

LSE_DA101_Data Analytics for Business: Final Assignment

Market Customer Analysis

Ryan Lin

Word Count: 1132

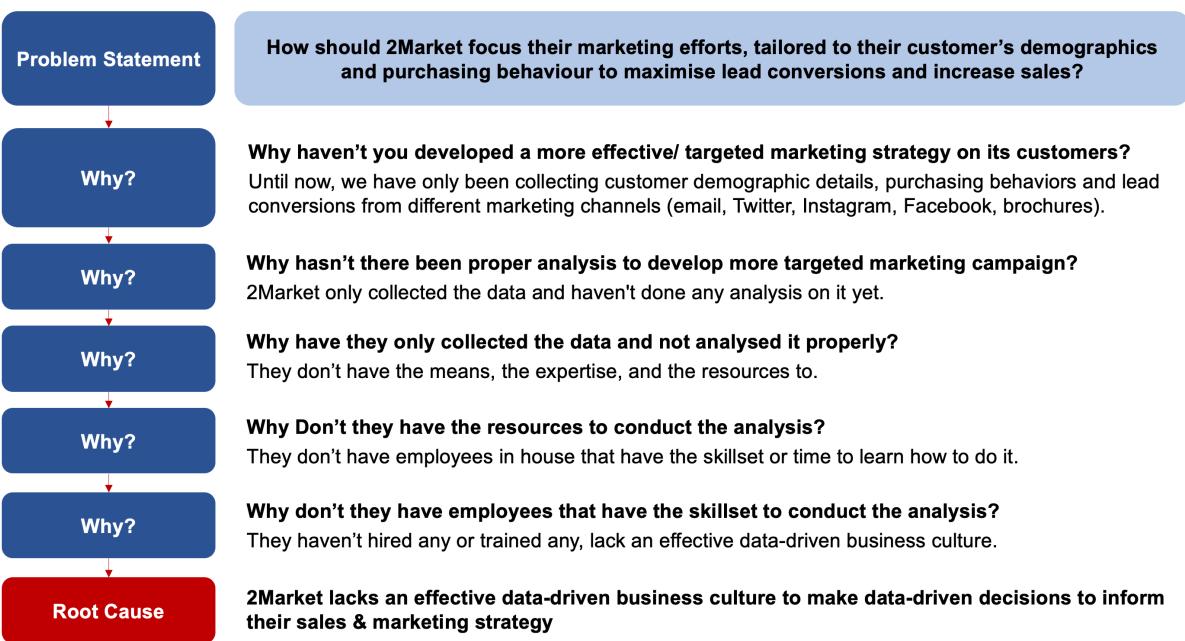
Introduction & Background

This report and the resulting findings are derived from two primary data sets provided by 2Market:

Dataset	Dataset Contents
Table A <code>marketing_data.csv</code>	<p>Customer Information/ Demographics</p> <ul style="list-style-type: none"> • Unique Customer ID, Age (Birth Year), Education, Marital Status, Income, Children at Home, Date of Registration, Country of Registration <p>Sales Data</p> <ul style="list-style-type: none"> • Amount (\$USD) Spent on Alcohol, Meat, Fish, Vegetables, Chocolates and Commodities <p>Marketing Data</p> <ul style="list-style-type: none"> • No. of Purchases Made In-store, Online, With a Discount • No. of Website Visits, No. of Successful Lead Conversions • Whether Customer Complained in the Last 2 years (Boolean), Whether Customer Accepted Last Campaign Offer (Boolean)
Table B <code>ad_data.csv</code>	<p>Customer Information</p> <ul style="list-style-type: none"> • Unique Customer ID <p>Successful Lead Conversion (Boolean)</p> <ul style="list-style-type: none"> • Bulk mails, Twitter Ad, Instagram Ad, Facebook Ad, Brochure Ad

Note: See meta data file titled `metadata_2Market.txt` for more information

Utilising the Five Why's structured thinking framework, the business problem and example root cause are outlined for this project below.



Note: As discourse was not possible with stakeholders, the most logical assumptions were made to answer each question.

One of the main limitations of this framework is that it's a linear method of communication for what is usually a non-linear event. There is also the tendency to isolate only a single root cause when there could be multiple. Nonetheless, by better understanding the business problem and their customers, 2Market will benefit in

delivering more personalised and targeted advertisements to increase customer spending, resulting in higher sales.

A variety of questions also need to be asked to better understand the project and the data provided. Additional areas to explore outside the scope of this project are also listed.

Questions Related to This Project

- Who are the key decision makers and the main stakeholders involved?
- Who will the findings be presented to?
- What sales and marketing KPIs are currently used?
- What does success look like?
- Who will ultimately be responsible for implementing the new marketing strategy based on the findings?

Questions Related to The Data

- How was the data collected and collated? How reliable is it?
- What is the timeframe of when this data was collected?
- Are there specific products that 2Market are more interested in selling or advertising? Perhaps products with higher margins?
- Why are there no decimal places for the amount spent for each product type? Was the data rounded?
- Why are there gaps in the sequential order of the unique customer ID's?

Additional Areas to Explore

- If the effectiveness of advertising channels differ with Age, Marital Status, Education or Income of 2Market customers.

Analytical Approach

The data was cleaned and transformed to ensure the highest data quality before analysis (see Appendix 1 for full approach). Additional columns were added to aid analysis, column headings modified to be more understandable, the data was checked for missing values or empty cells, any incorrect datatypes were corrected. Changes were also made to consolidate words with the same meaning or alternate expressions.

Without being able to clarify with the stakeholders, incorrect entries for certain fields were not deleted as the rest of the customers demographic and purchasing information could still be analysed. Similarly, suspected duplicates were highlighted and noted down and were not removed, since they are impossible to confirm without speaking with the stakeholders.

Example where only Customer ID is different: 47 suspected duplicates

ID	Year_Birth	Age	Education	Marital_Status	Income_(USD)	Kidhome	Teenhome	Registration_Dt	Recency	AmLit	AmVeg	AmNonVeg	AmPois	AmChocolate	AmComm	Total_Sales	NumDeals	NumWebBuy	NumWalkInPur	NumVists	Response	Compli	Country	Count_Success
5430	1957	66	Bachelor	Partnership	54,450	1	1	2012-09-14	0	454	0	171	8	19	32	684	12	9	8	8	0	0 SP	0	
8432	1957	66	Bachelor	Partnership	54,450	1	1	2012-09-14	0	454	0	171	8	19	32	684	12	9	8	8	0	0 SP	0	

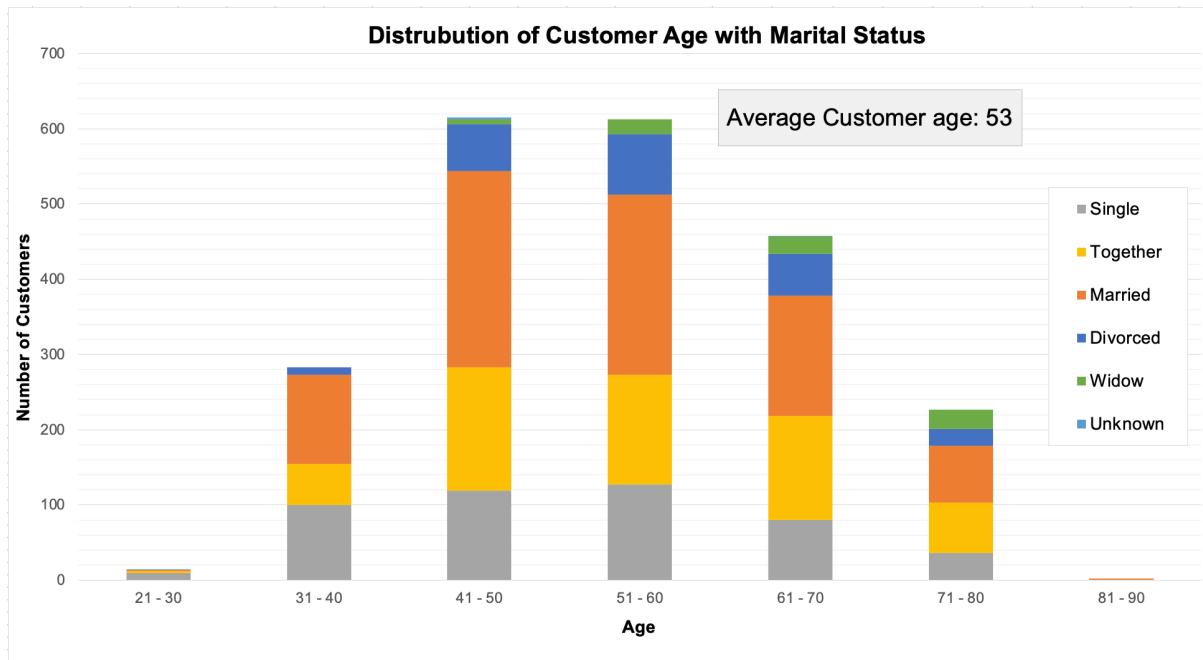
Examples where Customer ID and Registration Country is different: 180 suspected duplicates

ID	Year_Birth	Age	Education	Marital_Status	Income_(USD)	Kidhome	Teenhome	Registration_Dt	Recency	AmLit	AmVeg	AmNonVeg	AmPois	AmChocolate	AmComm	Total_Sales	NumDeals	NumWebBuy	NumWalkInPur	NumVists	Response	Compli	Country	Count_Success
4047	1955	68	PHD	Married	65,324	0	1	2014-01-11	0	384	0	102	21	32	5	544	3	6	9	4	0	0 US	0	
9477	1955	68	PHD	Married	65,324	0	1	2014-01-11	0	384	0	102	21	32	5	544	3	6	9	4	0	0 IND	0	

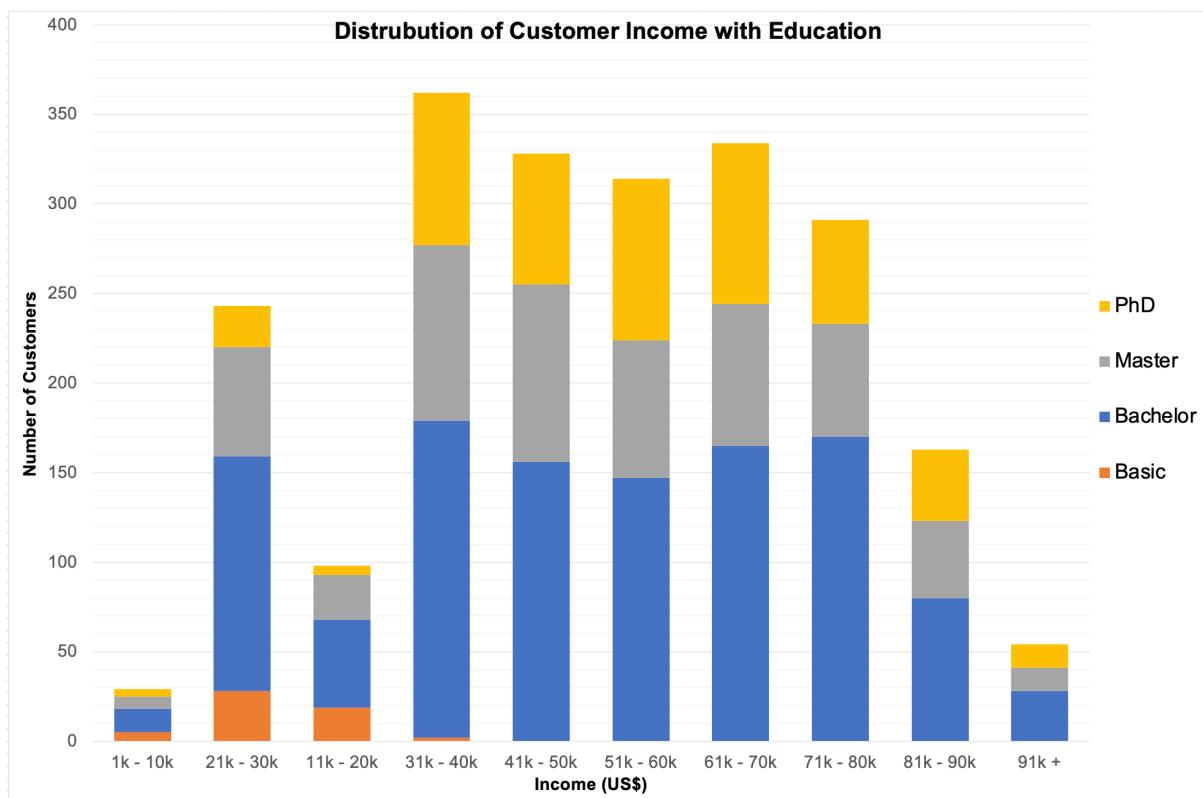
ID	Year_Birth	Age	Education	Marital_Status	Income_(USD)	Kidhome	Teenhome	Registration_Dt	Recency	AmLit	AmVeg	AmNonVeg	AmPois	AmChocolate	AmComm	Total_Sales	NumDeals	NumWebBuy	NumWalkInPur	NumVists	Response	Compli	Country	Count_Success
492	1974	49	PHD	YOLO	\$48,432.00	0	1	2012-10-18	3	322	3	50	4	3	42	424	5	7	6	8	0	0 CA	0	
11133	1974	49	PHD	YOLO	\$48,432.00	0	1	2012-10-18	3	322	3	50	4	3	42	424	5	7	6	8	1	0 IND	0	

The examples above have the same Age (Birth Year), Education, Marital Status, Income, Registration Date and more. On top of that, considering the incorrect entry "YOLO" for Marital Status in the last example, it is unlikely two separate customers with the same demographics and the exact same purchase amounts incorrectly entered "YOLO".

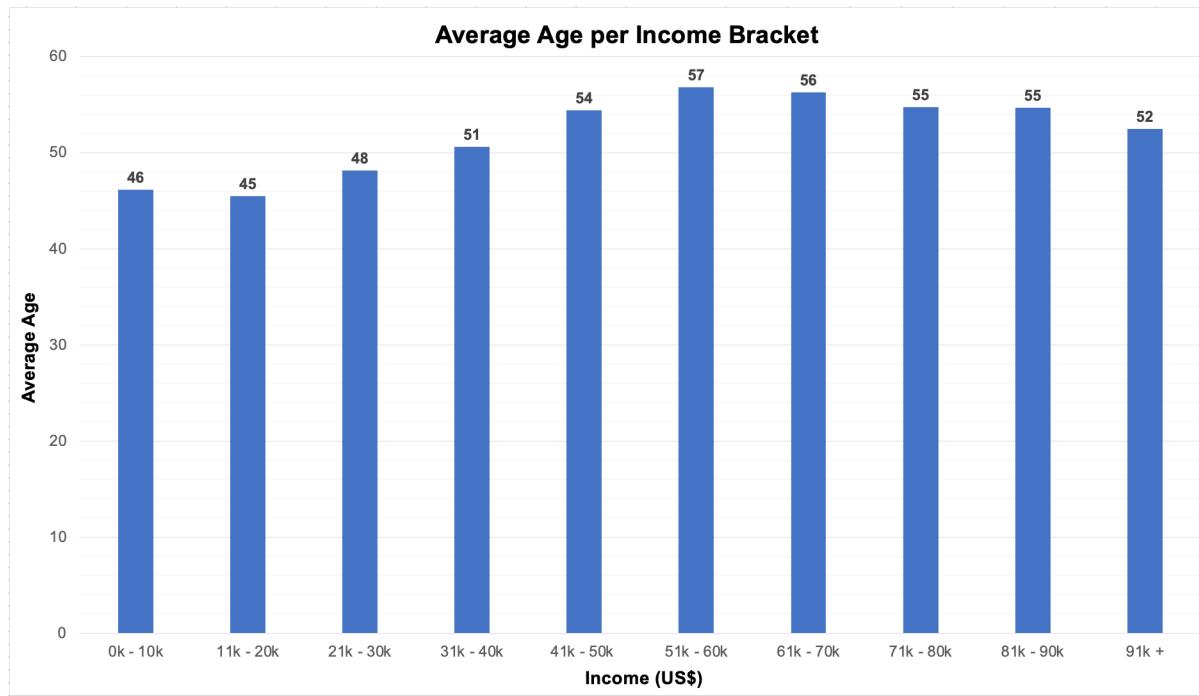
Utilising pivot tables in Excel, initial insights on customer demographics are presented.



The average customer age is 53, where the age bracket between 40 – 50 represents 55% of customers. There is a similar proportion of marital levels in all age groups, with an increase in divorcees and widows after the age of 40.



No significant relationship was found between age and income. There is a relatively even spread of ages which earn between US\$10,000 – 100,000. As to be expected, there is a general increase in average age as income increases, which then starts to plateau and decrease slightly after an income of US\$60,000.



The datasets were then imported, joined and queried in SQL (see syntax used in Appendix).

The total and average spend per country as well as the number of customers per country is shown below.

	Country character	total_spend numeric	average_spend numeric	cnt bigint
1	Spain	659557	603.44	1093
2	South Africa	211071	626.32	337
3	Canada	167403	629.33	266
4	Australia	85576	582.15	147
5	India	77806	529.29	147
6	Germany	73198	631.02	116
7	USA	67546	631.27	107
8	Montenegro	3122	1040.67	3

A major limitation in this data set when comparing between countries is the sample size bias which exists with customers from Spain accounting for 49% of all entries. As any SUM values will correlate to the number of entries that exists, count or percentage is shown to not mislead the viewer and the average value is investigated instead where possible.

In accordance with the central limit theorem, any values investigated with a sample size of less than 30 is considered insufficient in size and ignored. As in the case with Montenegro which only has 3 total customer entries.

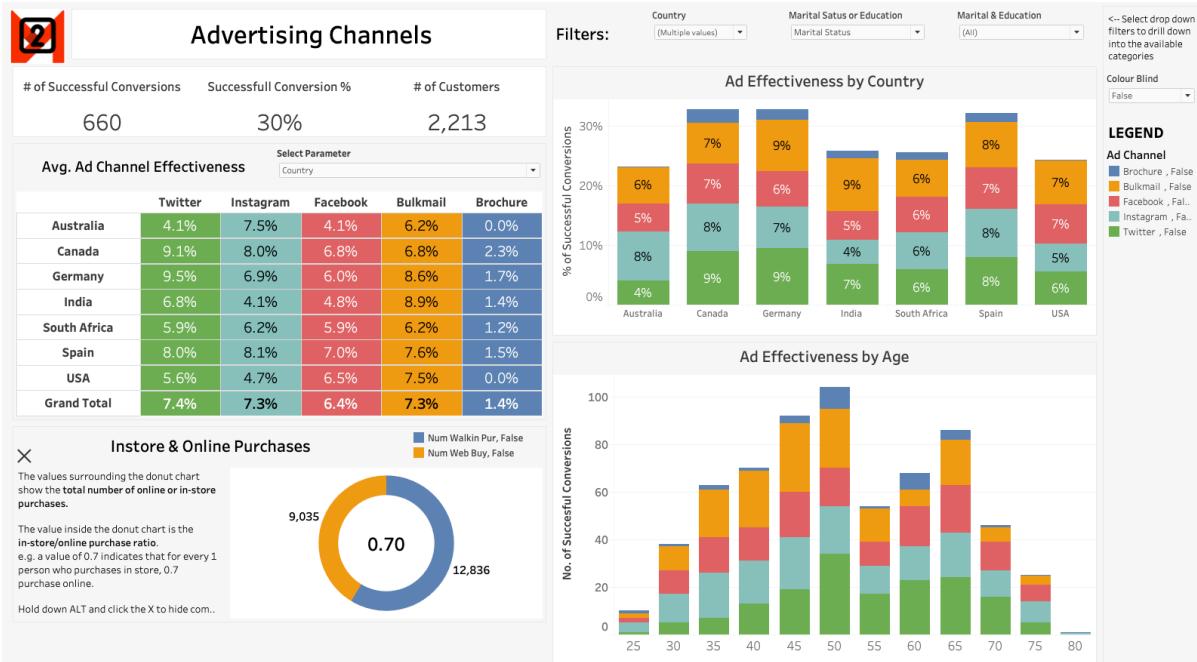
Additionally, to determine the most popular product per country, the amount spent per product was divided by the total amount spent per customer, which reflects the proportion each customer spends for each product category.

	Country character	avg_spend numeric	avg_liquor numeric	avg_vegetables numeric	avg_meat numeric	avg_fish numeric	avg_chocolates numeric	avg_commodities numeric
1	Montenegro	1040.67	576.33	2.67	272.33	75.33	40.67	73.33
2	USA	631.27	301.07	28.36	188.64	41.22	26.76	45.22
3	Germany	631.02	317.03	25.69	174.76	39.66	24.15	49.72
4	Canada	629.33	316.04	28.88	172.65	37.52	28.60	45.65
5	South Africa	626.32	314.30	26.52	173.29	40.56	26.76	44.89
6	Spain	603.44	307.77	25.88	163.23	36.74	27.57	42.25
7	Australia	582.15	290.83	25.10	151.89	37.73	28.09	48.52
8	India	529.29	246.50	25.77	161.42	32.78	21.91	40.91

The most spent on a product category on average is on alcohol, followed by meat, commodities, fish, vegetables and chocolate and this is consistent across all countries, marital status, education levels and whether there are children at home. The differences in average spend likely reflect the differential pricing that exists for the same products in different countries. For the purpose of this project, prices are assumed to be the same across countries

Dashboard design and development

The datasets were then brought into Tableau as a blend since there were no duplicates between the ID columns.

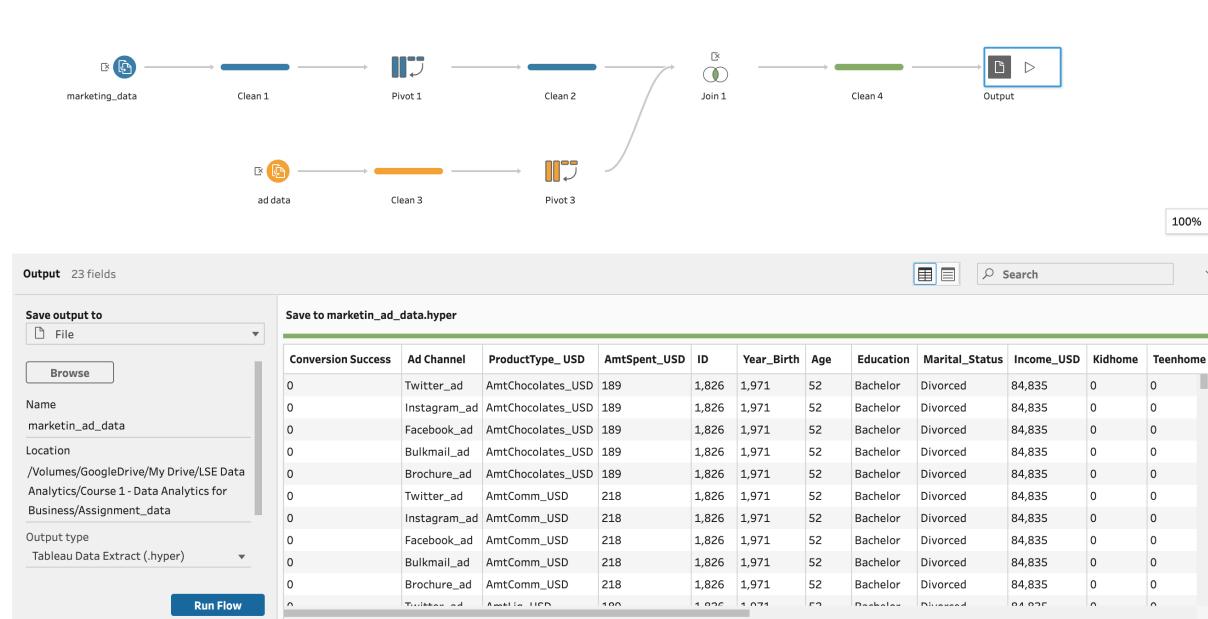


A dashboard was created for each area of focus, including only the most relevant insights and findings. For example, investigating customer demographics, the findings shown were how the demographics correlated with total and average spending.

As the dashboard is not intended for exploratory analysis, there are limited interactive elements in the default layout. However, if the user wants to drill down into the data, consistent filters were employed in each dashboard. A description was included containing instructions of how to use the filters, while also minimising unnecessary text to assist with screen readers. A consistent KPI card was also included in each dashboard for a quick overview of the data being viewed.

Visualisation type were chosen based of the [FT Visual Vocabulary Guide](#). All visualisations include clear titles, where necessary captions are used which provide additional information on how to view a chart. Contrasting colours were chosen to differentiate different categories, and a toggle to enable colour-blind colours in included. Sufficient contrast between charts and text's foreground and background were also ensured. Similar to above, Montenegro was excluded when investigating differences between countries. Incorrect entries where the customer's age was 120+ were also excluded in the visualisations.

In the future, it is also recommended that 2Market consider changing their data tables from column-oriented to row-oriented, which is what Tableau is optimised for. This would make it easier to work with the data in Tableau. An attempt was made to pivot the data in Tableau Prep. However, this method only allows for creating visualisations when a distinct count or percentage difference is used, otherwise there are issues with duplicates.



Patterns, Trends and Insights

The age group between 45 – 50 spend the most accounting for 41% of sales and make up 44% of the customer base.

	Country character	avg_spend numeric	weighted_liquor numeric	weighted_vegetables numeric	weighted_meat numeric	weighted_fish numeric	weighted_chocolates numeric	weighted_commodities numeric
1	Montenegro	1040.67	0.57	0.00	0.25	0.06	0.04	0.07
2	USA	631.27	0.46	0.05	0.27	0.07	0.05	0.11
3	Germany	631.02	0.43	0.05	0.25	0.07	0.05	0.14
4	Canada	629.33	0.46	0.05	0.24	0.07	0.05	0.12
5	South Africa	626.32	0.46	0.05	0.26	0.07	0.05	0.12
6	Spain	603.44	0.46	0.05	0.25	0.07	0.05	0.12
7	Australia	582.15	0.46	0.05	0.24	0.08	0.05	0.13
8	India	529.29	0.44	0.06	0.25	0.08	0.04	0.13

The proportion spent on each product type is very similar across all countries. Alcohol is by far the most popular product type, with almost 50% of customers spending allocated to it, followed by meat at around 25%. For marital status the proportion spent is similar. On average, Widows spend most, married couples the least. 43% of single customers total spending is on alcohol while other marital groups spend more between 46% - 50%.

	Marital_Status character	avg_spend numeric	weighted_liquor numeric	weighted_vegetables numeric	weighted_meat numeric	weighted_fish numeric	weighted_chocolates numeric	weighted_commodities numeric
1	Married	591.46	0.46	0.05	0.25	0.07	0.05	0.12
2	Partnership	608.85	0.47	0.04	0.25	0.07	0.05	0.12
3	Single	610.34	0.43	0.06	0.26	0.07	0.05	0.13
4	Divorced	610.63	0.48	0.05	0.24	0.07	0.05	0.12
5	Widow	727.96	0.50	0.04	0.23	0.07	0.05	0.11

1583 customers (71%) have either a child or teenager at home, compared to 633 without any children. Customers with children spend significantly less on average (\$407.87 vs \$1105.25) and spend a higher proportion on alcohol (49% vs 39%) and commodities (13% vs 9%).

	kidhome_count bigint	teenhome_count bigint	avg_spend numeric	weighted_liquor numeric	weighted_vegetables numeric	weighted_meat numeric	weighted_fish numeric	weighted_chocolates numeric	weighted_commodities numeric
1	1583	1583	407.87	0.49	0.05	0.23	0.07	0.05	0.13

	kidhome_count bigint	teenhome_count bigint	avg_spend numeric	weighted_liquor numeric	weighted_vegetables numeric	weighted_meat numeric	weighted_fish numeric	weighted_chocolates numeric	weighted_commodities numeric
1	633	633	1105.25	0.39	0.06	0.31	0.09	0.06	0.09

The successful conversion rate of the most effective advertising channels in each country are in the range 6% - 9% and for marital status in the range of 1 – 13%. The most effective ad channels in table below are highlighted in green. Overall,

Looking at the most effective advertising channel, Twitter highest average conversion rate followed by bulkmail, Instagram and Facebook. This is consistent with marital status, age and income as well.

Most Effective Ad Channel by Country

Country	bulkmail	twitter	instagram	facebook	brochure
Australia	6%	4%	8%	5%	0%
Canada	7%	9%	8%	7%	2%
Germany	9%	9%	7%	6%	2%
India	9%	7%	4%	5%	1%
South Africa	6%	6%	6%	6%	1%
Spain	8%	8%	8%	7%	1%
USA	7%	6%	5%	7%	0%

Most Effective Ad Channel by Marital Status

Marital Status	bulkmail	twitter	instagram	facebook	brochure
Divorced	9%	8%	6%	5%	2%
Married	7%	7%	8%	7%	1%
Partnership	6%	7%	8%	6%	2%
Single	8%	7%	7%	6%	1%
Widow	5%	13%	9%	7%	1%

Assuming only these five ad channels exist, it is recommended 2Market allocate their marketing budgets proportional to number of best performing ad channels in each country. For example, Australia 100% on Instagram, Germany 50% on Bulkmail and 50% on Twitter.

Recommendations for data to provide for more impactful analysis

To overcome data limitations

- Number of times product purchased
- Number of ads shown

For proper Customer Segmentation

- Order Frequency, Order Value and Order Quantity

To Determine Return on Investment

- Cost per click or any costs associated with running the ads.

Time Series Data

- To analyse any changes in purchasing habits or ad effectiveness over time

Recommended Future Actions and areas to explore

- Better data validation for the format of data types, standardisation of data types (such as Date Format). Explore automation possibilities for collection of this data.
- A/B testing on different ads towards different demographic categories
- If possible, to collect information on the successful conversions of ads shown to specific product types which would allow for more targeted marketing strategy based on product type of interest.
- Utilise Regression analysis and check for statistical significance of the findings (outside scope of this report)

Appendix

Appendix 1: Data Cleaning and Transformation Approach

Columns Added

- Age: using the function =YEAR(TODAY())- Year_Birth Column
- Total_Spent: SUM of amount spent on all product types
- Weighted Product Spend: a new column was inserted next to each product column which contained a weighted value of the amount spent on the product type divided by total spend to give the proportion each customer spends on each product type.

Heading/ Data Name Changes:

- Dt_Customer changed to Customer_Registration
- The abbreviations for Country were changed to the country full names using INDEX MATCH

1.Cleaning Data

- **Spell Check:** Ran a Spell Check in Excel → no spelling errors found
- Correcting Date Format: text to column to make the dates correct, issues with 13 for 2013, used the find and replace function to find all 12 → 2012, 13 → 2013, 14 → 2014
- **Remove Blank Space:** Removed any blank spaces using the TRIM function for the education, marital status and country columns

2.Data Completeness

- **Replacing Blank Cells With #N/A:** There were no blank cells found using Ctrl + g → Go to Special → Selecting Blanks
- **Incomplete or Incorrect Calculations:** No calculations present in any of the columns

3.Data Consistency

- **Consistent Headings:** US\$ added onto the end of the column heading for Income and products sales (amount of product sold) to indicate that amount spent is US\$ and not amount of times the product type was purchased.
- **Change Text Case:** Ensured all text results first letter are capitalised.
- **Converting Numbers Stored as Text:** Checked using the type TYPE function and helper columns, displaying each columns datatype values.
 - Converted Income column into number type and removed the \$ with find and replace.
 - The different data types for dates in the "Registration_Date" were cleaned and corrected using text to columns and DATE function in excel then pasting the values.
- **Change No. of Decimal Places:** no columns had values with decimal places.
- **Correct Date Structure:** After the changes mentioned above, there were no issues with data structure.

4. Data Uniqueness

- **Words With the Same Meaning & Words with Alternate representations:**
 - In the "Marital_Status" column, all entries of "Alone" were changed to "Single" with find and replace.
 - In the "Marital_Status" column, the entries "Together" were changed to "Partnership" assuming "Together refers to in a relationship, civil partnership or cohabiting.
 - In the "Education" column, 2N Cycle refers to a Master's degree, so all "2n Cycle entries were change to "Master".
 - In the "Education" column, we can infer that the entry "Graduation" refers to an undergraduate/ bachelor's degree as there already is Basic, Master's and PhD entry but no entry for undergraduate or bachelor's. All entries of "Graduation" were changed to "Bachelor".
 - In the "Education" Column "Basic" was left as it was as according to International Standard Classification of Education (ISCED): basic education comprises primary education (first stage of basic education) and lower secondary education.
- **Incorrect Data Entries:**
 - The incorrect entries of "YOLO" and "Absurd" were changed to "Unknown".
- **Data Errors:** No errors found using Ctrl + g → Go to Special → Formulas → Errors
- **Strange Year_Birth Values:** There were birth year entries of 1894, 1900 and 1901 which would indicate the age of 129, 123 and 122 respectively. This is highly unlikely however as we can't go back and clarify the data from stakeholders or their customers, these incorrect data entries were left unchanged but noted down to inform the stakeholders. They cells were also highlighted in red in Excel.

5.Duplicate Values

- Duplicate values were checked for using **Conditional Formatting > Highlight Cell Rules > Duplicate Values** in the ID column to ensure there were no duplicates with the same ID.
- Selecting all the data and using the Remove Duplicates feature in the Data tab also resulted in no duplicates found.
- As for duplicate entries between the two datasets, the 1-to-1 match of the two datasets was checked using INDEX MATCH in Excel.
- However, when selecting all columns except the ID column, then using the TEXTJOIN function and a helper column with conditional formatting to check for duplicates, and then filtering coloured cells, there were 94 entries (47 duplicates) where every column is exactly the same. These are likely to be duplicate entries but because the duplicate ID numbers vary so much (you would assume that if they were duplicates of customer details registered on the same day that the ID numbers would be close together) there is no way to rule out they are duplicates without confirming this with the stakeholders.
- Applying TEXTJOIN to all columns except ID and Country column, there are 360 rows (180 duplicates) where all columns are the same. Due to the fact that the ID and the country entries are different, these rows of data can not be excluded from the data set. If this was a real life business case situation, the data analyst should speak directly with stakeholders, including ones who collected the data to clarify.

6. Data Timeliness

- Completed check on data formats and data relevance
- The structure of the data in "Registration_Date" column was formatted from DD/MM/YYYY to YYYY-MM-DD to satisfy the DATE datatype when importing and creating a new table in SQL.
- **Date Relevance:** All the data provided corresponds to a single point in time.

Appendix 2: Syntax Used to Create, Join & Query the Data Tables in SQL

```

CREATE TABLE marketing_data(
    "ID" varchar(10) PRIMARY KEY,
    "Year_Birth" int,
    "Age" numeric(5),
    "Education" character(30),
    "Marital_Status" character(30),
    "Income_USD" money,
    "Kidhome" numeric(5),
    "Teenhome" numeric(5),
    "Registration_Date" Date,
    "Recency" numeric(5),
    "AmtLiq_USD" numeric(10),
    "AmtVege_USD" numeric(10),
    "AmtNonVeg_USD" numeric(10),
    "AmtPes_USD" numeric(10),
    "AmtChocolates_USD" numeric(10),
    "AmtComm_USD" numeric(10),
    "Total_Spend_USD" numeric(10),
    "NumDeals" numeric(5),
    "NumWebBuy" numeric(5),
    "NumWalkinPur" numeric(5),
    "NumVisits" numeric(5),
    "Response" numeric(5),
    "Complain" numeric(5),
    "Country" character(30),
    "Count_success" numeric(5)
);
-- Join of marketing_data and ad_data

CREATE TABLE full_data AS
SELECT
    m."ID", m."Year_Birth", m."Age", m."Education", m."Marital_Status", m."Income_USD", m."Kidhome", m."Teenhome",
    m."Registration_Date", m."Recency", m."AmtLiq_USD", m."AmtVege_USD", m."AmtNonVeg_USD", m."AmtPes_USD", m."AmtChocolates_USD",
    m."AmtComm_USD", m."Total_Spend_USD", 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 marketing_data m
INNER JOIN ad_data a
USING ("ID")
Order BY m."ID";

-- total spend per country with count

SELECT "Country",
SUM("Total_Spend_USD") AS Total_Spend,
ROUND(AVG("Total_Spend_USD"),2) AS Average_Spend,
COUNT("Total_Spend_USD") AS Cnt
FROM marketing_data
GROUP BY "Country"
Order BY Total_Spend DESC;

-- weighted average spend per country

SELECT
    "Country",
    ROUND(AVG("Total_Spend_USD"),2) AS AVG_Spend,
    ROUND(AVG("W_AmtLiq"),2) AS weighted_Liquor,
    ROUND(AVG("W_AmtVege"),2) AS weighted_Vegetables,
    ROUND(AVG("W_AmtNonVeg"),2) AS weighted_Meat,
    ROUND(AVG("W_AmtPes"),2) AS weighted_Fish,
    ROUND(AVG("W_AmtChocolates"),2) AS weighted_Chocolates,
    ROUND(AVG("W_AmtComm"),2) AS weighted_Commodities
FROM weight_marketing_data
GROUP BY "Country"
ORDER BY AVG_Spend DESC;

-- most popular products based on marital status (weighted)

SELECT
    "Marital_Status",
    ROUND(AVG("Total_Spend_USD"),2) AS AVG_Spend,
    ROUND(AVG("W_AmtLiq"),2) AS weighted_Liquor,
    ROUND(AVG("W_AmtVege"),2) AS weighted_Vegetables,
    ROUND(AVG("W_AmtNonVeg"),2) AS weighted_Meat,
    ROUND(AVG("W_AmtPes"),2) AS weighted_Fish,
    ROUND(AVG("W_AmtChocolates"),2) AS weighted_Chocolates,
    ROUND(AVG("W_AmtComm"),2) AS weighted_Commodities
FROM weight_marketing_data
GROUP BY "Marital_Status"
ORDER BY "Marital_Status" DESC;

CREATE TABLE public.ad_data(
    "ID" varchar(10) PRIMARY KEY,
    "Bulkmail_ad" numeric(2),
    "Twitter_ad" numeric(2),
    "Instagram_ad" numeric(2),
    "Facebook_ad" numeric(2),
    "Brochure_ad" numeric(2)
);

-- average spend per product per country

SELECT
    "Country",
    ROUND(AVG("Total_Spend_USD"),2) AS AVG_Spend,
    ROUND(AVG("AmtLiq_USD"),2) AS AVG_Liquor,
    ROUND(AVG("AmtVege_USD"),2) AS AVG_Vegetables,
    ROUND(AVG("AmtNonVeg_USD"),2) AS AVG_Meat,
    ROUND(AVG("AmtPes_USD"),2) AS AVG_Fish,
    ROUND(AVG("AmtChocolates_USD"),2) AS AVG_Chocolates,
    ROUND(AVG("AmtComm_USD"),2) AS AVG_Commodities
FROM marketing_data
GROUP BY "Country"
ORDER BY AVG_Spend DESC;

-- most effective social ad channel in each country

SELECT "Country",
    ROUND(AVG("Bulkmail_ad"),2) AS Bulkmail,
    ROUND(AVG("Twitter_ad"),2) AS Twitter,
    ROUND(AVG("Instagram_ad"),2) AS Instagram,
    ROUND(AVG("Facebook_ad"),2) AS Facebook,
    ROUND(AVG("Brochure_ad"),2) AS Brochure
FROM full_data
GROUP BY "Country"
ORDER BY "Country" ASC;

-- most effective ad channel for marital status

SELECT "Marital_Status",
    SUM("Bulkmail_ad") AS Bulkmail,
    SUM("Twitter_ad") AS Twitter,
    SUM("Instagram_ad") AS Instagram,
    SUM("Facebook_ad") AS Facebook,
    SUM("Brochure_ad") AS Brochure
FROM full_data
GROUP BY "Marital_Status"
ORDER BY "Marital_Status" ASC;

```

LSE_DA101_Data Analytics for Business: Final Assignment

```
-- most popular products with kids at home
```

```
SELECT COUNT("Kidhome") AS Kidhome_Count,
       COUNT("Teenhome") AS Teenhome_Count,
       ROUND(AVG("Total_Spend_USD"),2) AS AVG_Spend,
       ROUND(AVG("W_AmtLiq"),2) AS weighted_Liquor,
       ROUND(AVG("W_AmtVege"),2) AS weighted_Vegetables,
       ROUND(AVG("W_AmtNonVeg"),2) AS weighted_Meat,
       ROUND(AVG("W_AmtPes"),2) AS weighted_Fish,
       ROUND(AVG("W_AmtChocolates"),2) AS weighted_Chocolates,
       ROUND(AVG("W_AmtComm"),2) AS weighted_Commodities
  FROM weight_marketing_data
 WHERE "Kidhome" >=1 or "Teenhome" >=1
```

```
-- most popular products without kids at home
```

```
SELECT COUNT("Kidhome") AS Kidhome_Count,
       COUNT("Teenhome") AS Teenhome_Count,
       ROUND(AVG("Total_Spend_USD"),2) AS AVG_Spend,
       ROUND(AVG("W_AmtLiq"),2) AS weighted_Liquor,
       ROUND(AVG("W_AmtVege"),2) AS weighted_Vegetables,
       ROUND(AVG("W_AmtNonVeg"),2) AS weighted_Meat,
       ROUND(AVG("W_AmtPes"),2) AS weighted_Fish,
       ROUND(AVG("W_AmtChocolates"),2) AS weighted_Chocolates,
       ROUND(AVG("W_AmtComm"),2) AS weighted_Commodities
  FROM weight_marketing_data
 WHERE "Kidhome" =0 and "Teenhome" =0
```