

2Market Technical Report

Leonard Okhani

Background of the business

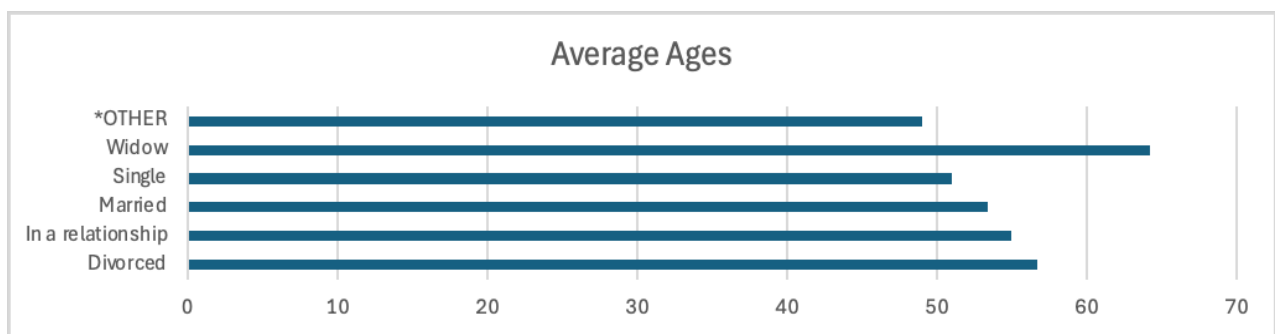
2Market is a global supermarket, which sells products both in-store and online. They want to gain better insights of their customer demographics, purchasing behavior and advertising channel effectiveness. These are the questions that are going to be answered during the investigations:

- Which customer demographics do the products appeal to the most?
- Do the sales of the 2Market products vary based on the demographics?
- How effective are the current advertising channels used by 2Market?

Analytical Approach

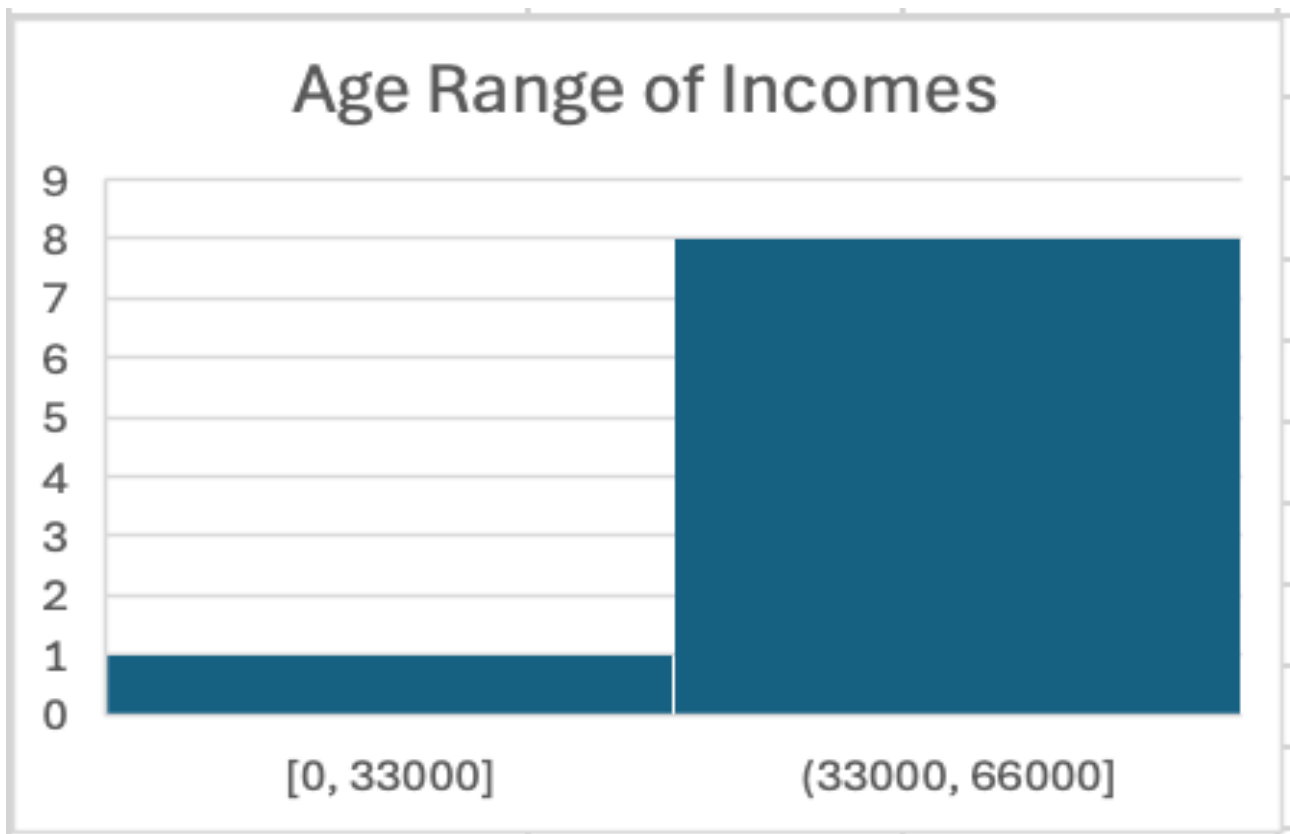
Excel:

Excel was used to clean the data by checking for missing, duplicate and outlier values and changing the data type for the date column. A new column, “Age” was added, and this was calculated by subtracting the “Year_Birth” values from the current year. The average age for each marital status is shown below in the graph:



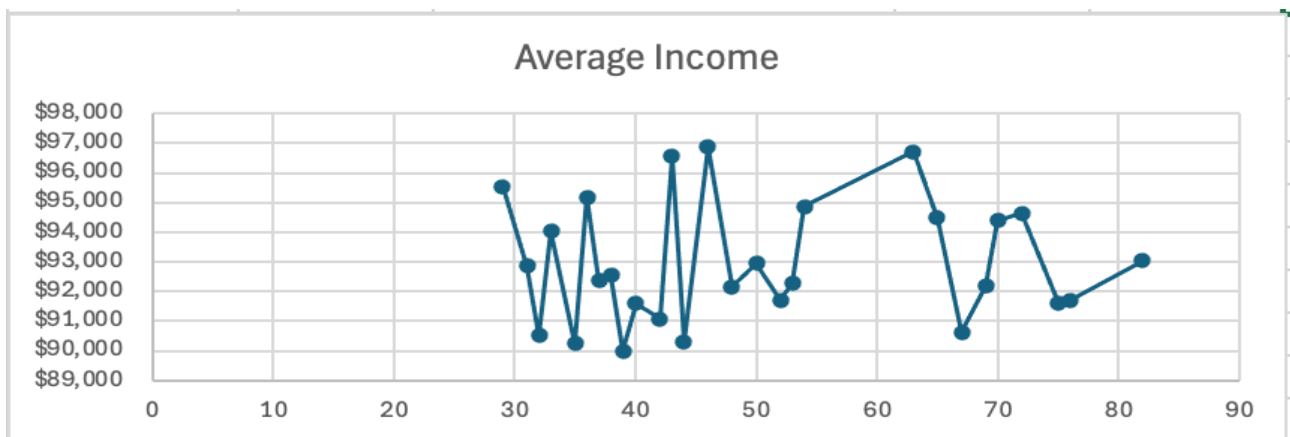
There were changes made to the categories. ‘Alone’ was added to ‘Single’. ‘Together’ was changed into ‘In a relationship’. Both ‘Absurd’ and ‘YOLO’ were put in a new category – ‘*OTHER’, due to both being uncommon statuses. A decision was made to keep the outlier values, such as “Age” being over 100 and the values from ‘Absurd’ and ‘YOLO’ due to increase of the variability it brings to the data.

To see if there was a trend of customer income with the progressing age of the customers, a histogram was made. The ages went in specific ranges. Below is the result of the histogram:



With this histogram being left-skewed, it shows that overall, as age increases, so does the average income.

By applying PivotTables, worked out the age range of customers who earn an income between \$90,000 and \$100,000. Below is the graph for the average income for per age of the customers:



Based on this graph, there is no relationship between the customers' ages and their incomes.

SQL:

Alongside Excel, SQL was also used to analyze the data to see if the customer demographics would affect the product sales. A database was created in Postgres and inside, the following syntax was designed:

```
Query Query History
1  CREATE TABLE public.marketing_data
2  (
3  "ID" BIGSERIAL PRIMARY KEY,
4  "Year_Birth" INTEGER,
5  "Education" VARCHAR(100),
6  "Marital_Status" VARCHAR(100),
7  "Income" INTEGER,
8  "Kidhome" INTEGER,
9  "Teenhome" INTEGER,
10 "Dt_Customer" DATE,
11 "Recency" INTEGER,
12 "AmtLiq" INTEGER,
13 "AmtVege" INTEGER,
14 "AmtNonVeg" INTEGER,
15 "AmtPes" INTEGER,
16 "AmtChocolates" INTEGER,
17 "AmtComm" INTEGER,
18 "NumDeals" INTEGER,
19 "NumWebBuy" INTEGER,
20 "NumWalkinPur" INTEGER,
21 "NumVisits" INTEGER,
22 "Response" INTEGER,
23 "Complain" INTEGER,
24 "Country" CHAR(5),
25 "Count_success" INTEGER,
26 "Age" INTEGER);
27
28 CREATE TABLE public.addata
29 (
30 "ID" BIGSERIAL PRIMARY KEY,
31 "Bulkmail_ad" INTEGER,
32 "Twitter_ad" INTEGER,
33 "Instagram_ad" INTEGER,
34 "Facebook_ad" INTEGER,
35 "Brochure_ad" INTEGER);
```

With the above syntax, two tables were generated - one with the data for the customer demographics and products sales and another one with the advertising channels data.

With the data from the first table, worked out the total spend per product per country with the and later the total spends per country:

```
Query Query History
1  SELECT "Country", SUM("AmtLiq") AS "Liq", SUM("AmtVege") AS "Vege",
2  SUM("AmtNonVeg") AS "NonVeg", SUM("AmtPes") AS "Pes", SUM("AmtChocolates") AS "Chocolates",
3  SUM("AmtComm") AS "Comm"
4  FROM marketing_data
5  GROUP BY "Country"
```

	Country character (5) 🔒	Liq bigint 🔒	Veg bigint 🔒	NonVeg bigint 🔒	Pes bigint 🔒	Chocolates bigint 🔒	Comm bigint 🔒
1	SP	336392	28288	178409	40153	30134	46181
2	CA	84066	7681	45925	9980	7607	12144
3	IND	36236	3788	23729	4818	3221	6014
4	AUS	42752	3689	22328	5546	4129	7132
5	US	32214	3034	20185	4411	2863	4839
6	ME	1729	8	817	226	122	220
7	SA	105918	8937	58398	13670	9019	15129
8	GER	36776	2980	20272	4601	2801	5768

Query	Query History	
1	SELECT "Country", SUM("AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtPes" + "AmtChocolates" +	
2	"AmtComm") AS "Total Spend"	
3	FROM marketing_data	
4	GROUP BY "Country"	
5	ORDER BY "Total Spend" DESC;	

	Country character (5) 🔒	Total Spend bigint 🔒
1	SP	659557
2	SA	211071
3	CA	167403
4	AUS	85576
5	IND	77806
6	GER	73198
7	US	67546
8	ME	3122

In addition, an investigation was made to see whether the types of marital statuses affected the products sales as customers could be buying products for their partners rather than themselves.

Query	Query History	
1	SELECT "Marital_Status", SUM("AmtLiq") AS "Liq", SUM("AmtVege") AS "Vege",	
2	SUM("AmtNonVeg") AS "NonVeg", SUM("AmtPes") AS "Pes", SUM("AmtChocolates") AS "Chocolates",	
3	SUM("AmtComm") AS "Comm"	
4	FROM marketing_data	
5	GROUP BY "Marital_Status"	

	Marital_Status character varying (100) 🔒	Liq bigint 🔒	Veg bigint 🔒	NonVeg bigint 🔒	Pes bigint 🔒	Chocolates bigint 🔒	Comm bigint 🔒
1	In a relationship	176715	14612	95374	22383	15031	24754
2	*OTHER	1355	175	725	419	67	492
3	Married	256976	21981	137888	30395	22926	36719
4	Widow	27902	2422	14085	3793	2878	4245
5	Single	137771	12852	87143	18285	12772	20478
6	Divorced	75364	6363	34848	8130	6222	10739

Furthermore, SQL was used to also see which products were popular based on the whether customers have kids, teens, both or none at home due to the possibility of young kids living with their parents affecting the products they buy.

Query	Query History
1	SELECT SUM("AmtLiq") AS "Liq", SUM("AmtVege") AS "Veg", SUM("AmtNonVege") AS "NonVeg",
2	SUM("AmtPes") AS "Pes", SUM("AmtChocolates") AS "Chocolates", SUM("AmtComm") AS "Comm"
3	FROM marketing_data
4	WHERE "Kidhome" = '0' AND "Teenhome" = '0'
5	UNION
6	SELECT SUM("AmtLiq") AS "Liq", SUM("AmtVege") AS "Veg", SUM("AmtNonVege") AS "NonVeg",
7	SUM("AmtPes") AS "Pes", SUM("AmtChocolates") AS "Chocolates", SUM("AmtComm") AS "Comm"
8	FROM marketing_data
9	WHERE "Kidhome" = '0' AND "Teenhome" > '0'
10	UNION
11	SELECT SUM("AmtLiq") AS "Liq", SUM("AmtVege") AS "Veg", SUM("AmtNonVege") AS "NonVeg",
12	SUM("AmtPes") AS "Pes", SUM("AmtChocolates") AS "Chocolates", SUM("AmtComm") AS "Comm"
13	FROM marketing_data
14	WHERE "Kidhome" > '0' AND "Teenhome" = '0'
15	UNION
16	SELECT SUM("AmtLiq") AS "Liq", SUM("AmtVege") AS "Veg", SUM("AmtNonVege") AS "NonVeg",
17	SUM("AmtPes") AS "Pes", SUM("AmtChocolates") AS "Chocolates", SUM("AmtComm") AS "Comm"
18	FROM marketing_data
19	WHERE "Kidhome" > '0' AND "Teenhome" > '0'

	Liq bigint 🔒	Veg bigint 🔒	NonVeg bigint 🔒	Pes bigint 🔒	Chocolates bigint 🔒	Comm bigint 🔒
1	308950	33090	234758	48500	33663	40661
2	53872	2698	19766	3723	2999	9305
3	41990	5156	25178	7487	4819	11084
4	271271	17461	90361	23695	18415	36377

- 1 – Customers without kids or teenagers.
- 2 – Customers with kids and teenagers.
- 3 – Customers with kids.
- 4 – Customers with teenagers.

The second part of the SQL analysis involved joining the marketing and advertising tables. As there were matching values between the two tables, the INNER JOIN was the best option to join them.

Query
Query History

```

1 CREATE TABLE advertising_marketing AS
2 SELECT *
3 FROM marketing_data AS mk
4 INNER JOIN addata AS ad
5 USING ("ID");

```

The new table was given the name advertising_marketing. The “ID” column was used to join the tables due to being the only common column between them - making it the primary key.

With the new generated table, another analysis was made to see which social media platform was the most effective in terms of advertising for each of the countries.

Query
Query History

```

1 SELECT "Country", SUM("Bulkmail_ad") AS "Bulkmail", SUM("Twitter_ad") AS "Twitter",
2 SUM("Instagram_ad") AS "Instagram", SUM("Facebook_ad") AS "Facebook",
3 SUM("Brochure_ad") AS "Brochure"
4 FROM advertising_marketing
5 GROUP BY "Country"

```

	Country character (5)	bulkmail bigint	twitter bigint	instagram bigint	facebook bigint	brochure bigint
1	SP	83	87	89	76	16
2	CA	18	24	21	18	6
3	IND	13	10	6	7	2
4	AUS	9	6	12	7	0
5	US	8	6	5	7	0
6	ME	1	0	0	0	0
7	SA	21	20	21	20	4
8	GER	10	11	8	7	2

Additionally, an inquiry was made to see which advertising channel was more popular based on the marital status type.

Query
Query History

```

1 SELECT "Marital_Status", SUM("Bulkmail_ad") AS "Bulkmail", SUM("Twitter_ad") AS "Twitter",
2 SUM("Instagram_ad") AS "Instagram", SUM("Facebook_ad") AS "Facebook",
3 SUM("Brochure_ad") AS "Brochure"
4 FROM advertising_marketing
5 GROUP BY "Marital_Status"

```

	Marital_Status character varying (100) 🔒	bulkmail bigint 🔒	twitter bigint 🔒	instagram bigint 🔒	facebook bigint 🔒	brochure bigint 🔒
1	In a relationship	37	42	44	32	12
2	*OTHER	0	0	1	1	0
3	Married	63	62	66	62	7
4	Widow	4	10	7	5	1
5	Single	39	32	31	30	5
6	Divorced	20	18	13	12	5

The next step was to discover how effective the advertising channels were in bringing in the revenues.

Query	Query History	
1	SELECT "Country",	
2	SUM(CASE WHEN "Brochure_ad" > 0 THEN "AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtChocolates"	
3	+ "AmtPes" + "AmtComm" ELSE 0 END) AS "Brochure_Spent",	
4	SUM(CASE WHEN "Bulkmail_ad" > 0 THEN "AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtChocolates"	
5	+ "AmtPes" + "AmtComm" ELSE 0 END) AS "Bulkmail_Spent",	
6	SUM(CASE WHEN "Facebook_ad" > 0 THEN "AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtChocolates"	
7	+ "AmtPes" + "AmtComm" ELSE 0 END) AS "Facebook_Spent",	
8	SUM(CASE WHEN "Instagram_ad" > 0 THEN "AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtChocolates"	
9	+ "AmtPes" + "AmtComm" ELSE 0 END) AS "Instagram_Spent",	
10	SUM(CASE WHEN "Twitter_ad" > 0 THEN "AmtLiq" + "AmtVege" + "AmtNonVege" + "AmtChocolates"	
11	+ "AmtPes" + "AmtComm" ELSE 0 END) AS "Twitter_Spent"	
12	FROM advertising_marketing	
13	GROUP BY "Country"	

	Country character (5) 🔒	Brochure_Spent bigint 🔒	Bulkmail_Spent bigint 🔒	Facebook_Spent bigint 🔒	Instagram_Spent bigint 🔒	Twitter_Spent bigint 🔒
1	SP	21808	55250	114018	147622	102271
2	CA	6397	16457	27650	30954	25261
3	IND	3736	9779	8963	9444	8129
4	AUS	0	9866	10480	18505	7915
5	US	0	3986	8806	7048	4698
6	ME	0	874	0	0	0
7	SA	4282	13287	31211	35738	26748
8	GER	3007	7949	9649	12551	11538

Dashboard Design and Development

To support 2Market, 3 dashboards were designed with each supporting the answers to the 3 questions asked above.

The first dashboard was designed with 5 visualizations to give insights of the most popular customers demographics. The visualizations include:

- 2 bar graphs – one to show the total number of customers in each age group and another to show the total number of customers with specific income ranges.
- A pie chart showing the proportions of the customers' education level.
- A map with the 8 countries, the number of customers in each and their median incomes.
- A highlight table, which shows the number of customers in each marital status.

The second dashboard was designed with 5 visualizations to give insights of which products are the most popular and if their sales vary on the customers demographics. The visualizations include:

- 3 text tables – two to show the amount spent per product and total per country and marital status respectively and one to show the amount spent per product and total with the conditions of the customers having children or not.
- A side-by-side bar to see which products were the most popular in each age group.
- A bar chart showing the total sales for each product.

The third dashboard was designed with 5 visualizations to give insights of how effective the advertisement channels were for improving the product sales. The visualizations include:

- 3 text tables – two show the number of successful conversions per channel and total per country and marital status respectively and one to show the amount spent per channel and total per country.
- A side-by-side bar to compare the advert sales and the total sales.
- A bar chart the showing the number of successful conversions per channel.

Insights and Recommendations

Customers Insights:

- Average age of customers: 54.
- Average age of customers with \$90K-\$100K income: 52.
- Country with the most customers: Spain.
- Country with the least customers: Montenegro.
- Marital status with the most customers: Married.
- Marital status with the least customers: *OTHER.

Products Insights:

- Most popular product (countries & statuses): Liquor.
- Least popular product (countries & statuses): Vegetables.
- Highest revenue-generating product: Liquor.
- Lowest revenue-generating product: Vegetables.

Advertisements Insights:

- Most successful converted channel: Twitter.
- Least successful converted channel: Brochure.
- Highest revenue-generating channel: Instagram.
- Lowest revenue-generating channel: Brochure.

Recommendations:

- Customers: As Spain has been the most successful country, 2Market should apply their success plans to other locations. In terms of customers, appeal to the ones who are aged 40-69, have no children, earn high-median incomes and has a partner/spouse.
- Products: Promote sales with liquor and meat and stop all sales with chocolate and vegetables.
- Advertising channels: Stop using Brochure immediately but continue with the other 4 advertising channels, particularly the social media platforms (Instagram, Facebook, Twitter). As Instagram is the channel that brought in the most revenues, it would be advantageous to make it the main advertising channel of the company.