

2MARKET DATA ANALYSIS PROJECT

Module 1

Identifying the Problem and Opportunities

Considering 2Market's aim from the analysis, we can conclude that we need to identify which advertising strategy works well in which country and which products are most and least in demand based on location and customers' demographic. In other words, which advertising channel to be used in which country, targeting what products and customers' demography.

Questions

- Who are the stakeholders? How do they aim to use the analysis?
- How long has 2Market been operating in different countries for?
- How much has 2Market spent on various channels of advertising by country, age group, etc.?

Assumptions

- The key stakeholders are Senior Marketing Executives and Board members.
- 2Marketing has spent equally on various advertising channels.

Additional Questions:

- How many units per product category are sold? To establish most popular category.

Define Goals

Basic goal for any organisation is to minimize costs and increase profitability. In this scenario, 2Market is aiming to minimize costs by determining the most effective advertising channel to effectively use its funds on marketing and get the best return on investment. Also, they aim to increase their sales by targeting the right products to advertise based on customers' demographics to attract more customers to shop with them, to increase their profits.

Module 2

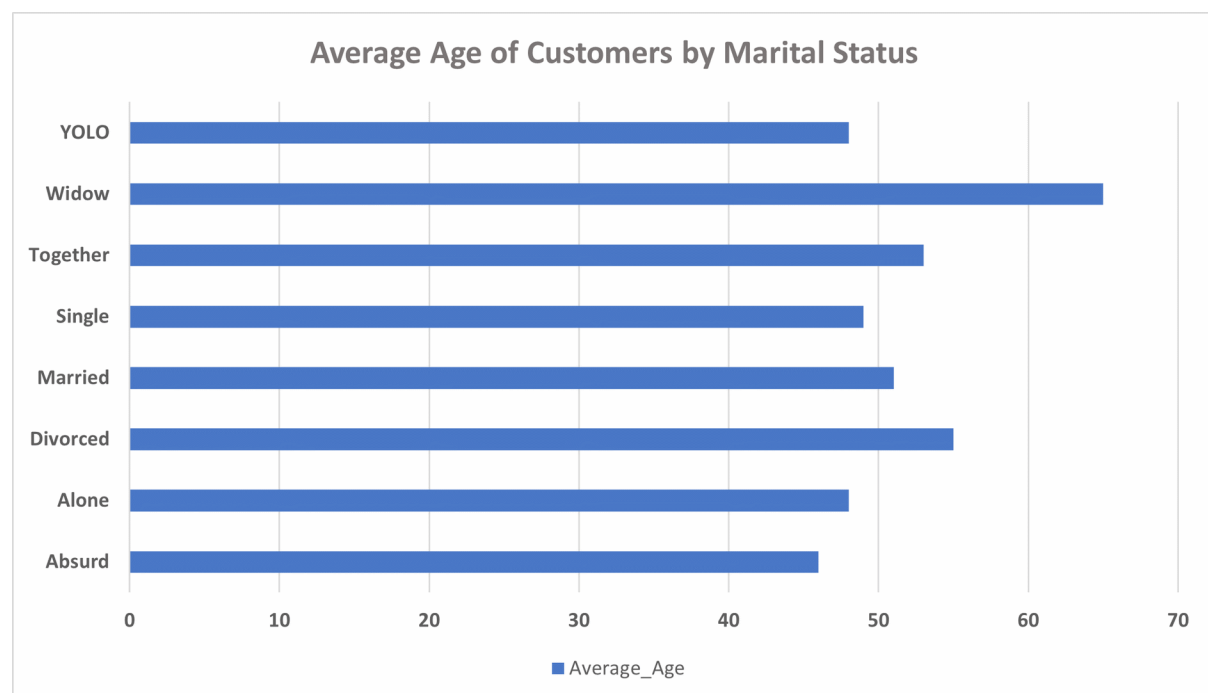
Data in both csv. files cleaned in Excel. Following tasks carried out to check the accuracy and completion of data (not necessarily in described order): Checked for duplicates (none removed as all had different customer ID), formatted data type and removed \$ from income column, trimmed columns, checked for spelling errors, changed country code for Spain, removed outliers, checked blanks, checked for error values, and formatted date column.

Also, added two columns for ease of interpretation and use of data: Customer age and Total spend by Customer.

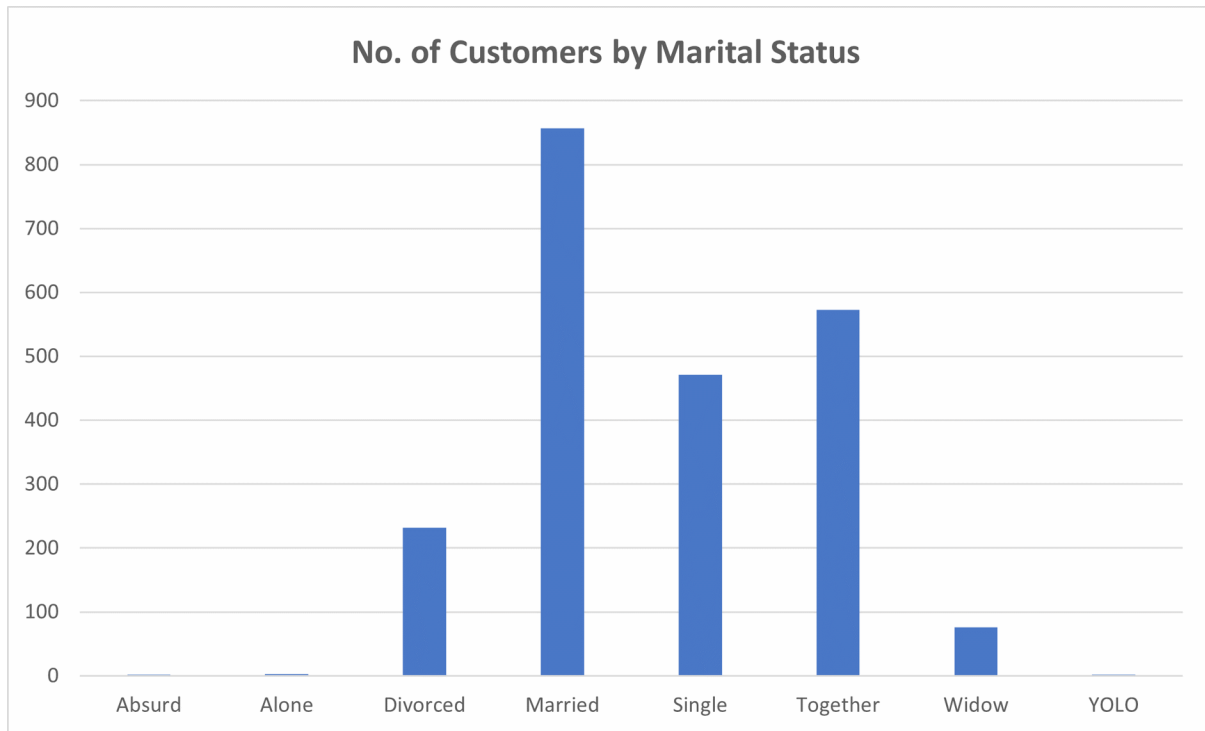
Following are some analyses based on customer's age, income and marital status:

	Year of birth	Avg_age	Number	Percentage
Average Age of Customers	1970	52		
Absurd	1976	46	2	0%
Alone	1974	48	3	0%
Divorced	1968	54	230	10%
Married	1971	51	854	39%
Single	1973	49	470	21%
Together	1969	53	568	26%
Widow	1960	62	76	3%
YOLO	1974	48	2	0%
Income between 90000 and 100000	1972	50	41	2%

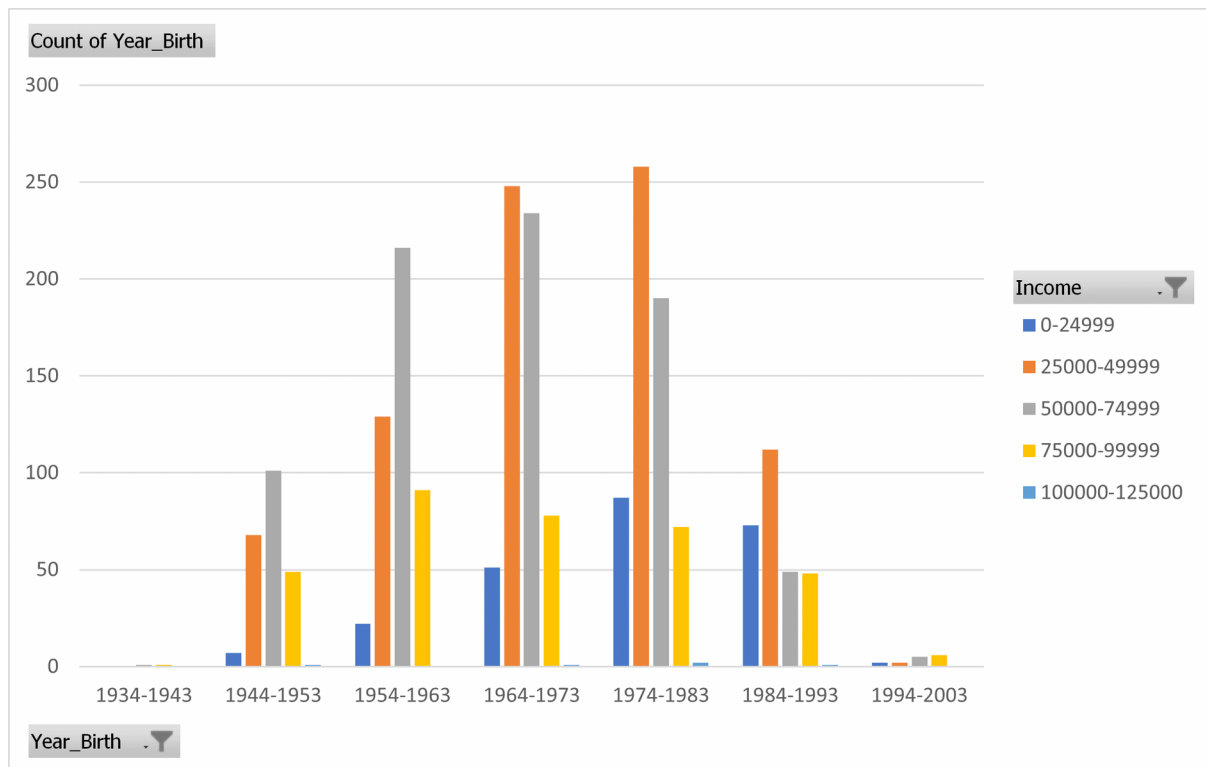
Average age of customers by marital status shown by using a column chart below for ease of understanding the variance. Highest Average age belongs to 'Widow' and Lowest to 'Single' (Ignoring Absurd, Alone and YOLO as due to very low customer count, technically outliers).



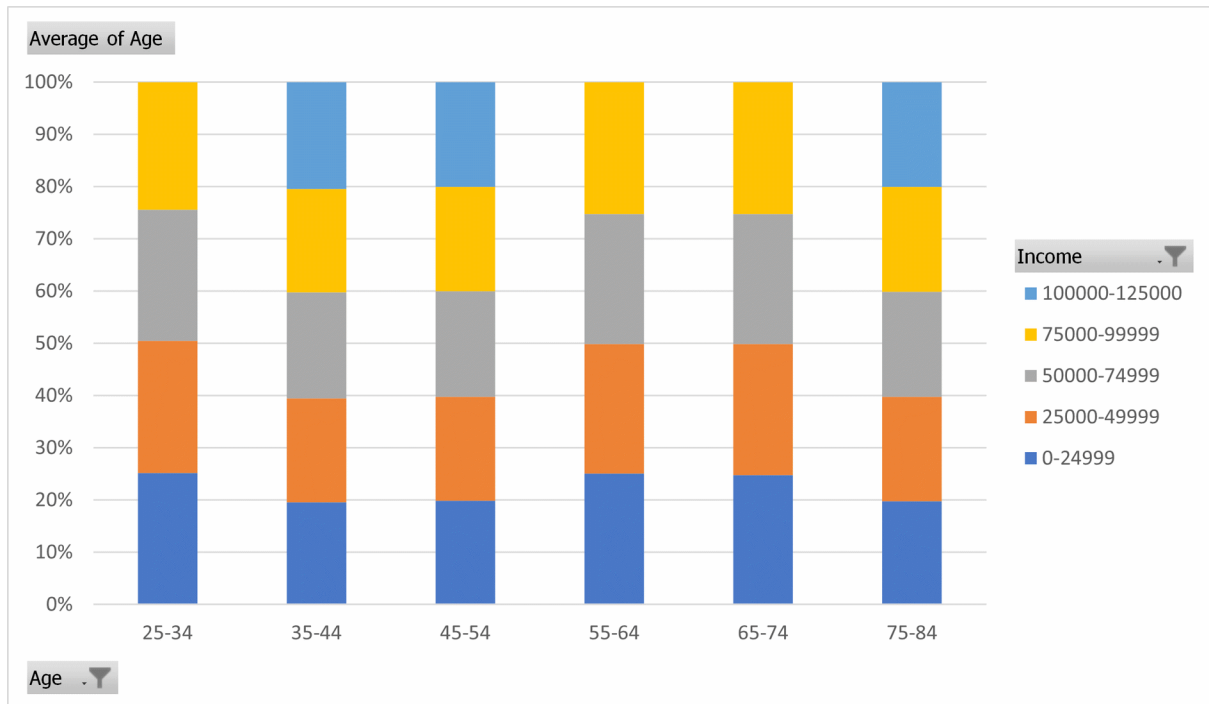
No of customers by marital status shown below using a bar chart for the ease of understanding the data and spread.



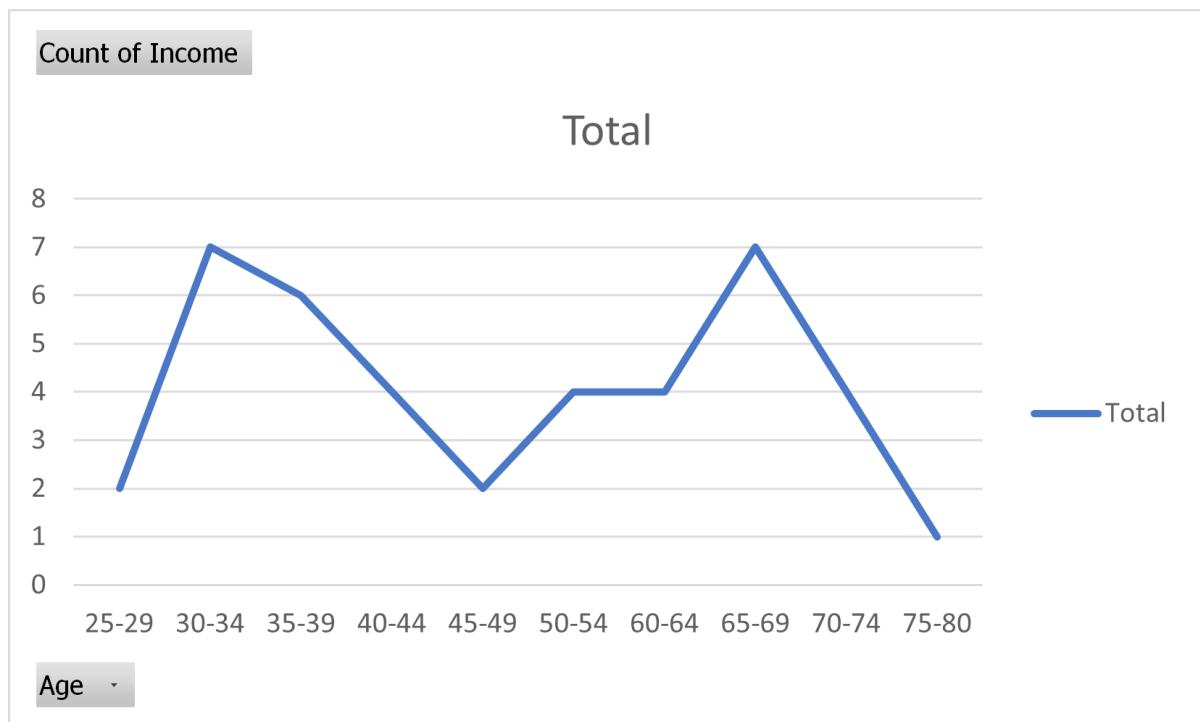
Side by side Bar chart showing No. of customer within various age groups and their income. Easily identifiable that most people are within 25000-49000 income but as age increases majority shifts to 50000-74999



Stacked bar graph showing income level by average age. Easily identifiable that higher income bracket only appears for average ages between 35-54 and 75-84. The lowest average age group has 75000-99999 as the highest income.



Line graph showing number of customers earning 90k-100k within various age groups. Easy to identify the peak and dip. Age doesn't play a factor for earning potential in this case.



Overall, average age of customers shopping is at higher end, 2Market may look to target more youth to expand its customer base.

Module 3

Created an age bin of size 5 for the ease of analysing data in tableau. Sales by month cannot be established as there is no data for sale date. Used bar charts to determine average age of customers and which age group spends most. Bar charts are simple to understand and convey the required message in this case. Customers within 45-49 age spend the most and that's the age group where most customers belong to as well (CY-2022).

Side by side bar chart is used to establish which product sells most in each age group, again keeping it simple and easy to extract information. Liquor is the most sold item across all age groups. For similar reasons, used map to present which country has the highest sale, using colour shading. 2Market is doing best in Spain and has least sales in Montenegro.

Module 4

1.

```
SELECT country, SUM(amttotal) as total_spend
FROM marketing_data
GROUP BY country
ORDER BY total_spend desc;
```

2.

```
SELECT country, SUM(amtliq) as liq, SUM(amtvege) as veg,
SUM(amtnonveg) as nonveg, SUM(amtpes) as fish,
SUM(amtchocolates) as choc,
SUM(amtcomm) as comm
FROM marketing_data
GROUP BY country;
```

3.

```
SELECT country, SUM(amtliq) as liq, SUM(amtvege) as veg,
SUM(amtnonveg) as nonveg, SUM(amtpes) as fish,
SUM(amtchocolates) as choc,
SUM(amtcomm) as comm
FROM marketing_data
GROUP BY country
ORDER BY liq desc;
```

4.

```
SELECT marital_status, SUM(amtliq) as liq,
SUM(amtvege) as veg, SUM(amtnonveg) as nonveg,
SUM(amtpes) as fish, SUM(amtchocolates) as choc,
SUM(amtcomm) as comm
FROM marketing_data
GROUP BY marital_status
ORDER BY liq desc;
```

5.

```
SELECT kidhome, teenhome, SUM(amtliq) as liq, SUM(amtvege),
SUM(amtnonveg),
SUM(amtcomm), SUM(amtpes), SUM(amtchocolates)
FROM marketing_data
GROUP BY kidhome, teenhome
ORDER BY liq desc;
```

Shows all possible combinations for whether there is a kid or teen at home or whether there is none at home.

Module 5

1.

```
SELECT country, SUM(twitter_ad) as twitter, SUM(facebook_ad) as facebook,  
SUM(instagram_ad) as instagram  
FROM marketing_data  
JOIN ad_data  
USING (id)  
GROUP BY country;
```

Canada - twitter, **India** – Twitter, **Australia** – Instagram, **Spain** – Instagram, **US** – Facebook, **Saudi Arabia** – Instagram, **Germany** – Twitter, **Montenegro** – None.

2.

```
SELECT marital_status, SUM(twitter_ad),  
SUM(facebook_ad), SUM(instagram_ad)  
FROM marketing_data  
JOIN ad_data  
USING (id)  
GROUP BY marital_status;
```

YOLO – none, **Widow** – Twitter, **Together** – Instagram, **Alone** – None, **Absurd** – Facebook/Instagram, **Married** – Instagram, **Divorced** – Twitter, **Single** – Twitter.

3.

```
SELECT country, SUM(count_success) as total_conv,  
SUM(twitter_ad), SUM(facebook_ad),  
SUM(instagram_ad), SUM(amtliq), SUM(amtvege),  
SUM(amtnonveg), SUM(amtpeps), SUM(amtchocolates),  
SUM(amtcomm), SUM(amttotal)  
FROM ad_data  
JOIN marketing_data  
USING (id)  
GROUP BY country  
ORDER BY total_conv desc;
```

Spain has highest conversion rate and the highest sales. General trend suggests countries with more successful conversion tends to have higher sales. Two countries which defy this logic are Saudi Arabia and Australia. They both have higher sales and lower conversion than the countries preceding them in ranking table of successful conversions.

Module 6

Bar chart, side by side bar chart and tables were used to create dashboard and analyse data to keep it simple and be able to extract information easily.

Apparently, all the product groups have equal number of orders placed, or every time a customer places an order they buy products from every product group. Generally, average unit price of Liquor is higher, that's the reason for higher turnover for liquor.

Twitter, Instagram, and Bulk mail have the highest average conversion rate by country and customer marital status. When analysing the advertising channel efficiency by age-Instagram, Facebook and Bulk mail seem to be more efficient than Twitter for certain age groups. 2Market will need to adopt an advertising strategy that will cater to various customer demographic in different countries. Dashboard has a chart with option to filter by country, that shows conversion rate both by age group and marital status. Can say goodbye to Brochure channel.