Course 1 Assignment: Exploratory analysis and presenting insights
Case Study: 2Market Global Supermarket

Section 1: Background/context of the business

Problem statement: 2Market is a global supermarket which sells products online and in-store. They have marketing and advertisement data stored which they want to analyse in order to answer the following questions:

- What are the demographics of their customers?
- Which of their advertising channels seem to be the most effective?
- Which products seem to sell the best and does this vary based on the demographic?

An additional question I'd like to ask of the data is:

- In which location do customers spend the money at 2Market and do they spend this money in-store or online? If the former, then 2Market might consider opening more stores within location to increase their revenue.

These questions are important to ask as they will quickly enable 2Market to reveal sales trends and other insights that can drive more informed decision-making within the business.

I would like to start by asking the 2Market team the following questions:

1. Who will I be presenting this data to? If it is to other Data Analysts, then they will be attuned to the jargon often used as an analyst. If it is to employees working outside this circle, then they will be less familiar with some of the concepts, and I will need to tailor my presentation accordingly. For this report presentation I will assume it is presented to the C-suite as they are the ones in charge of making the business decisions for the company.
2. Where was this data obtained from? It is important to know whether the data was gathered from a reliable source as this will impact the reliability of the data, and ultimately the reliability of the business insights that follow.

Section 2: Analytical approach

Section 2.1: Data Cleaning

2Market plans to roll out a marketing campaign in the coming weeks and need to better understand the customer demographics to inform their campaign strategy. The following steps were taken to clean the data in Excel:

1. Checking for Duplicates and Blanks in the data.
   Findings: No Duplicates or Blanks were identified.

2. Checking the Year Of Birth (YOB) Column.
   Findings: The three oldest customers were born in 1894, 1900 & 1901. The oldest person alive on the planet was born in 1907 and therefore the three dates stated previously are not accurate.
   Action: Remove the rows which include YOB's 1894, 1900 & 1901.

3. Checking the Education column.
   Findings: One of the options in this column is named as '2n Cycle' which is a typo.
   Action: Use Find and Replace to replaced '2n Cycle' with '2nd Cycle'.

4. Checking Marital Status column.
   Findings: 'Absurd' and 'YOLO' are not common categories for Marital Status. 'Alone' is another way of saying 'Single'. 'Married' and 'Together' can be categorised together as partners who live under the same roof.
   Actions: 'Absurd' and 'YOLO' occupy only 4 rows and can therefore be deleted. Use Find and Replace to replace 'Alone' with 'Single'. Use Find and Replace to replace both 'Married' and 'Together' with 'Couple'. 'Divorced' and 'Widow' can both remain as they are.

5. Checking that all number/text values are correctly formatted.
   Findings: All values in the Income column were held as Text. Some of the registration dates were formatted incorrectly and held as Text values.
   Actions: Use the Text to Column tool to convert the Income Text to Number values with the '$' sign as a Delimiter. Change column heading to 'Income ($)'. Check the income column values are held as Numbers using '=ISNUMBER()' and ensuring the argument comes back as 'TRUE'. Next, use the Text to Column tool to convert the incorrect registration dates to the correct date format (i.e. matching the existing correct registration dates).

Section 2.2: Analytical Approach

The following section describes the questions I asked of the data in Excel and the approach taken to answer them:

1. Average age of 2Market's customers.
   Approach: First we need to identify the age of each customer individually. To do this, create a new column headed 'Age'. Next, we need to subtract the customers YOB from the year today using the following formula: `=YEAR(TODAY())-[@[Year_Birth]]`
   To find the average age, use the function '=AVERAGE()'.
   Solution: The average age of customers shopping at 2Market is **53 years old**.

2. Average age of customers belonging to each type of marital status.
   Approach: Use the '=AVERAGEIFS()' function to find the average age belonging to each type of marital status as follows:

Z	AA
Marital_Status	**Average Age**
Single	=AVERAGEIFS(Table1[Age],Table1[Marital_Status],Z3)
Couple	=AVERAGEIFS(Table1[Age],Table1[Marital_Status],Z4)
Divorced	=AVERAGEIFS(Table1[Age],Table1[Marital_Status],Z5)
Widow	=AVERAGEIFS(Table1[Age],Table1[Marital_Status],Z6)
All	=AVERAGE(Table1[Age])

   Solution:
   The formulas above give the following outputs rounded to the nearest whole number:

Z	AA
Marital_Stat	**Average Age**
Single	50
Couple	53
Divorced	55
Widow	63
All	53

3. Average age of customers who earn a yearly income between $90,000 and $100,000
   Approach: Use the 'AVERAGEIFS()' function to find the average age of customers with yearly income between $90,000 and $100,000 as follows:

Income	Average Age
$90,000 > Income > $100,000	=AVERAGEIFS(Table1[Age],Table1[Income],">90000",Table1[Income],"<100000")

Solution: The formula above gives the following output rounded to the nearest whole number:

Income	Average Age
$90,000 > Income > $100,000	51

4. Which products sell the best and which country spends the most?

Using the '=SUMIF' function to produce the following table.

Country	Country Code	Alcohol ($)	Vegetables ($)	Meat ($)	Fish ($)	Chocolate ($)	Commodities ($)	Total ($)
Spain	SP	335637	28144	177847	40049	30070	45957	**657704**
South Africa	SA	105910	8937	58393	13663	9019	15127	**211049**
Canada	CA	83504	7611	45375	9777	7604	11939	**165810**
Australia	AUS	42281	3587	22203	5334	4068	6887	**84360**
India	IND	35899	3779	23671	4807	3214	5947	**77317**
Germany	GER	36776	2980	20272	4601	2801	5768	**73198**
USA	US	32214	3034	20185	4411	2863	4839	**67546**
Montenegro	ME	1729	8	817	226	122	220	**3122**
	Total ($)	673950	58080	368763	82868	59761	96684	**1340106**

The data shows that the best-selling product in every country is Alcohol, whilst the worst selling products are split between Vegetables and Chocolate. The country which spends the most amount of money at 2Market is Spain, whilst the country which spends the least amount of money at 2Market is Montenegro.

Section 3: Dashboard

Section 3.1 Dashboard Design

The dashboard is designed to clearly articulate any patterns, trends or insights into the business questions asked in Section 1. The demographics explored in the dashboard include; Product Sales by Country, Product Sales by Marital Status, Product Sales by Education and Product Sales by Age. The average age by country and number of successful advertising leads were displayed in tables. A Product Sales by Age bar chart was created to show how the amount of products bought varied with age.

Section 3.2 Dashboard Insights

1. What are the demographics of 2Market's customers?

The dashboard shows that the average age of customers purchasing products from 2Market is between 50 and 59. The Marital Status of customers spending the most at 2Market is Couples and they spend the most across all six product types. The Education of customers spending the most at 2Market is those who have graduated with a bachelor's degree and they spend the most across all 6 product types.

2. Which of their advertising channels seem to be the most effective?

The most successful advertising channel is Twitter with 164 successful leads. This is closely followed by Email on 163, Instagram on 160 and Facebook on 141. The least successful advertising channel was Brochures with only 30 leads.

3. Which products seem to sell the best and does this vary based on demographic?

The product which sells the best in all countries and across all ranges is Alcohol. This is followed by Meat being the second best seller.

4. Which location spends the most money at 2Market and do they spend this money in-store or online?

The location which spends the most money across all products is Spain. In Spain, the number of website purchases is 4,370 whilst the number of in-store purchases 6,326. The trend of having more in-store purchases than online purchases continues across every country. Therefore, 2Market may want to consider setting up more store locations to increase their

revenue. However, as Montenegro has very few sales relative to the other countries, 2Market may wish to consider shutting down their stores in this location as operational costs are most likely exceeding the sales of their products. However, more data would be required to confirm this.