

Assignment: Technical report

Problem Statement:

- Understanding 2Market's customer's demographic
- What is 2Market's most effective advertisement channel
- Understanding 2Market's best-selling product

Background

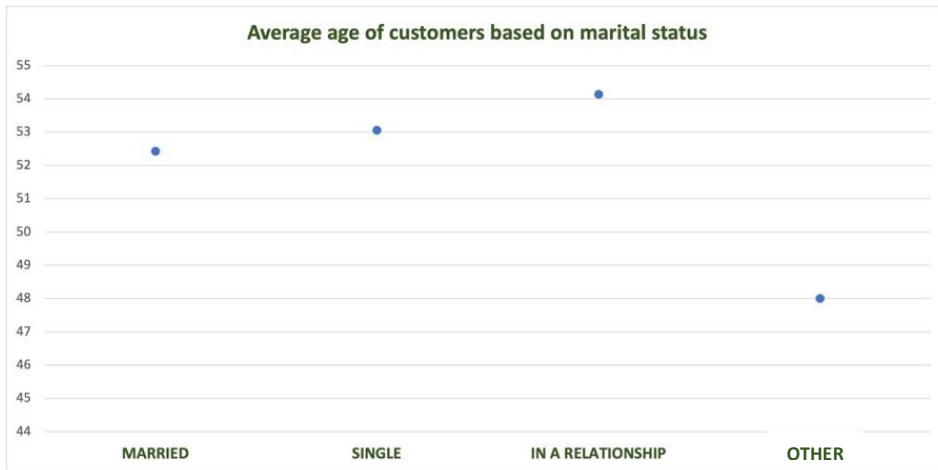
2Market is a global supermarket that wants to understand customer's purchase behaviours and effective marketing strategies to optimize its services in-stores globally and online. To be able to find significant insights on the data, I first listed relevant questions to answer during the analysis process that will help get a better and clearer understanding of the data. Using the 5 why method helped me to get to the questions listed below.

1. Average age of customers based on country/marital status
2. Does customer's income influence how much they spend on shopping and how it varies based on demographic?
3. If number of children and teenagers in a household influence customer's spending?
4. In-store to online shopping ratio?

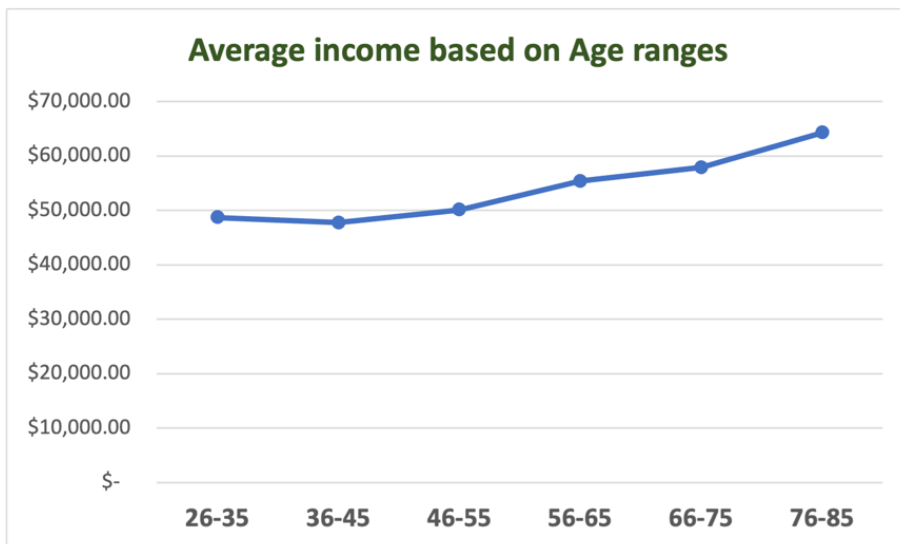
Analytical Approach

I initially started off by checking the accuracy of my data, this involved checking spellings and changing headers to more comprehensible titles that's easier for the reader to understand using 'find and replace' and 'spell check'. I then started by removing line breaks and parenthesis and removing blanks on data cells by using "=TRIM ()" function. I focused on the completeness of the data and looked for any missing data this was done by going on columns by pressing "control + G" to replace blank cells with "#N/A", luckily no blank cells were found. I used case sensitivity functions for text to check for data consistency and reviewed if correct data type is correct using the (=TYPE) function. I readjusted the data type that required changing using "text to columns", and formatted data type accordingly. I checked for data uniqueness using conditional formatting to remove errors and concluded by highlighting the table to remove duplicates. Lastly, I checked for the timeliness of the data and if data format is set correctly, I had to adjust date format using text to columns. After I've ensured that my data is clean, I proceeded to add filters to my table and making necessary changes to the data for better understandability and readability, this involved changing of marital status filters to more reasonable groups. I then proceeded to remove any outlier values by calculating IQR range of age and removed rows that don't fall within the relevant range (18 – 90 years). I used the functions "COUNTIF" and "SUMIF" to calculate the average income of different marital statuses and used pivot tables to create groups for ages and income to find if there's a correlation. I've used Excel as my main method of analysis due to the data being small and not being limited to having access to 2Market's relational database, therefore the use of SQL wasn't necessary.

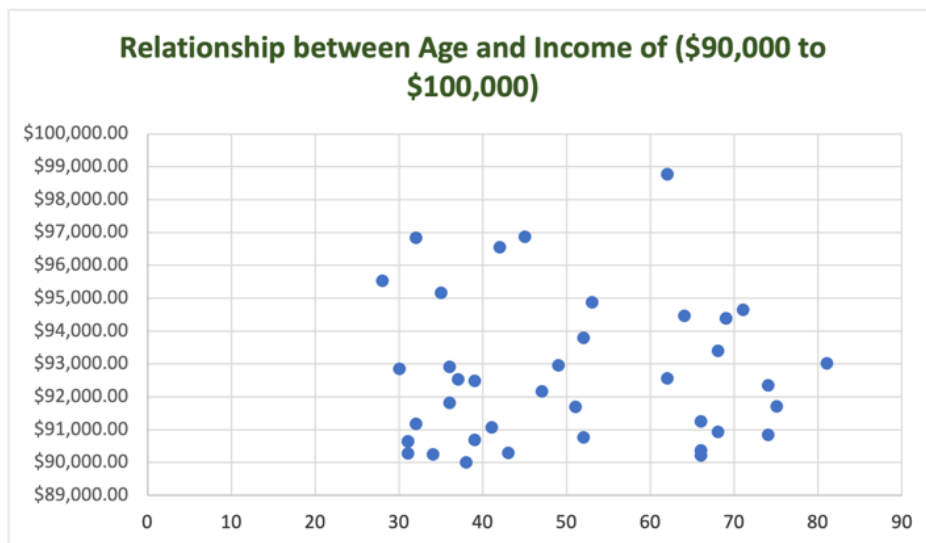
Insights gathered based on visualisations on excel:



Customer's average age doesn't range much based on marital status except for "other" which could be due to the limited number of customers (4) that fall into that category



There's a clear positive correlation between customer's age ranges and average income.



There is no correlation between customer's income ranging between \$90,000 and \$100,000 and age

Dashboard design and development:

I imported both my tables onto Tableau and connected them using joins. I had to go back on Excel and clean `ad_data` based on changes made on the `marketing_data` to remove null values. I've made three dashboards, with the first focusing on customer population in general and most effective advertisement. I started off by first looking for the customer's numbers across all countries and how does it vary, I used a map and gradient colours to emphasize which countries have the highest and lowest number of customers. I considered removing Montenegro as an outlier but due to the numbers of customer's being relatively low across countries, I've decided to keep it. I've then decided to use packed bubbles and gradient colours to show how customer's number vary across different marital statuses. I placed a parameter filter to interactively switch between marital status and country, to assess both demographics using one chart throughout all dashboards. I created a pie chart to highlight the difference between the instore purchases and online purchases in both demographics, I've also done the same to show the most effective advertising channel and placed in order of most to least effective and added legends to show the numbers and filters based on assigned colour of demographics. The second dashboard was to get a deeper understanding of our customer's demographic, that was done by displaying the difference between age across demographic using a table and highlighting the difference between average age across demographics using a scatter plot and connecting the dots to show correlations. I then created a bar chart to demonstrate the number of kids and teenagers per demographic and the average income to gather whether they have an influence on customer's spending. This all brings us to our final dashboard which demonstrates a better understanding of customer shopping spending and overall success to complain ratio. I've done this by making two separate bar graphs to show age range and demographic spending put them in descending order to figure what customers spend on. Lastly, I concluded my chart with 2 packed bubble charts to show overall lead conversions compared to complains. I used legends to help show exact number of each based on demographic to help stakeholders get a better understanding.

Patterns, Trends, and Insights:

From my analysis I've concluded that Spain has the highest number of customers making up more than 60% of the total number of customers globally, this explains why Spain makes up the most of 2Market's total revenue although in average Spanish customers spending falls in the bottom 3. Graphs also illustrates that all countries except for Montenegro roughly spend the same on average. This shows that there's a close correlation between average income and average spending as India has the lowest average income and spends the least on average. My analysis has also made me come to the realisation that age ranges are quite close throughout all demographics. Chart shows a higher instore to online purchases ratio, however the difference is relatively low therefore, optimizing website according to customers need can be relevant and focusing on marketing on social-media apps would be most effective. Lastly, is understanding best selling product, this was more difficult to conclude due to the lack of data on the number of products customers purchase, however based on the charts on average spending, Alcohol and meat products is constant throughout all demographics and age groups; therefore, making up most of 2Market's revenue.