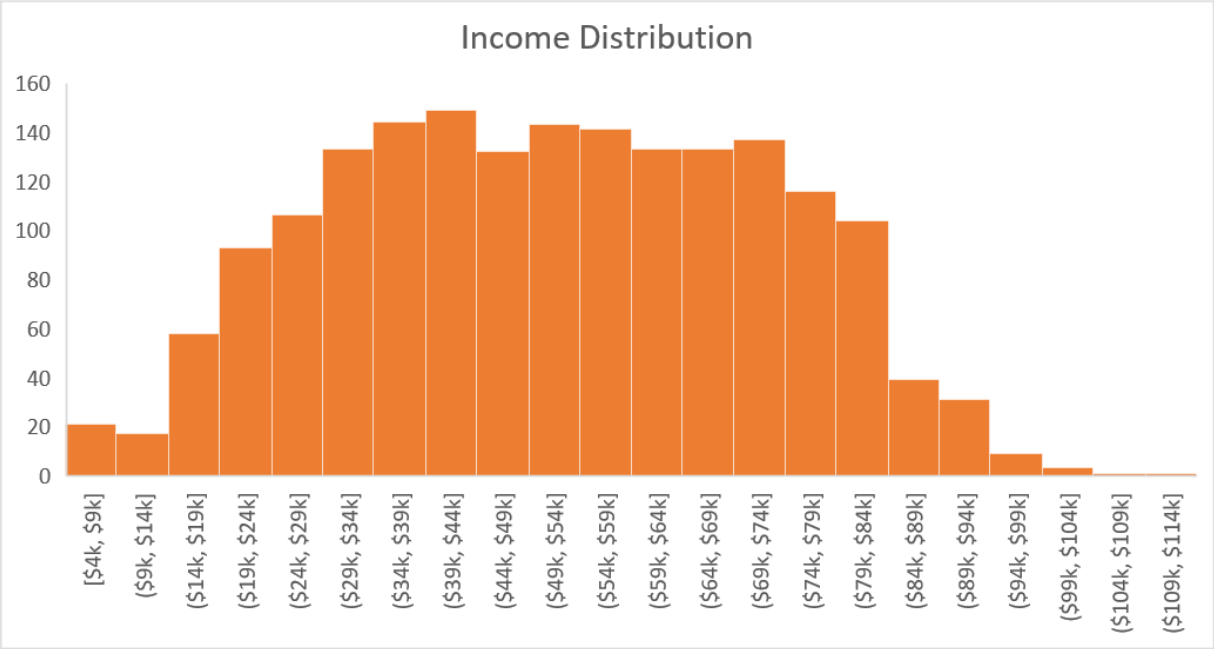
MODULE 2

Age Distribution



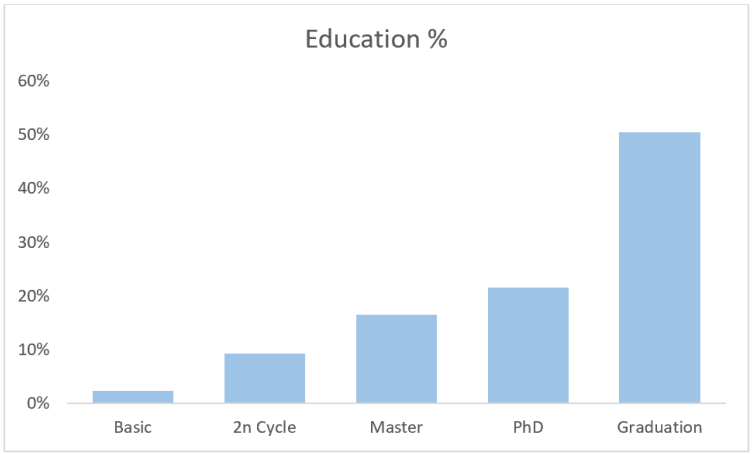
Mean	53.17
Median	52

The age bracket with the highest number of customers is 50-53.



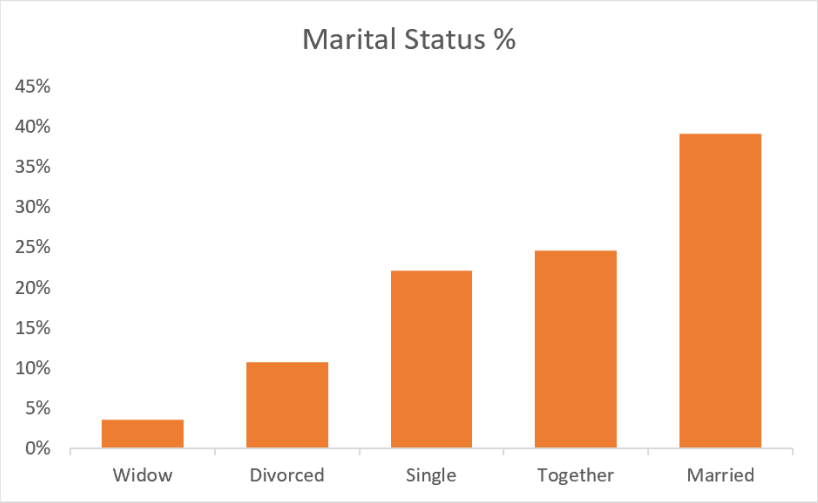
Descriptive Statistics:

Mean	**$51,891**
Standard Error	$480
Median	**$51,684**
Mode	$7,500
Standard Deviation	$20,611
Sample Variance	$424,817,626
Kurtosis	-0.824427683
Skewness	0.00801373
Range	$109,306
Minimum	$4,428
Maximum	$113,734
Sum	$95,687,580
Count	1844

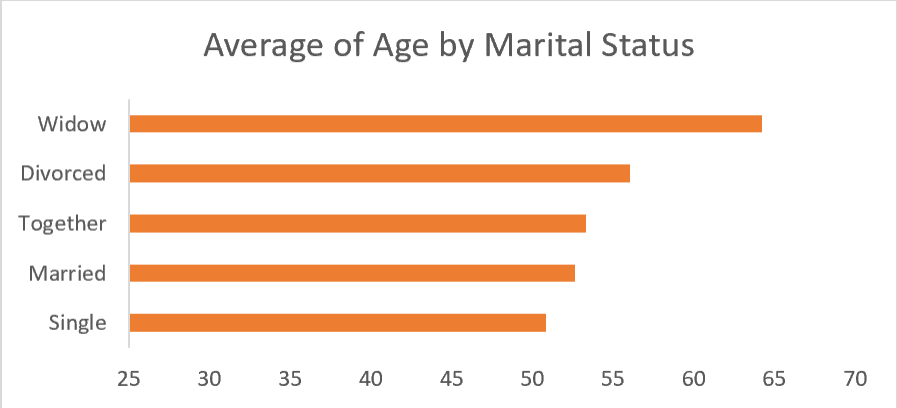


Average Age and Marital Status

Predictably, the highest average age customers had the marital status “Widow” (64.2) and “Divorced” (56.0). The lowest average age was in the marital status “Single” (50.8). Pivot tables and Pivot Charts were used for advanced aggregations.



Married and together form the majority of 2Market’s Customers.

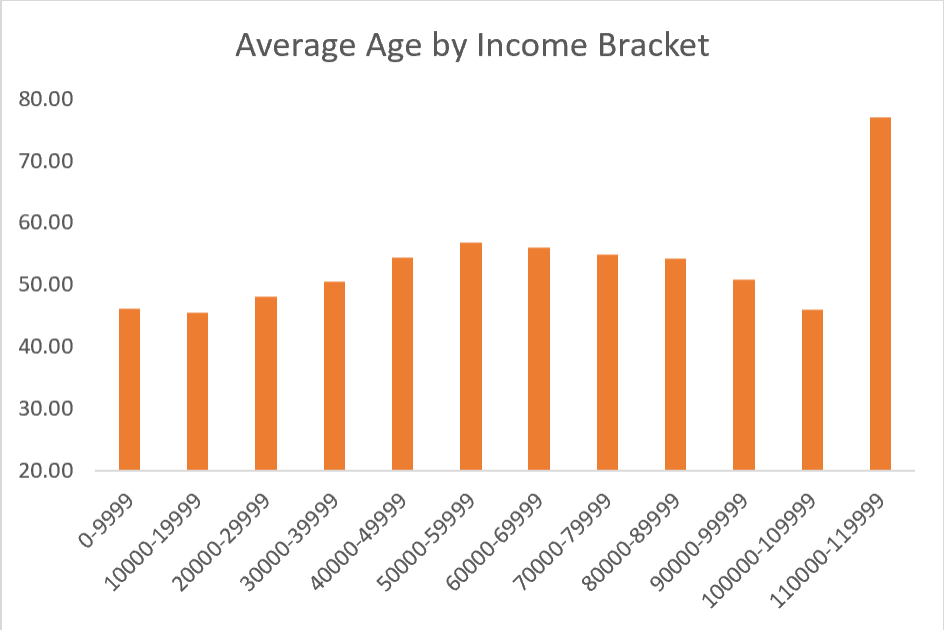


i. Income Bracket $90k to $100k

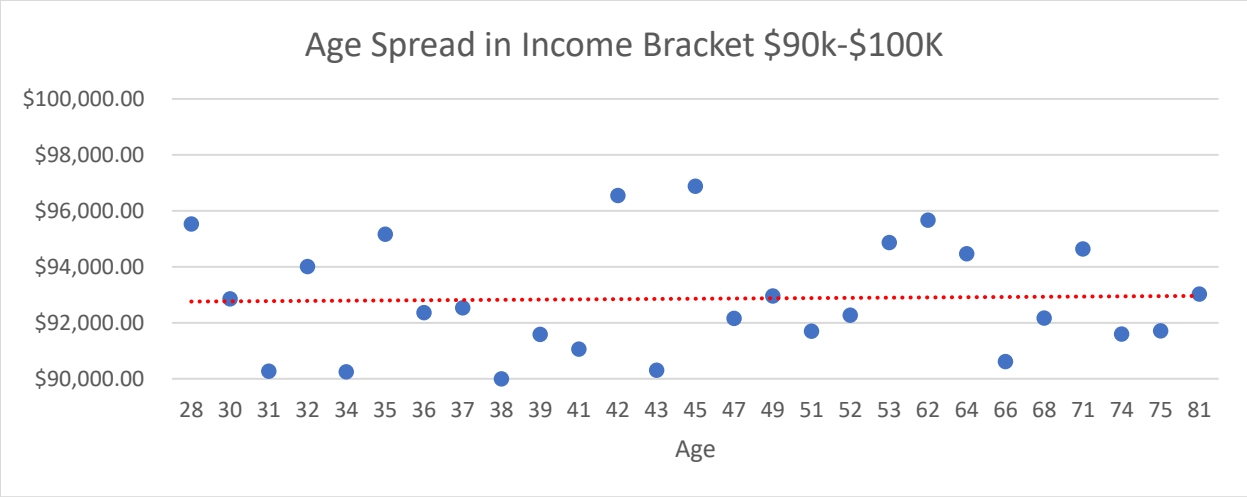
The average age of customers with a yearly income of between $90,000 and $100,000 was 50.5.

To find the average age of a single salary bracket, one might use a simple filter. To find multiple, the grouping function within a pivot table was utilised.

Salary Bracket	Average of Age
0-9999	45.6
10000-19999	44.8
20000-29999	47.7
30000-39999	50.6
40000-49999	54.1
50000-59999	56.7
60000-69999	56.3
70000-79999	55.2
80000-89999	54.2
90000-99999	50.5
100000-109999	46.0
110000-119999	77.0
Grand Total	**53.2**

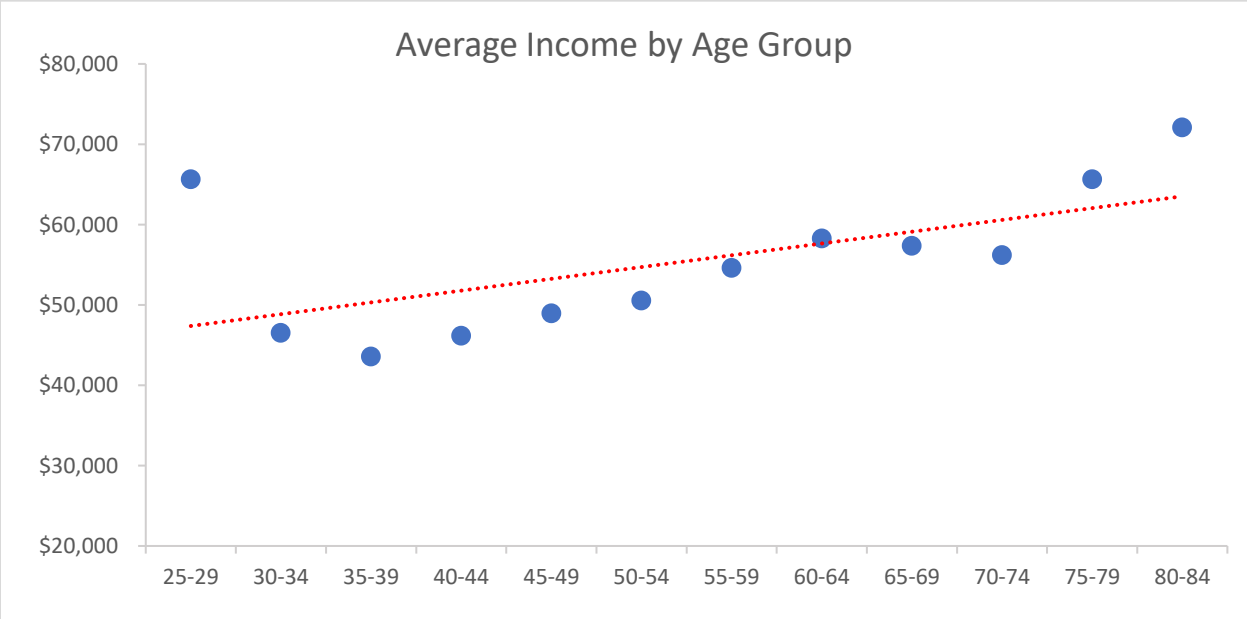


Age is evenly distributed within the income bracket $90,000 and $100,000. There is no correlation between income and age within this bracket.



ii. Income and Age Correlation

Customer age and average income are positively correlated. There is a notable outlier in the 25–29 year-old bracket.



MODULE 4

Table Creation

Two new tables were created in PgAdmin corresponding to the cleaned Marketing and Ad data files. The following syntax was used:

```
CREATE TABLE marketing_data(
    "ID" INT PRIMARY KEY,
    "Year_Birth" INT,
    "Age" INT,
    "Education" VARCHAR(50),
    "Marital_Status" VARCHAR(50)
    "Income" INT,
    "Kidhome" INT,
    "Teenhome" INT,
    "Dt_Customer" DATE,
    "Recency" INT,
    "AmtLiq" INT,
    "AmtVege" INT,
    "AmtNonVeg" INT,
    "AmtPes" INT,
    "AmtChocolates" INT,
    "AmtComm" INT,
    "Total_Sales" INT,
    "NumDeals" INT,
    "NumWebBuy" INT,
    "NumWalkinPur" INT,
    "NumVisits" INT,
    "Response" BOOL,
    "Complain" BOOL,
    "Country" VARCHAR(50),
    "Count_success" INT
);
```

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

Total Spend Per Country

The total spend per country was ascertained using the sum function and the output labelled as “Total_Spend”. The data was grouped by Country and ordered in descending order by “Total_Spend”.

The following syntax was used:

```
SELECT "Country", SUM("Total_Sales") AS "Total_Spend"
FROM public.marketing_data
GROUP BY "Country"
ORDER BY "Total_Spend" DESC;
```

The following output was produced:

	Country character varying (50)	Total_Spend bigint
1	Spain	568644
2	South Africa	169986
3	Canada	142927
4	Australia	65009
5	India	61956
6	Germany	58113
7	USA	53597
8	Montenegro	1258

Spain is 2Market's largest market in terms of total spend, by a considerable margin.

Total Spend per Product per Country

The total spend per product per country was ascertained using the sum function and the output labelled in a more obvious manner. The data was grouped by country and ordered by total spend. The following syntax was used:

```
SELECT "Country",
    SUM("AmtLiq") AS "Alcohol",
    SUM("AmtVege") AS "Vegetables",
    SUM("AmtNonVeg") AS "Meat",
    SUM("AmtPes") AS "Fish",
    SUM("AmtChocolates") AS "Chocolates",
    SUM("AmtComm") AS "Commodities",
    SUM("Total_Sales") AS "Total_spend"
FROM public.marketing_data
GROUP BY "Country"
ORDER BY "Total_spend" DESC;
```

The following output was produced:

	Country character varying (50)	Alcohol bigint	Vegetables bigint	Meat bigint	Fish bigint	Chocolates bigint	Commodities bigint	Total_spend bigint
1	Spain	289622	24374	154514	34481	26353	39300	568644
2	South Africa	85997	7251	45923	11092	7378	12345	169986
3	Canada	72205	6566	39267	8302	6571	10016	142927
4	Australia	31749	2859	17238	4353	3255	5555	65009
5	India	30152	2982	17927	3712	2670	4513	61956
6	Germany	29746	2485	16002	3694	2095	4091	58113
7	USA	25927	2697	14678	3707	2516	4072	53597
8	Montenegro	420	0	452	182	64	140	1258

Except for Montenegro (with only 1 customer in the dataset), total spend on alcohol exceeded all other products in all countries.

Meat was the most popular product in all countries (the USA equal with alcohol and commodities and Montenegro only having 1 record).

Most Popular Product based on Marital Status

The most popular products (those bought most often) in each Marital Status were ascertained. The data was filtered for 0 values, relabeled, grouped and ordered by the total count. The following syntax was used:

```
SELECT "Marital_Status",
    COUNT("AmtLiq") FILTER (WHERE "AmtLiq" != 0) AS "Alcohol",
    COUNT("AmtVege") FILTER (WHERE "AmtVege" != 0) AS "Vegetables",
    COUNT("AmtNonVeg") FILTER (WHERE "AmtNonVeg" != 0)AS "Meat",
    COUNT("AmtPes") FILTER (WHERE "AmtPes" != 0)AS "Fish",
    COUNT("AmtChocolates") FILTER (WHERE "AmtChocolates" != 0) AS "Chocolates",
    COUNT("AmtComm") FILTER (WHERE "AmtComm" != 0)AS "Commodities",
    COUNT("Total_Sales") FILTER (WHERE "Total_Sales" != 0) AS "Total_count"
FROM public.marketing_data
GROUP BY "Marital_Status"
ORDER BY "Total_count" DESC;
```

The following output was produced:

Marital_Status character varying (50)	Alcohol bigint	Vegetables bigint	Meat bigint	Fish bigint	Chocolates bigint	Commodities bigint	Total_count bigint
Married	716	589	721	592	585	702	721
Together	452	361	453	373	365	435	453
Single	405	349	408	334	340	397	408
Divorced	194	167	196	163	162	192	197
Widow	65	54	65	54	50	64	65

Meat is the most popular product in all marital statuses except “Widow” where it was joint top with alcohol.

Most Popular Products based on Kids / Teens at Home

Four alternatives were considered.

1. Both kids and teens at home - Alcohol and Meat were the joint most popular product.
2. Teens at home but no kids - Meat was the most popular product.
3. Kids at home but no teens - Alcohol and Meat were the joint most popular product.
4. No children at home - Meat was the most popular product.

The following syntax was used:

```
SELECT *
FROM (SELECT COUNT("AmtLiq") FILTER (WHERE "AmtLiq" != 0) AS "Alcohol",
    COUNT("AmtVege") FILTER (WHERE "AmtVege" != 0) AS "Vegetables",
    COUNT("AmtNonVeg") FILTER (WHERE "AmtNonVeg" != 0)AS "Meat",
    COUNT("AmtPes") FILTER (WHERE "AmtPes" != 0)AS "Fish",
    COUNT("AmtChocolates") FILTER (WHERE "AmtChocolates" != 0) AS "Chocolat
    COUNT("AmtComm") FILTER (WHERE "AmtComm" != 0)AS "Commodities"
FROM public.marketing_data
WHERE "Kidhome" != 0 AND "Teenhome" != 0)
UNION
SELECT *
FROM (SELECT COUNT("AmtLiq") FILTER (WHERE "AmtLiq" != 0) AS "Alcohol",
    COUNT("AmtVege") FILTER (WHERE "AmtVege" != 0) AS "Vegetables",
    COUNT("AmtNonVeg") FILTER (WHERE "AmtNonVeg" != 0)AS "Meat",
    COUNT("AmtPes") FILTER (WHERE "AmtPes" != 0)AS "Fish",
    COUNT("AmtChocolates") FILTER (WHERE "AmtChocolates" != 0) AS "Chocolat
    COUNT("AmtComm") FILTER (WHERE "AmtComm" != 0)AS "Commodities"
FROM public.marketing_data
WHERE "Kidhome" = 0 AND "Teenhome" != 0)
UNION
SELECT *
FROM (SELECT COUNT("AmtLiq") FILTER (WHERE "AmtLiq" != 0) AS "Alcohol",
    COUNT("AmtVege") FILTER (WHERE "AmtVege" != 0) AS "Vegetables",
    COUNT("AmtNonVeg") FILTER (WHERE "AmtNonVeg" != 0)AS "Meat",
    COUNT("AmtPes") FILTER (WHERE "AmtPes" != 0)AS "Fish",
    COUNT("AmtChocolates") FILTER (WHERE "AmtChocolates" != 0) AS "Chocolat
    COUNT("AmtComm") FILTER (WHERE "AmtComm" != 0)AS "Commodities"
FROM public.marketing_data
WHERE "Kidhome" != 0 AND "Teenhome" = 0)
UNION
SELECT *
FROM (SELECT COUNT("AmtLiq") FILTER (WHERE "AmtLiq" != 0) AS "Alcohol",
    COUNT("AmtVege") FILTER (WHERE "AmtVege" != 0) AS "Vegetables",
    COUNT("AmtNonVeg") FILTER (WHERE "AmtNonVeg" != 0)AS "Meat",
    COUNT("AmtPes") FILTER (WHERE "AmtPes" != 0)AS "Fish",
    COUNT("AmtChocolates") FILTER (WHERE "AmtChocolates" != 0) AS "Chocolat
    COUNT("AmtComm") FILTER (WHERE "AmtComm" != 0)AS "Commodities"
FROM public.marketing_data
WHERE "Kidhome" = 0 AND "Teenhome" = 0);
```

The following output was produced:

	Alcohol bigint	Vegetables bigint	Meat bigint	Fish bigint	Chocolates bigint	Commodities bigint
1	344	231	344	231	226	331
2	430	358	435	357	359	422
3	552	448	552	440	437	542
4	506	483	512	488	480	495

Module 5

Most Effective Method of Advertising by Social Media Lead Conversions

The marketing data file and ad data file were left joined. This enabled retention of all the data from the marketing file whilst adding matching records from the ad data.

Country

The following syntax was used to ascertain the total number of lead conversions by social media platform in each country:

```
SELECT m."Country",
    COUNT("Twitter_ad") FILTER (WHERE "Twitter_ad" = true) AS "Twitter",
    COUNT("Instagram_ad") FILTER (WHERE "Instagram_ad" = true) AS "Instagram",
    COUNT("Facebook_ad") FILTER (WHERE "Facebook_ad" = true) AS "Facebook"
FROM public.marketing_data m
LEFT JOIN public.ad_data a
USING ("ID")
GROUP BY "Country"
ORDER BY "Country" ASC;
```

The following output was produced:

	Country character varying (50)	Twitter bigint	Instagram bigint	Facebook bigint
1	Australia	4	7	5
2	Canada	23	19	14
3	Germany	9	5	7
4	India	9	5	4
5	Montenegro	0	0	0
6	South Africa	16	16	16
7	Spain	77	76	71
8	USA	4	3	6

Twitter was most effective in Canada, Germany, India and Spain. Facebook was most effective in the USA. Instagram was most effective in Australia. South Africans responded equally to all channels and Montenegro's sole customer did not provide a lead conversion.

Marital Status

The following syntax was used to ascertain the total number of lead conversions by social media platform in each Marital Status category:

```
SELECT m."Marital_Status",
    COUNT("Twitter_ad") FILTER (WHERE "Twitter_ad" = true) AS "Twitter",
    COUNT("Instagram_ad") FILTER (WHERE "Instagram_ad" = true) AS "Instagram",
    COUNT("Facebook_ad") FILTER (WHERE "Facebook_ad" = true) AS "Facebook"
FROM public.marketing_data m
LEFT JOIN public.ad_data a
USING ("ID")
GROUP BY "Marital_Status"
ORDER BY "Marital_Status" ASC;
```

The following output was produced:

	Marital_Status character varying (50)	Twitter bigint	Instagram bigint	Facebook bigint
1	Divorced	16	12	11
2	Married	55	53	55
3	Single	27	25	28
4	Together	34	34	24
5	Widow	10	7	5

Twitter was most effective in divorced and widowed customers. Married customers responded equally to Twitter and Facebook campaigns. Together customers responded equally to Twitter and Instagram campaigns. Facebook marginally beat Twitter in the single customer category.

The Effectiveness of Social Media Advertising with amount spent per Product per Country

Initially, the amount spent per product type per country was ascertained and then compared to the lead conversions of the various ad campaigns. The following syntax was used:

```
SELECT m."Country" AS country,
    SUM("AmtLiq") AS alcohol,
    SUM("AmtVege") AS veg,
    SUM("AmtNonVeg") AS meat,
    SUM("AmtPes") AS fish,
    SUM("AmtChocolates") AS chocolates,
    SUM("AmtComm") AS commodities,
    SUM("Total_Sales") AS total,
    COUNT("Twitter_ad") FILTER (WHERE "Twitter_ad" = true) AS "Twitter",
    COUNT("Instagram_ad") FILTER (WHERE "Instagram_ad" = true) AS "Instagram",
    COUNT("Facebook_ad") FILTER (WHERE "Facebook_ad" = true) AS "Facebook"
FROM public.marketing_data m
LEFT JOIN public.ad_data a
USING ("ID")
GROUP BY country
ORDER BY total DESC;
```

The following output was produced:

	country character varying (50)	alcohol bigint	veg bigint	meat bigint	fish bigint	chocolates bigint	commodities bigint	total bigint	Twitter bigint	Instagram bigint	Facebook bigint
1	Spain	289622	24374	154514	34481	26353	39300	568644	77	76	71
2	South Africa	85997	7251	45923	11092	7378	12345	169986	16	16	16
3	Canada	72205	6566	39267	8302	6571	10016	142927	23	19	14
4	Australia	31749	2859	17238	4353	3255	5555	65009	4	7	5
5	India	30152	2982	17927	3712	2670	4513	61956	9	5	4
6	Germany	29746	2485	16002	3694	2095	4091	58113	9	5	7
7	USA	25927	2697	14678	3707	2516	4072	53597	4	3	6
8	Montenegro	420	0	452	182	64	140	1258	0	0	0

It appears that lead conversion is positively correlated with amounts spent. However, this correlation might also be associated with larger customer numbers. I intend to further investigate this by charting the correlation and assessing the R squared values.