

DATA ANALYSIS REPORT

2Market Review & Analysis by Dhru Mistry

Introduction

In this report, I will talk about my process and findings behind my assignment I was tasked with by 2Market, which is a global supermarket that sells products online and instore, the main task is that 2Market want to understand their customer purchase behaviour, the way to do this successfully is to look at 3 critical areas: The demographic of their customers; Which advertising channels seem to sell the best and if that varies based on demographic. An additional question I would like to answer on top of this is whether the future looks good for 2Market or do they need to make some changes, I will try to do this with the help of analysing patterns and trends that I can find.

Analytical Approach

The first step I took was to look at the provided excel sheet with all the data about marketing data and try to clean it. This involved spell checking, searching for blanks + duplicate IDs followed by basic cleaning, which was using the trim function, proper text on all cells, upper country column. Next, I added an Age column as it will help understand the demographic even more. After this, the next step was simplifying the data, by getting rid of unnecessary values and combining columns. The next step was now to visualise the data; this was done through creating pivot tables. Next, was importing this data into SQL, this was done by creating 2 tables, marketing_data and ad_data next created all the columns by inputting all the headings followed by the data types then after creating the table, the data was imported. The next step was then to use SQL queries to format, transform, validate and analyse the data to determine which products sell the best and whether that varies based on demographic factors.

Findings and Analysis

From the marketing_data set, the main points we can take away is firstly, that there is strong correlation between age and income in that as a person's age gets higher, the income they get also increases ([Appendix 1](#)). There is, however, some anomalies within the data set for example, 3 people aged over 100 which is highly unreliable and secondly, some people in their early 30s earning a substantially higher amount than the next 30 years ahead of them with someone earning \$666k also. Next answering the questions that were asked, the average age per marital status ([Appendix 2](#)) is Widower being 65.24 yrs and the lowest being Single at 52.40 yrs to add to this the average age of customers who earn between \$90,000 and \$100,000 is 52.76 yrs. Some other observations I made was looking at average income per country, from this ME has the highest average at \$57,880 and the lowest is India with \$49,016. A question that does need to be asked is are these income figures before or after tax? As the data includes figures from different countries, tax brackets and percentage will differ and will affect the results massively. The last observation was looking at how education can affect the income and age correlation, from this I was able to see that the people who went through higher education (Graduated, Master, PhD)

had a stable salary that progressively increased with age. Whereas people who had Basic education face a very unstable income, constantly fluctuating with age ([Appendix 3](#)).

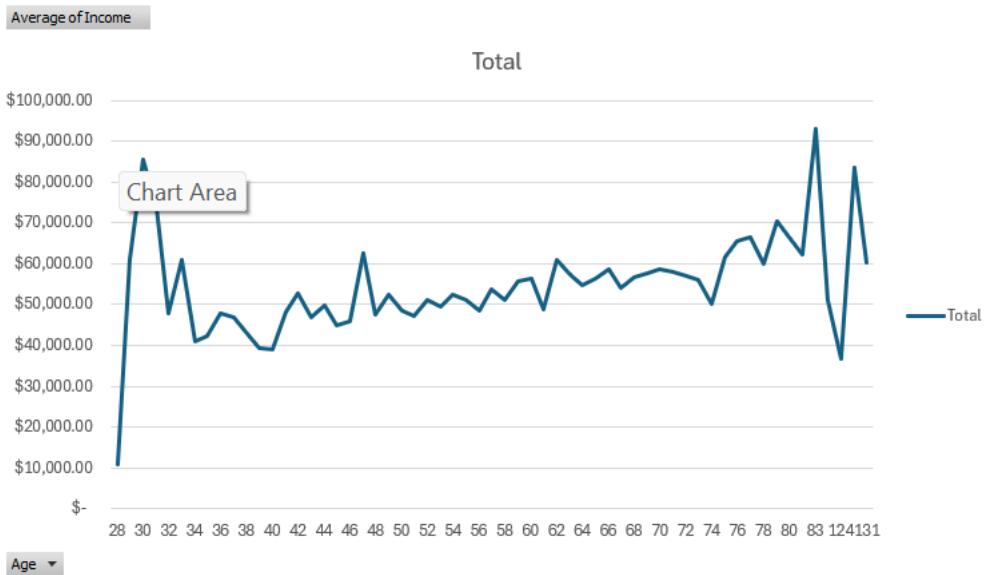
When it came to SQL, the answers we gained was that Spain has the highest total spending per country with \$659557 but this could be due to the fact out of all 2216 people registered, 1093 are from Spain for comparison the 2nd highest is South Africa with 337 people. [Appendix 4](#) was the query used to gather this information. Next was total spend per product per country, which is shown in [Appendix 5](#). Next, the most popular product in each country, each country follows the same trend of having the most popular products being liquids, meats and commodities followed very closely behind by fish, except for ME but is probably due to their only being 3 customers, the query used for this is shown in [Appendix 5](#) also. Next was most popular products depending on martial status, it follows the same trend as before of liquid being most popular followed by meat then commodities, the query is shown in [Appendix 6](#). Finally, the last question is most popular based on whether there is children in the home or not, as before, it follows the same trend of Liquids being most popular followed by meats and then commodities, a better insight into this though would be if we knew the religion of the families as some households may be vegetarian or have a restricted diet, the query for this is shown in [Appendix 7](#).

Dashboard Design & Development

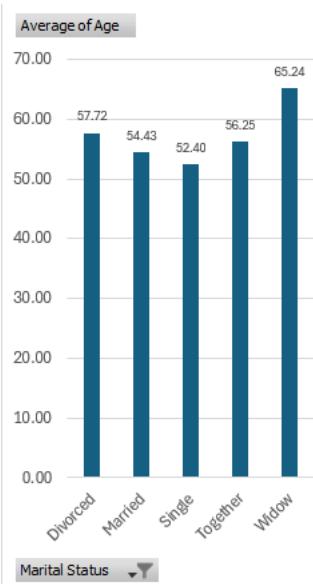
To visualise the data, tableau was used, to start both marketing data and ad data sets were imported, and a left outer join was used to connect them. The development started with planning what the critical pieces of data I wanted to include in my sheets and eventually, my dashboard. The first sheet was KPI, which is probably the most important pieces of data for the stakeholders. Next was total sales by region, which I used a geographical map, as it is aesthetically pleasing also easy & simple to read also, it will be used as an interactive filter within the dashboard as 'Country' is probably the most important demographic for the stakeholders and so being able to filter & sort by region is crucial helps making it easier to engage with the dashboard. Next was total spend by the date they registered and lastly, the sales of each individual product by region. Now it was time to produce the dashboard, when it came to the design elements the main idea was to keep it simple and have the data easily identifiable that's why I used shades of blue and a light grey background as they compliment each other well and aren't strong colours that'll put the person viewing it off.

Appendices

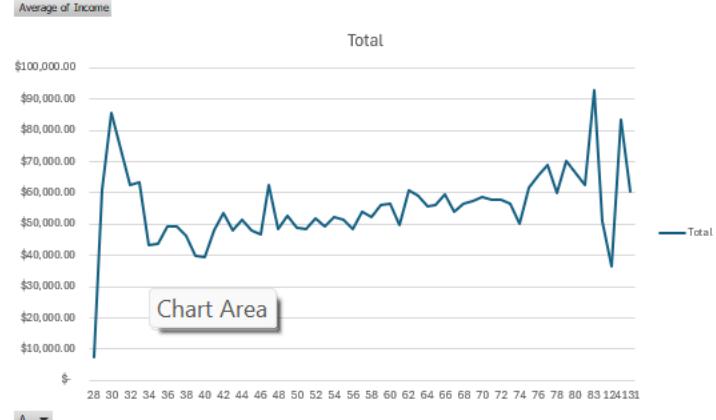
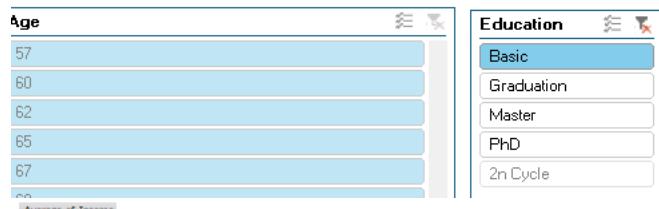
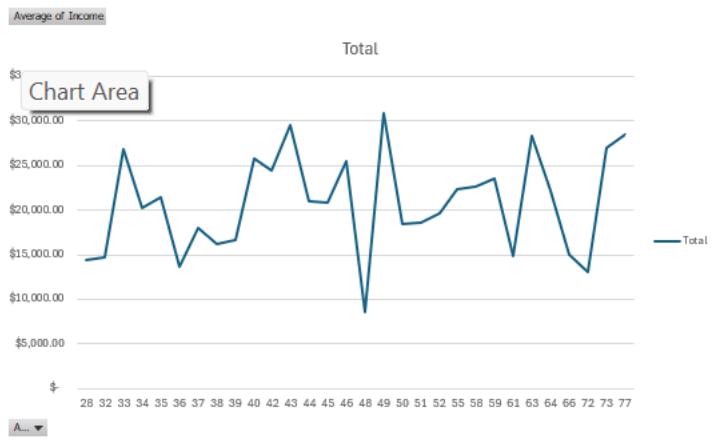
- Appendix 1



- Appendix 2



- Appendix 3



- Appendix 4

```
2
3 ▼ SELECT "Country", SUM("TotalAS")
4   FROM marketing_data
5   GROUP BY "Country";
6
```

Data Output Messages Notifications

A screenshot of a SQL query results table. The table has two columns: 'Country' (character varying) and 'sum' (numeric). The data shows the following rows:

	Country	sum
1	SP	659557.00
2	CA	167403.00
3	AUS	85576.00
4	IND	77806.00
5	US	67546.00
6	ME	3122.00
7	SA	211071.00
8	GER	73198.00

```
7 ▼ SELECT "Country", COUNT("ID")
8   FROM marketing_data
9   GROUP BY "Country";
10
```

Data Output Messages Notifications

A screenshot of a SQL query results table. The table has two columns: 'Country' (character varying) and 'count' (bigint). The data shows the following rows:

	Country	count
1	SP	1093
2	CA	266
3	AUS	147
4	IND	147
5	US	107
6	ME	3
7	SA	337
8	GER	116

- Appendix 5

```

11 v SELECT ("Country"),
12 ROUND(SUM("ASOnLiq"),2) AS ASOnLiq,
13 ROUND(SUM("ASOnVeg"),2) AS ASOnVeg,
14 ROUND(SUM("ASOnNonveg"),2) AS ASOnNonveg,
15 ROUND(SUM("ASOnPes"),2) AS ASOnPes,
16 ROUND(SUM("ASOnChocolates"),2) AS ASOnChocolates,
17 ROUND(SUM("ASOnComm"),2) AS ASOnComm,
18 ROUND(SUM("TotalAS"),2) AS TotalAS
19 FROM marketing_data
20 GROUP BY "Country"
21 ORDER BY TotalAS DESC;

```

Data Output Messages Notifications

	Country character varying	asonliq numeric	asonveg numeric	asonnonveg numeric	asonpes numeric	asonchocolates numeric	asoncomm numeric	totalas numeric
1	SP	336392.00	28288.00	178409.00	40153.00	30134.00	46181.00	659557.00
2	SA	105918.00	8937.00	58398.00	13670.00	9019.00	15129.00	211071.00
3	CA	84066.00	7681.00	45925.00	9980.00	7607.00	12144.00	167403.00
4	AUS	42752.00	3689.00	22328.00	5546.00	4129.00	7132.00	85576.00
5	IND	36236.00	3788.00	23729.00	4818.00	3221.00	6014.00	77806.00
6	GER	36776.00	2980.00	20272.00	4601.00	2801.00	5768.00	73198.00
7	US	32214.00	3034.00	20185.00	4411.00	2863.00	4839.00	67546.00
8	ME	1729.00	8.00	817.00	226.00	122.00	220.00	3122.00

- Appendix 6

```

23 v SELECT ("MaritalStatus"),
24 ROUND(SUM("ASOnLiq"),2) AS ASOnLiq,
25 ROUND(SUM("ASOnVeg"),2) AS ASOnVeg,
26 ROUND(SUM("ASOnNonveg"),2) AS ASOnNonveg,
27 ROUND(SUM("ASOnPes"),2) AS ASOnPes,
28 ROUND(SUM("ASOnChocolates"),2) AS ASOnChocolates,
29 ROUND(SUM("ASOnComm"),2) AS ASOnComm,
30 ROUND(SUM("TotalAS"),2) AS TotalAS
31 FROM marketing_data
32 GROUP BY "MaritalStatus"
33 ORDER BY TotalAS DESC;
34

```

Data Output Messages Notifications

	MaritalStatus character varying	asonliq numeric	asonveg numeric	asonnonveg numeric	asonpes numeric	asonchocolates numeric	asoncomm numeric	totalas numeric
1	Married	256976.00	21981.00	137888.00	30395.00	22926.00	36719.00	506885.00
2	Together	176715.00	14612.00	95374.00	22383.00	15031.00	24754.00	348869.00
3	Single	139126.00	13027.00	87868.00	18704.00	12839.00	20970.00	292534.00
4	Divorced	75364.00	6363.00	34848.00	8130.00	6222.00	10739.00	141666.00
5	Widow	27902.00	2422.00	14085.00	3793.00	2878.00	4245.00	55325.00

- Appendix 7

```
34
35 v SELECT ("NumOfChildren"),
36   ROUND(SUM("ASOnLiq"),2) AS ASOnLiq,
37   ROUND(SUM("ASOnVeg"),2) AS ASOnVeg,
38   ROUND(SUM("ASOnNonveg"),2) AS ASOnNonveg,
39   ROUND(SUM("ASOnPes"),2) AS ASOnPes,
40   ROUND(SUM("ASOnChocolates"),2) AS ASOnChocolates,
41   ROUND(SUM("ASOnComm"),2) AS ASOnComm,
42   ROUND(SUM("TotalAS"),2) AS TotalAS
43 FROM marketing_data
44 GROUP BY "NumOfChildren"
45 ORDER BY TotalAS DESC;
```

Data Output Messages Notifications

	NumOfChildren numeric	asoniq numeric	asonveg numeric	asonnonveg numeric	asonpes numeric	asonchocolates numeric	asoncomm numeric	totalas numeric
1	0	308950.00	33090.00	234758.00	48500.00	33663.00	40661.00	699622.00
2	1	299933.00	21747.00	110820.00	29943.00	22523.00	45274.00	530240.00
3	2	59133.00	3269.00	21504.00	4694.00	3456.00	10586.00	102642.00
4	3	8067.00	299.00	2981.00	268.00	254.00	906.00	12775.00