



2 MARKET

ANALYSIS REPORT

June 2024



Prepared By:

Beata Faitli

LSE Career Accelerator



EXECUTIVE SUMMARY

This comprehensive analysis for 2Market, a global supermarket chain, examines customer demographics, purchase behaviour, product popularity, and the effectiveness of marketing channels to guide strategic improvements. Utilizing data from 2012 to 2014, the study uncovers critical insights while acknowledging limitations such as the age of the data and the absence of cost and profitability information. Key findings reveal that alcoholic beverages are the most popular product category, with Twitter and Instagram emerging as the most effective marketing channels. Recommendations include targeted marketing campaigns, enhanced advertising strategies, product line optimization, and customer retention programs. Implementing these strategies is anticipated to boost sales and sustain 2Market's competitive edge in the dynamic supermarket industry. The analysis, visualized through Tableau, underscores the importance of leveraging customer data for strategic decision-making, even as market conditions evolve.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
TABLE OF CONTENTS.....	2
2 MARKET Business Analysis Insights Report	3
Introduction:.....	3
Methodology:.....	3
Analysis and Results:.....	4
Contextualization within the industry or market:	7
Conclusion:.....	7
Recommendations:	7
Areas for future research:	7
REFERENCES.....	7
APPENDIX 1 – 5 Why’s analysis of 2Market project	8
Additional questions for the 2Market Team	9
APPENDIX 2 – Ethical considerations	11
APPENDIX 3 – Data cleaning methodology	13
Summary of Data Cleaning Steps:.....	13
Assumptions Made:.....	13
APPENDIX 4 – Data dictionary	16
APPENDIX 5 – Initial insights from MsExcel exploratory analysis	17
Initial customer demographic insights:.....	17
Age insights:	17
Marital status insights	18
Income:	18
Sales insights.....	20
APPENDIX 6 – pgAdmin SQL code syntax.....	22
APPENDIX 7 – pgAdmin insights tables	43
APPENDIX 8 – Tableau dashboard design.....	47
APPENDIX 9 – Summary of key insights.....	50
APPENDIX 10 - Recommendations	51
Potential Impacts and Benefits	51
APPENDIX 11 – Future research	52
APPENDIX 12 – REFERENCES.....	53

2 MARKET Business Analysis Insights Report

Introduction:

2Market is a global supermarket chain that sells products both online and in-store. The analysis project aimed to understand customer demographics, purchase behaviour, product popularity and if that differs based on demographics, and the effectiveness of current marketing channels. The aim of the analysis is to provide data-driven insights to enhance marketing strategies, optimize product offerings, and ultimately increase sales.

This analysis covers data from 30/07/2012 to 29/06/2014. Key limitations include the absence of cost/profitability data, outdated information, GDPR compliance issues, and limited stakeholder input. Further details are in [Appendix 1 and 2](#).

Methodology:

Data Description:

- **Sources** – Provided by 2Market in three files marketing_data.csv, ad_data.csv and metadata_2Market.csv
- **Data collection methods** – not available.
- **Tools and techniques used for analysis:**
 - MsExcel for initial cleaning and exploratory analysis, pgAdmin (SQL) for deeper insights into product popularity and advertising effectiveness, and Tableau for comprehensive dashboard presentation.

Brief data overview:

- Demographics: customers year of birth, income, marital status, geographic location, the number of children/teens.
- Purchase information: Various product categories and marketing insights for Bulk mail, Twitter, Instagram, Facebook and Brochures, discounts, number of purchases from website or instore.

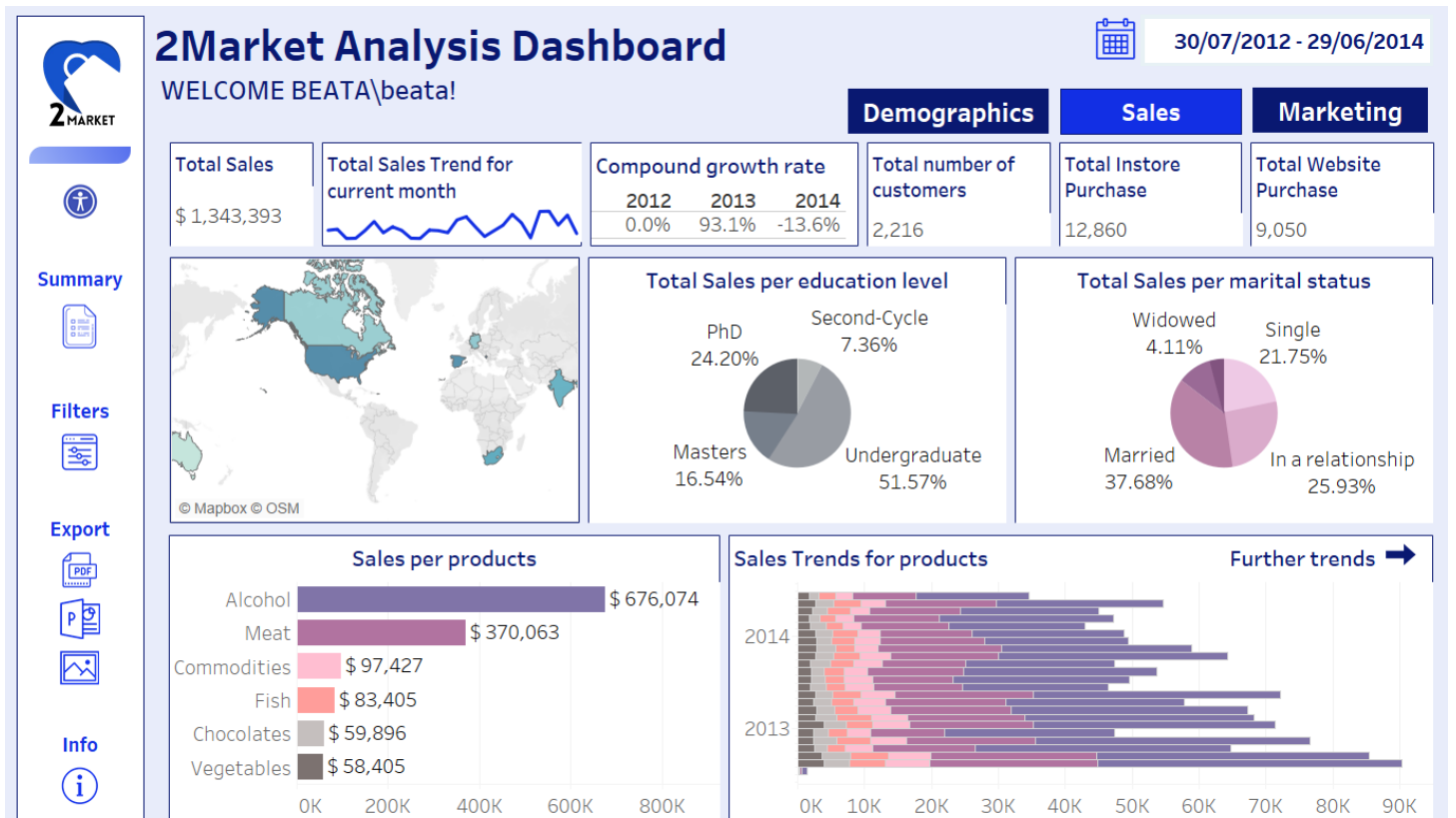
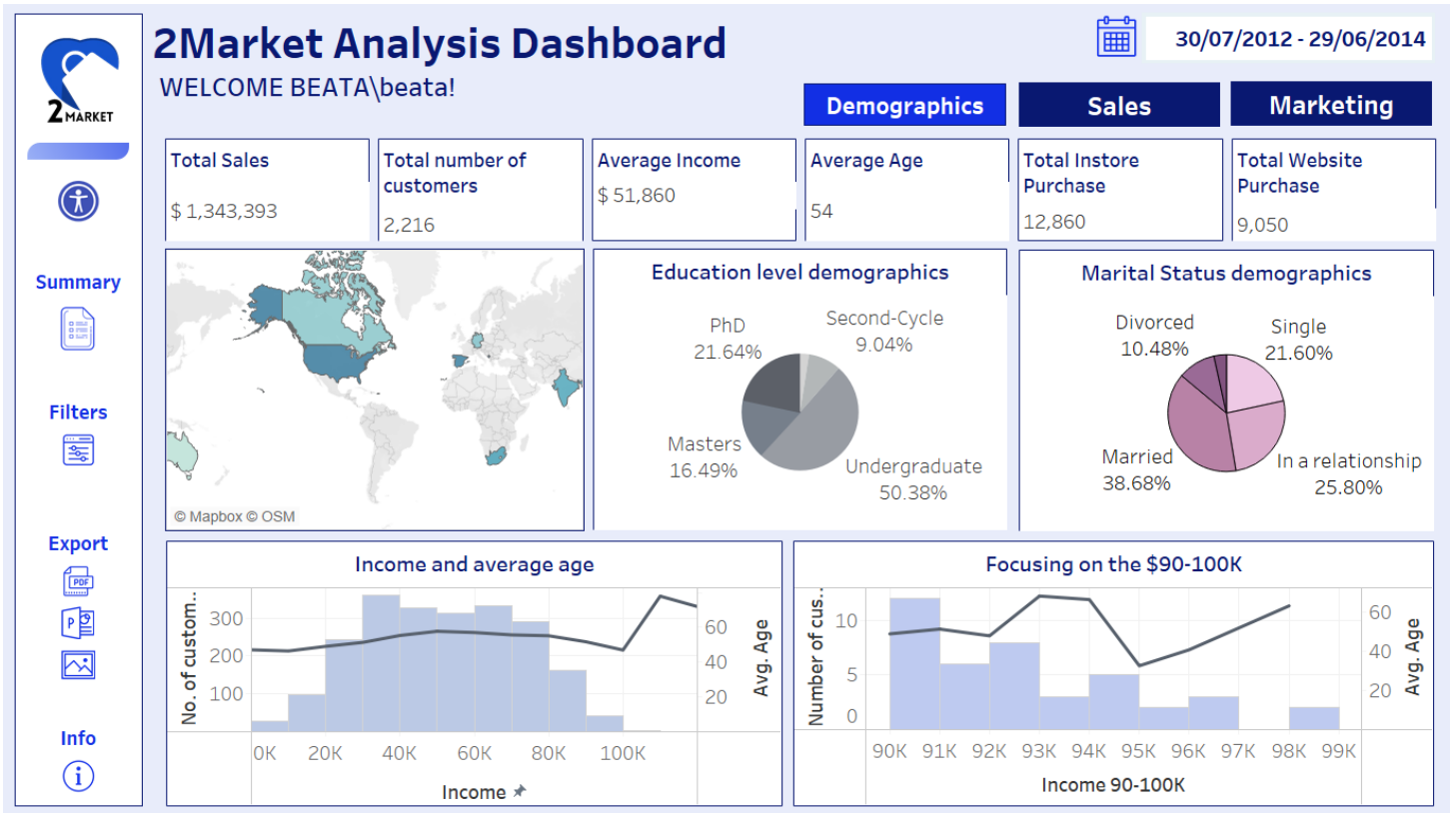
Assumptions, variables and data cleaning steps

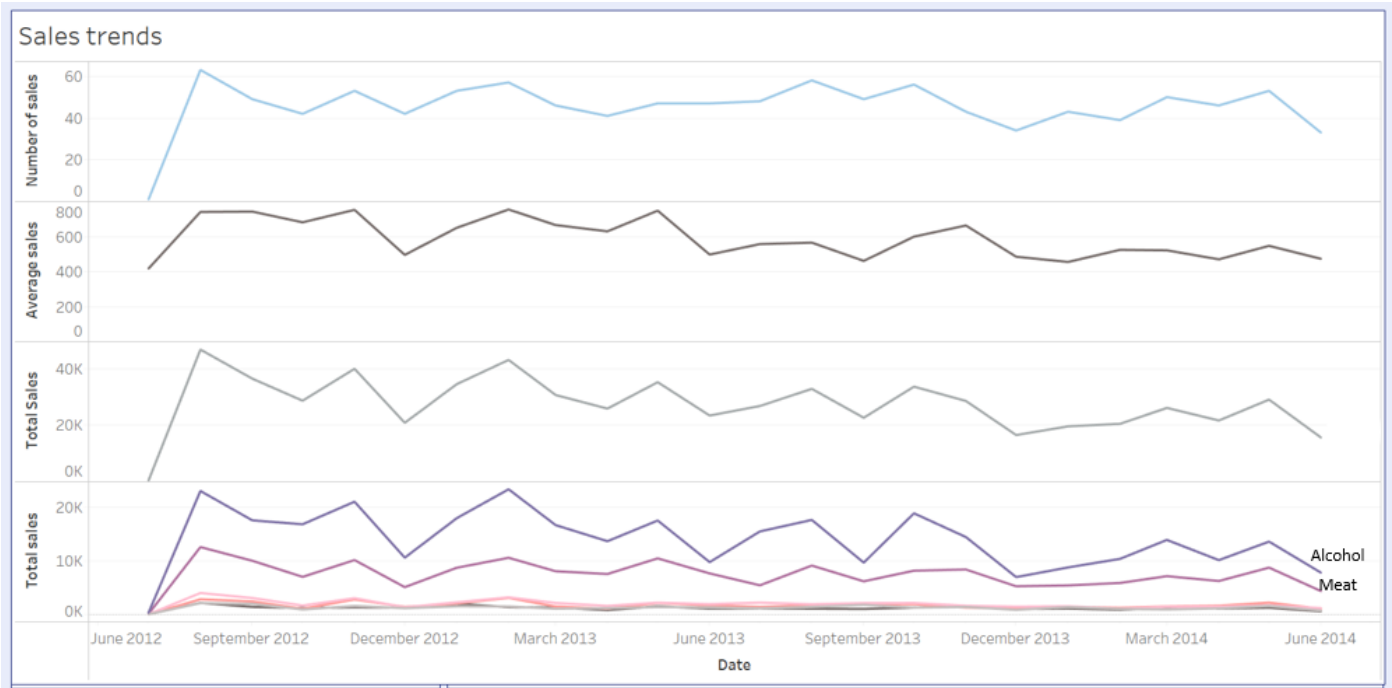
- see details in [Appendix 3 and 4](#).
- **Data cleaning in Excel:**
 - The data was cleaned by adhering to the six principles of data quality as per the Data Management Association's (DAMA) Data Management Body of Knowledge (DMBOK) framework. These included data accuracy, completeness, data validity, consistency, uniqueness and timeliness.
 - The data cleaning process began with importing the CSV files into Excel. Initial steps included checking for missing values, correcting data entry errors, and standardizing formats. Outliers were identified using the IQR method, and appropriate decisions were made on whether to include or exclude these values based on their impact on the analysis. Data was then structured into a clean format for analysis.


Analysis and Results:

- **Ms Excel Analysis:**
 - Exploratory analysis involved calculating descriptive statistics and using pivot tables to identify patterns in customer demographics and purchase behaviour. This preliminary analysis provided a foundation for deeper insights. Results of this can be found under the Descriptive_statistics_results tab in the marketing_clean_tableau_1 file and in [Appendix 5](#).
 - **In summary**, There was data available for 2216 customer purchases, the average age of customers is 54 years, the average income of customers is \$ 52,247. The highest number of customers have a salary of £30-39.9K. 39% of customers are married, 50% of customers have undergraduate level of education. This demographic insight suggests targeted marketing strategies should focus on middle-aged, higher-income segments.
- **pgAdmin (SQL) Analysis:**
 - Data was imported into pgAdmin for advanced querying. SQL was used to analyse product popularity across different demographics and assess the effectiveness of various marketing channels. Complex queries were written to join tables, filter data, and aggregate results, providing a nuanced understanding of the data.
 - The most popular product type is Alcoholic beverages and consistent across customer demographics. The outlier values does not impact on these results. Therefore in the Tableau visualization all sales data is displayed. However, additional filters could be applied depending on stakeholder preferences to show outlier values. This trend indicates a potential for expanding the range of alcoholic products and related promotions.
 - Advertisement effectiveness varies across customer demographics: Twitter has the most conversions, Instagram leads in average sales. There were 87 customers, who was influenced by more than one platform.
 - The SQL syntax can be found in [Appendix 6](#) and the results tables in [Appendix 7](#).
- **Tableau Insights:**
 - Comprehensive view of demographics, sales, and marketing effectiveness. Further detailed insights can be achieved through application of the filters, such as looking at the data for specific time or customer demographic. For additional details refer to the 2Market_analysis_2.twbx file and the presentation file.
 - Detailed rationale for **dashboard design** can be found in [Appendix 8](#).
 - **Trends:** Highest sales were seen from married, undergraduate-educated customers, Spain has the highest sales volume in terms of geographic area. A noticeable decline in sales was observed over the analysed period, despite a stable customer base. This suggests potential issues with customer retention or market saturation, necessitating strategic adjustments to stimulate growth.
 - **Marketing Insights:** Higher average sales from returning customers. Advert channel uptake is low but results in higher average sales per purchase. These insights highlight the importance of leveraging social media platforms more effectively, with tailored strategies for each channel.
- **Challenges and Rationale:**
 - Addressing inconsistent data formats required meticulous cleaning in Excel. Ensuring optimized performance of SQL queries was achieved by indexing key columns and employing efficient joins. The strategic use of Excel facilitated initial data exploration and cleaning, SQL enabled robust data manipulation and querying, and Tableau was utilized for comprehensive data visualization and dashboard development.


- See the main views of the dashboard below.







2Market Analysis Dashboard


30/07/2012 - 29/06/2014

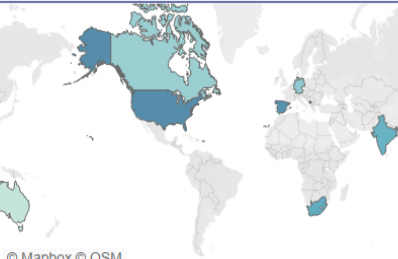
WELCOME BEATA\beata!

Demographics

Sales

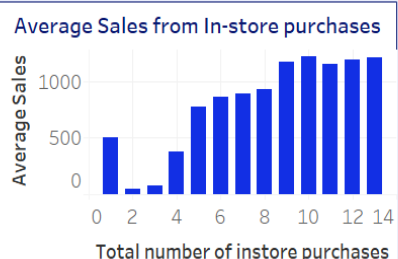
Marketing

<p style="text-align: center; font-weight: bold;">Total Sales</p> <p style="text-align: center;">\$ 1,343,393</p>	<p style="text-align: center; font-weight: bold;">Total number of customers</p> <p style="text-align: center;">2,216</p>	<p style="text-align: center; font-weight: bold;">Average Income</p> <p style="text-align: center;">\$ 51,860</p>	<p style="text-align: center; font-weight: bold;">Average Age</p> <p style="text-align: center;">54</p>	<p style="text-align: center; font-weight: bold;">Total Instore Purchase</p> <p style="text-align: center;">12,860</p>	<p style="text-align: center; font-weight: bold;">Total Website Purchase</p> <p style="text-align: center;">9,050</p>
---	--	---	---	--	---




© Mapbox © OSM

Average Sales from In-store purchases



Total number of instore purchases

Average Sales from website purchases



Total number of website purchases

Marketing channels

Sum of Brochure	30
Sum of Bulkmail	163
Sum of Facebook	142
Sum of Instagram	162
Sum of Twitter	164

Successful lead conversions

Average sales per unsuccessful (0) or successful (1) lead conversions

Channel	Unsuccessful (0)	Successful (1)
Brochure *	\$597	\$1,308
Bulkmail	\$598	\$721
Facebook	\$547	\$1,484
Instagram	\$527	\$1,616
Twitter	\$565	\$1,138

Contextualization within the industry or market:

- 2Market operates in a highly competitive global supermarket industry. Supermarkets face constant pressure to differentiate themselves through personalized marketing, targeted advertising, and efficient product offerings. By leveraging customer data and analytics, companies can enhance customer experience, streamline operations, and drive sales growth. The insights from this analysis help 2Market stay competitive by identifying key demographic trends, product preferences, and the effectiveness of marketing channels, which are vital for developing informed business strategies

Conclusion:

- Summary of key insights – see [Appendix 9](#).
- The analysis reveals that while certain demographics consistently drive sales, product preferences are skewed towards alcoholic beverages.
- The effectiveness of marketing channels varies, with Twitter and Instagram showing higher impacts.
- Despite a stable customer base, declining sales trends indicate a need for strategic adjustments. Advertising, although underutilized, results in higher average sales per purchase, highlighting an opportunity for increased focus on marketing efforts.

Recommendations:

- See detailed recommendations and their potential impact in [Appendix 10](#).
 - **Targeted Marketing Campaigns** - Focus on demographics driving the highest sales.
 - **Enhanced Advertising Strategy** - Increase investment in Twitter and Instagram.
 - **Product Line Optimization** - Focus on popular product categories like alcoholic beverages.
 - **Customer Retention Programs** - Enhance programs to retain returning customers.
 - **Data-Driven Decision Making** - Use customer data for strategic planning, increase connected data such as profitability and costs.

Areas for future research:

- **Limitations and Future Directions:**
 - **Data Limitations:** Lack of cost and profitability data.
 - **Timeliness:** Data is over 10 years old, possibly outdated customer demographics and marketing strategies.
 - **Ethical Considerations:** Compliance with GDPR.
- Future research should address these limitations by incorporating recent data and considering profitability alongside sales volume. More details are in Appendix 11.

REFERENCES

- See references under [Appendix 12](#).

APPENDIX 1 – 5 Why's analysis of 2Market project

Problem statement and initial questions:

2Market is a global supermarket chain that sells products both online and in-store. They aimed to understand customer demographics, purchase behaviour and the effectiveness of current marketing channels in order to enhance marketing strategies, optimize product offerings, and ultimately increase sales.

1. Who is 2Market and what are they trying to achieve?

2Market is a global supermarket chain that sells products both online and in-store. They aim to understand their customer purchase behaviour better to enhance marketing strategies, optimize product offerings, and ultimately increase sales.

2. Why are they trying to achieve this?

By gaining insights into customer demographics, the effectiveness of advertising channels, and product sales variations across different demographics, 2Market seeks to improve customer targeting, allocate marketing budgets more efficiently, and tailor product offerings to meet the needs of various customer segments.

3. How will this data analytics project support 2Market's objectives?

This project will provide actionable insights through the analysis of customer demographics, advertising effectiveness, and product sales data. These insights will help 2Market make informed decisions to boost customer engagement, improve sales performance, and achieve a higher return on marketing investments.

5 Whys analysis of the problem statement:

Problem: Sales are not optimized, and marketing strategies are not yielding maximum results.

1. Why are sales not optimized, and marketing strategies are not yielding maximum results?
Because 2Market does not have a detailed understanding of their customer demographics, the effectiveness of different advertising channels, and how product sales vary by demographic.
2. Why does 2Market not have a detailed understanding of these factors?
Because they have not yet thoroughly analysed the available data on customer demographics, advertising performance, and product sales.
3. Why has 2Market not thoroughly analysed this data?
Because the data has not been fully integrated, and comprehensive analysis methods have not been applied to extract actionable insights.
4. Why has the data not been fully integrated, and comprehensive analysis methods not applied?
Because there may be a lack of structured approach and resources dedicated to data analytics, including tools and expertise to perform such analysis.
5. Why is there a lack of structured approach and resources dedicated to data analytics?
Because there may be insufficient awareness or emphasis within the organization on the value of data-driven decision-making and potential benefits of investing in data analytics capabilities.

Summary of Root Cause

The root cause identified is a lack of structured approach and resources dedicated to data analytics, stemming from insufficient organizational awareness of the value of data-driven decision-making. To address this, 2Market needs to invest in integrating their data, applying comprehensive analysis methods, and fostering a culture that values insights derived from data analytics.

Recommendations to Address the Root Cause

- **Develop a Structured Data Analytics Plan:**
Define clear objectives and key performance indicators (KPIs) for the data analytics project. Outline a step-by-step approach to data integration and analysis.
- **Invest in Data Analytics Tools and Expertise:**
Acquire or develop the necessary tools (e.g., SQL databases, visualization tools like Tableau or Power BI, statistical software like R or Python).
Hire or train staff with expertise in data analytics and business intelligence.
- **Integrate Data Sources:**
Combine customer demographic data, sales data, and advertising performance data to create a comprehensive dataset for analysis.
- **Perform Comprehensive Analysis:**
Conduct demographic analysis to understand customer segments.
Analyse the effectiveness of different advertising channels.
Evaluate product sales variations across different demographics.
- **Promote a Data-Driven Culture:**
Educate the organization on the benefits of data-driven decision-making.
Regularly share insights and success stories from data analytics projects to demonstrate value.

By addressing these root causes, 2Market can optimize their sales and marketing strategies, ultimately leading to improved performance and increased revenue.

Additional questions for the 2Market Team

1. **Scenario Context:**
 - Can you provide more details about any recent changes in your marketing strategy or business operations that might impact customer behaviour?
 - Are there specific regions or demographics you are particularly interested in analysing for targeted marketing efforts?
2. **Data Provided:**
 - Is there any additional data available that might be relevant to this analysis, such as customer feedback or loyalty program details?
 - Is there data available with regards to cost and profitability?
 - The data provided appears to be over 10 years old, is there any more recent data available?
 - Would you like us to consider further data, such as competitor analysis?
 - Would you like the currency to remain in dollar format or to be converted to pounds?
3. **Audience Understanding:**
 - Who will be the primary audience for the findings of this analysis (e.g., marketing team, senior management, product managers)?
 - What are the key metrics or insights that the audience is most interested in?

Additional questions in relation to the data:

- What would you like us to consider as the customer's age? Age in the year of analysis in 2024 or when the customer came in contact with 2Market?
- How was the data entered from different countries? Manually in one area or streamed from different countries as there appears to be a large number of similar entries. Could this have been due to administrative errors?
- How do customer preferences vary across different regions? Are there specific products that are particularly popular or unpopular in certain countries?
- What is the relationship between the frequency of website visits and the amount spent on various product categories?

More advanced techniques to answering these questions:

Regional Preferences:

- Approach: Segment the data by country and analyse the sales data for different product categories within each region. Use visualization techniques such as heat maps or bar charts to illustrate the differences in preferences.
- Tools: Use SQL for data segmentation, and visualization tools like Tableau or Power BI for creating visuals.

Frequency vs. Spending:

- Approach: Conduct a correlation analysis between the number of website visits and the amount spent on different product categories. Use regression analysis to identify any significant relationships.
- Tools: Utilize statistical software like R or Python (with libraries such as pandas, seaborn, and scikit-learn) to perform the correlation and regression analysis.

APPENDIX 2 – Ethical considerations

The analysis of 2Market's data entails significant ethical considerations, particularly in light of the General Data Protection Regulation (GDPR), which governs data privacy and protection in the European Union. Although the data used spans from 2012 to 2014, prior to the GDPR's enactment in 2018, its principles are crucial for ensuring ethical data handling practices. Given that some of the customers are based within the EU.

GDPR Overview and Principles: The GDPR emphasizes transparency, accountability, and the individual's right to privacy. Key principles include data minimization, accuracy, storage limitation, integrity, confidentiality, and lawful processing (European Parliament, 2016). Organizations must obtain explicit consent for data processing, provide individuals with access to their data, and ensure secure data storage and transfer.

Key GDPR Considerations for 2Market:

1. Consent and Data Subject Rights:

- **Consent:** For any future data collection and analysis, 2Market must ensure that customers provide informed and explicit consent. This consent should be specific, granular, and easy to withdraw.
- **Right to Access and Erasure:** Customers have the right to access their data and request its deletion. 2Market must establish clear processes to handle such requests efficiently.

2. Data Anonymization:

The historical data used for this analysis must be anonymized to protect customer identities. Anonymization helps mitigate risks of re-identification and complies with the GDPR's emphasis on privacy protection. The dataset did not include identifiable information, such as name, date of birth, address, etc.

- **Personal vs. sensitive data** - Under the GDPR, personal data refers to any information that can identify an individual directly or indirectly. This includes names, email addresses, identification numbers, location data, and online identifiers. Sensitive data, a special category of personal data, requires even stricter protection due to its nature. This category includes data revealing racial or ethnic origin, political opinions, religious beliefs, trade union membership, genetic data, biometric data, health information, and data concerning a person's sex life or sexual orientation. Whilst the data included marital status, it did not refer to the sensitive information of sexual orientation (European Parliament, 2016).

3. Data Security:

- Ensuring robust data security measures, including encryption and regular security audits, is essential to protect customer data from breaches and unauthorized access. Article 32 of the GDPR mandates implementing appropriate technical and organizational measures to safeguard data (European Parliament, 2016).

4. Data Minimization and Purpose Limitation:

- Data collected should be limited to what is necessary for the specific analysis purpose. Avoid collecting excessive or irrelevant data points. The GDPR advocates for data minimization to reduce the risk of misuse and enhance privacy.

5. Historical Data Use:

- While the GDPR applies to data collected post-2018, ethical considerations suggest applying similar standards to historical data. This includes respecting the privacy of individuals whose data was collected and ensuring it is not used in ways that could



harm them. Consent for the use of the provided historical data was implied for the purpose of this assignment.

Implications for 2Market: Adhering to GDPR principles not only ensures legal compliance but also builds customer trust, which is essential for long-term business success. By prioritizing data ethics, 2Market can demonstrate its commitment to respecting customer privacy and maintaining high standards of data governance.

In conclusion, integrating GDPR principles into 2Market's data practices, even for historical data, will help the company navigate ethical challenges and align with modern data protection standards, fostering a trustworthy relationship with its customers.

APPENDIX 3 – Data cleaning methodology

Data cleaning steps and assumptions made during analysis.

This section outlines the data cleaning steps and assumptions made during the analysis. For detailed steps and assumptions, refer to the "steps_risks_and_assumptions" tab in the marketing_clean_tableau_1 Excel file and/or the table inserted below.

Summary of Data Cleaning Steps:

1. Data Import and Documentation:

- Imported marketing_data.csv and ad_data.csv into Excel using the "Get Data from CSV" option.
- Created a sheet to document cleaning and analysis steps and assumptions.
- Adhered to the Data Management Association's (DAMA) Data Management Body of Knowledge (DMBOK) principles: data accuracy, completeness, validity, consistency, uniqueness, and timeliness.

2. Duplicate and Missing Data Handling:

- Initially found no missing data or duplicates, but manual checks identified potential duplicates with similar attributes except for customer ID and/or country.
- Created additional columns to highlight potential duplicates, which require further stakeholder clarification.

3. Date and Currency Standardization:

- Changed US date format (MM/DD/YYYY) to UK format (DD/MM/YYYY) for stakeholder understanding.
- Kept currency in dollars but removed the currency sign for numerical calculations in pgAdmin and Tableau.

4. Data Normalization and Name Consistency:

- Adjusted category names for consistency:
 - Education: "2n Cycle" to "Second-cycle," "Graduation" to "Undergraduate," "Master" to "Masters."
 - Marital Status: "YOLO, absurd, Alone" to "Single," "Together" to "In a relationship," "widow" to "Widowed."
- Applied name consistency across the dataset.

5. Outlier Identification and Exclusion:

- Excluded customers born before 1920 (age > 104).
- Excluded income outliers above \$117,820, as identified by the IQR method, from visualizations and calculations to prevent skewing results.
- Analysed sales insights in pgAdmin, excluding outliers from product types, this showed that results are not impacted by the outlier values. Consequently, included total values in visualizations in Tableau for key insights as insights are unaffected by outliers for product types.

Assumptions Made:

- **Unique IDs:** Assumed all customers had unique IDs, with no duplicates.



- **Manual Data Inspection:** Rows with identical attributes except for ID and country were assumed to be system-generated rather than manually entered, hence included in visualizations. Duplicate filters are available in the Tableau dashboard for further exclusion.
- **Salary Ranges:** Assumed salaries below \$20,000 were annual, not monthly entries.
- **Customer Age:** Displayed customer ages as of 2024, though actual ages at the time of purchase (2012-2014) differ by 10-12 years. No stakeholder preference available to displaying current ages.
- **Purchase Counts:** Assumed both web and in-store purchases could not be zero; in such cases, in-store purchase counts were adjusted to 1.

By following these steps, it was ensured the dataset was clean, consistent, and ready for analysis, while making informed assumptions to address data ambiguities.

Data Quality Category	Looking for	Solution	Method / Shortcut	Dataset cleaned using the method/shortcut	Assumptions made on data
Data Accuracy	Spelling errors	Spell check	Review > Spell Check	No spelling errors	
	Values out of range	Set rule for desired range.	Data > Data Validation		Year of birth to be assumed above 1920 as age today would be over 104, 3 outliers removed from dataset, dates below this in data would be too old to be valid
	Anomalies or outliers	Identify and review outliers, IQR=Q3-Q1, Lower limit=Q1-1.5*IQR, Upper limit=Q3+1.5*IQR	Conditional Formatting > Highlight Cells Rules > More Rules or Box plots Select Text to Columns > Finish TRIM() Find & Replace to remove characters e.g., \$ as text	No blank cells, blank spaces removed from text columns using TRIM In all amounts spent 0 changed to N/A as it will distort calculations, using aggregate function on these fields	Salaries > 117820 removed as outliers, decided not to analyse outliers for other columns at this stage An assumption was made that salaries below 20K were in fact annual salaries and not monthly income mistaken for annual income
	Incorrect or invalid data types	Multiple methods available		data types were correct	
	Blank cells or spaces	Trim	=TRIM() formula		
	Incorrect use of nulls	Replace with #N/A	F5 > Special > Blanks > Type > #N/A		
	Incorrect calculations	Test calculations	Run test calculations in a spare column	No Errors in raw data	
	Mistypes and other format errors	Find & Replace	Find & Replace	Education - 2n Cycle replaced with Second-cycle, Graduation replaced with Undergraduate, Master - replaced with Masters	
	Inconsistent units of measurement	Standardize units	Find & Replace or Formula to convert units	Income changed to \$ currency format	
	Unnecessary data	Remove irrelevant data	Filter and delete unnecessary columns/rows	recency - irrelevant data as the number of days since last purchase would depend when this data was queried and it does not appear to be uniform between the year of purchase and the data appears to be over 10 years old	
Data Completeness	Partial or incomplete data	Review and fill missing data	Manual review	n/a	
	Incorrect or invalid calculations	Test calculations	Run test calculations in a spare column	no calculations in raw data	
	Missing values	Replace missing values	Find & Replace "" with "NA"	no missing values	
	Redundant columns	Remove redundant columns	Manually review and delete	recency column	
	Aggregated data	Break down aggregated data into meaningful parts	Text to Columns	data was imported correctly, no aggregate data	
Data Validity	Entries with the same spelling but in different case	Change text to lower case, upper case, or proper case	=LOWER(text), =UPPER(text), =PROPER(text)	Manually checked - no issues format left in PROPER text and Phd left as it is	
	Entries with different spelling	Find & Replace	Column filter > Find & Replace	No issues	
	Different words but with the same meaning	Find & Replace	Find & Replace	Marital status - "VOLO", "absurd", "Alone" changed to Single, "Together" changed to "in a relationship", "widow" changed to "Widowed"	
	Words with alternate representations	Find & Replace	Find & Replace	n/a	
	Highlight errors	Conditional Formatting	Home > Conditional Formatting > New Rule > Format only cells that contain > Errors > Format > Fill Tab	No Errors in raw data	
	Delete errors	F5 Special	F5 > Special > Formulas > Deselect all except errors > OK. Delete or replace errors		
	Remove duplicates	Multiple options available	Home > Conditional Formatting > Highlight Cell Rules > Duplicate Values Filter to spot duplicate identifiers	There were data that appeared the same, but only differed in country code or ID, I did not treat this data as duplicate. Similar data that was only different by customer ID or country was highlighted in the duplicate columns and using colour.	Similar data from different countries and only different in ID were treated as different, assuming that data streams were coming in from different locations. In reality would have further asked stakeholders in relation to this, if data was manually entered in one location, they could have been potentially duplicate entries.
	Data integrity constraints	Apply constraints to maintain data integrity	Data Validation rules		
	Incorrect formats for dates/numbers	Standardize date and number formats	Format Cells	date was set as US format month/day/year, used =Textsplit() and put the data back using =Date to be in day/month/year format and in 20* not 19*	
	Cross-field validation	Validate fields in relation to each other	Formulas to check cross-field consistency	Assumed that each customer has spent and amount. Checked subtotal of Amt_spent, there is no blank transaction string identified	
Data Uniqueness	Duplicate data	Visually inspect with condition formatting + remove duplicate data	check system generated duplicates against raw data	See above	See above
	Inconsistent coding schemes	Standardize coding schemes	Find & Replace or Use a reference table		
	Data normalization	Normalize data for consistency	Transform data to a consistent format		
Data Timeliness	Dates being correct	Format dates	Format > Format Cells > Date or Custom	see above	
	Columns related to a point in time	Date relevance	List of date-related columns		
	Historical data relevance	Archive outdated data	Move old data to separate archive files	n/a	Potential ethical issues under GDPR
	Time-series data accuracy	Verify and correct time-series data	Graphical analysis and trend checking	n/a	
Data Consistency	Precision (e.g., same number of decimal places)	Change the number of decimal places	Highlight column > Format Cells or Decimal Place Button	all income was a whole number so changed to 0 decimals, all Amt * spent was also whole number so used 0 decimals	
	Structure of data (e.g., IP addresses)	Correct the structure	Find & Replace	n/a	
	Case sensitivity	Change text to lower case, upper case, or proper case	=LOWER(text), =UPPER(text), =PROPER(text)	see above	
	Data type (e.g., text vs numbers)	Check data types	Copy column headers > =TYPE() (2 = TEXT, 1 = NUMBER)	checked using =TYPE, only date columns needed adjusting	
	Convert numbers stored as text into numbers	Text to Columns	Select Text to Columns in the Data tab > Finish	n/a	
	Change to decimals from full stops	Multiple options available	Find & Replace "", with "." Format Cells > Custom format > Specify number format	not needed to change	
	Consistent abbreviations	Standardize abbreviations	Find & Replace	abbreviations were the same as per metadata	
	Consistent text formats	Enforce consistent text formats	LOWER(), UPPER(), PROPER() functions		
General Practices	Documenting data cleaning steps	Maintain a data cleaning log	Document all changes in a separate log file or worksheet	yes	
	Regular audits and reviews	Schedule periodic data audits	Set reminders for regular checks	Not applied to dataset	
	Use of macros for repetitive tasks	Automate repetitive cleaning tasks	Record and run macros	Not applied to dataset	
	Error handling procedures	Implement procedures to handle errors	Error handling protocols and documentation	Not applied to dataset	
	Back up original data before cleaning	Always back up data	Create a copy of the original dataset	Not applied to dataset	
	Training and development	Regularly train team on data cleaning best practices	Conduct training sessions and share resources	Not applied to dataset	
Advanced Techniques	Using Power Query for data transformation	Leverage Power Query for complex data transformations	Data > Get & Transform > Power Query Editor	Not applied to dataset	
	Integrating external validation tools	Use third-party tools for additional validation	Integrate with tools like OpenRefine or Trifacta	Not applied to dataset	
	Creating data cleaning pipelines	Develop automated pipelines for data cleaning	Use ETL (Extract, Transform, Load) tools	Not applied to dataset	
	Analysis steps				
	Customers age was determined as their age in the year of analysis in 2024 using =YEAR(NOW())-[@Year_Birth]. I also analysed the data to show insights using the customers age when they came into contact with 2Market using year of contact-year of birth. In real life would have gone back to stakeholders and asked, which age is more useful to them as the data is over 10 years old and demographics may have changed since then.				
	statistical analysis performed see in statistical_analysis tab				
	IQR, Q1 and Q3 calculated for age, year, and income, outliers highlighted for further analysis				
	Pivot analysis of stakeholder questions				
	Data was further transferred into Tableau and pgAdmin for further analysis - 0 customer id was changed 11192 in pgAdmin, the header was adjusted for every column to be in lower case for easier analysis.				
Ethical considerations:					
As the data presented is over 10 years old there could be ethical implications for its use as they not necessarily confirm to the current GDPR rules. This would be a stakeholder question to explore, weather consent was gained from customers to use this historical dataset.					

APPENDIX 4 – Data dictionary

This can also be found in the marketing_clean_tableau_1 excel file under the data_dictionary tab

Attribute name	Interpretation of column	Data type	Possible values	Mandatory	Changes made
ID	Unique customer ID	Number	>=0	Yes	for analysis in pgAdmin 0 primary key was not an option. In SQL 0 ID was changed to 11192
Year_Birth	Customer's year of birth	Number	>= 1920 or too old to be valid	Yes	
Education	Educational qualification of the customer	Text	"Basic", "Masters", "Second-cycle", "Undergraduate", "PhD"	Yes	Education - 2n Cycle replaced with Second-cycle, Graduation replaced with Undergraduate, Master - replaced with Masters
Marital_Status	Customer's marital status	Text	"Divorced", "In a relationship", "Married", "Single", "Widowed"	Yes	Marital status-"YOLO", "absurd", "Alone" changed to Single, "Together" changed to "In a relationship", "widow" changed to "Widowed"
Income	Customer's annual income	Accounting \$	> 0	Yes	for analysis in pgAdmin it was reverted back to number. Salaries >117820 excluded as outliers,
Kidhome	Number of kids the customer has	Number	>=0	Yes	
Teenhome	Number of teenagers the customer has	Number	>=0	Yes	
Dt_Customer	Date of customer's registration with the company	Date	original data is in month/day/year format	Yes	an additional column was created called Date d/m/y to display date in day/month/year format
Recency	Number of days since customer's last purchase	Number	>=0	Yes	Irrelevant data
AmtLiq	Amount spent on alcoholic beverages	Number	>=0. 0, if none spent	Yes	
AmtVege	Amount spent on vegetables	Number	>=0. 0, if none spent	Yes	
AmtNonVeg	Amount spent on meat items	Number	>=0. 0, if none spent	Yes	
AmtPes	Amount spent on fish products	Number	>=0. 0, if none spent	Yes	
AmtChocolates	Amount spent on chocolates	Number	>=0. 0, if none spent	Yes	
AmtComm	Amount spent on commodities	Number	>=0. 0, if none spent	Yes	
NumDeals	Number of deals purchased made with a discount	Number	>=0. 0, if no deals	Yes	
NumWebBuy	Number of purchases made from the website	Number	>=0. 0, if no purchases	Yes	
NumWalkinPur	Number of in-store purchases	Number	>=0. 0, if no purchases	Yes	Both in-store purchase and website purchase cannot be 0, in this case, in-store purchase was changed to 1
NumVisits	Number of website visits per month	Number	>=0. 0, if no visit	Yes	
Response	Boolean. If the customer had accepted the last campaign's offer (1) or not (0)	Number	"0" or "1"	Yes	
Complain	Boolean. If the customer had complained in the last 2 years (1) or not (0)	Number	"0" or "1"	Yes	
Country	Customer's location:	Text	"AUS", "CA", "GER", "IND", "ME", "SA", "SP", "US"	Yes	
	AUS - Australia				
	CA - Canada				
	GER - Germany				
	IND - India				
	ME - Montenegro				
	SA - South Africa				
	SP - Spain				
	US - United States of America				
Count_success	Total number of successful lead conversions	Number	>=0. 0, if no conversion	Yes	
Calculated columns					
Age in 2024	Customers current age in the year of analysis, calculated as 2024-Year of date d/m/y	Number	< 104	No	
Date d/m/y	display date in day/month/year format	Date		No	
Duplicate_all	Boolean. If the data was same for every row except ID, (0) if unique, (1) if same	Number	"0" or "1"	No	
Duplicate_country	Boolean. If the data was same for every row except ID and country, (0) if unique, (1) if same	Number	"0" or "1"	No	
Sales total	calculated as a sum of AmtLiq, AmtVege, AmtNonVege, AmtPes, AmtChocolates and AmtComm	Accounting \$	>0 (there was no total sales =0 in dataset)	No	
Age at purchase	Customers current age in the year of analysis, calculated as Year of date d/m/y-Year of birth	Number	<104	No	

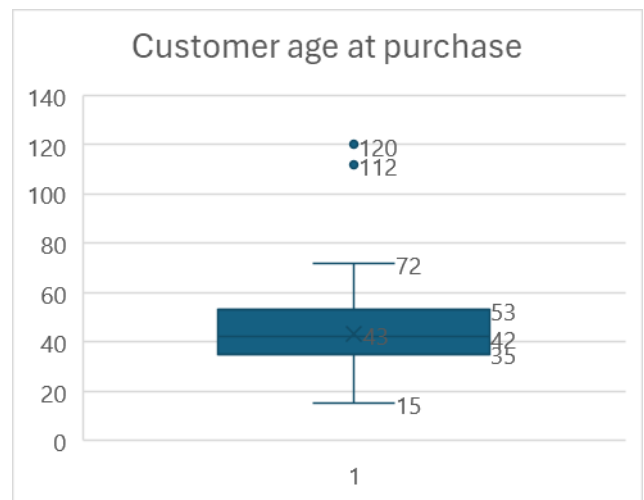
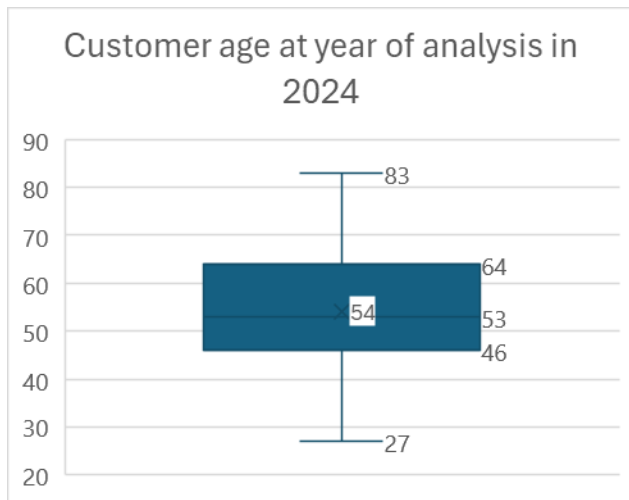
APPENDIX 5 – Initial insights from MsExcel exploratory analysis

Initial customer demographic insights:

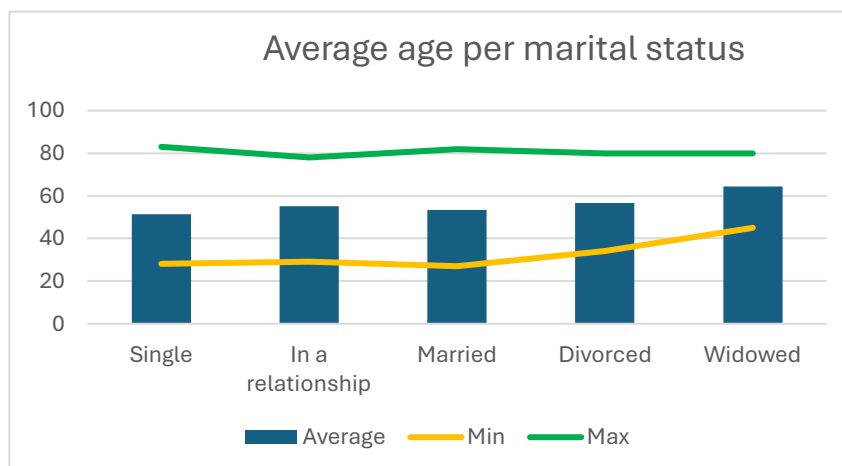
From the initial analysis the following insights were gathered in relation to the age, salary and marital status of the customers.

Age insights:

The customers age was calculated and it ranged from 27 years to 83 years, with an average age of 54 years (The invalid ages were excluded). 50% of the customers were between 35 and 53 years, at the time of purchase, this would be different and approximately 10 years lower. See charts below. A further question would have been important for stakeholders on what age they would have liked to focus, given that the data appears to be over 10 years old and demographics would have changed since data collection. However, for the purposes of this analysis, it was assumed that the stakeholders wanted to focus on their customers as they are in current day and further insights were based on this.

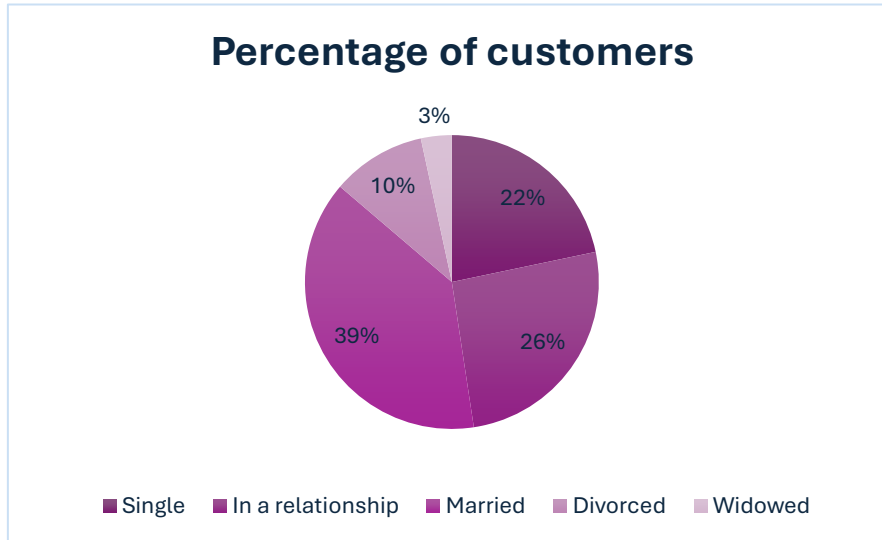


The following chart highlights the average of the customers per marital status. Which shows the highest average age was amongst the widowed group and the lowest in the single group. 64 and 51 respectively. It also shows that the minimum age increases in the divorced and widowed groups.



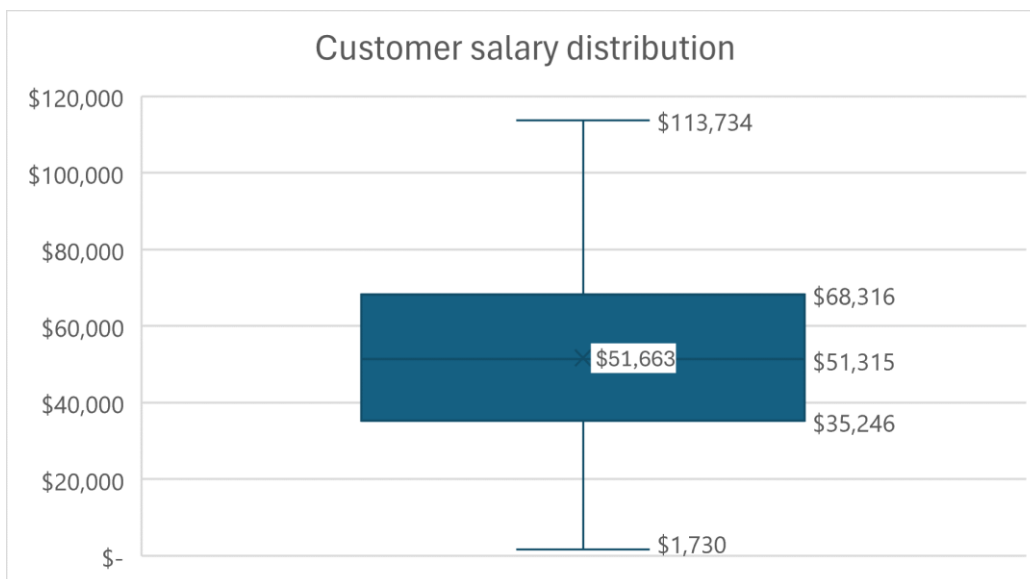
Marital status insights

The following charts shows further insights in relation to the marital status: Which shows that 39% of customers were married and 26% in a relationship.

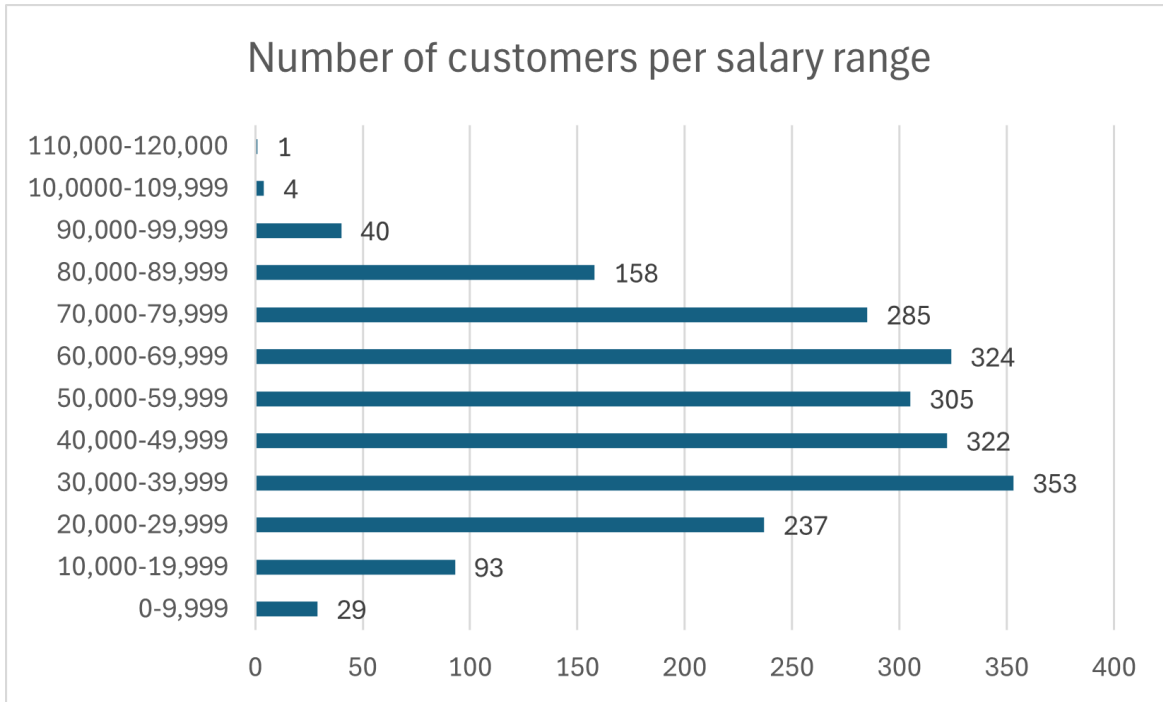


Income:

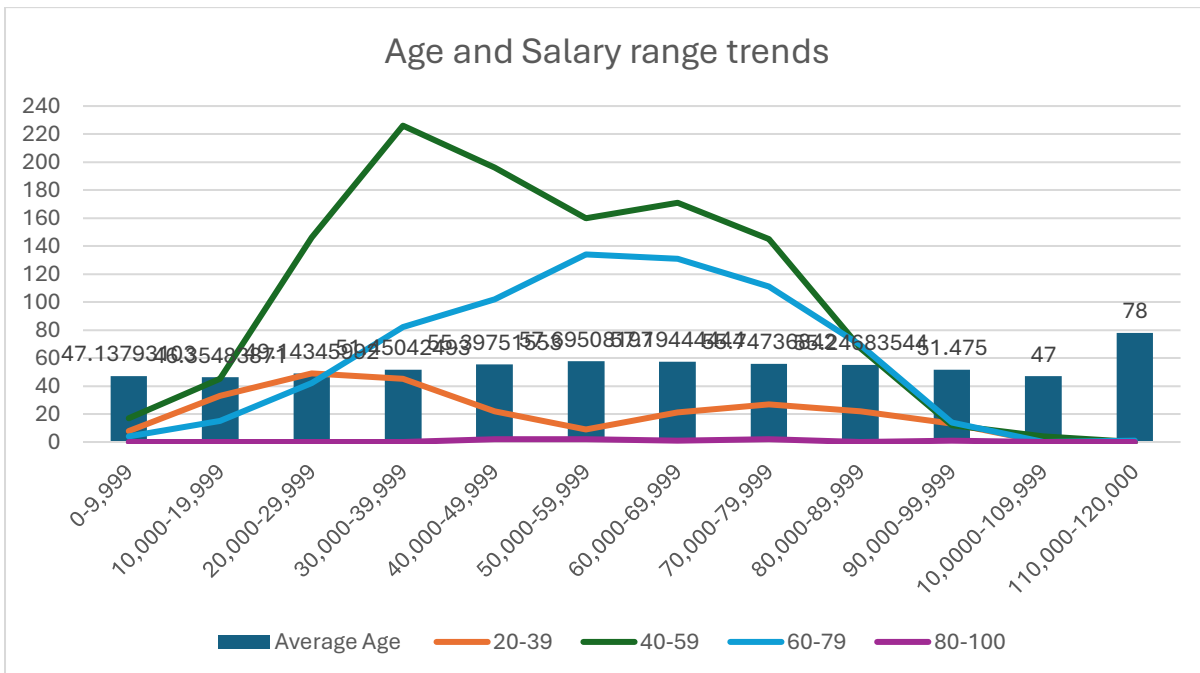
Next I focused on the insights from the income of customers. (Salaries above \$ 117,820, were removed as outliers as per the IQR upper limit, with the highest original data being 666,666). The analysed data for salary ranged between \$1,730 and \$113,734 with the average salary of \$51,663. In summary 50% of customers income was between \$35,246 and \$68,316.



The following chart highlights the number of customers in the different income categories.

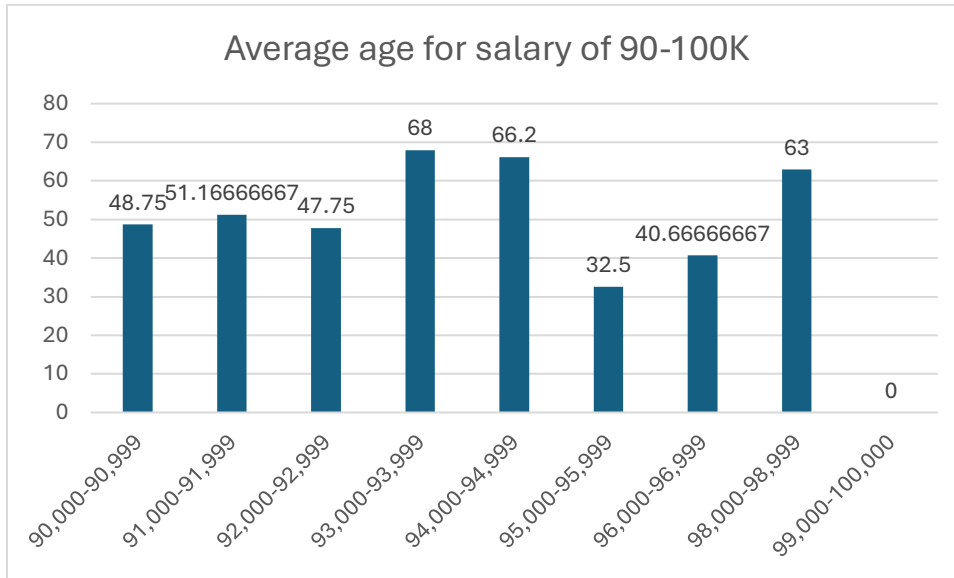


The highest number of customers has a salary of £30-39.9K.



In terms of the average age the lowest average age was in incomes between \$10-19.9K, and the highest in income categories of \$110-120K, 46 and 78 respectively.

Next we examined the income bracket of \$90-100K. The following chart shows the average age within this group. Which interestingly shows a variation of average age between 33 and 68, with no values in 99-100K.

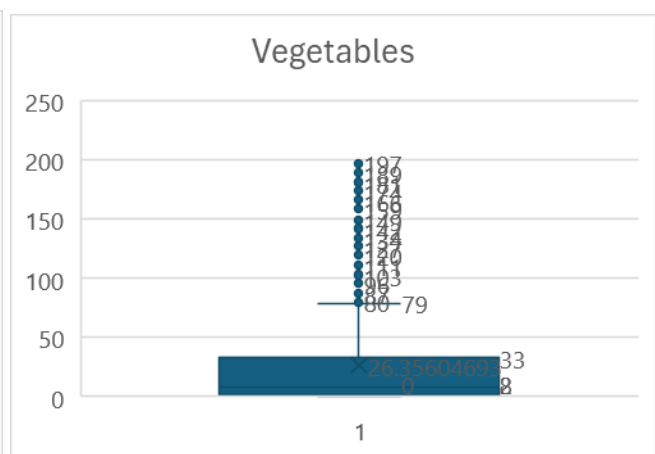
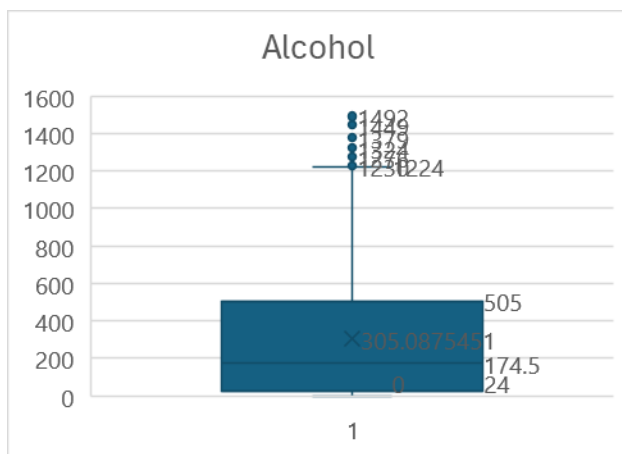


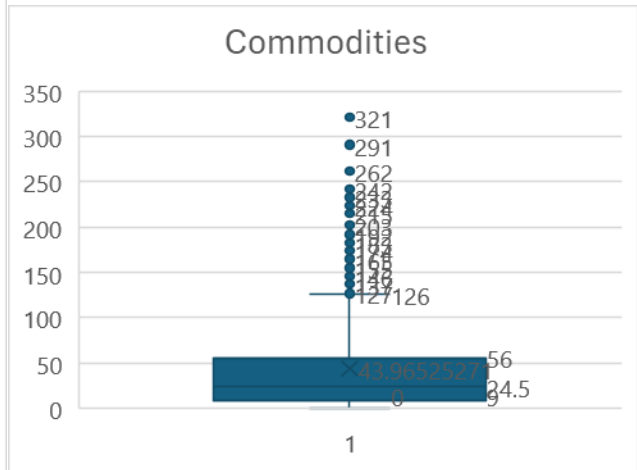
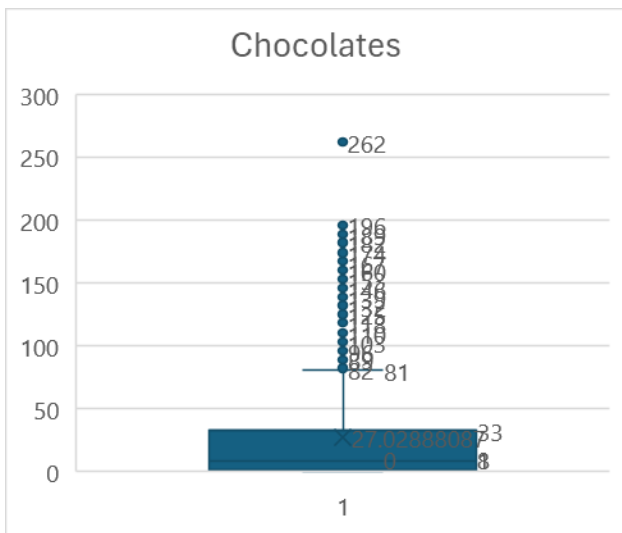
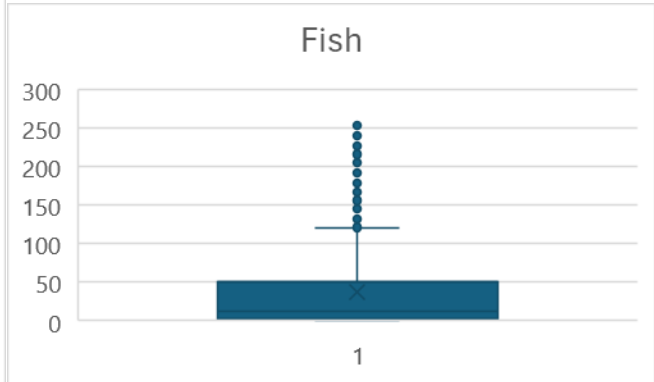
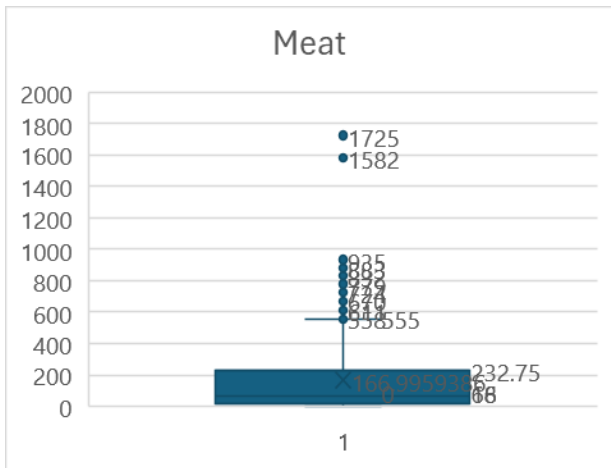
Sales insights

Total sales was calculates as the sum of all 6 products. (No 0 total sales were found, therefore, every customer contact resulted at least one product purchase)

The descriptive statistical analysis was performed on the different product types.

The box plots show that for each product category there would be outlier values.

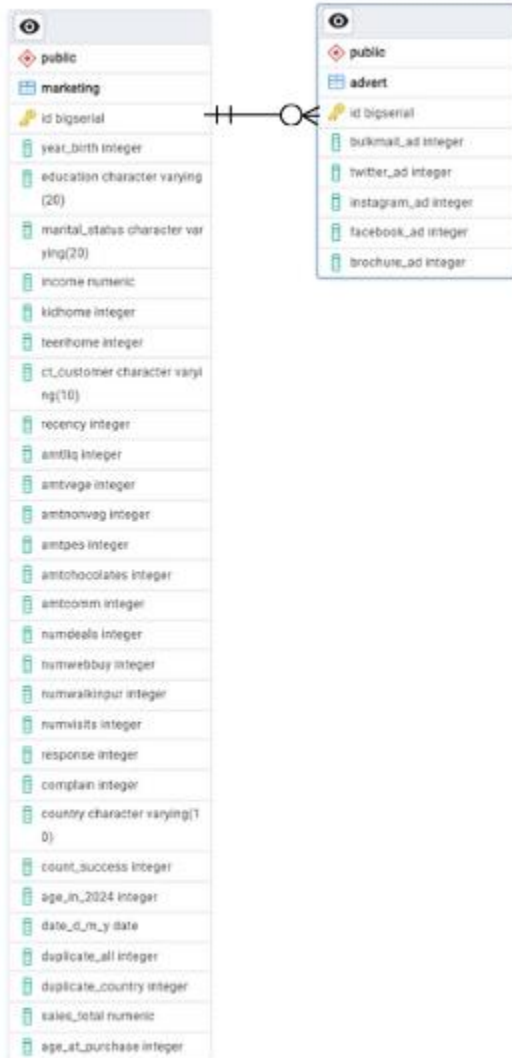




Following this exploratory analysis, the data was further explored using pgAdmin and Tableau.

APPENDIX 6 – pgAdmin SQL code syntax

pgAdmin 2Market ERD



There could have been more than one advert campaign, so the same customer ID could have been represented more than once in the ad_data, however this was not the case, and there was no duplicate customer ID there.

-----Create the tables----- (data was imported from ad_data and marketing_data csv files)-----

```
CREATE TABLE IF NOT EXISTS public.marketing
(
  id bigserial NOT NULL,
  year_birth integer,
  education character varying(20),
  marital_status character varying(20),
  income numeric,
  kidhome integer,
  teenhome integer,
  ct_customer character varying(10),
  recency integer,
  amtliq integer,
  amtvege integer,
  amtnonveg integer,
  amtpes integer,
  amtchocolates integer,
  amtcomm integer,
  numdeals integer,
  numwebbuy integer,
  numwalkinpur integer,
  numvisits integer,
  response integer,
  complain integer,
  country character varying(10),
  count_success integer,
  age_in_2024 integer,
  date_d_m_y date,
  duplicate_all integer,
  duplicate_country integer,
  sales_total numeric,
  age_at_purchase integer,
  CONSTRAINT marketing_pkey PRIMARY KEY (id)
)
```

TABLESPACE pg_default;

SELECT * FROM public.advert

```
CREATE TABLE IF NOT EXISTS public.advert
(
  id bigserial NOT NULL,
  bulkmail_ad integer,
  twitter_ad integer,
  instagram_ad integer,
  facebook_ad integer,
  brochure_ad integer)

```

-----To answer the question - What is the total spend per country?-----

```
CREATE VIEW total_spend_per_country AS
SELECT
  country,
  SUM(amtliq + amtvege + amtnonveg + amtpes + amtchocolates + amtcomm) AS total_spend
FROM
  public.marketing
GROUP BY
  country
ORDER BY
  total_spend DESC;
```

SELECT * FROM total_spend_per_country;

-----To answer the question - What is the total spend per country? when accounting for outliers-----

```
CREATE VIEW total_spend_per_country_excluding_outliers AS (
-- Calculate the IQR and filter outliers for each product type
WITH iqr_filtered AS (
-- Liquor
SELECT
  country,
```



```

    amtliq AS amtliq_filtered,
    NULL::numeric AS amtvege_filtered,
    NULL::numeric AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    NULL::numeric AS amtcomm_filtered
FROM
    public.marketing
WHERE amtliq BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtliq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtliq) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtliq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq))
    FROM public.marketing
)
)

UNION ALL

-- Vegetables
SELECT
    country,
    NULL::numeric AS amtliq_filtered,
    amtvege AS amtvege_filtered,
    NULL::numeric AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    NULL::numeric AS amtcomm_filtered
FROM
    public.marketing
WHERE amtvege BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtvege) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
    FROM public.marketing
)
)

UNION ALL

-- Meat
SELECT
    country,
    NULL::numeric AS amtliq_filtered,
    NULL::numeric AS amtvege_filtered,
    amtnonveg AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    NULL::numeric AS amtcomm_filtered
FROM
    public.marketing
WHERE amtnonveg BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtnonveg) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
    FROM public.marketing
)
)

UNION ALL

-- Fish
SELECT
    country,
    NULL::numeric AS amtliq_filtered,
    NULL::numeric AS amtvege_filtered,
    NULL::numeric AS amtnonveg_filtered,
    amtpes AS amtpes_filtered,

```

```

NULL::numeric AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
  public.marketing
WHERE amtpes BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtpes) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
  FROM public.marketing
)
)

UNION ALL

-- Chocolates
SELECT
  country,
  NULL::numeric AS amtliq_filtered,
  NULL::numeric AS amtvege_filtered,
  NULL::numeric AS amtnonveg_filtered,
  NULL::numeric AS amtpes_filtered,
  amtchocolates AS amtchocolates_filtered,
  NULL::numeric AS amtcomm_filtered
FROM
  public.marketing
WHERE amtchocolates BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtchocolates) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
  FROM public.marketing
)
)

UNION ALL

-- Commodities
SELECT
  country,
  NULL::numeric AS amtliq_filtered,
  NULL::numeric AS amtvege_filtered,
  NULL::numeric AS amtnonveg_filtered,
  NULL::numeric AS amtpes_filtered,
  NULL::numeric AS amtchocolates_filtered,
  amtcomm AS amtcomm_filtered
FROM
  public.marketing
WHERE amtcomm BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtcomm) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
  FROM public.marketing
)
)

-- Calculate total spend per country from the filtered data
SELECT
  country,
  SUM(COALESCE(amtliq_filtered, 0) + COALESCE(amtvege_filtered, 0) + COALESCE(amtnonveg_filtered, 0) +
COALESCE(amtpes_filtered, 0) + COALESCE(amtchocolates_filtered, 0) + COALESCE(amtcomm_filtered, 0)) AS total_spend
FROM
  iqr_filtered
GROUP BY
  country
ORDER BY
  total_spend DESC);
-- View the result

```

```
SELECT * FROM total_spend_per_country_excluding_outliers;
```

-----To answer the question - What is the total spend per product per country?-----

```
CREATE VIEW total_spend_per_country_product_type AS
SELECT
  country,
  'liquor' AS product_type, SUM(amtliq) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'vegetables' AS product_type, SUM(amtvege) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'meat' AS product_type, SUM(amtnonveg) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'fish' AS product_type, SUM(amtptes) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'chocolates' AS product_type, SUM(amtchocolates) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'commodities' AS product_type, SUM(amtcomm) AS total_spend
FROM
  public.marketing
GROUP BY
  country
ORDER BY
  country, total_spend DESC;
```

```
SELECT * FROM total_spend_per_country_product_type
```

-----accounting for outliers---What is the total spend per country per product type?-----

```
CREATE VIEW total_spend_per_country_product_type_without_outliers AS (
-- Calculate the IQR and filter outliers
WITH iqr_filtered AS (
  -- Liquor---
  SELECT
    country, amtliq AS amount, 'liquor' AS product_type
  FROM
    public.marketing
  WHERE amtliq BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
    (ORDER BY amtliq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq))
    FROM public.marketing
  ) AND (
```

```

SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtliq) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtliq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq))
FROM public.marketing
)

UNION ALL

-- Vegetables---
SELECT
country, amtvege AS amount, 'vegetables' AS product_type
FROM
public.marketing
WHERE amtvege BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtvege) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
FROM public.marketing
)

UNION ALL

-- Meat---
SELECT
country, amtnonveg AS amount, 'meat' AS product_type
FROM
public.marketing
WHERE amtnonveg BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtnonveg) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
FROM public.marketing
)

UNION ALL

-- Fish--
SELECT
country, amtpes AS amount, 'fish' AS product_type
FROM
public.marketing
WHERE amtpes BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtpes) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
FROM public.marketing
)

UNION ALL

-- Chocolates---
SELECT
country, amtchocolates AS amount, 'chocolates' AS product_type
FROM
public.marketing
WHERE amtchocolates BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtchocolates) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
FROM public.marketing
)

```

```

UNION ALL

-- Commodities
SELECT
  country, amtcomm AS amount, 'commodities' AS product_type
FROM
  public.marketing
WHERE amtcomm BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
  (ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtcomm) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
  GROUP (ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
  FROM public.marketing
)
)

-- Calculate total spend per country and product type----
SELECT
  country,
  product_type,
  SUM(amount) AS total_spend
FROM
  iqr_filtered
GROUP BY
  country, product_type
ORDER BY
  country, total_spend DESC);

-- View the result----

SELECT * FROM total_spend_per_country_product_type_without_outliers;

-----To answer the question - Which products are the most popular in each country? ---- REMOVING THE OUTLIERS WOULD NOT
CHANGE THE OVERALL RESULTS ----

CREATE VIEW most_popular_product_type_per_country AS
WITH product_spend AS (
SELECT
  country,
  'liquor' AS product_type, SUM(amtliq) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'vegetables' AS product_type, SUM(amtvege) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'meat' AS product_type, SUM(amtnonveg) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'fish' AS product_type, SUM(amtpes) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,

```

```

'chocolates' AS product_type, SUM(amtchocolates) AS total_spend
FROM
  public.marketing
GROUP BY
  country
UNION ALL
SELECT
  country,
  'commodities' AS product_type, SUM(amtcomm) AS total_spend
FROM
  public.marketing
GROUP BY
  country
)
SELECT
  country,
  product_type,
  total_spend
FROM (
  SELECT
    country,
    product_type,
    total_spend,
    ROW_NUMBER() OVER (PARTITION BY country ORDER BY total_spend DESC) AS rn
  FROM
    product_spend
) AS ranked
WHERE
  rn = 1
ORDER BY
  country;

```

SELECT * FROM most_popular_product_type_per_country

-----Which products are the most popular based on marital status?-----

```

CREATE VIEW most_popular_product_type_per_marital_status AS
WITH product_spend AS (
SELECT
  marital_status,
  'liquor' AS product_type, SUM(amtliq) AS total_spend
FROM
  public.marketing
GROUP BY
  marketing.marital_status
UNION ALL
SELECT
  marital_status,
  'vegetables' AS product_type, SUM(amtvege) AS total_spend
FROM
  public.marketing
GROUP BY
  marketing.marital_status
UNION ALL
SELECT
  marital_status,
  'meat' AS product_type, SUM(amtnonveg) AS total_spend
FROM
  public.marketing
GROUP BY
  marketing.marital_status
UNION ALL
SELECT
  marital_status,
  'fish' AS product_type, SUM(amtpes) AS total_spend
FROM
  public.marketing
GROUP BY
  marketing.marital_status
UNION ALL
SELECT
  marital_status,
  'chocolates' AS product_type, SUM(amtchocolates) AS total_spend

```

```

FROM
  public.marketing
GROUP BY
  marketing.marital_status
UNION ALL
SELECT
  marital_status,
  'commodities' AS product_type, SUM(amtcomm) AS total_spend
FROM
  public.marketing
GROUP BY
  marketing.marital_status
)
SELECT
  marital_status,
  product_type,
  total_spend
FROM (
  SELECT
    marital_status,
    product_type,
    total_spend,
    ROW_NUMBER() OVER (PARTITION BY marital_status ORDER BY total_spend DESC) AS rn
  FROM
    product_spend
) AS ranked
WHERE
  rn = 1
ORDER BY
  marital_status;

```

```
SELECT * FROM most_popular_product_type_per_marital_status;
```

-----Which products are the most popular based on marital status? excluding the outliers-----

```

CREATE VIEW most_popular_product_type_per_marital_status_excluding_outliers AS (
-- Calculate the IQR and filter outliers for each product type
WITH iqr_filtered AS (
  -- Liquor
  SELECT
    marital_status,
    amtliq AS amtliq_filtered,
    NULL::numeric AS amtvege_filtered,
    NULL::numeric AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    NULL::numeric AS amtcomm_filtered
  FROM
    public.marketing
  WHERE amtliq BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtliq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq))
    FROM public.marketing
  ) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtliq) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtliq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtliq))
    FROM public.marketing
  )
)
UNION ALL
-- Vegetables
SELECT
  marital_status,
  NULL::numeric AS amtliq_filtered,
  amtvege AS amtvege_filtered,
  NULL::numeric AS amtnonveg_filtered,
  NULL::numeric AS amtpes_filtered,
  NULL::numeric AS amtchocolates_filtered,
  NULL::numeric AS amtcomm_filtered
FROM
  public.marketing
WHERE amtvege BETWEEN (

```

```

SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtvege) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
FROM public.marketing
)

UNION ALL

-- Meat
SELECT
marital_status,
NULL::numeric AS amtliq_filtered,
NULL::numeric AS amtvege_filtered,
amtnonveg AS amtnonveg_filtered,
NULL::numeric AS amtpes_filtered,
NULL::numeric AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
public.marketing
WHERE amtnonveg BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtnonveg) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
FROM public.marketing
)

UNION ALL

-- Fish
SELECT
marital_status,
NULL::numeric AS amtliq_filtered,
NULL::numeric AS amtvege_filtered,
NULL::numeric AS amtnonveg_filtered,
amtpes AS amtpes_filtered,
NULL::numeric AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
public.marketing
WHERE amtpes BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
FROM public.marketing
) AND (
SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtpes) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
FROM public.marketing
)

UNION ALL

-- Chocolates
SELECT
marital_status,
NULL::numeric AS amtliq_filtered,
NULL::numeric AS amtvege_filtered,
NULL::numeric AS amtnonveg_filtered,
NULL::numeric AS amtpes_filtered,
amtchocolates AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
public.marketing
WHERE amtchocolates BETWEEN (
SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
FROM public.marketing
) AND (

```



```

SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtchocolates) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
FROM public.marketing
)

UNION ALL

-- Commodities
SELECT
    marital_status,
    NULL::numeric AS amtliq_filtered,
    NULL::numeric AS amtvege_filtered,
    NULL::numeric AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    amtcomm AS amtcomm_filtered
FROM
    public.marketing
WHERE amtcomm BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtcomm) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
    FROM public.marketing
)
)

-- Calculate the total spend per product type per marital status from the filtered data
, product_spend AS (
    SELECT
        marital_status,
        'liquor' AS product_type,
        SUM(COALESCE(amtliq_filtered, 0)) AS total_spend
    FROM
        iqr_filtered
    GROUP BY
        marital_status

    UNION ALL

    SELECT
        marital_status,
        'vegetables' AS product_type,
        SUM(COALESCE(amtvege_filtered, 0)) AS total_spend
    FROM
        iqr_filtered
    GROUP BY
        marital_status

    UNION ALL

    SELECT
        marital_status,
        'meat' AS product_type,
        SUM(COALESCE(amtnonveg_filtered, 0)) AS total_spend
    FROM
        iqr_filtered
    GROUP BY
        marital_status

    UNION ALL

    SELECT
        marital_status,
        'fish' AS product_type,
        SUM(COALESCE(amtpes_filtered, 0)) AS total_spend
    FROM
        iqr_filtered
    GROUP BY
        marital_status

```

```

UNION ALL

SELECT
    marital_status,
    'chocolates' AS product_type,
    SUM(COALESCE(amtchocolates_filtered, 0)) AS total_spend
FROM
    iqr_filtered
GROUP BY
    marital_status

UNION ALL

SELECT
    marital_status,
    'commodities' AS product_type,
    SUM(COALESCE(amtcomm_filtered, 0)) AS total_spend
FROM
    iqr_filtered
GROUP BY
    marital_status
)

-- Select the most popular product type per marital status
SELECT
    marital_status,
    product_type,
    total_spend
FROM (
    SELECT
        marital_status,
        product_type,
        total_spend,
        ROW_NUMBER() OVER (PARTITION BY marital_status ORDER BY total_spend DESC) AS rn
    FROM
        product_spend
) AS ranked
WHERE
    rn = 1
ORDER BY
    marital_status);

-- View the result
SELECT * FROM most_popular_product_type_per_marital_status_excluding_outliers;

-----To answer the question - Which products are the most popular based on whether or not there are children or teens in the home?
CREATE VIEW popular_product_type_family_status AS
WITH categorized_data AS (
    SELECT
        CASE
            WHEN kidhome > 0 OR teenhome > 0 THEN 'With Children/Teens'
            ELSE 'Without Children/Teens'
        END AS home_status,
        amtliq, amtvege, amtnonveg, amtpes, amtchocolates, amtcomm
    FROM
        public.marketing
),
product_spend AS (
    SELECT
        home_status,
        'liquor' AS product_type, SUM(amtliq) AS total_spend
    FROM
        categorized_data
    GROUP BY
        home_status
    UNION ALL
    SELECT
        home_status,
        'vegetables' AS product_type, SUM(amtvege) AS total_spend
    FROM
        categorized_data
    GROUP BY

```

```

    home_status
UNION ALL
SELECT
    home_status,
    'meat' AS product_type, SUM(amtnonveg) AS total_spend
FROM
    categorized_data
GROUP BY
    home_status
UNION ALL
SELECT
    home_status,
    'fish' AS product_type, SUM(amtptes) AS total_spend
FROM
    categorized_data
GROUP BY
    home_status
UNION ALL
SELECT
    home_status,
    'chocolates' AS product_type, SUM(amtchocolates) AS total_spend
FROM
    categorized_data
GROUP BY
    home_status
UNION ALL
SELECT
    home_status,
    'commodities' AS product_type, SUM(amtcomm) AS total_spend
FROM
    categorized_data
GROUP BY
    home_status
)
SELECT
    home_status,
    product_type,
    total_spend
FROM (
    SELECT
        home_status,
        product_type,
        total_spend,
        ROW_NUMBER() OVER (PARTITION BY home_status ORDER BY total_spend DESC) AS rn
    FROM
        product_spend
) AS ranked
WHERE
    rn = 1
ORDER BY
    home_status;

```

SELECT * FROM popular_product_type_family_status;

-----To separate it children or teenagers or no kids-----

```

CREATE VIEW popular_product_type_family_status_children_or_teens AS
WITH categorized_data AS (
    SELECT
        CASE
            WHEN kidhome > 0 THEN 'With Children'
            WHEN teenhome > 0 THEN 'With Teenagers'
            ELSE 'Without Children/Teenagers'
        END AS home_status,
        amtliq, amtvege, amtnonveg, amtptes, amtchocolates, amtcomm
    FROM
        public.marketing
),
product_spend AS (
    SELECT
        home_status,
        'liquor' AS product_type, SUM(amtliq) AS total_spend

```



```
FROM
  categorized_data
GROUP BY
  home_status
UNION ALL
SELECT
  home_status,
  'vegetables' AS product_type, SUM(amtvege) AS total_spend
FROM
  categorized_data
GROUP BY
  home_status
UNION ALL
SELECT
  home_status,
  'meat' AS product_type, SUM(amtrnonveg) AS total_spend
FROM
  categorized_data
GROUP BY
  home_status
UNION ALL
SELECT
  home_status,
  'fish' AS product_type, SUM(amtpes) AS total_spend
FROM
  categorized_data
GROUP BY
  home_status
UNION ALL
SELECT
  home_status,
  'chocolates' AS product_type, SUM(amtchocolates) AS total_spend
FROM
  categorized_data
GROUP BY
  home_status
UNION ALL
SELECT
  home_status,
  'commodities' AS product_type, SUM(amtcomm) AS total_spend
FROM
  categorized_data
GROUP BY
  home_status
)
SELECT
  home_status,
  product_type,
  total_spend
FROM (
  SELECT
    home_status,
    product_type,
    total_spend,
    ROW_NUMBER() OVER (PARTITION BY home_status ORDER BY total_spend DESC) AS rn
  FROM
    product_spend
) AS ranked
WHERE
  rn = 1
ORDER BY
  home_status;
```

```
SELECT * FROM popular_product_type_family_status_children_or_teens;
```

-----To answer the question - Which products are the most popular based on whether or not there are children or teens in the home?--
excluding outliers-----

```
CREATE VIEW popular_product_type_family_status_excluding_outliers AS (
WITH iqr_filtered AS (
  SELECT
    CASE
```

```

    WHEN kidhome > 0 THEN 'With Children'
    WHEN teenhome > 0 THEN 'With Teenagers'
    ELSE 'Without Children/Teenagers'
END AS home_status,
amtlq AS amtlq_filtered,
NULL::numeric AS amtvege_filtered,
NULL::numeric AS amtnonveg_filtered,
NULL::numeric AS amtpes_filtered,
NULL::numeric AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
    public.marketing
WHERE amtlq BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtlq) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtlq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtlq))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtlq) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtlq) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtlq))
    FROM public.marketing
)

UNION ALL

-- Vegetables
SELECT
    CASE
        WHEN kidhome > 0 THEN 'With Children'
        WHEN teenhome > 0 THEN 'With Teenagers'
        ELSE 'Without Children/Teenagers'
    END AS home_status,
    NULL::numeric AS amtlq_filtered,
    amtvege AS amtvege_filtered,
    NULL::numeric AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    NULL::numeric AS amtcomm_filtered
FROM
    public.marketing
WHERE amtvege BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtvege) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtvege) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtvege))
    FROM public.marketing
)

UNION ALL

-- Meat
SELECT
    CASE
        WHEN kidhome > 0 THEN 'With Children'
        WHEN teenhome > 0 THEN 'With Teenagers'
        ELSE 'Without Children/Teenagers'
    END AS home_status,
    NULL::numeric AS amtlq_filtered,
    NULL::numeric AS amtvege_filtered,
    amtnonveg AS amtnonveg_filtered,
    NULL::numeric AS amtpes_filtered,
    NULL::numeric AS amtchocolates_filtered,
    NULL::numeric AS amtcomm_filtered
FROM
    public.marketing
WHERE amtnonveg BETWEEN (
    SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))
    FROM public.marketing
) AND (
    SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtnonveg) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtnonveg) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtnonveg))

```

```

FROM public.marketing
)

UNION ALL

-- Fish
SELECT
CASE
  WHEN kidhome > 0 THEN 'With Children'
  WHEN teenhome > 0 THEN 'With Teenagers'
  ELSE 'Without Children/Teenagers'
END AS home_status,
NULL::numeric AS amtliq_filtered,
NULL::numeric AS amtvege_filtered,
NULL::numeric AS amtnonveg_filtered,
amtpes AS amtpes_filtered,
NULL::numeric AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
public.marketing
WHERE amtpes BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(OORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtpes) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(OORDER BY amtpes) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtpes))
  FROM public.marketing
)
)

UNION ALL

-- Chocolates
SELECT
CASE
  WHEN kidhome > 0 THEN 'With Children'
  WHEN teenhome > 0 THEN 'With Teenagers'
  ELSE 'Without Children/Teenagers'
END AS home_status,
NULL::numeric AS amtliq_filtered,
NULL::numeric AS amtvege_filtered,
NULL::numeric AS amtnonveg_filtered,
NULL::numeric AS amtpes_filtered,
amtchocolates AS amtchocolates_filtered,
NULL::numeric AS amtcomm_filtered
FROM
public.marketing
WHERE amtchocolates BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtchocolates) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtchocolates) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtchocolates))
  FROM public.marketing
)
)

UNION ALL

-- Commodities
SELECT
CASE
  WHEN kidhome > 0 THEN 'With Children'
  WHEN teenhome > 0 THEN 'With Teenagers'
  ELSE 'Without Children/Teenagers'
END AS home_status,
NULL::numeric AS amtliq_filtered,
NULL::numeric AS amtvege_filtered,
NULL::numeric AS amtnonveg_filtered,
NULL::numeric AS amtpes_filtered,
NULL::numeric AS amtchocolates_filtered,
amtcomm AS amtcomm_filtered
FROM

```

```

public.marketing
WHERE amtcomm BETWEEN (
  SELECT PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm) - 1.5 * (PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
  FROM public.marketing
) AND (
  SELECT PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY amtcomm) + 1.5 * (PERCENTILE_CONT(0.75) WITHIN
GROUP (ORDER BY amtcomm) - PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY amtcomm))
  FROM public.marketing
)
)
), product_spend AS (
  SELECT
    home_status,
    'liquor' AS product_type,
    SUM(COALESCE(amtliq_filtered, 0)) AS total_spend
  FROM
    iqr_filtered
  GROUP BY
    home_status

  UNION ALL

  SELECT
    home_status,
    'vegetables' AS product_type,
    SUM(COALESCE(amtvege_filtered, 0)) AS total_spend
  FROM
    iqr_filtered
  GROUP BY
    home_status

  UNION ALL

  SELECT
    home_status,
    'meat' AS product_type,
    SUM(COALESCE(amtnonveg_filtered, 0)) AS total_spend
  FROM
    iqr_filtered
  GROUP BY
    home_status

  UNION ALL

  SELECT
    home_status,
    'fish' AS product_type,
    SUM(COALESCE(amtpeps_filtered, 0)) AS total_spend
  FROM
    iqr_filtered
  GROUP BY
    home_status

  UNION ALL

  SELECT
    home_status,
    'chocolates' AS product_type,
    SUM(COALESCE(amtchocolates_filtered, 0)) AS total_spend
  FROM
    iqr_filtered
  GROUP BY
    home_status

  UNION ALL

  SELECT
    home_status,
    'commodities' AS product_type,
    SUM(COALESCE(amtcomm_filtered, 0)) AS total_spend
  FROM

```

```

iqr_filtered
GROUP BY
  home_status
)
SELECT
  home_status,
  product_type,
  total_spend
FROM (
  SELECT
    home_status,
    product_type,
    total_spend,
    ROW_NUMBER() OVER (PARTITION BY home_status ORDER BY total_spend DESC) AS m
  FROM
    product_spend
) AS ranked
WHERE
  m = 1
ORDER BY
  home_status);

```

```
SELECT * FROM popular_product_type_family_status_excluding_outliers;
```

```
-----WEEK 5-----
```

```
SELECT * FROM public.advert;
```

```
SELECT * FROM public.marketing;
```

```
-----check duplicates in advert----
```

```
SELECT id,
COUNT (*) AS count
FROM public.advert
```

```
GROUP BY id
```

```
HAVING COUNT (*)>1
```

```
-----check duplicates in marketing----
```

```
SELECT country, income,
COUNT (*) AS count
FROM public.marketing
```

```
GROUP BY id, country, income
```

```
HAVING COUNT (*)>1
```

```
-----no duplicate id's were identified
```

```
-----Which social media platform (Twitter, Instagram, or Facebook) is the most effective method of advertising in each country?--
```

```
----- (In this case, consider the total number of lead conversions as a measure of effectiveness).-----
```

```
CREATE VIEW effective_social_media_per_country AS (
```

```
SELECT
```

```
  country,
```

```
  CASE
```

```
    WHEN SUM(twitter_ad) >= SUM(instagram_ad) AND SUM(twitter_ad) >= SUM(facebook_ad) THEN 'Twitter'
```

```
    WHEN SUM(instagram_ad) >= SUM(twitter_ad) AND SUM(instagram_ad) >= SUM(facebook_ad) THEN 'Instagram'
```

```
    ELSE 'Facebook'
```

```
  END AS most_effective_platform
```

```
FROM
```

```
  marketing m
```

```
JOIN
```

```
  advert a ON m.id = a.id
```

```
GROUP BY
```

```
  country
```

```
ORDER BY country ASC);
```

```
SELECT * FROM effective_social_media_per_country;
```

```
-----Which social media platform is the most effective method of advertising based on marital status?
```

```
----- (In this case, consider the total number of lead conversions as a measure of effectiveness).
```

```
CREATE VIEW effective_social_media_per_marital_status AS
```

```
  (SELECT
```

```
    marital_status,
```

```
    CASE
```

```
      WHEN SUM(twitter_ad) >= SUM(instagram_ad) AND SUM(twitter_ad) >= SUM(facebook_ad) THEN 'Twitter'
```

```
      WHEN SUM(instagram_ad) >= SUM(twitter_ad) AND SUM(instagram_ad) >= SUM(facebook_ad) THEN 'Instagram'
```

```
      ELSE 'Facebook'
```

```
    END AS most_effective_platform
```

```
FROM
```

```
  marketing m
```

```
JOIN
```



```

    advert a ON m.id = a.id
GROUP BY
    marital_status
ORDER BY
    CASE marital_status
        WHEN 'Single' THEN 1
        WHEN 'In a relationship' THEN 2
        WHEN 'Married' THEN 3
        WHEN 'Divorced' THEN 4
        WHEN 'Widowed' THEN 5
        ELSE 6 -- Handle any other cases not explicitly ordered
    END);
SELECT * FROM effective_social_media_per_marital_status;
-----Which social media platform(s) seem to be the most effective per country? -----
-----In this case, assume that purchases were in some way influenced by lead conversions from a campaign)-----
CREATE VIEW campaign_take_on_per_country AS
    (SELECT
        country,
        CASE
            WHEN SUM(CASE WHEN twitter_ad = 1 THEN response ELSE 0 END) >= SUM(CASE WHEN instagram_ad = 1 THEN response
            ELSE 0 END)
            AND SUM(CASE WHEN twitter_ad = 1 THEN response ELSE 0 END) >= SUM(CASE WHEN facebook_ad = 1 THEN response
            ELSE 0 END)
            THEN 'Twitter'
            WHEN SUM(CASE WHEN instagram_ad = 1 THEN response ELSE 0 END) >= SUM(CASE WHEN twitter_ad = 1 THEN response
            ELSE 0 END)
            AND SUM(CASE WHEN instagram_ad = 1 THEN response ELSE 0 END) >= SUM(CASE WHEN facebook_ad = 1 THEN
            response ELSE 0 END)
            THEN 'Instagram'
            ELSE 'Facebook'
        END AS most_effective_platform
    FROM
        marketing m
    JOIN
        advert a ON m.id = a.id
    GROUP BY
        country
    ORDER BY country ASC);
SELECT * FROM campaign_take_on_per_country;
-----Compare if campaign changed the platform effectiveness-----
SELECT s.country, s.most_effective_platform AS no_campaign, c.most_effective_platform AS campaign_take_on
FROM effective_social_media_per_country s
JOIN campaign_take_on_per_country c ON s.country=c.country
-----Considering product_types and sales-----
CREATE VIEW effective_product_type_social_platform AS (
    WITH total_spent AS (
        SELECT
            m.country,
            a.twitter_ad,
            a.instagram_ad,
            a.facebook_ad,
            SUM(CASE WHEN m.amtlq > 0 THEN m.amtlq ELSE 0 END) AS total_amtliq,
            SUM(CASE WHEN m.amtvege > 0 THEN m.amtvege ELSE 0 END) AS total_amtvege,
            SUM(CASE WHEN m.amtnonveg > 0 THEN m.amtnonveg ELSE 0 END) AS total_amtnonveg,
            SUM(CASE WHEN m.amtpes > 0 THEN m.amtpes ELSE 0 END) AS total_amtpes,
            SUM(CASE WHEN m.amtchocolates > 0 THEN m.amtchocolates ELSE 0 END) AS total_amtchocolates
        FROM
            marketing m
        JOIN
            advert a ON m.id = a.id
        GROUP BY
            m.country, a.twitter_ad, a.instagram_ad, a.facebook_ad
    )
    SELECT
        country,
        CASE
            WHEN SUM(total_amtliq) >= SUM(total_amtvege) AND SUM(total_amtliq) >= SUM(total_amtnonveg)
            AND SUM(total_amtliq) >= SUM(total_amtpes) AND SUM(total_amtliq) >= SUM(total_amtchocolates) THEN 'Alcohol'
            WHEN SUM(total_amtvege) >= SUM(total_amtliq) AND SUM(total_amtvege) >= SUM(total_amtnonveg)
            AND SUM(total_amtvege) >= SUM(total_amtpes) AND SUM(total_amtvege) >= SUM(total_amtchocolates) THEN 'Vegetables'
            WHEN SUM(total_amtnonveg) >= SUM(total_amtliq) AND SUM(total_amtnonveg) >= SUM(total_amtvege)
            AND SUM(total_amtnonveg) >= SUM(total_amtpes) AND SUM(total_amtnonveg) >= SUM(total_amtchocolates) THEN 'Meat'
            WHEN SUM(total_amtpes) >= SUM(total_amtliq) AND SUM(total_amtpes) >= SUM(total_amtvege)

```

```

        AND SUM(total_amtpes) >= SUM(total_amtnonveg) AND SUM(total_amtpes) >= SUM(total_amtchocolates) THEN 'Fish'
        ELSE 'Chocolate'
    END AS most_effective_product_type,
    CASE
        WHEN SUM(twitter_ad) >= SUM(instagram_ad) AND SUM(twitter_ad) >= SUM(facebook_ad) THEN 'Twitter'
        WHEN SUM(instagram_ad) >= SUM(twitter_ad) AND SUM(instagram_ad) >= SUM(facebook_ad) THEN 'Instagram'
        ELSE 'Facebook'
    END AS most_effective_platform
FROM
    total_spent
GROUP BY
    country
ORDER BY country ASC);
SELECT * FROM effective_product_type_social_platform;
-----to answer the average revenue from each marketing channel-----

```

```
CREATE VIEW total_sales_per_channel AS
```

```
SELECT
    m.id,
    m.sales_total,
    a.bulkmail_ad,
    a.twitter_ad,
    a.instagram_ad,
    a.facebook_ad,
    a.brochure_ad
```

```
FROM
```

```
    marketing m
```

```
JOIN
```

```
    advert a ON m.id = a.id;
```

```
SELECT * FROM total_sales_per_channel;
```

```
-----step 2-----
```

```
SELECT
```

```
    'Bulkmail' AS channel,
    ROUND(AVG(sales_total), 2) AS average_sales,
    ROUND(SUM(sales_total), 2) AS total_sales,
    COUNT(CASE WHEN bulkmail_ad = 1 THEN 1 ELSE NULL END) AS successful_conversions
```

```
FROM
```

```
    total_sales_per_channel
```

```
WHERE
```

```
    bulkmail_ad = 1
```

```
UNION ALL
```

```
SELECT
```

```
    'Twitter' AS channel,
    ROUND(AVG(sales_total), 2) AS average_sales,
    ROUND(SUM(sales_total), 2) AS total_sales,
    COUNT(CASE WHEN twitter_ad = 1 THEN 1 ELSE NULL END) AS successful_conversions
```

```
FROM
```

```
    total_sales_per_channel
```

```
WHERE
```

```
    twitter_ad = 1
```

```
UNION ALL
```

```
SELECT
```

```
    'Instagram' AS channel,
    ROUND(AVG(sales_total), 2) AS average_sales,
    ROUND(SUM(sales_total), 2) AS total_sales,
    COUNT(CASE WHEN instagram_ad = 1 THEN 1 ELSE NULL END) AS successful_conversions
```

```
FROM
```

```
    total_sales_per_channel
```

```
WHERE
```

```
    instagram_ad = 1
```

```
UNION ALL
```

```
SELECT
```

```
    'Facebook' AS channel,
    ROUND(AVG(sales_total), 2) AS average_sales,
    ROUND(SUM(sales_total), 2) AS total_sales,
    COUNT(CASE WHEN facebook_ad = 1 THEN 1 ELSE NULL END) AS successful_conversions
```

```

FROM
    total_sales_per_channel
WHERE
    facebook_ad = 1

UNION ALL

SELECT
    'Brochure' AS channel,
    ROUND(AVG(sales_total), 2) AS average_sales,
    ROUND(SUM(sales_total), 2) AS total_sales,
    COUNT(CASE WHEN brochure_ad = 1 THEN 1 ELSE NULL END) AS successful_conversions
FROM
    total_sales_per_channel
WHERE
    brochure_ad = 1;
-----To find out if customers responded to more than one ad-----
SELECT
    m.id,
    m.country,
    CASE
        WHEN a.bulkmail_ad = 1 AND a.twitter_ad = 1 THEN 'Bulkmail, Twitter'
        WHEN a.bulkmail_ad = 1 AND a.instagram_ad = 1 THEN 'Bulkmail, Instagram'
        WHEN a.bulkmail_ad = 1 AND a.facebook_ad = 1 THEN 'Bulkmail, Facebook'
        WHEN a.bulkmail_ad = 1 AND a.brochure_ad = 1 THEN 'Bulkmail, Brochure'
        WHEN a.twitter_ad = 1 AND a.instagram_ad = 1 THEN 'Twitter, Instagram'
        WHEN a.twitter_ad = 1 AND a.facebook_ad = 1 THEN 'Twitter, Facebook'
        WHEN a.twitter_ad = 1 AND a.brochure_ad = 1 THEN 'Twitter, Brochure'
        WHEN a.instagram_ad = 1 AND a.facebook_ad = 1 THEN 'Instagram, Facebook'
        WHEN a.instagram_ad = 1 AND a.brochure_ad = 1 THEN 'Instagram, Brochure'
        WHEN a.facebook_ad = 1 AND a.brochure_ad = 1 THEN 'Facebook, Brochure'
        ELSE 'Unknown Combination'
    END AS advertising_channels
FROM
    marketing m
JOIN
    advert a ON m.id = a.id
WHERE
    m.response = 1
    AND (a.bulkmail_ad + a.twitter_ad + a.instagram_ad + a.facebook_ad + a.brochure_ad) > 1
ORDER BY
    m.id;
-----display ad channels in different columns-----
SELECT
    m.id,
    m.country,
    CASE WHEN a.bulkmail_ad = 1 THEN 'Bulkmail' ELSE NULL END AS bulkmail,
    CASE WHEN a.twitter_ad = 1 THEN 'Twitter' ELSE NULL END AS twitter,
    CASE WHEN a.instagram_ad = 1 THEN 'Instagram' ELSE NULL END AS instagram,
    CASE WHEN a.facebook_ad = 1 THEN 'Facebook' ELSE NULL END AS facebook,
    CASE WHEN a.brochure_ad = 1 THEN 'Brochure' ELSE NULL END AS brochure
FROM
    marketing m
JOIN
    advert a ON m.id = a.id
WHERE
    m.response = 1
    AND (a.bulkmail_ad + a.twitter_ad + a.instagram_ad + a.facebook_ad + a.brochure_ad) > 1
ORDER BY
    m.id;

```

APPENDIX 7 – pgAdmin insights tables

Most popular product type per demographics:

The below results tables highlight that the most popular product type is alcoholic beverages in every marital status group.

Most popular products per marital status SQL results		
Excluding outlier values		
marital_status	product_type	total_spend
Divorced	liquor	72609
In a relationship	liquor	163512
Married	liquor	237782
Single	liquor	130071
Widowed	liquor	25142
Including outlier values		
marital_status	product_type	total_spend
Divorced	liquor	75364
In a relationship	liquor	176706
Married	liquor	256976
Single	liquor	139126
Widowed	liquor	27902

The below table show that alcoholic beverages remain the most popular regardless of the number of children in the household.

home_status	product_type	total_spend
With Children	liquor	93088
With Teenagers	liquor	257716
Without Children/Teenagers	liquor	278312

The below tables highlight that in every country the most popular product is alcoholic beverages.

Including outlier values			Excluding outlier values		
country	product_type	total_spend	country	product_type	total_spend
AUS	liquor	42752	AUS	liquor	40021
AUS	meat	22328	AUS	meat	17526
AUS	commodities	7132	AUS	commodities	4247
AUS	fish	5546	AUS	fish	3113

AUS	chocolates	4129		AUS	chocolates	2161	
AUS	vegetables	3689		AUS	vegetables	1822	
CA	liquor	84066		CA	liquor	74604	
CA	meat	45925		CA	meat	26645	
CA	commodi- ties	12144		CA	commodi- ties	6854	
CA	fish	9980		CA	fish	5305	
CA	vegetables	7681		CA	vegetables	3537	
CA	chocolates	7607		CA	chocolates	3244	
GER	liquor	36776		GER	liquor	35465	
GER	meat	20272		GER	meat	13436	
GER	commodi- ties	5768		GER	commodi- ties	3800	
GER	fish	4601		GER	fish	2362	
GER	vegetables	2980		GER	vegetables	1560	
GER	chocolates	2801		GER	chocolates	1527	
IND	liquor	36236		IND	liquor	34928	
IND	meat	23729		IND	meat	15257	
IND	commodi- ties	6014		IND	commodi- ties	4283	
IND	fish	4818		IND	fish	2596	
IND	vegetables	3788		IND	vegetables	1951	
IND	chocolates	3221		IND	chocolates	1882	
ME	liquor	1729		ME	liquor	1729	
ME	meat	817		ME	meat	817	
ME	fish	226		ME	chocolates	122	
ME	commodi- ties	220		ME	commodi- ties	80	
ME	chocolates	122		ME	fish	44	
ME	vegetables	8		ME	vegetables	8	
SA	liquor	105909		SA	liquor	97832	
SA	meat	58398		SA	meat	38950	
SA	commodi- ties	15129		SA	commodi- ties	8849	
SA	fish	13670		SA	fish	6110	
SA	chocolates	9019		SA	chocolates	4254	
SA	vegetables	8937		SA	vegetables	3982	
SP	liquor	336392		SP	liquor	312323	
SP	meat	178409		SP	meat	111920	
SP	commodi- ties	46181		SP	commodi- ties	30212	
SP	fish	40153		SP	fish	22732	
SP	chocolates	30134		SP	vegetables	14153	
SP	vegetables	28288		SP	chocolates	13845	
US	liquor	32214		US	liquor	32214	
US	meat	20185		US	meat	13807	

US	commodities	4839		US	commodities	3269	
US	fish	4411		US	fish	2318	
US	vegetables	3034		US	chocolates	1682	
US	chocolates	2863		US	vegetables	1227	

Advertisement insights:

The below table show the most effective social media platforms per country

country	most_effective_platform
AUS	Instagram
CA	Twitter
GER	Twitter
IND	Twitter
ME	Twitter
SA	Instagram
SP	Instagram
US	Facebook

This slightly differs if a campaign influences the results

country	most_effective_platform
AUS	Instagram
CA	Instagram
GER	Facebook
IND	Twitter
ME	Twitter
SA	Instagram
SP	Instagram
US	Instagram

Social media for different marital status

marital_status	most_effective_platform
Single	Twitter
In a relationship	Instagram
Married	Instagram
Divorced	Twitter
Widowed	Twitter

The below table shows the sales insights for the different media channels. Twitter has the highest number of conversions, whilst Instagram leads to the highest average sales.

To show the sales for ad channels			
channel	average_sales	total_sales	successful_conversions
Bulkmail	720.54	117448	163
Twitter	1137.56	186560	164
Instagram	1616.43	261862	162
Facebook	1484.35	210777	142
Brochure	1307.67	39230	30

A further exploration highlighted that there were 87 customers who was influenced by more than one advertisement channel.

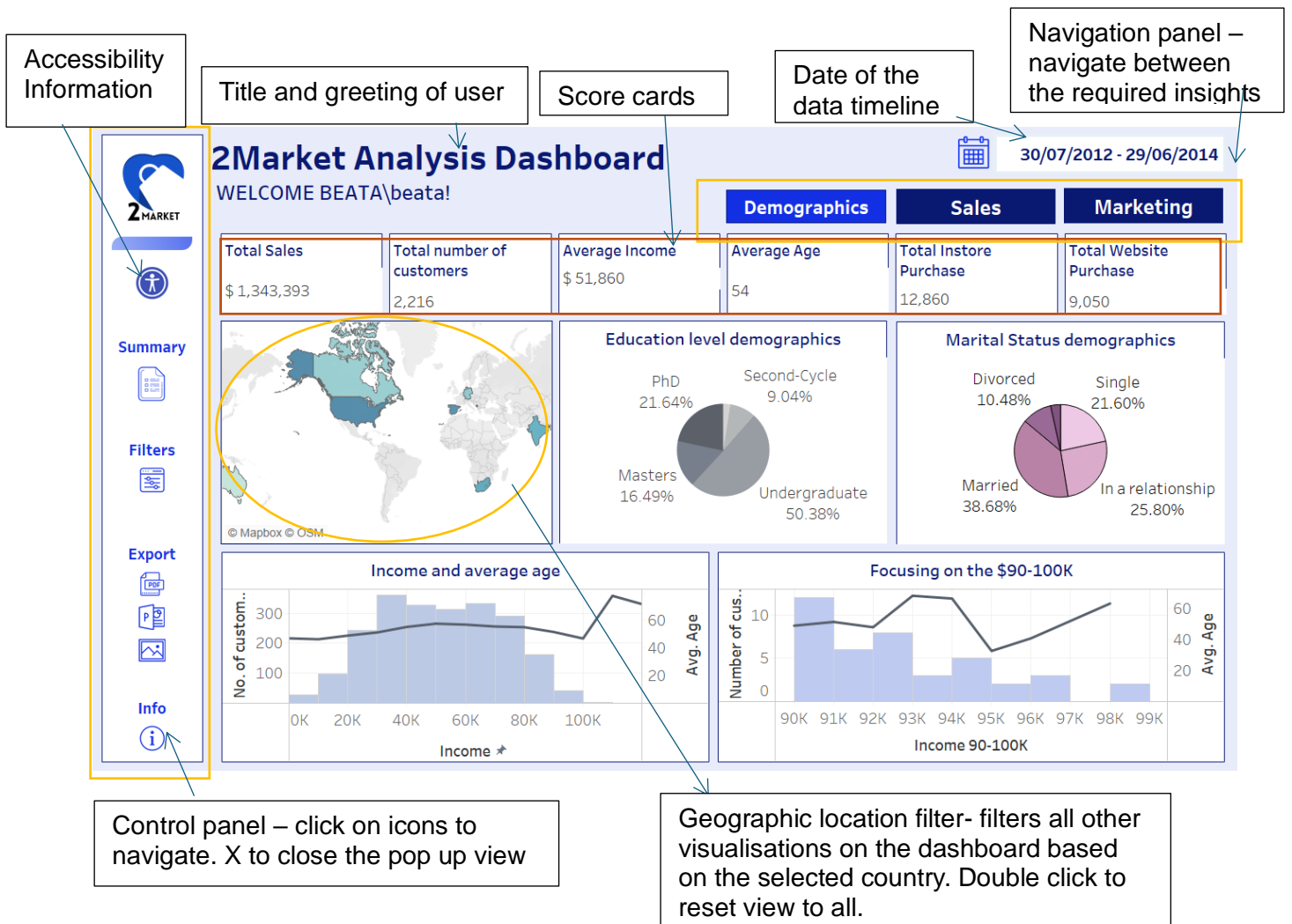
APPENDIX 8 – Tableau dashboard design

Please refer to the 2Market_analysis_2.twbx file and the presentation file for additional information.

Dashboard Design and Development:

- Rationale:** The dashboard was designed to provide a comprehensive yet intuitive view of the analysis results. Key principles included clarity, ease of use, and interactivity. The design was developed with a specific layout in mind. There is a control panel on the left side of the dashboard which includes interactive elements. There is a navigation panel on the right side of the dashboard to navigate between demographics, sales and marketing insights.

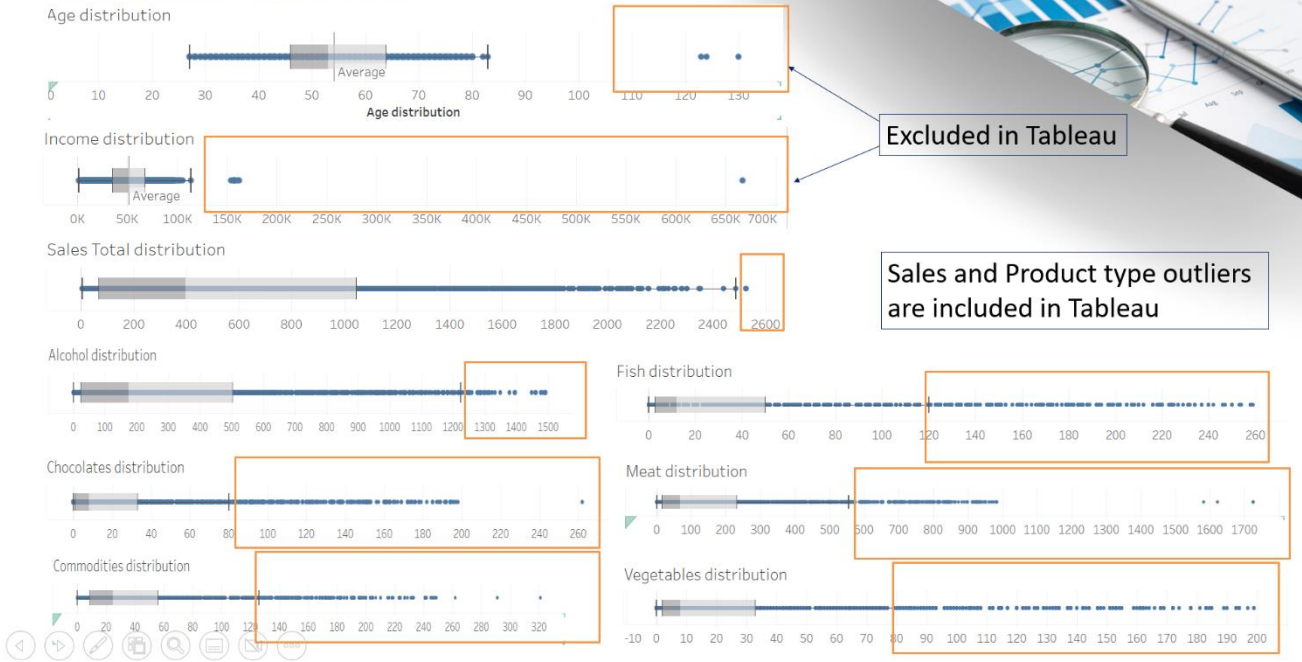
See below the main functions of the dashboard.



- Visualization Types:** Different types of visualizations were used to represent data effectively. Bar charts were used to display product popularity, as they provide a clear comparison of different categories. Line charts were chosen to show sales trends over time, making it easy to identify patterns and fluctuations. Pie charts illustrated demographic distributions, offering a quick visual representation of proportions. Each visualization type was selected based on its ability to convey the relevant information clearly and concisely.

- **Colour Scheme:** A consistent colour scheme was applied to differentiate between various data points and ensure visual harmony. High-contrast colours were used to enhance readability and accessibility for users with visual impairments. The colour scheme adhered to 2Market's branding guidelines to ensure a professional and cohesive appearance (A fictitious company logo was used to tie in the colour scheme, based on the logo and possible company branding). The colour blue was chosen as the main colour and, shades of grey, purples and pinks as the visualisation colours.
- **Size and Layout:** The dashboard layout was designed to be clean and organized, with key metrics and insights prominently displayed. Interactive elements such as filters and highlights and navigation bars were incorporated to allow users to customize their views and drill down into specific data segments.
- **Interactivity:** Interactivity was a crucial aspect of the dashboard design. Filters and drill-down capabilities were integrated to allow users to customize their view based on specific parameters such as time period, product category, and geographic location. Interactive elements enable users to explore the data dynamically, facilitating a more engaging and insightful experience. Tooltips and hover actions were included to provide additional context and details without cluttering the visualizations.
- **Accessibility:** Whilst creating the dashboard great care was taken to ensure accessibility of the dashboard, following the Web Content Accessibility Guidelines (WCAG) 2.1. Accessibility features included high-contrast colour options, text alternatives for visual elements, and keyboard navigation support, ensuring the dashboard is usable by individuals with disabilities. Each viz has a description, which is also displayed as alt text. Additionally, if ctrl+E is pressed, this will also display a description of the individual visualisation. The dashboard was tested for compatibility with screen readers, ensuring that users with disabilities can navigate and interpret the data effectively. At times it was difficult to keep the whole axis title within the view, and abbreviation was used. This is a limitation of the dashboard.
- **Feedback Loop:** In reality stakeholder feedback would have been integrated into the design process, with iterative improvements would have been made based on user input to ensure the dashboard met the needs of its intended audience." With Tableau's built in functionality user performance can be monitored and comments can enhance dashboard performance to meet stakeholders needs.
- **Outlier value treatment in Tableau**

Outlier values



User Group and Access/Security Considerations

- The dashboard was designed with specific user groups in mind, including marketing executives, product managers, and senior management. Each user group has distinct data needs and levels of access. For instance, marketing executives may focus on advertising effectiveness, while product managers may prioritize sales trends and product popularity.
- Access and security considerations are paramount. User roles and permissions should be defined to ensure that sensitive data is protected. Role-based access control (RBAC) needs to be implemented when the dashboard is being published, allowing users to view only the data relevant to their roles. Sensitive information, such as customer demographics and financial data, needs to be restricted to authorized personnel only. Data encryption and secure login protocols would be employed to safeguard the dashboard from unauthorized access and potential data breaches.

By incorporating these considerations, the Tableau dashboard was designed to be a powerful tool for data-driven decision-making, enabling 2Market to leverage insights for strategic enhancements while ensuring user accessibility and data security.

APPENDIX 9 – Summary of key insights

1. Customer Demographics:

- The average age of customers is 54 years.
- The average income of customers is \$52,247.
- The highest number of customers have an income of \$30-40K.
- 39% of customers are married.
- 50% of customers have an undergraduate level of education.

2. Product Popularity:

- Alcoholic beverages are the most popular product category, regardless of customer demographics.

3. Marketing Channel Effectiveness:

- No single advertising platform is universally effective; effectiveness varies by demographics.
- 87 customers were influenced by more than one advertising platform.
- Twitter has the highest number of conversions, while Instagram leads to the highest average sales.

4. Sales Trends:

- For age, the highest number of customers are between 50-60 years old and they also lead to the highest volume of sales.
- For income, whilst the highest number of customers have an income of 30-40K, they represent a low sales volume and the highest sales volume comes from customers, who have an income of 70-80K.
- There is a declining trend in total and average sales, despite a stable number of customers.
- Returning customers show higher average sales both in-store and online, with a slight decline above 20 online purchases.

5. Advertisement Insights:

- Advert channel uptake remains low, but average sales per purchase are significantly higher for each ad channel.

APPENDIX 10 - Recommendations

- **Targeted Marketing Campaigns:**
 - Develop personalized marketing campaigns targeting demographics with higher conversion rates, such as married individuals and those with undergraduate education.
- **Enhanced Advertising Strategy:**
 - Increase investment in Instagram and Twitter advertising, as they show the highest average sales and conversions.
 - Experiment with multi-channel marketing to leverage the influence on customers reached through more than one platform.
- **Product Line Optimization:**
 - Focus on promoting and expanding the alcoholic beverage category, which is the most popular among customers.
 - Explore opportunities to introduce premium alcoholic beverages to increase sales volume and profit margins.
- **Customer Retention Programs:**
 - Implement loyalty programs targeting returning customers to maintain their higher average sales rates.
 - Offer incentives for frequent online shoppers to counteract the slight decline observed after 20 purchases.
- **Data-Driven Decision Making:**
 - Continuously analyse customer data to adjust marketing strategies and product offerings based on evolving trends.
 - Utilize Tableau and other visualization tools to keep stakeholders informed and engaged with real-time data insights.
 - Connect additional data points to Tableau/pgAdmin database, such as profitability, costs, orders, customer feedback, loyalty program etc. to further add to the benefits of data-driven decision making. Whilst Tableau is great at overall visualisations, SQL can aid to answer key stakeholder questions with targeted querying.

Potential Impacts and Benefits

- **Increased Sales:** By targeting high-performing demographics and optimizing product offerings, 2Market can boost overall sales.
- **Higher Customer Engagement:** Personalized marketing campaigns can enhance customer satisfaction and loyalty.
- **Better ROI on Advertising:** Focusing on the most effective channels will improve the return on marketing investments.
- **Sustained Competitive Edge:** Data-driven strategies will enable 2Market to stay ahead in a competitive market.

APPENDIX 11 – Future research

- **Profitability Analysis:**
 - Integrate cost and profitability data to complement sales volume analysis, ensuring strategies also maximize profits.
- **Current Data and Trends:**
 - Collect and analyse more recent data to validate and update insights, considering changes in customer behaviour and market dynamics.
- **GDPR Compliance:**
 - Assess the impact of GDPR on data collection and usage, ensuring ethical and legal compliance in future analyses.
- **Advanced Analytics:**
 - Utilize machine learning models to predict customer behaviour and optimize marketing strategies.
 - Explore sentiment analysis of customer feedback to identify areas for improvement and innovation.

APPENDIX 12 – REFERENCES

1. Data Management Association's (DAMA) Data Management Body of Knowledge (DMBOK) – Retrieved from <https://technicpub.com/dmbok/>
2. Data Protection and Privacy Legislation Worldwide. Retrieved from - <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
3. European Parliament. (2016). General Data Protection Regulation (GDPR). Retrieved from [GDPR.eu](https://gdpr.eu) .
4. Web Content Accessibility Guidelines (WCAG) 2.1. Retrieved from - <https://www.w3.org/TR/WCAG21/>