# Report: Assignment 1

# Context of the business

* 2Market is a global supermarket that sells both online and in-store. They analyse marketing and ad data to identify the most effective advertising channels for specific products and customer profiles. This helps them tailor promotions and improve sales.

Questions to understand the context better

* Which countries is 2Market investing in, and how mature are these markets?
* Are there sales targets for products/countries?,
* How was the budget divided across different media channels?"

Questions to understand data better

* Are sales made through delivery apps included in online sales?
* What is the unit of measurement used in the products' KPIs (i.e. AmtLiq. AmtVege,?

# Analytical Approach

## The description of KPIs in the metadata\_2market file. For the analysis, we are using two datasets: marketing\_data.csv and ad\_data.csv.

Module 2: Visualise data in Excel

1)Clean and transform the data

I applied CTRL+G>Special>Blanks to ensure no missing or empty cells. **No blank cells** found.

For duplicate values, most columns allow duplicates **except** for **ID**. To check, I used “Conditional Formatting”.

A screenshot of a computer error

Description automatically generated

2) Analyse the cleaned data

* 1. **What is the average age of 2Market’s customers?**

53 years old.

* I created a new column, “Age”, by subtracting 2023-"Year\_Birth".
* Then, I created a copy of the dataset and used the AVG formula to calculate the average age
* A screenshot of a spreadsheet

  Description automatically generated
  1. **What is the average age of the customers belonging to each type of marital status?**

Avg. age by marital status:

- "Absurd": 47 yo

- "Widow": 63 yo

No errors/hid. rows. Used Avg. function directly in Pivot Table.A screenshot of a computer

Description automatically generatedA screenshot of a graph

Description automatically generated

* 1. **What is the average age of customers with an income US$90,000 - US$100,000?**

**51 years** . However, in the income range of **US$90,000 -US$100,000,** only 41 observations were found (1.85% of the customer panel).

A screenshot of a graph

Description automatically generated

**c.1) Adjust the “Income” column to transform it into INTEGER**

I replaced “,” with empty values (“ ”) using CTRL+F

Then, removed the “$”

**c.2) Creation of a Pivot Table, grouping “Income” by ranges**

A screenshot of a computer

Description automatically generated

3) FINDINGS AND CHARTS

**AGE AND MARITAL STATUS**

The highest average age is for "Widow" (63 years old), and the lowest is for "Absurd" (47). However, "Absurd", "Alone", and "YOLO" have a minimal sample size. The groups with the most customers are "Married" (38.7%, average age = 52), "Together" (25.9%, average age = 54), and "Single" (21.3%, average age = 50). There are three outliers aged >120 years, which should be removed from the sample. The age range of 44-53 years old has the most customers:

A number of numbers on a white background

Description automatically generated

* **Relation Between Average Age and Marital Status**

A screenshot of a graph

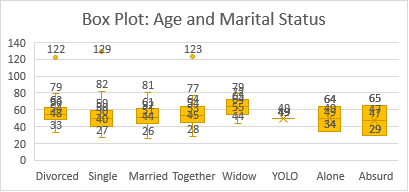
Description automatically generated

* **Histogram To Understand Age Ranges Concentration**

A graph with numbers and a red line

Description automatically generated

* **BOX PLOT: Age and Marital Status**



**CUSTOMER INCOME AND PROGRESSING AGE**

It is expected to find income outliers due to social inequality and wealth concentration.

* Income outliers are expected due to wealth concentration. There is a correlation between age and income for those earning less than £10000 to £79999, with age ranging from 46 to 57. After this point, only 9.8% of customers have incomes above the Income Q3. The maximum for income ($666.7 K) is higher than the upper limit (£118.3k) and attributed to only one 45 yo customer. Eight customers have incomes higher than the upper limit, aged between 40-50 years old, except for one 73-year-old, the second richest in the sample.
* **Customers with incomes higher than the upper limit**

A screenshot of a computer

Description automatically generated

* **Outliers and abnormalities in age and income:**

A table with numbers and letters

Description automatically generated

* **Relationship Between Age and Income**

A screenshot of a data

Description automatically generated

A screenshot of a computer

Description automatically generated

**What is the relationship between income and the age of the customers who earn between US$90,000 and US$100,000?**

The correlation (0.212625) is low but positive.

A graph with orange dots

Description automatically generated

**Module 4: Query the data with SQL**

1. CREATING TABLE IN SQL FOR marketing\_data.csv

A screenshot of a computer

Description automatically generated

A computer screen shot of a computer code

Description automatically generated

2)CREATING TABLE IN SQL FOR ad\_data.csv

A screenshot of a computer code

Description automatically generated

3)QUESTIONS

* the total spend per country

A screenshot of a computer

Description automatically generated

* **the total spend per product per country**

A screenshot of a computer

Description automatically generated

* **which products are the most popular in each country**

A screenshot of a computer

Description automatically generated

* **which products are the most popular based on marital status**

A screenshot of a computer

Description automatically generated **- which products are the most popular based on whether or not there are children or teens in the home.**

A screenshot of a computer

Description automatically generated

Module 5: Join and analyse tables in SQL

1. **JOINING** marketing\_data.csv and ad\_data.csv

LEFT JOIN returns all the rows from the left table (marketing\_data) and the matching rows from the right table (ad\_data).

A screenshot of a computer

Description automatically generated

1. QUESTIONS

* Which social media platform) is the most effective method of advertising in each country?

A screenshot of a computer code

Description automatically generated

* Which social media platform is the most effective method of advertising based on marital status? A screenshot of a computer code

  Description automatically generated

A screenshot of a computer

Description automatically generated

* Which social media is the most effective per country? A screenshot of a computer screen

  Description automatically generated

A screenshot of a computer

Description automatically generated

After this step, I exported the last table to a CSV file and calculated the lead conversion in excel.

I divided each product type by each of the ad channels'. Example: Facebook Efficiency for Fish = (Fish)/(Facebook\_ture\_Count)

# Dashboard Design

The following structure organises the dashboards. The rationale was going from Macro to Micro, identifying the 5 Whys along three central questions:

1) Which are the most important countries to focus go-to-market efforts?

2) What is the consumer profile, and how does it vary among countries and products purchased?

3) How big are online sales and web visits among countries? How does ad efficiency vary among countries, ad channels and ages?

1) Countries Dashboards

* Countries\_Topline Numbers
* Countries\_Product Spent

I used the map function in countries, emphasising the market size with the “size” formatting. I inserted filters for deep dives in specific countries and a colour code by-product to help navigation.

2)Customer Profile Dashboards

* Consumer Demographics
* Income Deep Dive by Country
* Age\_Deep Dive by Country

I used bins for age and income to create ranges. As they play, more weight given to income and age are crucial in purchasing and ad conversion.

3)Online Sales and Ad Dashboard

* Online/Promo KPIs per Country
* Online vs Ad Efficiency per Country

I displayed online market size and KPIs across countries, created a calculation for % of online sales and divided ad channel conversion metrics by age.A screenshot of a computer

Description automatically generated

# Findings

# As I created a summarised presentation, here is the slide with the key findings.

# A close-up of a list of information Description automatically generated