2MARKET CUSTOMERS - EXPLORATORY ANALYSIS AND INSIGHTS

Author: Neetu Thomas

4 September 2022

This page intentionally left blank

Contents

Abbreviations

MC	Marketing Campaign
SQL	Structured Query Language
KPI	Key Performance Indicator

1 Introduction

2Market is a global supermarket and is going to roll out a marketing campaign soon. They want to inform their campaign strategy through analysis of their customer database. The goal is to understand their customer purchase behaviour to find the best-selling products, the best advertising channel and whether these are based on customer demographics. 2market aims to focus the marketing campaign on the right target audience and improve their sales and thereby increase their revenue.

2market has two customer datasets. The first one is the marketing dataset that contains customer demographics and as well as their purchase history of various products. The second one is the advertising dataset which contains the number of purchases influenced by the different advertising channels.

2 Analytical Approach

2.1 Analytical approach in Excel

The data set marketing_data.csv and ad_data.csv was cleaned and validated in excel to ensure the data adheres to the characteristics of data quality. Data checked for:

- Missing values - None found
- Duplicates -47* found and removed
- Accuracy - Age>100 is assumed as outliers and 3 records removed
- - 'YOLO' and 'Absurd' marital status categories not considered

* Initially no duplicates were found. On further sorting and filtering, found 201 records have the same values in all columns except ID column out which 47 are from the same country. Excluded 47 records.

Pivot table function used to aggregate and analyse the data.

Subtotal formula gave Average age of customers as 52.

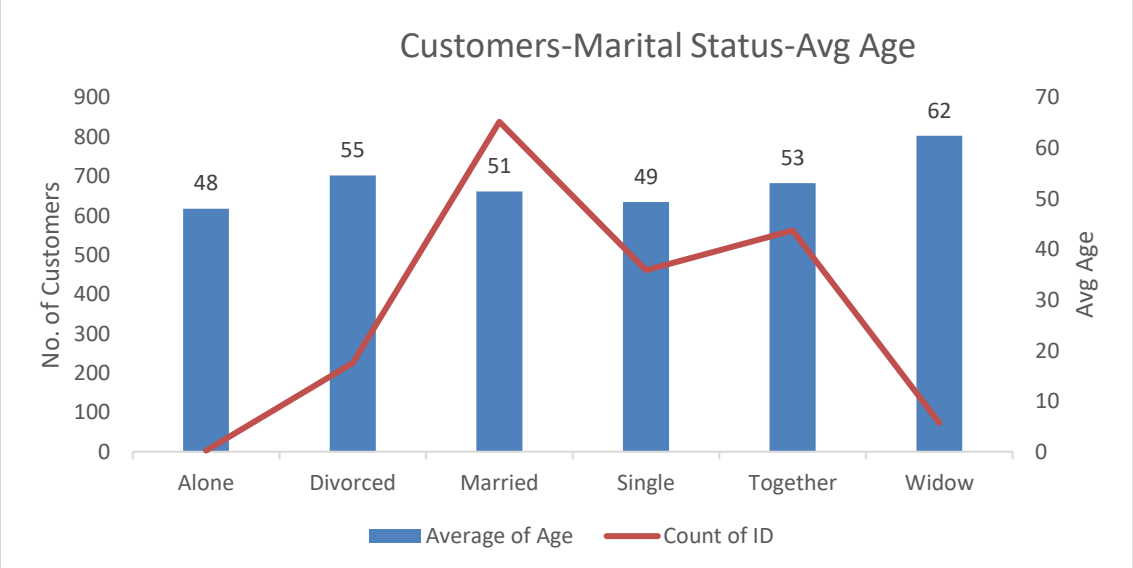


Figure 1- Average age based on Marital Status

Highest average age is 62 in the 'Widow' category. Lowest is 48 of Alone 'category. Most of the customers belong to the Divorced, Married, Single, together groups.

Age bins created to analyse the trends in income of the customers- Linear increase in income with progressing age of customers above 30 years seen.

Age group below 30 years had the third highest average income

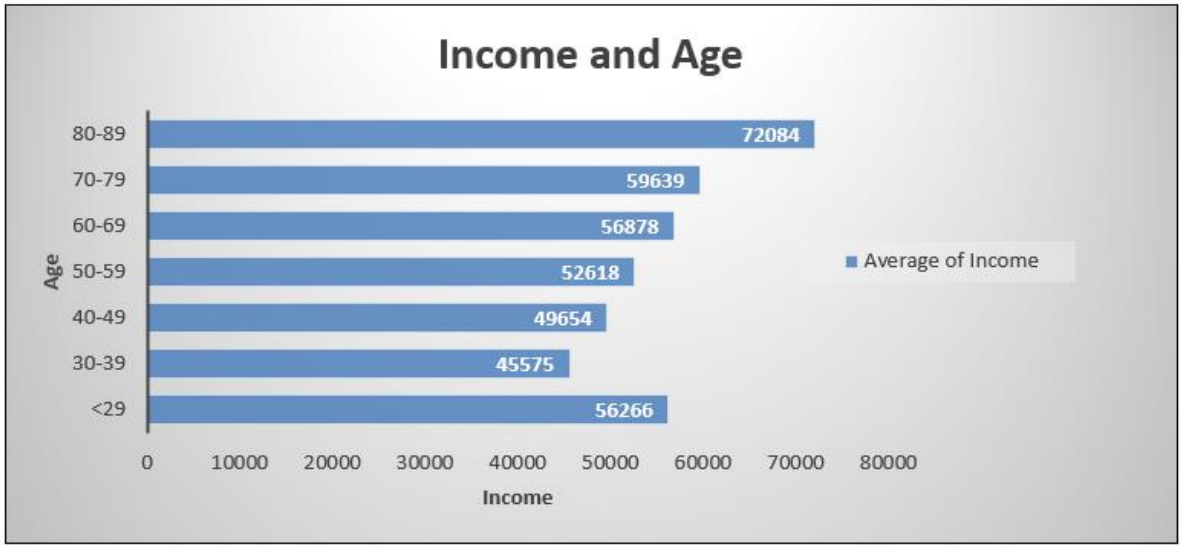


Figure 2- Average income in different age ranges

The average age based on the annual income, the $90000-$100000 range, age is 50.

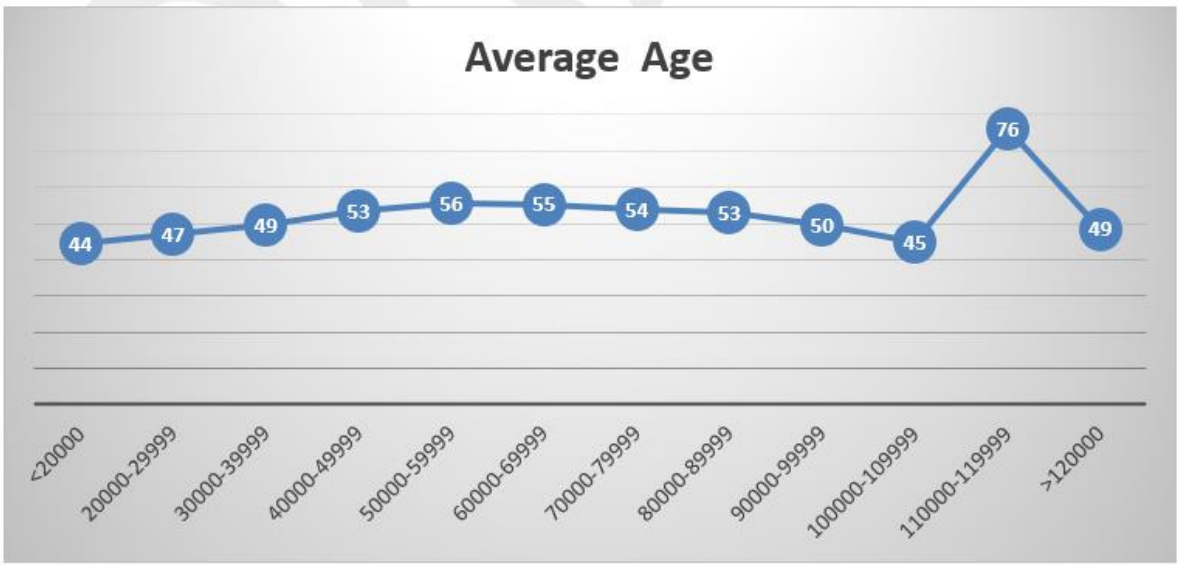


Figure 3 – Average age across income bands

The total and average spent in different age groups and income bands need to be explored further to derive trends and insights.

2.2 Analytical approach in SQL:

The datasets analysed in SQL to determine the best-selling products and whether this is based on demographics.

Created Tables with the columns names corresponding to the csv files and imported using pgAdmin into the PostgreSQL public database server.

The total spent per country determined.

```
SELECT "Country",SUM("AmtLiq" + "AmtVege" + "AmtNonVeg" + "AmtPes" + "AmtChocolates" + "AmtComm")AS total_spent,COUNT("ID")
FROM public.marketing_data
GROUP BY "Country"
ORDER BY total_spent DESC;
```

	Country character varying (5)	total_spent bigint	count bigint
1	SP	634298	1052
2	SA	209323	334
3	CA	167077	263
4	AUS	84970	146
5	IND	76507	145
6	GER	73198	116
7	US	67546	107
8	ME	3122	3

Figure 4 – Total Sales across countries

Further analysis required to calculate the average each customer spends and see whether there is any difference across countries

Alcoholic Beverages are the most popular product across all countries and not dependent on demographics

	Country character varying (5)	alcoholic_beverages bigint	meat_products bigint	fish_products bigint	vegetables bigint	chocolates bigint	commodities bigint
1	SP	323544	171786	38312	26964	29005	44687
2	SA	105026	57977	13621	8887	8921	14891
3	CA	83893	45855	9963	7678	7592	12096
4	AUS	42424	22204	5534	3680	4105	7023
5	GER	36776	20272	4601	2980	2801	5768
6	IND	35889	23344	4662	3588	3092	5932
7	US	32214	20185	4411	3034	2863	4839
8	ME	1729	817	226	8	122	220

Figure 5- Products and sales across countries

Spending habits of customers with & without kids at home analysed-

1) Customers with no kids or teenagers at home have a higher spending per month.
2) Chocolates is more popular among the no kids-no teens groups.

	alcoholic_beverages bigint	vegetables bigint	meat_products bigint	fish_products bigint	chocolates bigint	commodities bigint	total_spent bigint	count_of_kids numeric	count_of_teens numeric
1	301718	32053	229423	47054	32772	39750	682770	0	0
2	254616	16488	84944	22194	17459	33972	429673	0	609
3	39676	4774	24174	7180	4631	10520	90955	489	0
4	44115	2335	16207	3397	2674	8123	76851	357	357
5	12287	621	4004	1011	574	1711	20208	0	60
6	5796	270	2312	180	212	606	9376	21	42
7	2246	29	661	86	42	298	3362	56	28
8	1041	249	715	228	137	476	2846	34	0

Figure 6 – Sales based on customers with/ without children

To understand which advertising channel seems to be the most effective based on customer demographics, the marketing_data and ad_data tables are joined for in depth analysis.

INNER JOIN is performed on these 2 tables

```
SELECT "Country", SUM("AmtLiq" + "AmtVege" + "AmtNonVeg" + "AmtPes" + "AmtChocolates" + "AmtComm")AS total_spent,
       SUM("Twitter_ad")AS twitter_ad, SUM("Instagram_ad") AS instagram_ad,SUM("Facebook_ad")AS facebook_ad,
       SUM("Twitter_ad" + "Instagram_ad"+"Facebook_ad")AS total_socialmedia_ads
FROM public.marketing_data m
JOIN public.ad_data a ON m."ID" = a."ID"
GROUP BY "Country"
ORDER BY total_spent DESC;
```

	Country character varying (5)	total_spent bigint	twitter_ad numeric	instagram_ad numeric	facebook_ad numeric	total_socialmedia_ads numeric
1	SP	634298	85	86	74	245
2	SA	209323	20	21	20	61
3	CA	167077	24	21	18	63
4	AUS	84970	6	12	7	25
5	IND	76507	10	6	7	23
6	GER	73198	11	8	7	26
7	US	67546	6	5	7	18
8	ME	3122	0	0	0	0

Figure 7- Social Media ads and country sales

Further analysis required to determine influence of social media ads age wise and country wise

3 Dashboard Design and Development

To analyse the trends and patterns the marketing and advertising datasets were loaded to Tableau and the 2 tables was connected using a left join with live connection.

Metrics - age, marital status, income, no of kids and country used to analyse

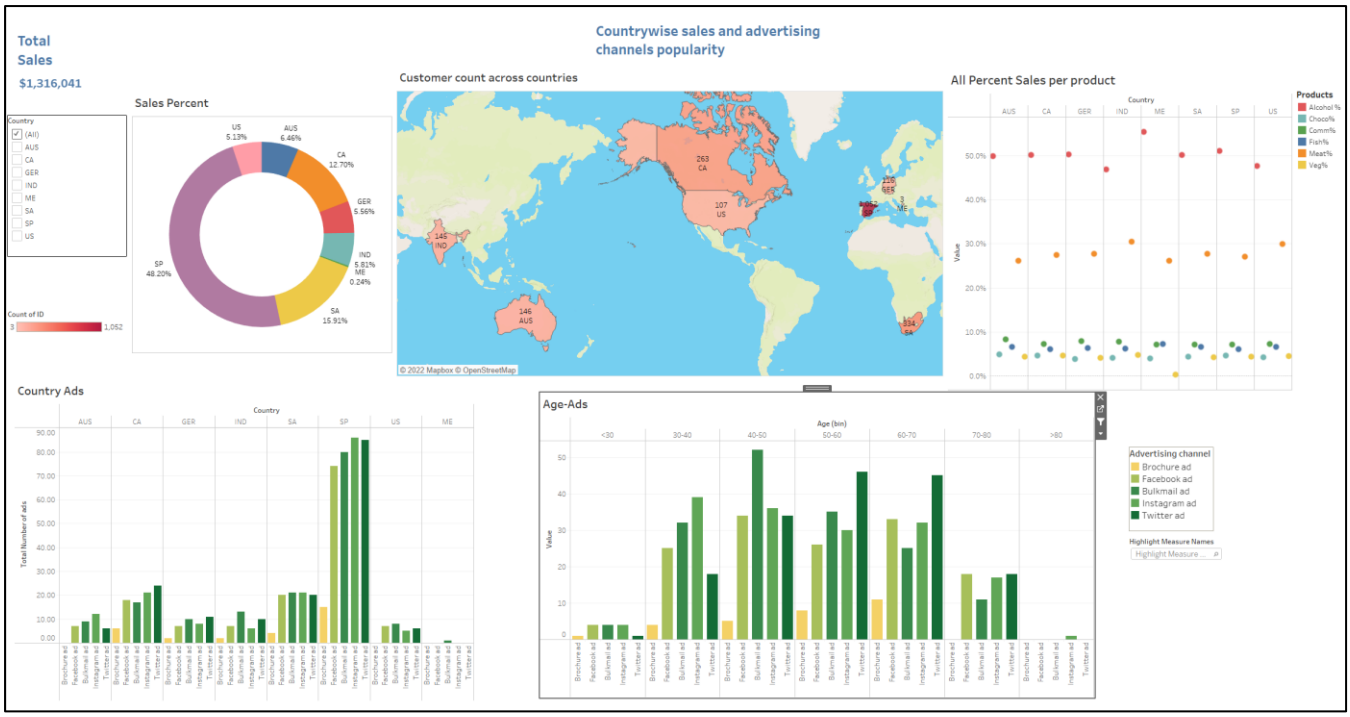
The spending habits of the customers

Effective advertising channel

Two dashboards created to graphically represent the data

Dashboard 1 - Customer demographics and Purchase Behaviour

Dashboard 2 – Revenue & effective advertising channel



3.1 Data Aggregation/design:

- Marital status 'Absurd' and 'YOLO' excluded.
- All labels customized and formatted to improve readability.
- Calculated field created as required (Example: 'Total_spent', '%Alcohol' etc).
- Age bins created.
- Group function used to combine the kids/teen category.
- Map location for SA corrected to South Africa from Saudi Arabia.

The views are interactive and clicking on the pictorial image further filters the dataset.

The dashboard contains visualisations in the form of column clusters, trend lines, butterfly pie and donut charts. Interactive features have been included such as filters for age, country.

Where relevant the colours in column clusters have varying intensities depending on how high/low the values are when compared to the dataset.

The views are placed in a logical way to enable the user to get the answer he wants after filtering the data set. The dashboards allow for easy slicing and dicing of the data to look for trends and make data driven decisions.

4 Patterns and trends

Below patterns and trends derived from Dashboard 1&2

- Majority of 2Market customers are married and from Spain.
- We can also see that above 30 years there is a linear increase of the average income with age
- Higher the income band, the more they spend in the supermarket.
- The highest total sales are to customers in the age 40-70 age range.
- Revenue generated from chocolate sales was higher from customers with no kids at home.
- Average revenue from a customer with children is lower than that from a customer without children
- The most popular product is alcohol followed by meat and the least popular amongst most customers is vegetables. The trend does not change with different marital status / countries
- Spain brings in the highest revenue followed by South Africa and then Canada. This can be attributed to the number of customers that 2 Market has in each of these countries.
- Brochure ads is the least effective.
- Bulk mail is still popular.
- Twitter is the most successful ad channel across all countries and ages.

5 Insights/ Recommendations/Conclusion

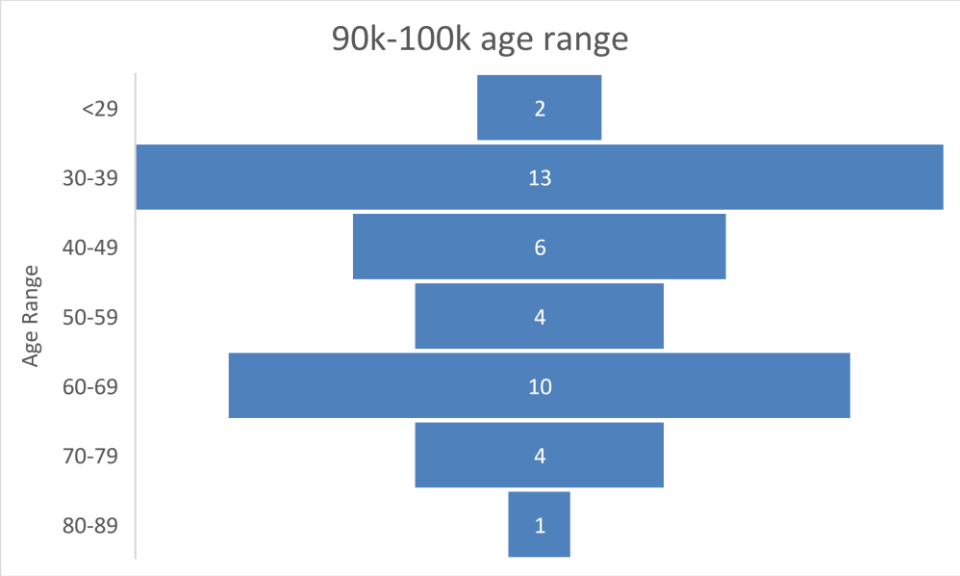
- Alcohol sale is the highest followed by meat irrespective of customer demographics
- The average of the total spent by customers that have kids is significantly less than those with no children – Recommend that the marketing campaign (MC) highlights the goods that this group of people usually spend on
- It is suggested that the next MC looks to improve the customers from India (largest population) and US (largest consumer market)
- Revenue from customers below 30 years low. Recommend MC not to focus on this age group
- MC to be mainly through bulk email and Twitter as these are most effective
- Depending on the business needs in-depth analysis of online and instore sales to be carrier out
- Key milestones and KPI's to be defined and agreed for the project
- Data collection formats and definitions to be agreed upon (eg: marital status, Amt spent - per week/per month/annual)

Appendix

Excel- More Analysis

Pivot table: showing average age based on marital status and the no of customers in each category. YOLO and ABSURD are excluded for further analysis

Row Labels	Average of Age	Count of ID
Absurd	46	2
Alone	48	3
Divorced	55	225
Married	51	838
Single	49	461
Together	53	562
Widow	62	73
YOLO	48	2
Average Age	**52**	**2166**



SQL- More Analysis

The total spent on various products based on marital status

```
SELECT "Marital_Status",SUM("AmtLiq") AS Alcoholic_beverages,SUM("AmtVege")AS Vegetables,SUM("AmtNonVeg")AS Meat_products,
       SUM("AmtPes")AS Fish_products,SUM("AmtChocolates")AS Chocolates,SUM("AmtComm")AS Commodities,count("Marital_Status")
FROM public.marketing_data
WHERE "Marital_Status"  NOT IN('Absurd','YOLO')
GROUP BY "Marital_Status"
ORDER BY "Marital_Status"
```

	Marital_Status character varying	alcoholic_beverages bigint	vegetables bigint	meat_products bigint	fish_products bigint	chocolates bigint	commodities bigint	count bigint
1	Alone	554	12	79	23	21	81	3
2	Divorced	73079	6074	33626	7758	5958	10588	225
3	Married	251717	21491	136121	29683	22453	35998	838
4	Single	135057	12608	85321	17889	12541	20183	461
5	Together	172332	14231	92922	21927	14722	23970	562
6	Widow	27401	2228	13646	3631	2739	4144	73

Joining of marketing_data and ad_data tables. ID is the common attribute in both tables and they are joined using an INNER JOIN

Marital Status and social media relationship

```
SELECT "Marital_Status", SUM("AmtLiq" + "AmtVege" + "AmtNonVeg" + "AmtPes" + "AmtChocolates" + "AmtComm")AS total_spent,
       SUM("Twitter_ad")AS twitter_ad, SUM("Instagram_ad") AS instagram_ad,SUM("Facebook_ad")AS facebook_ad
FROM public.marketing_data m
JOIN public.ad_data a ON m."ID" = a."ID"
WHERE "Marital_Status" NOT IN('YOLO','Absurd')
GROUP BY "Marital_Status"
ORDER BY total_spent DESC;
```

	Marital_Status character varying	total_spent bigint	twitter_ad numeric	instagram_ad numeric	facebook_ad numeric
1	Married	497463	61	65	61
2	Together	340104	42	43	32
3	Single	283599	31	31	30
4	Divorced	137083	18	12	11
5	Widow	53789	10	7	5
6	Alone	770	0	0	0