2Market Technical Report

By:
Geoff Widdowson
BEng (hons) IPS

Introduction/Problem Statement

I am looking to further understand 2Market's customer purchase behavior with in regards to customer demographics, most effective advertising and best-selling products. As such we are exploring and hoping to answer the following:

- ***What are the customer demographics?***
    - ***Who are the customers?***

- ***What are the best selling products?***
    - ***How profitable are these?***
    - ***How do other demographics impact sales?***
      Particularly in regard to:
        - ***Location/ Geography***
        - ***Marital status/ Children***
        - ***Income/ Education***

- ***What is most effective advertising channel?***

I aim to confirm the best-selling products and how these vary based on customer demographics as well as gain insight into the impact of advertising and marketing across these demographics. While gaining a deeper understanding of purchase behavior I hope to help 2Market better target advertising and marketing in the future!

Analytical Approach

I begun my analysis by first cleaning and organizing the data by bring the 2 separate tables of marketing and advertising csv files together to get a full picture of everything for my analysis. I did this Initially using Excel and subsequently looking at how SQL could improve upon this before using tableau to help with the final analysis and ultimately visualization.

Initially I set out to check for duplicates while also remove any outliers or anomalies. I then went onto creating an age column as well as ensuring the correct formatting of dates and currency. I found anomalies such as customers older than 120 or devilish salary of $666,666. While in martial status I noticed the use of YOLO and Alone (which I changed to single).

In looking at customers demographics I also discovered a large number of duplicate anomalies by location when grouping earnings. Where the duplicates were exact copies I eliminated them completely but where the location was not duplicate I worked on the assumption that this was one customer could but the location data was invalid and thus I amended the location in these customers to NA to highlight this fact. As such I found 296 customers with geographical anomalies and of these 144 had location amended to keep the financial numbers clean, particularly with regard to the impact on location!

Given all this in total I removed 207 outliers and duplicates to obtain the cleanest possible set I could.

I also chose to exclude information relating to the date the customer joined and instead focus some of my analysis on customer Recency as this will give a better indication of future customer engagement when it come to projecting sales into the future.

Having spent time looking at creating tables in SQL and working with possible approaches in using this for the analysis ultimately I imported my data into tableau to visualize the data. As such I felt that this approach has made my end results a far easier to digest for stakeholders.

Dashboard Design

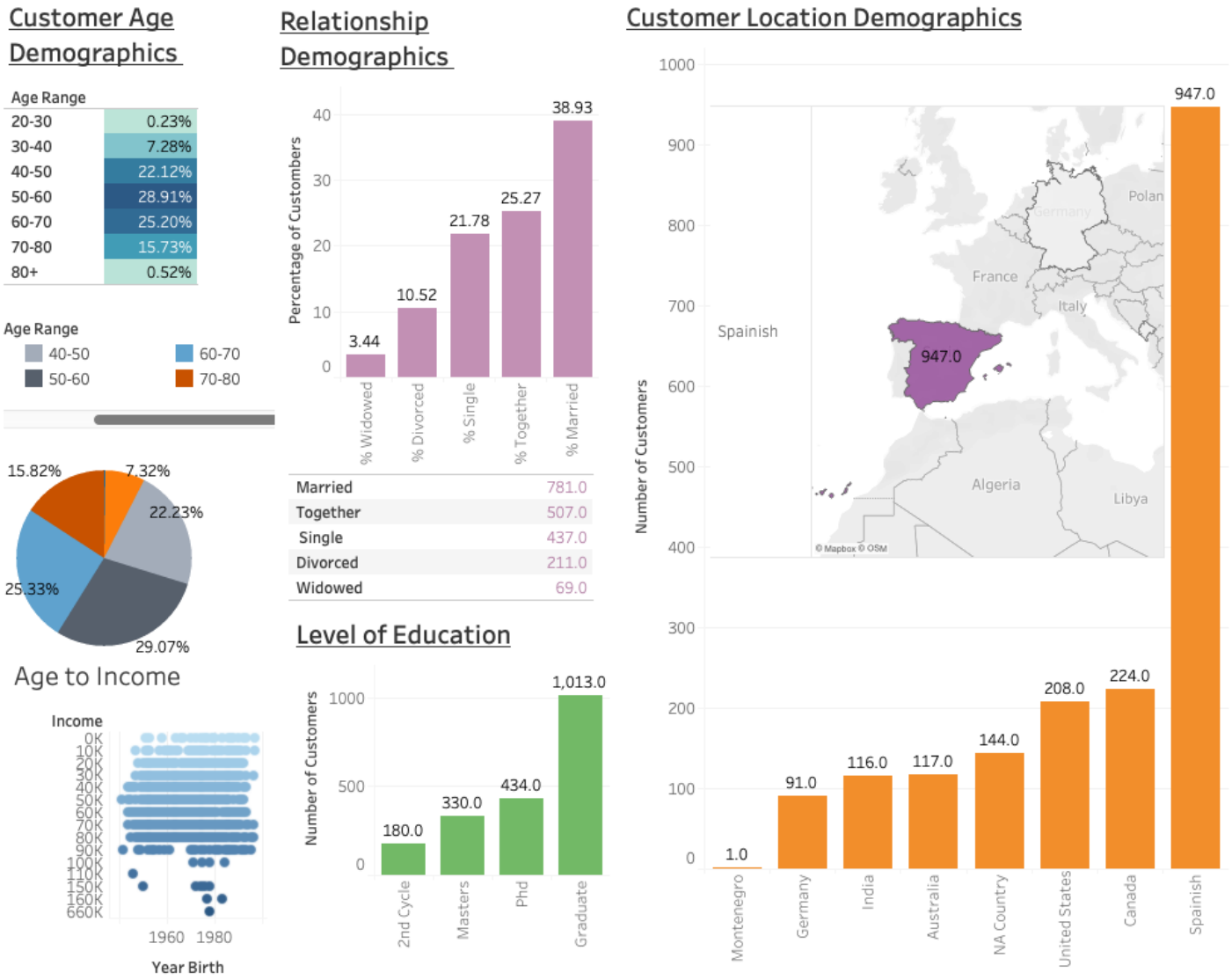
I chose to use a variety of visualizations including tables, charts, maps, and scatter diagrams. The majority of these are bar graphs as it was the most simple way to visualize the data in a clean and digestible way to stakeholders.

I tried to consolidate colours, fonts, formatting and positioning as best I could within the confines of tableau. At the same time I considered adding branding and tried to keep KPI information prominent in the design.

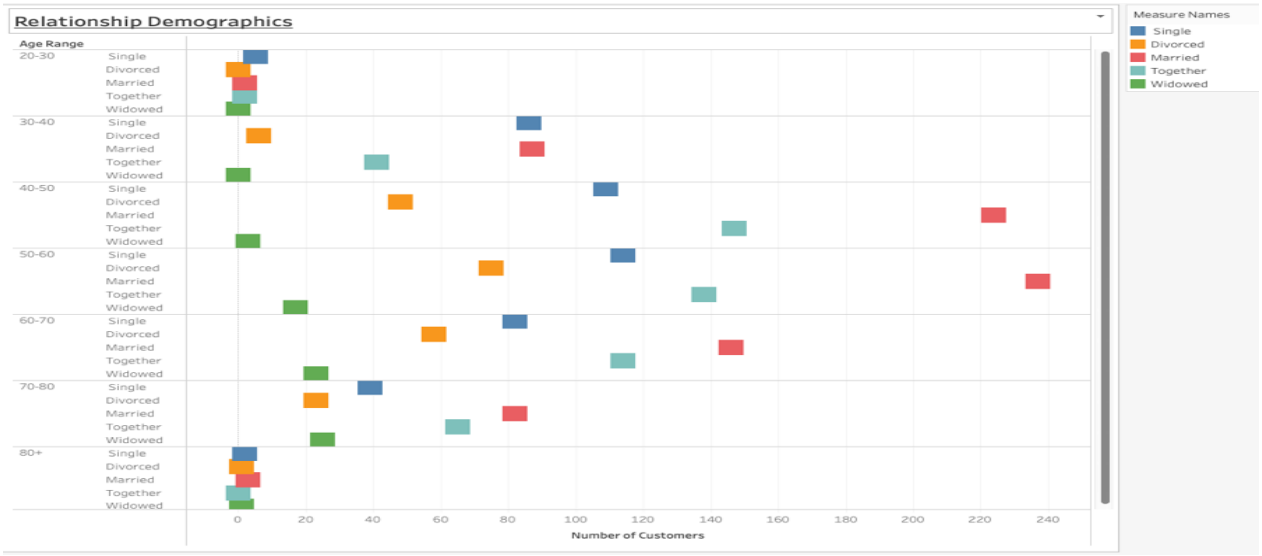
Patterns, trends, and insights:

Through my analysis I explored a number of patterns and trends while I gained a number of insights.

Who are the Customers?

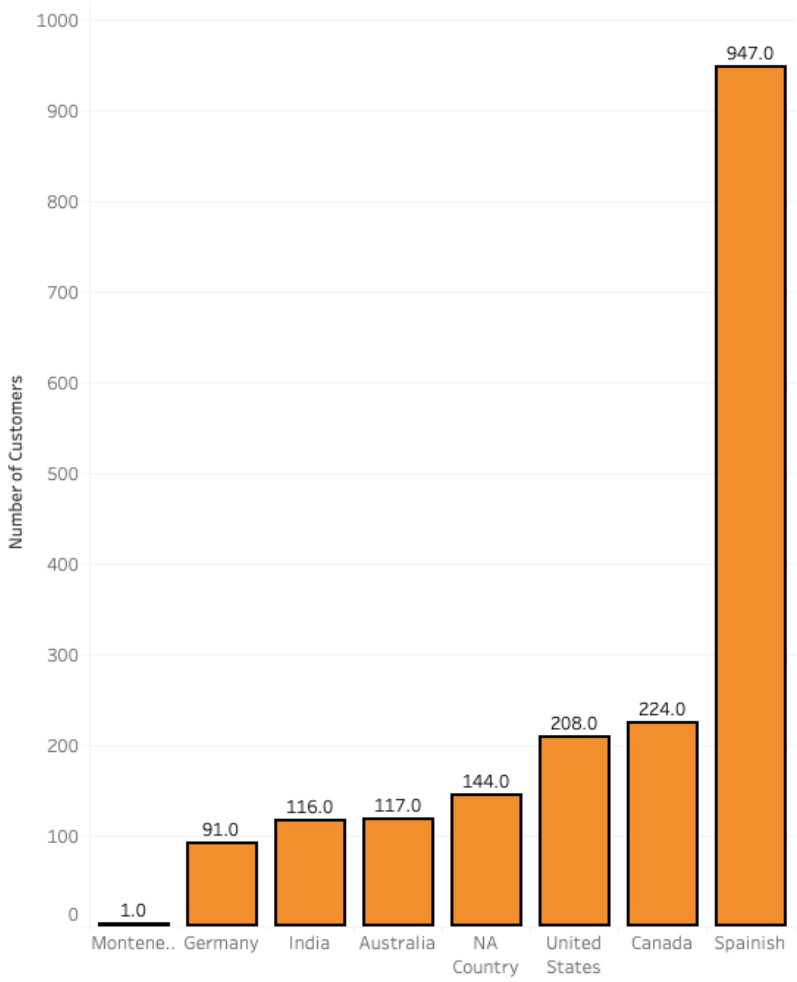


Initially I looked at customer demographics based around the number of customers and who these people are. Looking at age I discovered the vast majority (over 90%) were over 40 in age with the largest proportion (over 50%) being aged 50-60. I also discover that customers were well educated and over 60% were in a relationship. At the same time customers were also earning well with a fairly even distribution of wages across all ages and incomes.



With the help of using excel and SQL possibly the most interesting insights to customer demographics I gained early on through my data cleaning were in relation to geography..
Interestingly the impact of unclean location data can be seen here where we do not confidently know where the 4th largest customer group is
based?
More importantly however is the fact that Spain has the most customers by far! (almost 50%)

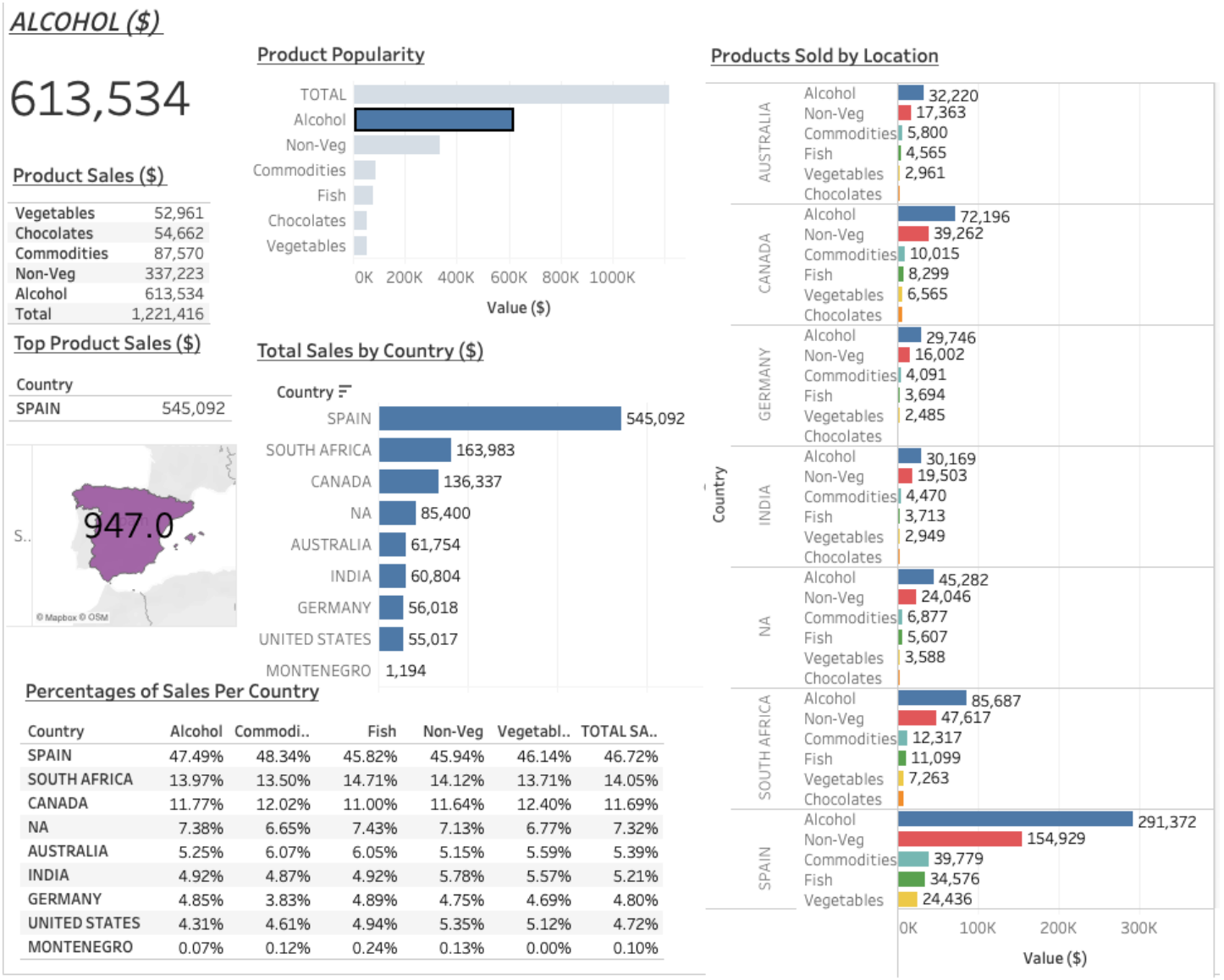
Customer Location Demographics



Products and Sales

In looking at products and sales I found that Spain accounted for 46.72% of the total sales and alcohol was the most popular product accounting for over 50% of the total product sales with customers spending $613,534 on booze!

- Spain's alcohol purchasing accounted for 47.49% with a value of $291,372 quite something for only 947 customers!
- NA location accounted for 7.32% of total sales
- Total product sales $1,221,416



Percentages of Sales Per Country

Country	Alcohol	Commodi..	Fish	Non-Veg	Vegetabl..	TOTAL SA..
SPAIN	47.49%	48.34%	45.82%	45.94%	46.14%	46.72%
SOUTH AFRICA	13.97%	13.50%	14.71%	14.12%	13.71%	14.05%
CANADA	11.77%	12.02%	11.00%	11.64%	12.40%	11.69%
NA	7.38%	6.65%	7.43%	7.13%	6.77%	7.32%
AUSTRALIA	5.25%	6.07%	6.05%	5.15%	5.59%	5.39%
INDIA	4.92%	4.87%	4.92%	5.78%	5.57%	5.21%
GERMANY	4.85%	3.83%	4.89%	4.75%	4.69%	4.80%
UNITED STATES	4.31%	4.61%	4.94%	5.35%	5.12%	4.72%
MONTENEGRO	0.07%	0.12%	0.24%	0.13%	0.00%	0.10%

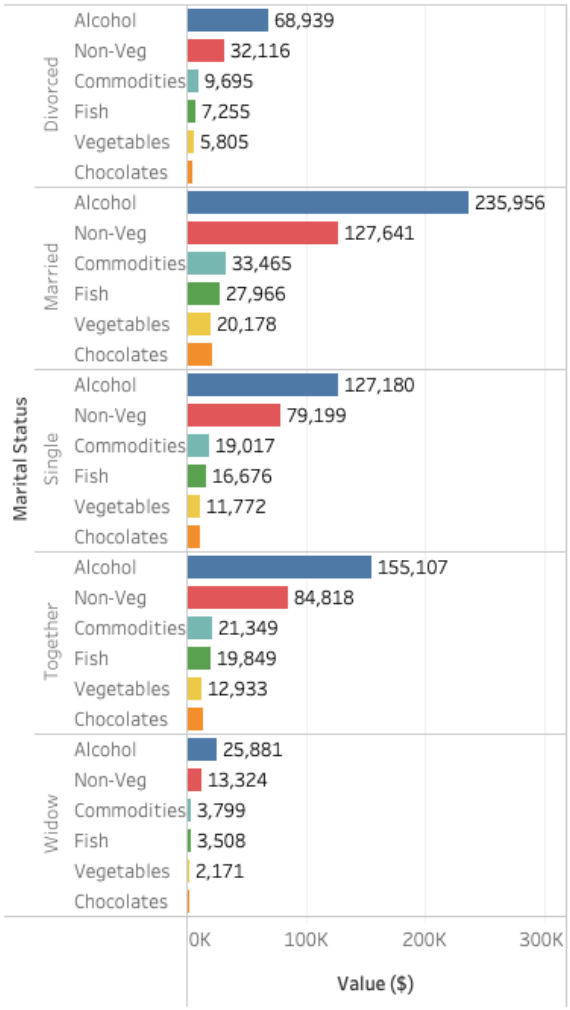
Looking further into products purchased based upon demographics I discovered:

- Couples together/ married accounted for over 50% alcohol sales with a value exceeding $600K!

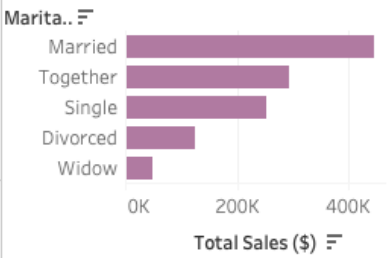
- People with Incomes of $50-80K accounted for the majority of purchases.

- Less kids and teens = more spending power (particularly with alcohol)

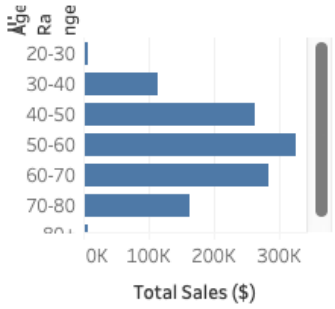
Products Sold by Relationship



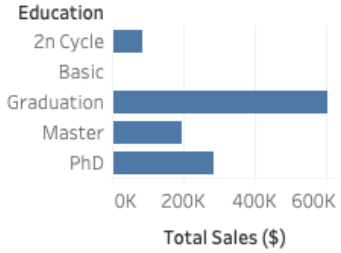
Total Sales by Relationship



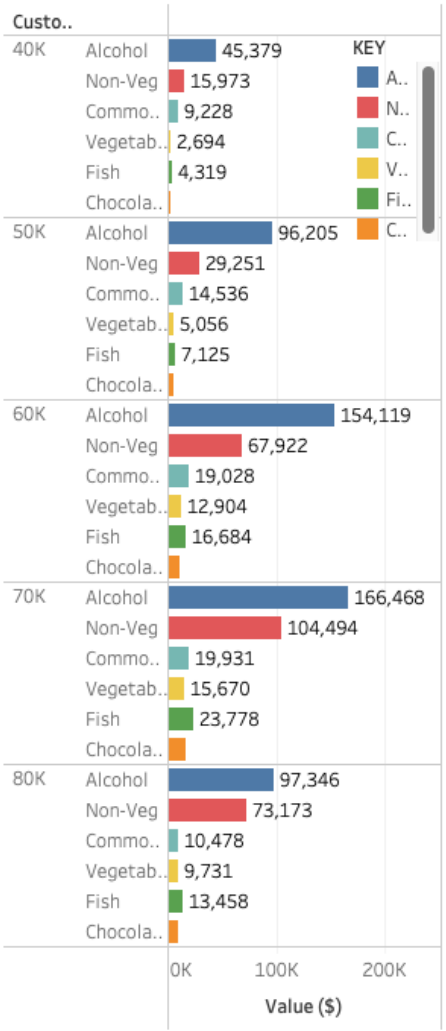
Total Sales by Age



Total Sales by Education

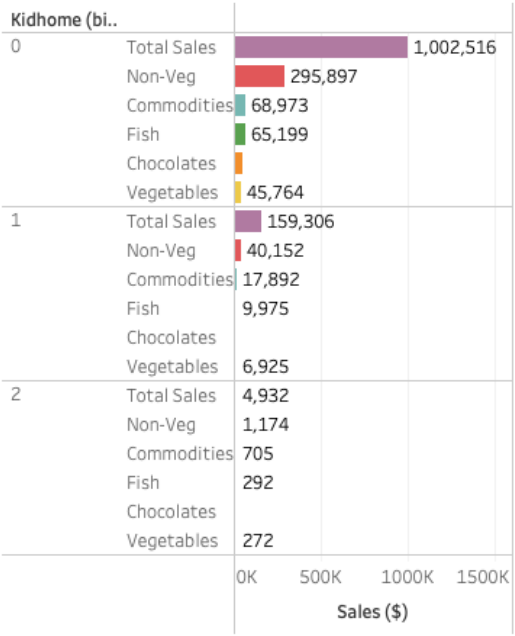


Product Sales by Income

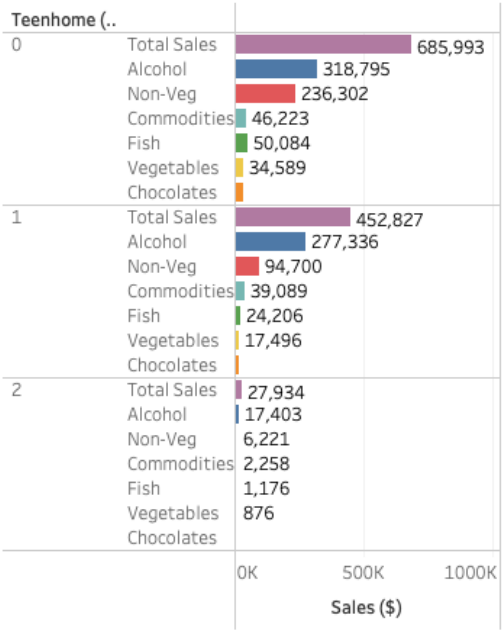


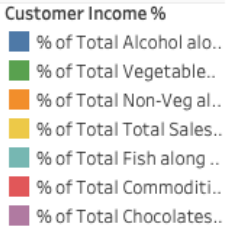
% Total Product Sales by Customer Income

Products/ Total Sales based on Number Kids @ Home

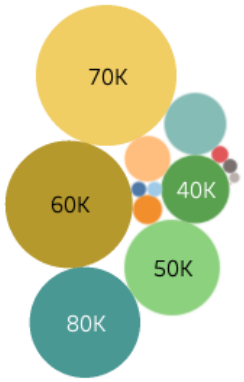


Products/ Total Sales based on Number Teens @ Home





% Total Sales by Customer Income (in US Dollars $)



Advertising

I continued my analysis to then look into advertising and the impact this had on sales.

More purchasing takes place instore while there are significantly more conversions through online.

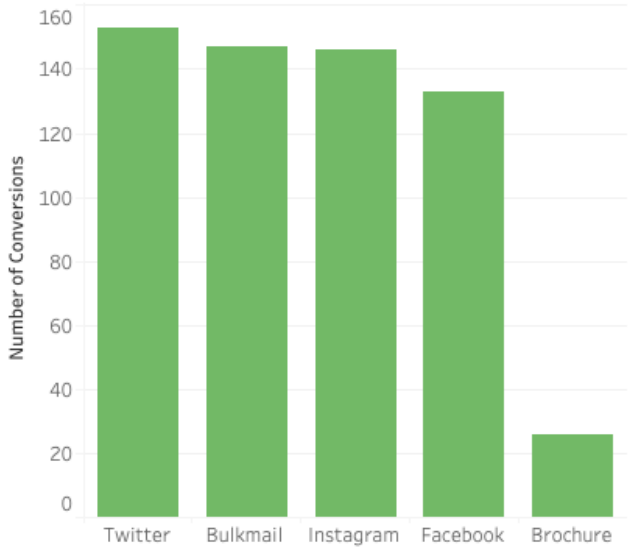
It should also be noted that significantly more conversions took place with customers who were educated than not.

Advertising Conversions

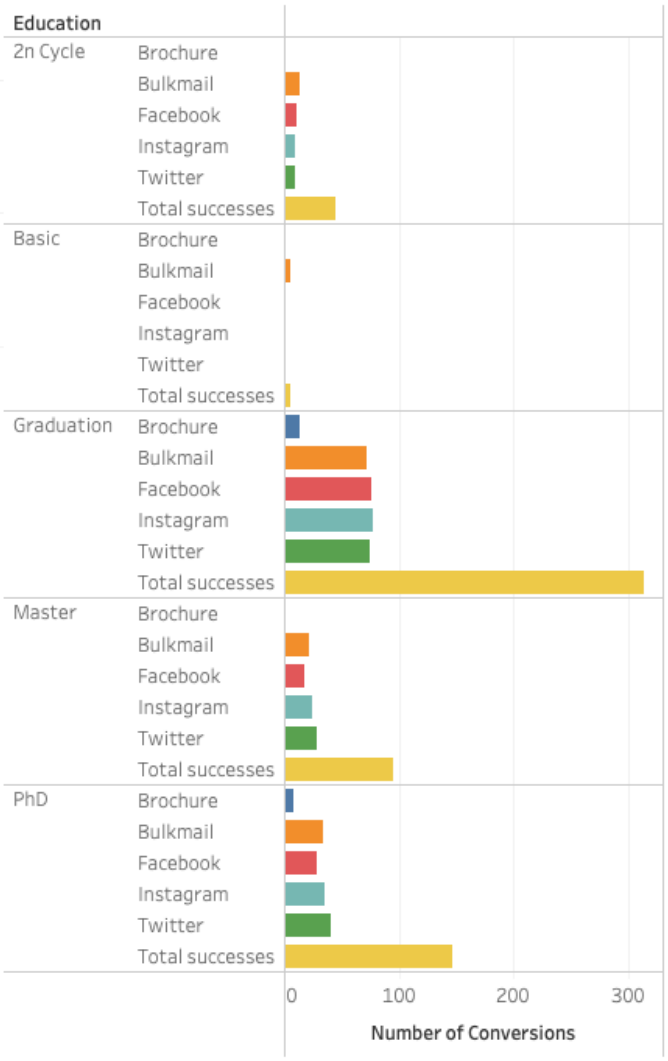
Total successes	605.0
Twitter	153.0
Bulkmail	147.0
Instagram	146.0
Facebook	133.0
Brochure	26.0

Instore vs Online

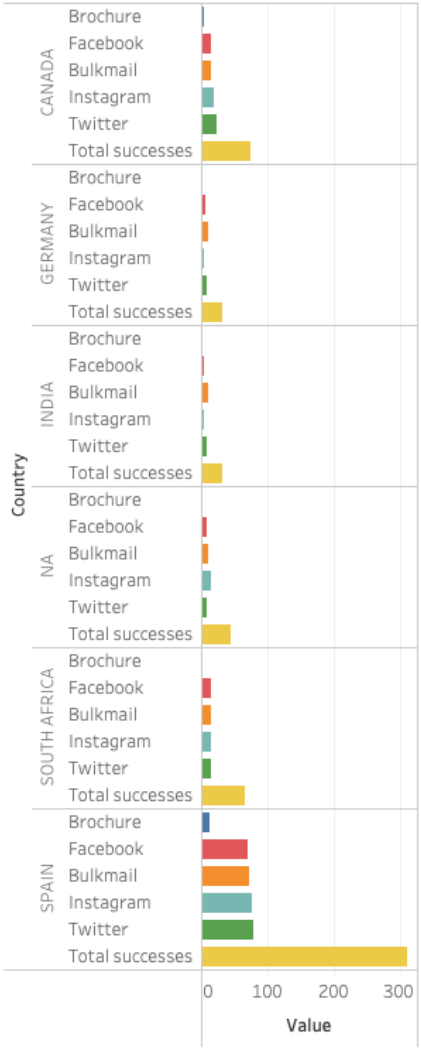
Advertising Conversions to Customer



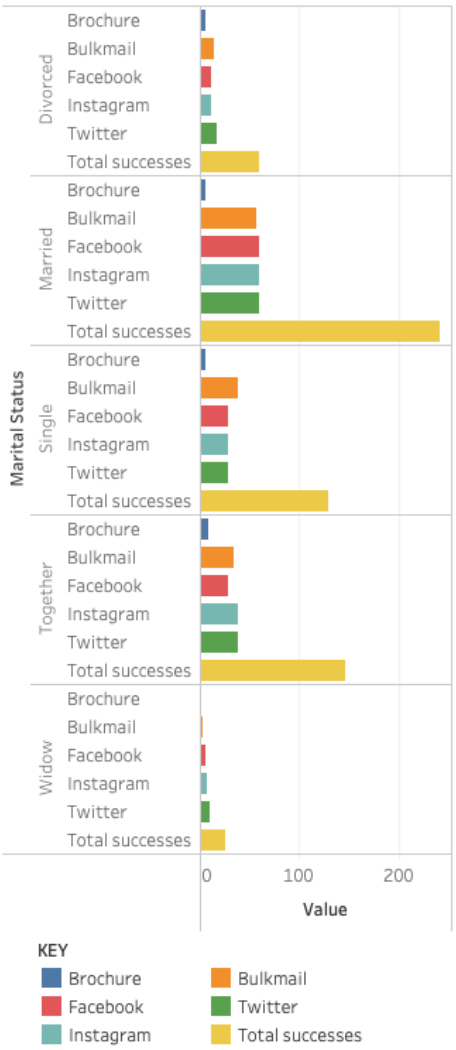
Educational Advertising Conversions



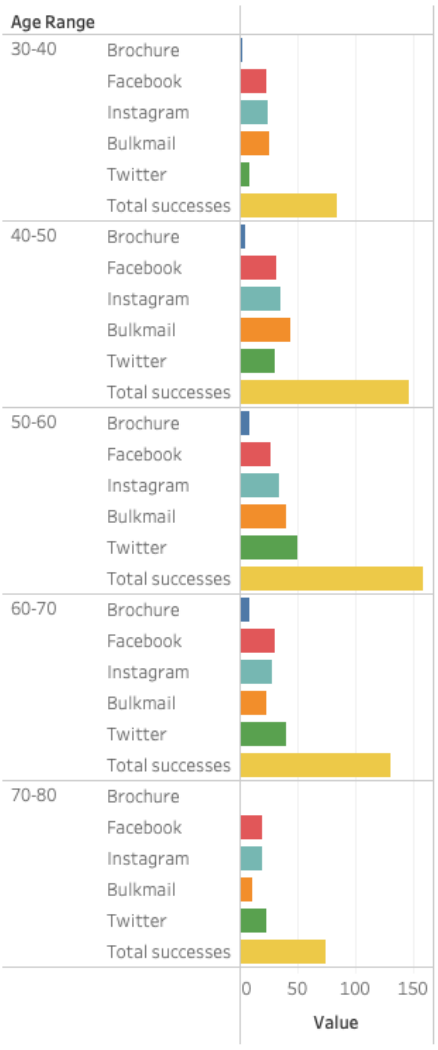
Geographic Advertising Conversions



Marital Advertising Conversions



Age based Advertising Conversions



Recommendations and Further Actions

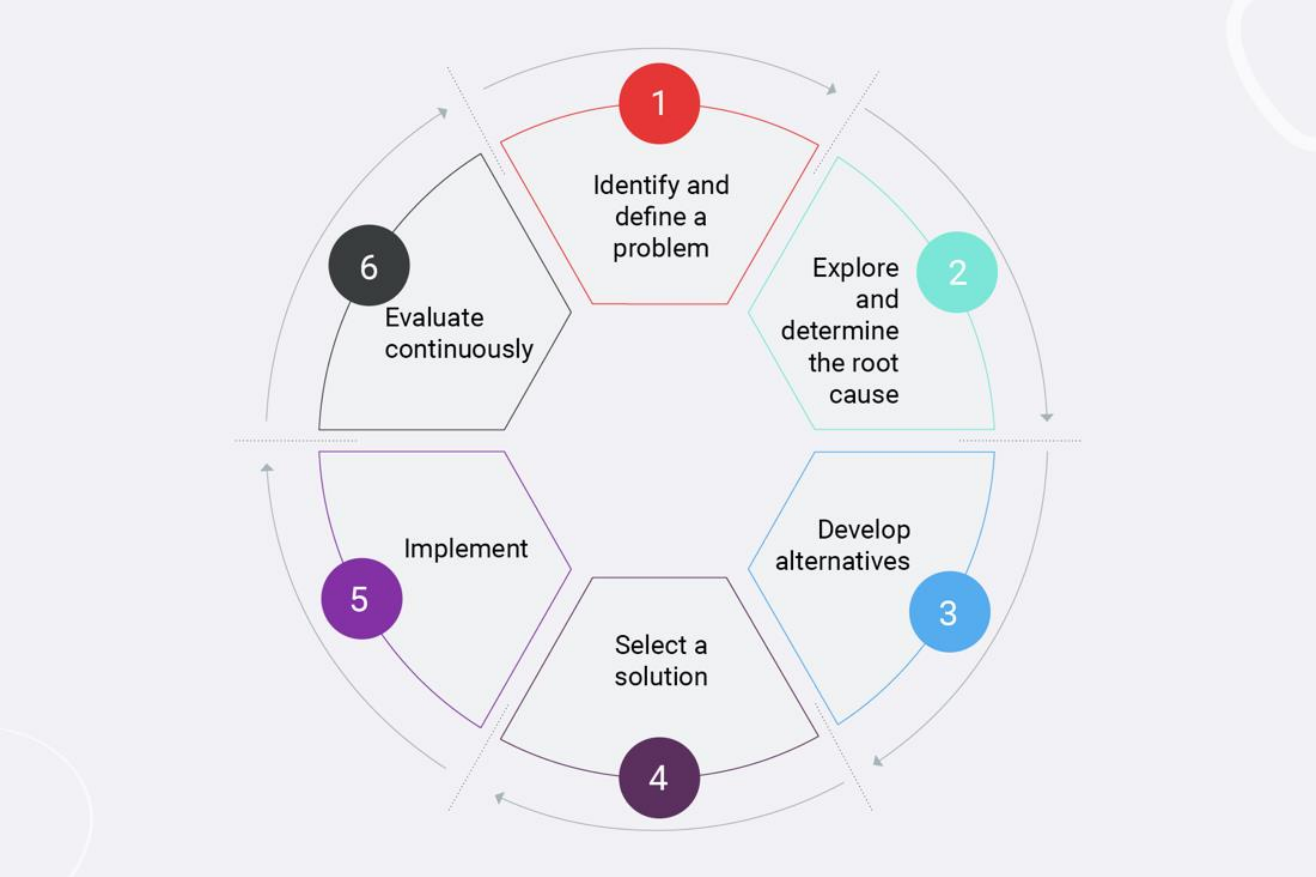
In conclusion I would suggest that 2Market focus more online advertising in the areas where they have less market share.

I would also suggest that they look to tighten up the collection of data with regards to geographical anomalies as this will continue to impact reports if not addressed.

If I had more time I would look to dig further into the percentages of the data set and see if sql could give me any further insight.

Appendix:

Problem solving:



SQL Table creation