2Market Demographic and Market Analysis

The client, 2Market, an online and in-store global supermarket has requested an analysis of their customers' demographic to launch a targeted marketing campaign. They have asked to investigate:

- Customers' demographic.
- Most effective advertising channels.
- Best-selling products per demographic.

From this problem statement, the deriving questions were:

- How diversified is their customers' demographic and their relative shopping habits.
- Were there products that were purchased significantly more or less than others.
    - If so, could they be paired or promoted alongside other products that performed better to increase overall sells through the marketing campaign.
- How to best harness advertising channels in relation to demographic.

To answer these questions, it is necessary to get acquainted with the data set. The first step in the analytical process after defining the problem question is to clean the data. So, I checked for accuracy and completeness in the marketing data set.

- The date format turned out to be the trickiest part:
    - Different formatting made harder to use the TEXTSPLIT function in a straightforward way.
    - Had to change OS setting to format the date as 20XX and not 19XX when the DATE function was used.
- Removed duplicates with Find and Replace.
- Created an 'Age' column by subtracting the year of birth to the current year with the TODAY function.
- Removed incorrect values.
- Noticed some outliers in the 'Age' column. I didn't delete those, but I did exclude them from the calculations of the average age as to not skew the data too much but kept them for later calculations.
    - Later realised they didn't contribute to the analysis and were excluded.
- Changed SP (Spain) to ES, so that Tableau was able to recognise the country.

After the data cleaning process, I began looking at the average age across marital status and the count of people across age group and marital status. It showed the overall average at 53 with the most populous age group in the 45-54 range and 'Married' as the most frequent marital status.

Wanting to explore the relationship between demographic and successful leads of the marketing campaigns, I performed more complex analysis in SQL by creating the tables on the database and importing the clean files. While importing the ad_data data set, I did not import the values as Boolean but as numeric. Boolean values required more complex syntax to be counted than numeric values, as such, I simplified my analysis by bypassing the Boolean values directly:

- Calculated the countries total spend in descending order alongside the breakdown of the total spending per product.
    - Spain generated the biggest total, followed by Saudi Arabia and Canada.

    - With alcohol as the most bought product across both countries, age group and marital status.
- Looked into the social media leads, performed an inner join calculation that showed:
    - Twitter led with a 164 total leads, followed closely by Instagram at 162 and then, Facebook at 142.
    - When looking at leads in relation to having kids or teens at home, it showed a change in ranking with Twitter still decisively leading but showing Facebook as the second most effective lead over Instagram.

With the cleaned data sets, I explored various visualisations that helped strengthening the analysis. The visualisations had already provided a first glimpse into the count of age groups and marital status as well as best selling products before being confirmed in SQL. With Tableau, I have investigated the demographics, countries, products and social media leads, in particular:

- The first visualisation in the dashboard from top left represents the average age per marital status in a bar char format with a colour gradient that shows how the count of people in each status, with 'married' as the most populated category with the darkest blue and the 'widow' as the least. The label shows the average income per category and shows the range between the five categories, varying from $51569 for 'Single' to $53003 for 'Married'.
    - The legend represents the count.
    - It has a filter for the marital status to look more into each category.
- The second visualisation on the top right shows the best-selling products across age groups. With alcohol as the top one across groups, followed by meat and, commodities, along to the label showing the sum spent per category.
    - The legend shows the products, and it is based on the colourblind palette provided by Tableau so that it could be read by a wider audience and the label with the sum was specified as part of accessibility concerns.
    - The filter applied to it allows to investigate the other products as well, if one wished to compare more in-depth.
    - This visualisation is itself a filter to all the other visualisations in the dashboard, allowing to not only compare products by products but also, relatively to demographic and location.
- The third visualisation is the one bottom right is a spatial visualisation with the best-selling products per country. It employed a dual-axis solution to represent more than one product.
    - There are two legends: one for the colour gradient referring to alcohol and the red symbol for meat. The colours were chosen to resemble the ones from the previous visualisation and provide continuity.
    - It has a filter for the 'Country' category to examine them more closely.
- The last visualisation represents the social media leads per marital status, showing how closely Twitter and Instagram are in terms of successful leads.
    - It has a legend but no filter. The colours are from the same palette as the products visualisations but not the same to not confuse but still provide continuity.

In conclusion, the analysis showed the majority of the demographic falling in the 'Married' category, in the age group ranging between 45-54 with the overall average age of 53. Across the various categories (marital status, with kids/teens at home, age group and country), alcohol is the most sold products, followed by meat and commodities. The marketing campaign could pair alcohol with other less selling products to increase overall sells. The largest spending country is Spain, with Canada and Saudi Arabia following. The difference between Spain and the other countries is quite large and is

worth looking at more deeply to fill the gap. The social media channel with the most range and effectiveness is Twitter with Instagram as a very close second. If the campaign were to focus on parents specifically, Twitter should be paired with Facebook for greater reach and success, but it is worth looking more into how much people with kids or teens at home make up of the overall demographic.

Appendix:

Countries Total Spending + Products

```
SELECT "Country", SUM("AmtLiq" + "AmtVege" + "AmtNonVeg" +"AmtPes" + "AmtChocolates" +"AmtComm" ) AS total_spend,
SUM("AmtLiq") AS alcohol_total, SUM("AmtVege") AS veggie_total, SUM("AmtNonVeg") AS meat_total, SUM("AmtPes") AS
fish_total, SUM("AmtChocolates") AS chocolate_total, SUM("AmtComm") AS commodities_total
FROM public."marketing_data"
GROUP BY "Country"
ORDER BY total_spend DESC
```

Output:

	Country character	total_spend numeric	alcohol_total numeric	veggie_total numeric	meat_total numeric	fish_total numeric	chocolate_total numeric	commodities_total numeric
1	ES	598980	306224	25610	161960	35994	27542	41650
2	SA	188500	94249	7961	52406	12270	8004	13610
3	CA	157474	79415	7201	43237	9133	7287	11201
4	AUS	74849	36833	3222	19859	5066	3547	6322
5	IND	67810	31558	3287	20908	4257	2904	4896
6	US	66030	31116	3009	19951	4362	2811	4781
7	GER	66010	32834	2811	18527	4189	2529	5120
8	ME	2132	940	8	675	214	113	182

Successful Social Media Leads per Marital Status:

```
SELECT md."Marital_Status", SUM(ad."Twitter_ad") AS twitter_leads, SUM(ad."Instagram_ad") AS instagram_leads,
SUM(ad."Facebook_ad") AS facebook_leads
FROM public."marketing_data" md
JOIN public."ad_data" ad
USING("ID")
GROUP BY md."Marital_Status"
```

Output:

Marital_Status "char"	twitter_leads bigint	instagram_leads bigint	facebook_leads bigint
W	10	7	5
T	38	39	28
M	58	59	58
A	0	1	1
S	29	28	29
D	17	12	11

Social Media Leads with Teen Home:

```
SELECT md."Marital_Status", SUM(ad."Twitter_ad") AS twitter_leads, SUM(ad."Instagram_ad") AS instagram_leads,
SUM(ad."Facebook_ad") AS facebook_leads
FROM public."marketing_data" md
JOIN public."ad_data" ad
USING("ID")
GROUP BY md."Teenhome", md."Marital_Status"
HAVING md."Teenhome" >= 1 AND md."Marital_Status" != 'Alone'
```

Output:

Marital_Status "char"	twitter_leads bigint	instagram_leads bigint	facebook_leads bigint
D	0	1	0
T	22	6	5
M	32	3	13
S	11	2	4
W	0	0	0
D	8	2	3
S	1	0	0
M	4	2	0
W	6	2	0
T	0	0	1

*did not manage to group them properly but summed them myself.