Background

2Market are in the early stages of planning a new marketing campaign. Through analysing a dataset from the previous campaign, we can recommend the most effective advertising channel to use. Furthermore, having a better understanding of customer demographics and best-selling products will help *2Market* to develop an advertising strategy to increase average product sales of target groups.

How will the effectiveness of advertising channel be measured?

What relationships can we find between demographics and products sales?

What products should be advertised?

Analytical Approach

Data cleaning processes began with changing column data types to correct format (e.g., Number, Date), henceforth:

1. Age value was created using calculation from Birth Year.
2. Marital status contained several values not considered useful, these were marked (#N/A) and consolidated using a lookup table e.g., “Alone” edited to be “Single”.
3. Abbreviated country names were changed to full names for ease of presenting.
4. New columns for Amount Total, Total Children and Total Ads created to allow for easier analysis.
5. Marking outliers (see Fig. 1) and duplicates (see Fig. 2) for filtering.

Excel provided for basic analysis of the cleaned dataset beginning with Total Customers (2005), Average Age (53.13), Average Income ($52,045) and Average Sale (607.62). The most common demographics are identified as:

- Education: Graduates (1012), PhD (424)
- Marital Status: Together & Married (1289)
- Country: Spain (997), South Africa (300)

Using PivotTables to manipulate the data, charts were created showing demographic trends. Fig. 3 illustrates the relationship between level of education, age and income, the trend is a higher level of education will correlate with increased age and income. Similarly, Fig. 4 shows there is a strong positive relationship between marital status, age and income. While there is no direct relationship between age and income (Fig. 1), another scatterplot (Fig. 5) shows a clear correlation between income and total sales per customer. Above average income demographic groups include:

- Widows ($56,957)
- USA ($53,525)
- PhD ($56,176)

This information will help to identify target groups for the upcoming advertising campaign.

SQL was used to combine the advertising data (contained in a separate Excel worksheet) with the sales data which allows us to easily examine any relationships between advertising channel, product sales and demographics. Fig. 7 shows the SQL statements. As a first step, a comparison between customer groups exposed to a single advertising campaign would be interesting, using customers with no adverts as the Control group. To achieve this, a new column String_ad was created to allow for easy filtering (Fig. 8). Preliminary results showed (Fig. 8 & Fig. 9):

1. The Control group had an income lower than the average.
2. The Instagram group had the top sales.
3. Social media advertising campaigns targeted high income customers.

The data (removed outliers, duplicates and errors) was exported as CSV file and imported into Tableau for further analysis and visualisation (Fig. 10).

Dashboard Design & Development

The dashboard design is based on solving the business problem and presenting this information to the audience with an engaging narrative. In this case, the audience is the Marketing Team, so a detailed knowledge of the business will be assumed.

Accessibility issues considered include:

- Good contrast between text and background
- Selection of colours with consideration to colour blind people(blues/orange/grey)
- Explanatory headers, dimensions, tooltips and captions on charts
- Easy navigation

At the top of each dashboard page a summary of key statistics was provided for easy reference, along with a brief explanation of the purpose of the page. Choice of visualisation was determined by the data:

- Histograms were chosen to represent the distribution of age and income
- A map shows the count of customers by country
- Scatterplot to show relationships between two numerical variables (Income vs. Average Sale)
- Bar charts were chosen to represent categorical data (Marital Status, Education, Product) by a numerical value (Average Sales).

Interactivity was included on the dashboard to add more value to each section, allowing the user to apply filters and show different sets of data by product and/or demographic.

While the dashboard is specifically designed for the provided dataset, it can be updated to inform decision-making from a live data feed.

Patterns, Trends & Insights

We want to know which advertising channel is most effective. The measure of effectiveness will be the average sales. To measure effectiveness of different advertising channels we can compare average sales between groups:

Control Group (No Adverts)	Test Groups (Brochure, Email, Twitter, Facebook, Instagram)

The percentage difference between average sales of the control group and the test groups will be compared. Fig. 9 illustrates the average income between the groups. To ensure a fair comparison all customers with income lower than $60,000 were filtered out, the results below are shown in Fig. 13.

1. Instagram was the most effective advertising channel, followed by Facebook.
2. Email was surprisingly effective at 21.8% increase compared to the control group.
3. Twitter did not produce effective results.

Top selling products are alcohol, meat and commodities (see Fig. 14). Meat and fish are of particular interest, as these items are perishable and have a high spoil rate, and hence a higher profit.

Analysis (See Fig. 15, Fig. 16, Fig. 17) shows:

1. Combining Instagram with any other channel led to the most sales.
2. Impressive increases in our alcohol sales across all social media channels.
3. Sales in alcohol increased by 98.90% compared to the control group when combining Instagram and Twitter.
4. Sales of meat and fish were best increased through Instagram and Facebook.

Product sales by demographic was analysed for customers with income above $50,000:

Education (see Fig. 18)

- Masters group 9.95% higher average alcohol sales than Graduation
- PHD group 30.35% higher average alcohol sales than Graduation

Marital Status (see Fig. 19)

- Average sales of fish and meat highest in the Singles group
- High earning Singles may have more disposable income

Country (see Fig. 20)

- Alcohol sales strongest in Germany & Spain
- Sales of meat in India most impressive.

Key recommendations:

1. Collect all social media and email account data from customers.
2. Combine social media channels rather than using a single channel.
3. Advertising should be targeted at above average income groups.
4. Meat and fish sales increased using Instagram. Instagram can best be combined with Facebook.
5. Use Instagram for alcohol promotions (must verify age > 18). Combine with Twitter for good results.
6. Email is a cost-effective channel.

Appendix

Fig. 1: Outliers

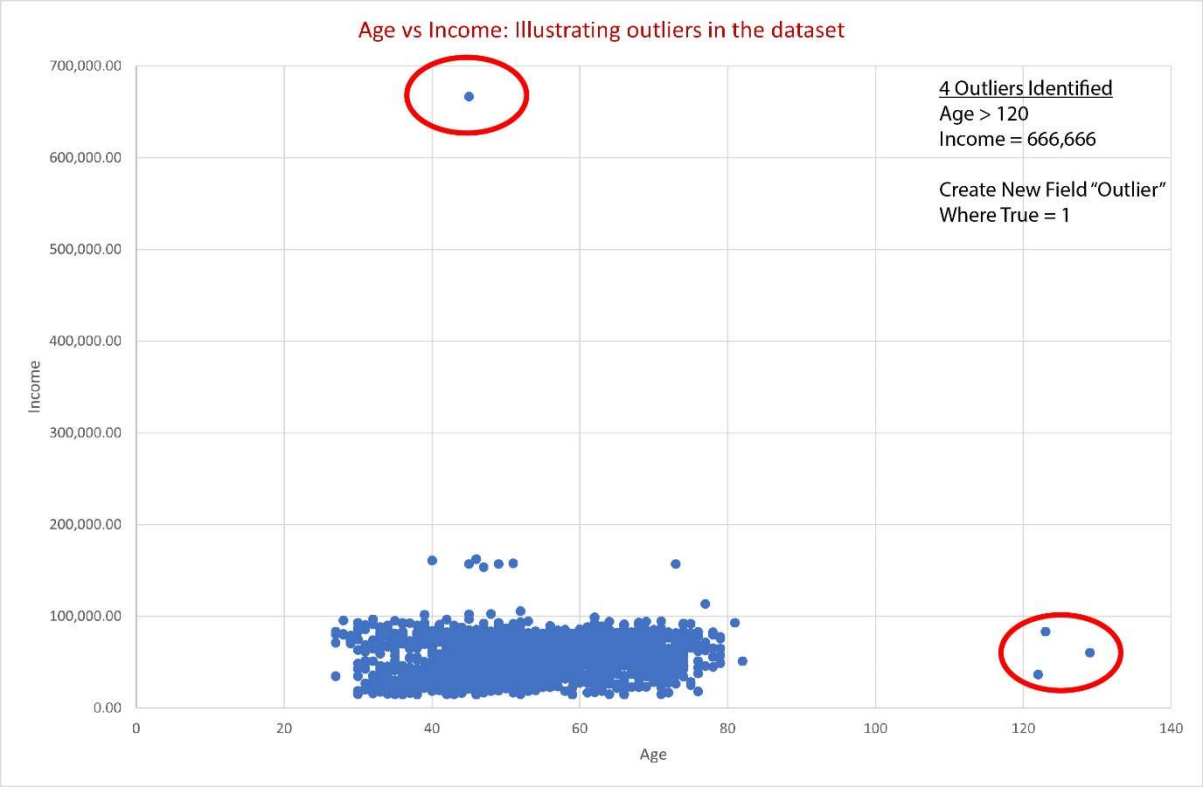

Fig. 2: Duplicates

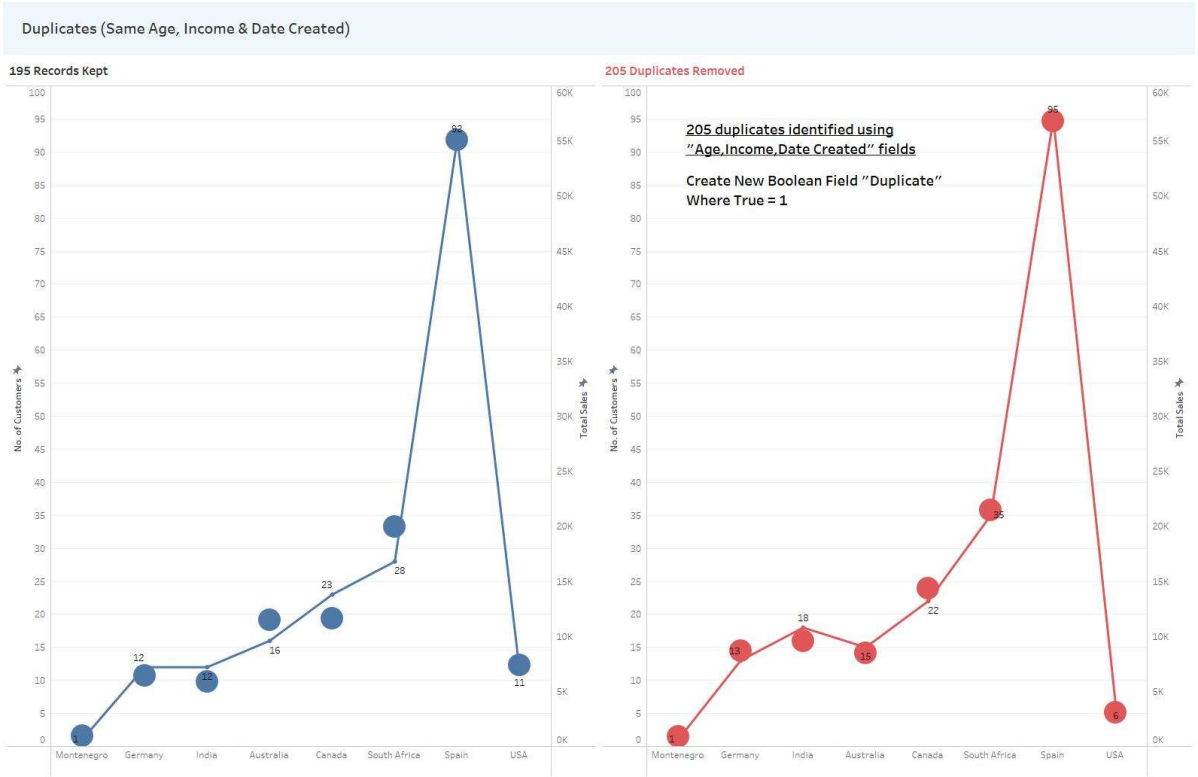

Fig. 3: Education Trends

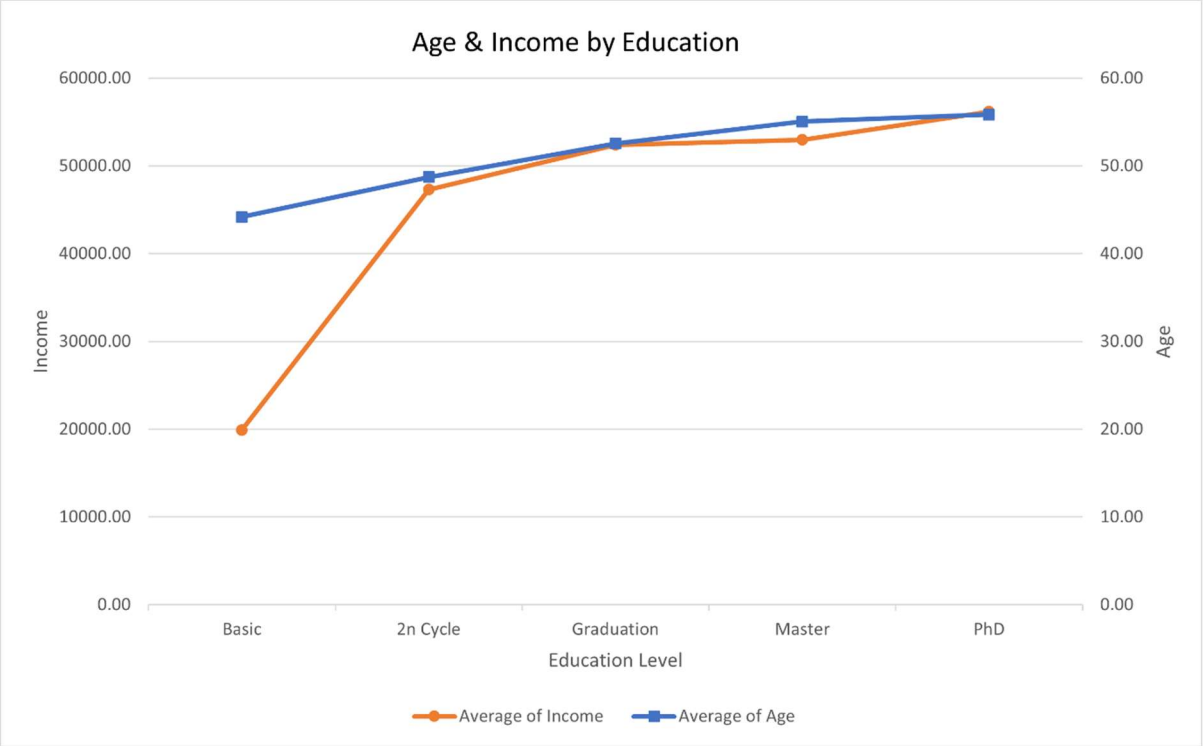

Fig. 4: Marital Status Trends

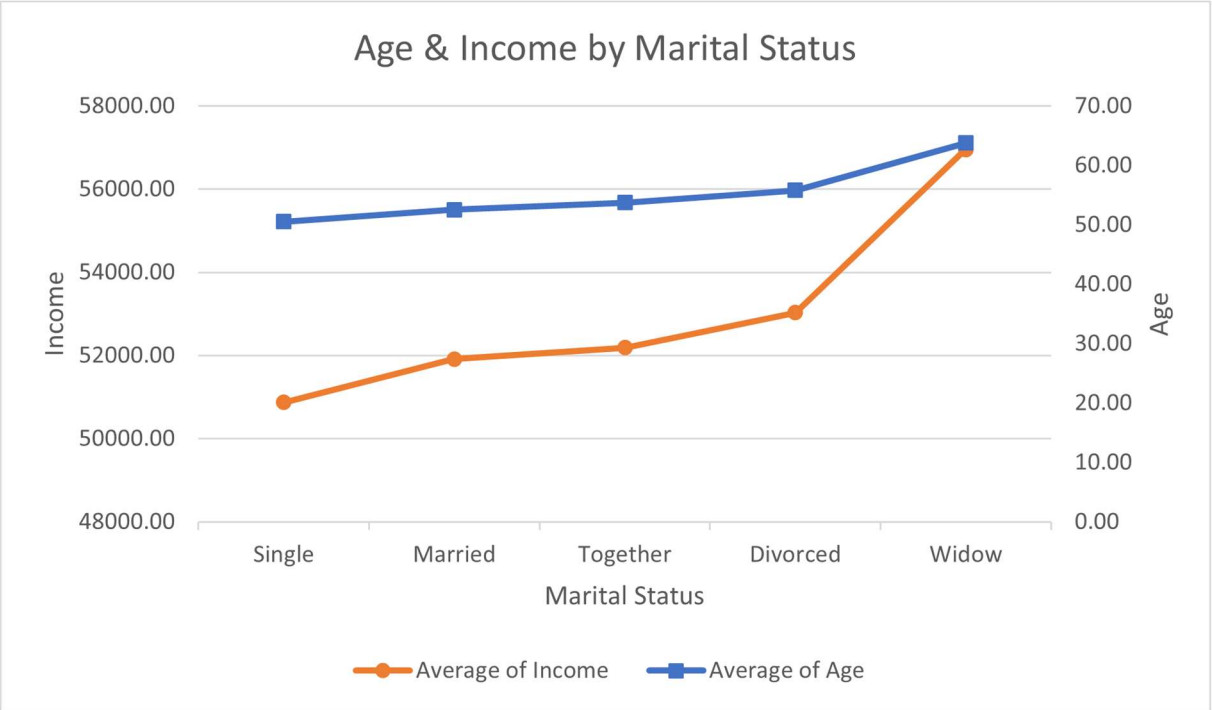

Fig. 5: Income vs Total Sales

Fig. 6: Age vs Income (between $90,000 – $100,000)

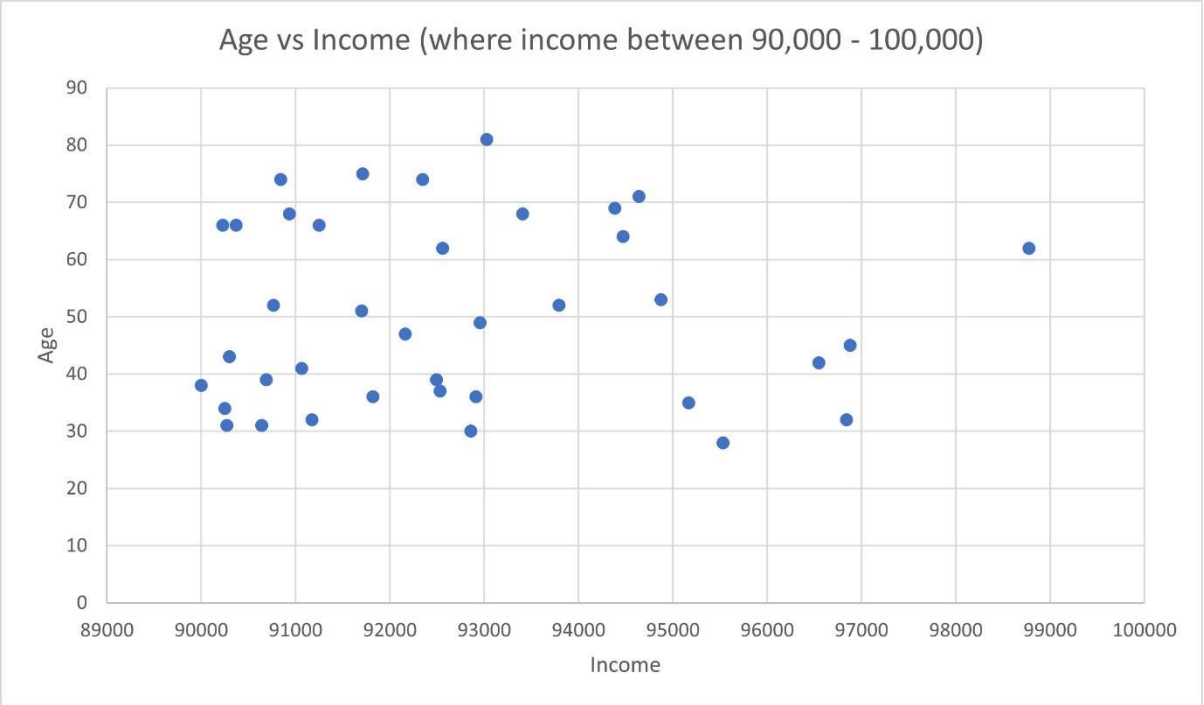

Fig. 7: Create 2Market Database, Create Tables & Join (SQL)

```sql
-- Create Database
CREATE DATABASE "2Market";

-- Create Table script for 2Market Database
-- Create Table "Marketing_Data"
CREATE TABLE "Marketing_Data"(
    "ID" INTEGER PRIMARY KEY,
    "Year_Birth" NUMERIC(4,0),
    "Age" NUMERIC(3,0),
    "Education" VARCHAR(20),
    "Marital_Status" VARCHAR(20),
    "Income" NUMERIC(6,0),
    "Duplicate" NUMERIC(1,0),
    "Duplicate Combined" NUMERIC(1,0),
    "TotalChildren" NUMERIC(1,0),
    "Kidhome" NUMERIC(1,0),
    "Teenhome" NUMERIC(1,0),
    "Dt_Customer" DATE,
    "Recency" NUMERIC(2,0),
    "AmtTotal" NUMERIC(4,0),
    "AmtLiq" NUMERIC(4,0),
    "AmtVege" NUMERIC(4,0),
    "AmtNonVeg" NUMERIC(4,0),
    "AmtPes" NUMERIC(4,0),
    "AmtChocolates" NUMERIC(4,0),
    "AmtComm" NUMERIC(4,0),
    "NumDeals" NUMERIC(2,0),
    "CountPurchase" NUMERIC(2,0),
    "NumWebBuy" NUMERIC(2,0),
    "NumWalkinPur" NUMERIC(2,0),
    "NumVisits" NUMERIC(2,0),
    "Response" NUMERIC(1,0),
    "Complain" NUMERIC(1,0),
    "Outlier" NUMERIC(1,0),
    "Country" VARCHAR(20),
    "Count_success" NUMERIC(1,0)
);
-- Create Table "Ad_Data"
CREATE TABLE "Ad_Data"(
    "ID" INTEGER PRIMARY KEY,
    "Bulkmail_ad" NUMERIC(1,0),
    "Twitter_ad" NUMERIC(1,0),
    "Instagram_ad" NUMERIC(1,0),
    "Facebook_ad" NUMERIC(1,0),
    "Brochure_ad" NUMERIC(1,0),
    "Total_ad" NUMERIC(1,0)
);

/*
-- Table Join statements for "Marketing_Data" & "Ad_Data"
*/

-- select all
SELECT *
FROM public."Marketing_Data"
INNER JOIN public."Ad_Data" USING ("ID")
ORDER BY "ID";
```

Fig. 8: Create New Column “String_ad” (SQL)

```sql
/*
-- Create new column in "Ad_Data" for text listing all adverts shown to customer
*/

-- add column "String_ad" to hold all advertising campaigns for each customer
ALTER TABLE public."Ad_Data"
ADD COLUMN "String_ad" VARCHAR(100);

-- write script to add "Email" to customer data for "Bulkmail_ad = 1"
UPDATE public."Ad_Data"
SET "String_ad" = 'Email'
WHERE "Bulkmail_ad" = 1
RETURNING *;

-- write script to add "Twitter" to customer data for "Twitter_ad = 1"
UPDATE public."Ad_Data"
SET "String_ad" = CONCAT("String_ad",',Twitter')
WHERE "Twitter_ad" = 1
RETURNING *;

-- write script to add "Instagram" to customer data for "Instagram_ad = 1"
UPDATE public."Ad_Data"
SET "String_ad" = CONCAT("String_ad",',Instagram')
WHERE "Instagram_ad" = 1
RETURNING *;

-- write script to add "Facebook" to customer data for "Facebook_ad = 1"
UPDATE public."Ad_Data"
SET "String_ad" = CONCAT("String_ad",',Facebook')
WHERE "Facebook_ad" = 1
RETURNING *;

-- write script to add "Brochure" to customer data for "Brochure_ad = 1"
UPDATE public."Ad_Data"
SET "String_ad" = CONCAT("String_ad",',Brochure')
WHERE "Brochure_ad" = 1
RETURNING *;

-- Remove comma from 1st character in "String_ad"
-- find
SELECT "String_ad" = LTRIM("String_ad",',')
FROM public."Ad_Data"
WHERE LEFT("String_ad",1) = ',';
-- remove
UPDATE public."Ad_Data"
SET "String_ad" = LTRIM("String_ad",',')
WHERE LEFT("String_ad",1) = ',';

```

Fig. 9: Query Social Media & Email vs Control Group (No Adverts) for All Customers (SQL)

```sql
-- select all adverising campaigns with single exposure
-- control group with no advert
-- and those exposed to all 3 social media platforms (Facebook,Instagram,Twitter)
-- Filter out "Duplicates", "Outliers" & "Marital_Status = N/A
SELECT  "String_ad" AS "Ad_Campaigns",
        COUNT("ID") AS "TotalCustomers",
        ROUND(AVG("Age"),2) AS "AvgAge",
        ROUND(AVG("Income"),2) AS "AvgIncome",
        ROUND(AVG("AmtTotal"),2) AS "AvgSpend",
        ROUND(AVG("AmtLiq"),2) AS "AvgAlcohol",
        ROUND(AVG("AmtVege"),2) AS "AvgVegetables",
        ROUND(AVG("AmtNonVeg"),2) AS "AvgMeat",
        ROUND(AVG("AmtPes"),2) AS "AvgFish",
        ROUND(AVG("AmtChocolates"),2) AS "AvgChocolates",
        ROUND(AVG("AmtComm"),2) AS "AvgCommodities"
FROM public."Marketing_Data"
INNER JOIN public."Ad_Data" USING ("ID")
WHERE   "Duplicate" = '0'
        AND "Outlier" = '0'
        AND "Marital_Status" NOT LIKE ('%N/A')
        AND ("String_ad" IN ('Twitter,Instagram,Facebook') OR "Total_ad" < '2')
GROUP BY "Ad_Campaigns"
ORDER BY "AvgSpend" DESC;
```

Data Output | Messages | Notifications

	Ad_Campaigns character varying (100)	TotalCustomers bigint	AvgAge numeric	AvgIncome numeric	AvgSpend numeric	AvgAlcohol numeric	AvgVegetables numeric	AvgMeat numeric	AvgFish numeric	AvgChocolates numeric	AvgCommodities numeric
1	Twitter,Instagram,Facebook	21	49.76	86753.48	1701.81	873.90	64.05	542.48	78.62	74.43	68.33
2	Instagram	50	53.94	81857.14	1614.76	789.12	60.04	539.82	83.42	70.44	71.92
3	Facebook	48	56.98	74405.42	1352.42	644.85	48.48	427.19	103.02	57.58	71.29
4	Twitter	78	57.58	59363.26	795.17	564.85	19.41	132.35	25.82	17.63	35.12
5	Brochure	1	61.00	57091.00	577.00	464.00	5.00	64.00	7.00	0.00	37.00
6	Email	117	50.90	43569.56	508.41	281.50	19.38	108.21	24.17	15.56	59.60
7	[null]	1588	53.03	48604.24	481.51	224.88	23.73	136.99	33.27	23.99	38.64

Fig. 10: Query Social Media & Email vs Control Group (No Adverts) for Customers Income GT 50000 (SQL)

```sql
-- filter by age and income (age > 50, income > 50000)
-- to be around average of dataset
-- this way we can compare dataset more accurately
-- filter out "Duplicates", "Outliers" & "Marital_Status = N/A
SELECT  "String_ad" AS "Ad_Campaigns",
        COUNT("ID") AS "TotalCustomers",
        ROUND(AVG("Age"),2) AS "AvgAge",
        ROUND(AVG("Income"),2) AS "AvgIncome",
        ROUND(AVG("AmtTotal"),2) AS "AvgSpend",
        ROUND(AVG("AmtLiq"),2) AS "AvgAlcohol",
        ROUND(AVG("AmtVege"),2) AS "AvgVegetables",
        ROUND(AVG("AmtNonVeg"),2) AS "AvgMeat",
        ROUND(AVG("AmtPes"),2) AS "AvgFish",
        ROUND(AVG("AmtChocolates"),2) AS "AvgChocolates",
        ROUND(AVG("AmtComm"),2) AS "AvgCommodities"
FROM public."Marketing_Data"
INNER JOIN public."Ad_Data" USING ("ID")
WHERE   "Duplicate" = '0'
        AND "Outlier" = '0'
        AND "Marital_Status" NOT LIKE ('%N/A')
        AND "Age" > '50'
        AND "Income" > '60000'
        AND ("String_ad" IN ('Twitter,Instagram,Facebook') OR "Total_ad" < '2')
GROUP BY "Ad_Campaigns"
ORDER BY "AvgSpend" DESC;
```

Data Output | Messages | Notifications

	Ad_Campaigns character varying (100)	TotalCustomers bigint	AvgAge numeric	AvgIncome numeric	AvgSpend numeric	AvgAlcohol numeric	AvgVegetables numeric	AvgMeat numeric	AvgFish numeric	AvgChocolates numeric	AvgCommodities numeric
1	Twitter,Instagram,Facebook	8	63.50	83021.75	1688.88	849.88	72.50	577.75	76.25	58.88	53.63
2	Instagram	29	63.48	81580.45	1614.34	806.07	60.90	515.86	90.03	74.66	66.83
3	Facebook	32	64.13	75367.28	1358.13	677.31	44.56	416.41	96.31	53.91	69.63
4	Email	15	63.40	71002.53	1146.80	604.33	51.27	273.67	58.87	40.87	117.80
5	Twitter	27	61.22	68438.44	1122.63	754.15	31.52	220.48	43.70	30.89	41.89
6	[null]	322	62.34	71564.48	1062.24	502.61	54.28	312.69	73.15	53.38	66.15

Fig. 11: Create New Table with Required Columns (SQL)

```sql
/*
-- Create new table from query using only fields I require for analysis (final version of dataset)
*/
CREATE TABLE final_data AS
SELECT  md."ID",
        md."Age",
        md."Education",
        md."Marital_Status",
        md."Country",
        md."Income",
        md."TotalChildren",
        md."Dt_Customer",
        md."AmtTotal",
        md."AmtLiq",
        md."AmtVege",
        md."AmtNonVeg",
        md."AmtPes",
        md."AmtChocolates",
        md."AmtComm",
        md."NumDeals",
        md."CountPurchase",
        md."NumWebBuy",
        md."NumWalkinPur",
        md."NumVisits",
        md."Count_success",
        ad."Bulkmail_ad",
        ad."Twitter_ad",
        ad."Instagram_ad",
        ad."Facebook_ad",
        ad."Brochure_ad",
        ad."Total_ad",
        ad."String_ad",
        md."Duplicate",
        md."Outlier"
FROM public."Marketing_Data" md
INNER JOIN public."Ad_Data" ad USING ("ID")
ORDER BY "ID";
```

Fig. 12: Remove Duplicates, Outliers & Marital Status=#N/A and Export CSV (SQL)

```sql
-- search & remove duplicates, outliers and marital status = n/a from data
SELECT *
FROM public."final_data"
WHERE "Duplicate" = '1';

DELETE FROM public."final_data"
WHERE "Duplicate" = '1';

SELECT *
FROM public."final_data"
WHERE "Outlier" = '1';

DELETE FROM public."final_data"
WHERE "Outlier" = '1';

SELECT *
FROM public."final_data"
WHERE "Marital_Status" LIKE ('%N/A');

DELETE FROM public."final_data"
WHERE "Marital_Status" LIKE ('%N/A');

-- remove columns "Duplicate" and "Outlier"
ALTER TABLE public."final_data"
DROP COLUMN "Duplicate",
DROP COLUMN "Outlier";

-- select all records from new clean table
SELECT *
FROM public."final_data";

-- update no adverts value text
SELECT *
FROM public."final_data"
WHERE "String_ad" IS NULL;

UPDATE public."final_data"
SET "String_ad" = 'No Adverts'
WHERE "String_ad" IS NULL;

-- export to CSV
COPY public."final_data" TO 'D:\Luke\Desktop\LSE\Assignment 1\LSE_DA101_Assignment_data\marketing_data_final.csv' DELIMITER ',' CSV HEADER;
```

Data Output | Messages | Notifications

	ID integer	Age numeric (3)	Education character varying (20)	Marital_Status character varying (20)	Country character varying (20)	Income numeric (6)	TotalChildren numeric (1)	Dt_Customer date	AmtTotal numeric (4)	AmtLiq numeric (4)	AmtVege numeric (4)
1	1	61	Graduation	Single	Canada	57091	0	2014-06-15	577	464	5
2	48	58	Graduation	Together	Spain	55761	1	2014-04-24	184	136	1
3	125	64	2n Cycle	Together	India	53083	2	2013-05-15	271	215	7
4	146	62	PhD	Single	Spain	76045	0	2013-11-15	1323	760	53
5	158	77	PhD	Together	Spain	71604	0	2013-11-17	1196	345	53
6	175	36	Graduation	Married	Spain	71952	1	2013-01-10	1443	656	80
7	193	26	Basic	Married	Spain	14421	0	2014-02-17	16	0	0
8	195	50	Graduation	Single	Spain	38808	1	2012-08-26	246	125	17

Total rows: 1000 of 2005 | Query complete 00:00:00.151

Fig. 13: Control group vs Test groups Table (Income Greater Than $60,000)

Average Sales by Marketing Campaign

89.20% of customers in social media groups above average income ($52,045)

	No. of Customers	Avg. Age	Avg. Income	Alcohol	Chocolates	Commodities	Fish	Meat	Vegetables	% Total
No Adverts	493	56	72,531	486	55	66	76	334	56	0.00%
Email	22	57	70,521	694	43	114	65	329	62	21.80%
Twitter	36	57	68,095	736	32	46	45	220	33	3.63%
Facebook	46	57	75,232	661	60	72	107	442	50	29.68%
Instagram	49	54	82,331	786	71	73	85	547	61	51.18%

Fig. 14: Top Selling Products

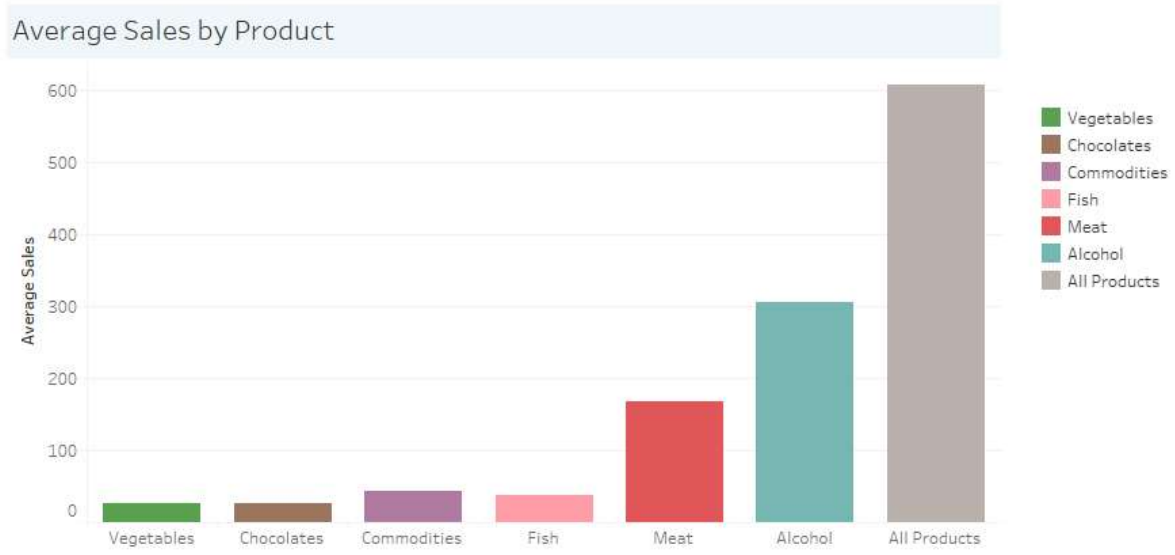

Fig. 15: Comparison of Avg. Sales by Control group and Social Media groups

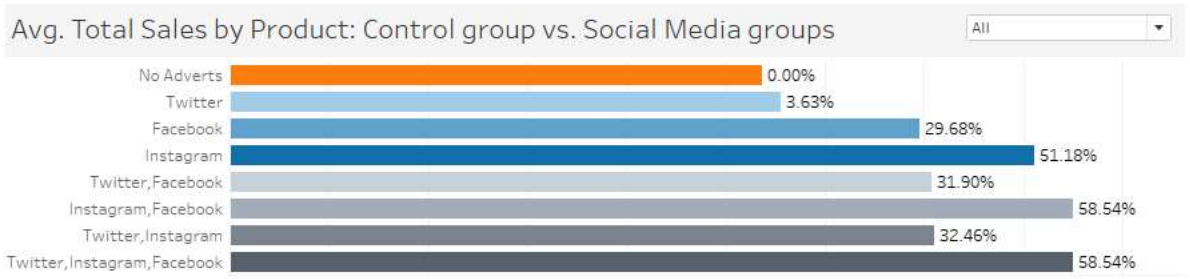

Fig. 16: Comparison of Avg. Sales of Alcohol by Control group and Social Media groups

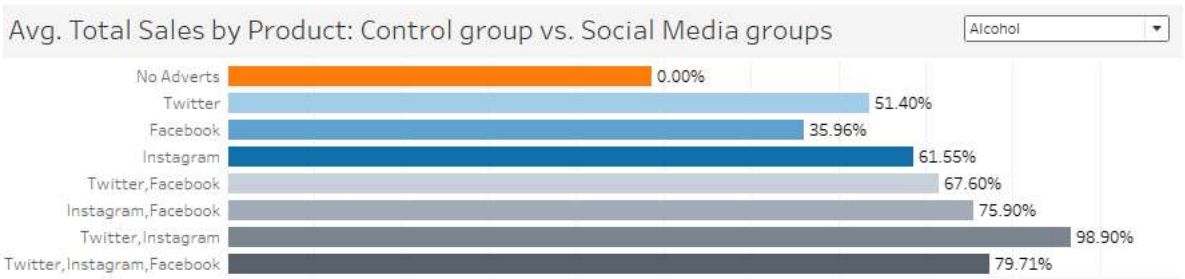

Fig. 17: Comparison of Avg. Sales of Meat by Control group and Social Media groups

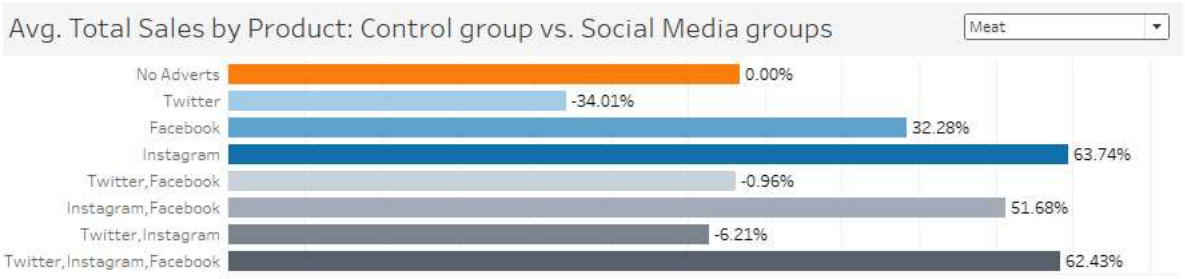

Fig. 18: Average Sales of Alcohol by Education (Income Greater Than $50,000)

Average Sales of Alcohol by Education (income greater than $50,000)

Education	No. of Customers	Avg. Age	Avg. Income	Alcohol
2n Cycle	53	53	66,293	327
Graduation	370	55	66,578	388
Master	120	58	66,853	427
PhD	187	58	67,462	506

% of Difference from Graduation group

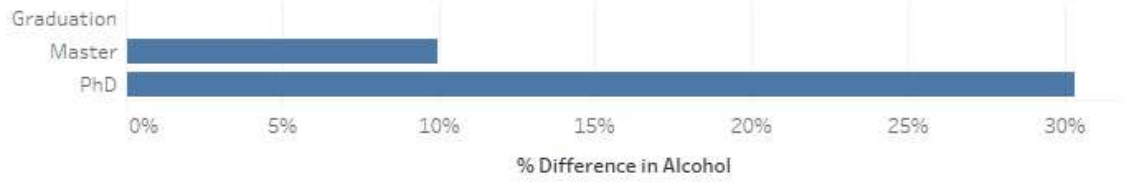

Fig. 19: Sales of Fish & Meat by Marital Status (Income Greater Than $50,000)

Product Sales by Marital Status (Income Greater Than $50,000)

Marital ..	No. of Custome..	Avg. Age	Avg. Income	% Fish	% Meat
Single	142	54	68,026	0.00%	0.00%
Together	187	56	67,470	-15.41%	-16.29%
Married	289	56	65,892	-26.46%	-22.17%
Divorced	82	56	67,453	-20.61%	-19.92%
Widow	30	65	64,483	-6.54%	-24.54%

Average Sales of Meat & Fish by Marital Status (Income Greater Than $50,000)

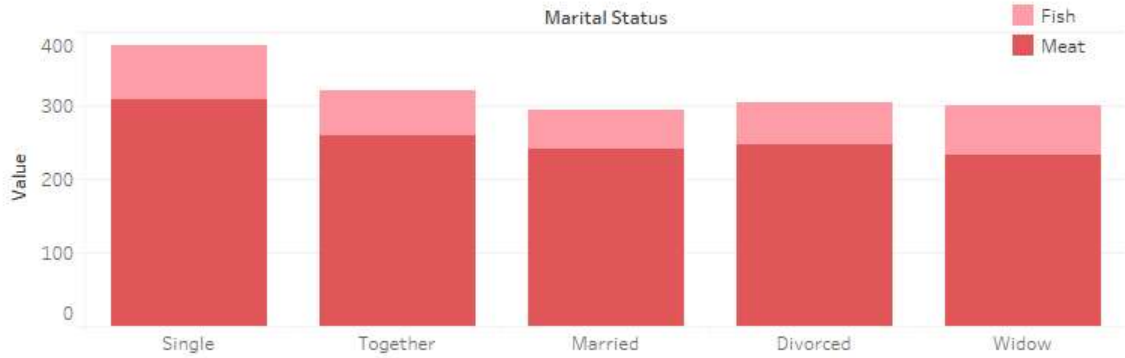

Fig. 20: Average Sales of Meat & Alcohol by Country (Income Greater Than $50,000)

Sales of Meat & Alcohol by Country (Income Greater Than $50,000)

Country	No. of Customers	Avg. Age	Avg. Income	Meat	Alcohol
India	42	50	67,447	323	391
Spain	340	57	66,765	250	430
Australia	54	57	65,281	224	422
Canada	91	56	68,403	275	420
Germany	36	54	67,368	288	433
South Africa	119	56	66,120	255	407
USA	46	56	67,084	267	383

% of Sales of Alcohol & Meat Comparative to India (Income Greater Than $50,000)