Technical Report

2 Market is a global supermarket, offering a wide range of products catering to customers of all ages. Recent observations have revealed a notable trend indicating certain age groups of customers are not as interested in their products and a significant difference in sales performance between Spain and other countries. This report aims to conduct an in-depth analysis focusing on demographic profiles, sales trends across various regions, and related factors to gain insight and develop strategic plans.

To ensure data accuracy in the Excel files "***marketing_data***" and "***ad_data***", empty cells were checked using the ***Find*** function and found none. Next, duplicate entries were removed using ***Remove Duplicates***. While no duplicates were found based on specific columns like ID, Year_Birth, Education, Marital Status, and Income, the situation of customers might have had multiple IDs was noticed and another check was run without the ID column. Then 201 duplicate entries were found and removed.

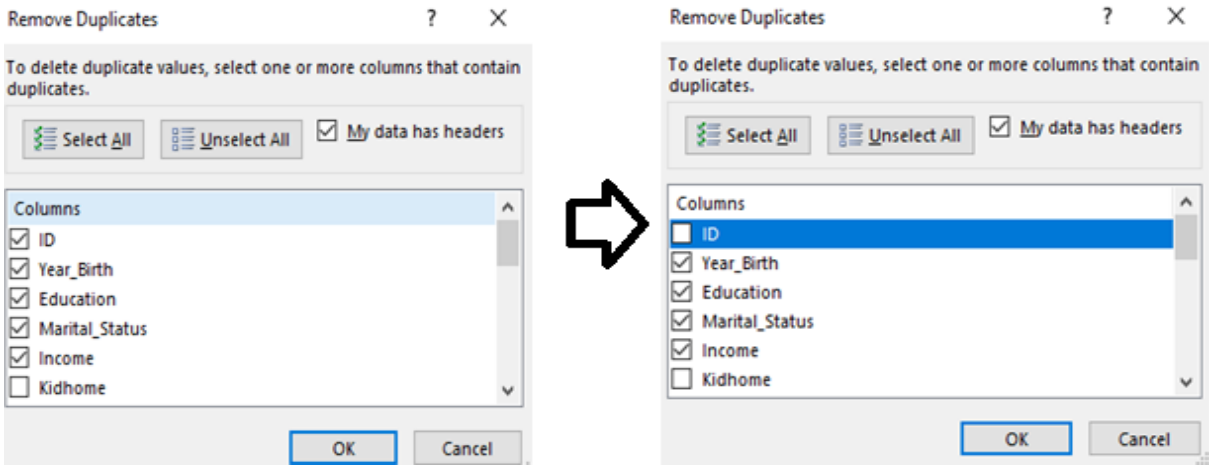

Figure 1:Steps to remove duplicates.

For the analysis, SQL and Excel were used to examine the limited appeal to certain age customers. To display spending across different age groups, a table was constructed with two columns in SQL: "***Age_group***" and "***Total_Spend***". The "***Age_group***" column was created by categorised the "***Ages***" column into bins of 10, ranging from 20-29 to 80-89 (lines 2 to 9).Then customers over 89 were excluded due to a small sample size and labelled them as ***'out of range'*** (line 10). Subsequently, "***Age_group***" was assigned as the name for the newly created bins. For the "***Total_Spend***" column, the ***SUM*** function was employed to compute the total spending on all products, naming the result as "***Total_Spend***" (line 13). Finally, ***'marketing_data***' from the public section was specified using ***FROM***, combined the results of "***Total_Spend***" into "***Age_Group***" using ***GROUP BY***, and arranged the outcome from smallest to largest using ***ORDER BY***.

Query Query History

```
1  SELECT
2      CASE
3          WHEN "Age" BETWEEN 20 AND 29 THEN '20-29'
4          WHEN "Age" BETWEEN 30 AND 39 THEN '30-39'
5          WHEN "Age" BETWEEN 40 AND 49 THEN '40-49'
6          WHEN "Age" BETWEEN 50 AND 59 THEN '50-59'
7          WHEN "Age" BETWEEN 60 AND 69 THEN '60-69'
8          WHEN "Age" BETWEEN 70 AND 79 THEN '70-79'
9          WHEN "Age" BETWEEN 80 AND 89 THEN '80-89'
10         ELSE 'Out of range'
11     END AS "Age_group",
12
13 SUM("AmtLiq"+"AmtVege"+"AmtNonVeg"+"AmtPes"+"AmtChocolates"+"AmtComm") AS "Total_Spend"
14
15 FROM public.marketing_data
16 GROUP BY "Age_group"
17 ORDER BY "Total_Spend" ASC;
```

Figure 2:SQL to show Age and Spending.

A table was created in Excel to view the relationship between total sales number in different age groups and average income in various age groups. The formula employed for this analysis is shown below. This formula calculates the average income (Column F) if age (Column C) falls within the range. The $ ensures that when the equation is copied, the cell references remain fixed, allowing for consistent reference to the specified cell values.

Age Groups	Average Income
20-29	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">=20",$C$2:$C$2016, "<=29")
30-39	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">=30",$C$2:$C$2016, "<=39")
4049	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">= 40",$C$2:$C$2016, "<=49")
50-59	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">=50",$C$2:$C$2016, "<=59")
60-69	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">=60",$C$2:$C$2016, "<=69")
70-79	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">=70",$C$2:$C$2016, "<=79")
80-89	=AVERAGEIFS($F$2:$F$2016, $C$2:$C$2016, ">=80",$C$2:$C$2016, "<=89")

Figure 3:Average Income formula in Excel.

Having previously identified the two groups with the highest income and the lowest sales spending using both SQL code and Excel formulas, an investigation of the relationship between age groups and successful lead conversion across different platforms was conducted. In SQL, a similar process was followed to create bins called "***Age_group***.". Subsequently, an inner join was created to merge both the "***marketing_data***" and "***ad_data***" into a single table. To aid SQL in identifying the source table for each column, column names were prefixed with "***m.***" for "***marketing_data***" and "***a.***" for "***ad_data***" (line 14). ***LEFT JOIN*** was utilised to combine "***ad_data***" with "***marketing_data***" using the common column "***ID***" found in both tables (line15-16). Finally, the same ***GROUP BY*** method was applied as before, but this time ordered the results by "***Age_group***" in ascending order to display age groups from smallest to largest (line 19-20).

Query Query History

```sql
SELECT
    CASE
        WHEN m."Age" BETWEEN 20 AND 29 THEN '20-29'
        WHEN m."Age" BETWEEN 30 AND 39 THEN '30-39'
        WHEN m."Age" BETWEEN 40 AND 49 THEN '40-49'
        WHEN m."Age" BETWEEN 50 AND 59 THEN '50-59'
        WHEN m."Age" BETWEEN 60 AND 69 THEN '60-69'
        WHEN m."Age" BETWEEN 70 AND 79 THEN '70-79'
        WHEN m."Age" BETWEEN 80 AND 89 THEN '80-89'
        ELSE 'Out of range'
    END AS "Age_group",
    COUNT(*) AS "Total_Successful_Conversion"

FROM public."marketing_data" m
LEFT JOIN
    public."ad_data" a ON m."ID" = a."ID"
GROUP BY
     "Age_group"
ORDER by
    "Age_group" ASC;

```

Figure 4:SQL to show Age and Successful lead conversion using "marketing_data" and "ad_data".

Similarly, SQL was used to analyse differences in sales performances between Spain and other countries. The goal was to show total sales across different countries. To achieve this, the total sales numbers were aggregated for all products, denoting the sum as "***Total_Sales***" and used ***GROUP BY*** to display sales by country in descending order.

Query Query History

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

Figure 5:SQL to show total sales in different countries.

Additionally, the relationship between successful lead conversions and sales in different countries was investigated. This involved assigning 1 for value ***TRUE*** within all the successful conversion across platforms and summing them up to get "***Total_successful_conversion***". Then, the "***marketing_data***" was combined with "***ad_data***" was combined using LEFT JOIN, filtering the "***Country***" column to exclude empty cells (line 12).

Query Query History

```sql
SELECT
    m."Country",
    SUM(CASE WHEN a."Bulkmail_ad" THEN 1 ELSE 0 END)+SUM(CASE WHEN a."Twitter_ad" THEN 1 ELSE 0 END)+
    SUM(CASE WHEN a."Instagram_ad" THEN 1 ELSE 0 END)+SUM(CASE WHEN a."Facebook_ad" THEN 1 ELSE 0 END)+
    SUM(CASE WHEN a."Brochure_ad" THEN 1 ELSE 0 END) AS "Total_Successful_Conversion"

FROM
    public."ad_data" a
LEFT JOIN
    public."marketing_data" m ON a."ID" = m."ID"
WHERE
    m."Country" IS NOT NULL
GROUP BY
    m."Country"
ORDER BY
    "Total_Successful_Conversion" DESC;

```

Figure 6:SQL to show successful lead conversion in different countries using "marketing_data" and "ad_data".

The main purpose of this dashboard was clear, interactive, and user-friendly. The layout was simple and structured into 3 sections: main title, an area to display graphs, and a section for guidelines or interactivity.

In designing the dashboard, a title was created to represent the main theme of the dashboard while offering a glimpse of what the audience can expect. A larger and bold font size was chosen to draw the audience's attention and enhance the overall impact of the title.

As for visualisation type, 2 bar charts and 3 scatter plots were built to show insights about the business problems. Bar charts were chosen for their ability to present information clearly and allow easy and direct comparison. Scatter plots show relationship between two variables by identifying their patterns, trends, and correlation. Then, calculated field function was used to create two new columns "***Total Successful Ad***" and "***Total Sales***" for further analysis. Furthermore, age groups were divided into bins of 10 for better visualisation.

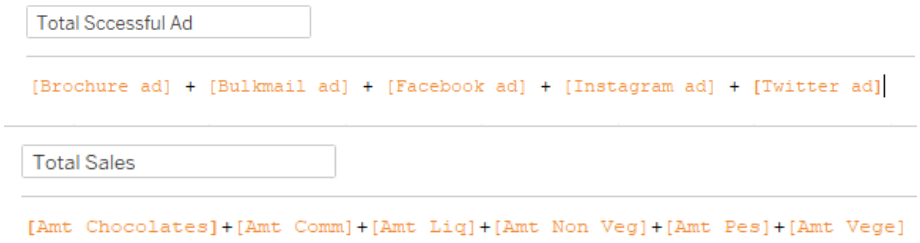

Figure 7: Formulas to create new inputs using Calculated Field in Tableau

Regarding colour choice, a grey background was chosen with yellow texts for graph titles, black for normal texts and blue for graph colour. The purposes of these colours' selection were colour-blind-friendly and create a visual contrast, drawing viewer's attention to the graphs while maintain great readability.

For interactivity, two filters were built for countries and age which are linked to relevant graphs. The goal was to simplify navigation, increase user engagement and enable them to explore the data. Additionally, a text box was inserted to show all the countries options in their full from to give the viewer a better understanding on the graphs and filter.

The analysis reveals server insights. Firstly, Age groups 80-89 and 20-29 display the lowest product sales numbers despite having the highest incomes. This suggests a negative correlation between these two variables among these demographics.

Secondly, Age groups 80-89 and 20-29 has the lowest total successful lead conversion rates, despite their higher income levels. Combine this insight to the first one, a positive correlation can be concluded between sales number and successful lead conversion. This indicates ineffectiveness in the marketing strategies for these demographics and potentially result to the lower spending habit.

Lastly, the data highlighted that Spain leads in sales, followed by South Africa and Canada which also lead in the most successful lead conversion in the same order. This demonstrates a positive correlation between sales and lead conversions, backing up the earlier statement regarding to the issue of the business.