2Market Report

Point 1

Background/context of the business: Describe the business problem, the questions you'd like to ask to better understand the project and / or data, and any additional questions you'd like to try answer by analysing the data.

2 Market is a super market chain with physical and online presence in various countries around the globe. They have presented us with a primary, internal dataset with the request to further analyse and understand :

i) The demographic segments of their customers
ii) Which advertising channels seem to be the most effective
iii) Which products seem to sell the best and if that varies based on demographic
iv) Behavioural segments in terms of frequent category purchasers
v) Online vs offline market share
vi) Any other insight or trend that can be revealed by further studying their data

Presently, 2Market, has both physical and online presence. Questions or clarifications that arise after our initial approach into the dataset are mainly focused around following points:

i) Is the dataset representative of 2Market's both online and "brick and mortar" clientele?
ii) Is the sample balanced or weighted (and if it is, based on what population metrics)
iii) Our demographics data is based on :
    a. *age*,
    b. *ethnicity (country)*,
    c. *marital status*,
    d. *income*,
    e. *education*,
    f. Is there a reason why sex and employment is missing from our demographics? (it would help in making associations with spending patterns and online marketing lead conversion)
iv) Do we consider the amount spent to be basis a daily, weekly, monthly or even annual basis or a calculated average? Also is the currency USD? (view that we refer to global business)
v) We have the recency and monetary value (how much money a customer spends on purchases), but do we purposefully missing the frequency of purchases from both online and physical marketplace? (we have the number of purchases made from the website along with the number of in-store purchases though both are not period defined – as point (iv) on amount spent?
vi) We make assumptions on statistical margin of error

Point 2

Analytical approach: Describe the approach taken to clean and analyse data in Excel and SQL, including any insights or observations. Include detailed descriptions of your approach and any decisions you made during analysis along with an explanation of your rationale.

Our dataset has been duly processed and cleaned according to best practises for:

1) spelling errors (eg 2n Cycle -> 2nd Cycle)
2) values out of range (eg cases in Age and Income Dimensions)
3) incorrect or invalid data types (Compared data types between original data dictionary provided and actual dataset values ie. Sheet "Data_Dictionary)
4) blank cells or spaces
5) incorrect use of nulls

We have also processed certain fields in an effort to further assist our analysis:

- Column (Year_Birth) has been removed and replaced with Column (Age) and column (Dt_Customer) has been replaced with (Years_as_Customer) (both columns basis 2022) in order to have more meaningful data for comparison
- Inserted column total spent (SUM of amounts spent for six categories per custom_id)
- Obvious data outliers have been removed (representative examples as "Income=666,666", "Age>120") from marketing data.csv and related records also removed from ad_data.csv (ensuring to get homogeneous results when INNER joining both tables)
- In worksheet utilizing average values, we made sure to apply the analytics model of "Average with 95% Confidence Level"
- Marital Status Options:
    - 'Absurd' & 'YOLO' changed to Single
    - 'Alone' changed to Single
    - Reasoning behind those changes was not to delete the data but change data type by grouping almost identical categories
- Renamed columns AmtLiq, AmtVege, AmtNonVege, AmtPes, AmtChocolates, AmtComm to Alcohol, Vegetables, Meat, Fish, Chocolates, Commodities (for sake of visual clarity during our comparisons – also edited Tableau axis aliases where needed)
- Excluded results from country: Montenegro (ME) in some of the worksheet as sample was just from 3 customer ids.
- Created income and age groups in Tableau. Aim was to be able to draw results from specific income or age groups as well to satisfy 2Market shareholders on their request to understand more about customers who meet the following criteria:
    - Marital status: Married
    - Age: Between 45 and 50
    - Income: Between US$90,000 and US$100,000

- For the above point and for facilitating the analysis we have grouped the available incomes in bands of 10,000 USD starting from min income up to 10k (ie 10,000 USD per year) and ending to individuals with income above 100,000 USD per annum. In the same spirit, we have grouped ages to groups 25-34, 35-44, 45-54, 55-64 and over 65 years old
- Worksheet comparisons trying to analyse whether high income groups are big spenders and whether younger or older age groups respond more positive to 2Market's online presence
- It proved easier to interpret aggregations using Postgres (SQL) script even though same results can be also achieved via the imported Excel file. For the queries/worksheets involving the data campaign, the results were fetched via custom SQL queries.

Point 3

Dashboard design and development: Describe the rationale behind your dashboard design and development. Consider decisions related to elements such as visualisation type, colour, size, interactivity, accessibility, and layout.

Our strategy is to group all the worksheets related to each topic (ie either demographic or revenue or campaign data analysis) and then include the latter in the same dashboard ensuring that our audience's attention remains on the topic being discussed for each dashboard (simplicity by expressing clear purpose and unified data on each dashboard).

Then our story will be unfolded by analysing every aspect of the three dashboards presented separately but all interconnected by the same business logic.

Most of the charts chosen are depicted via horizontal, stacked or side-by-side bars. We cannot include chart type diagrams as we don't have measurement of data over specific period but we do have considerable types of data to compare between them (as such privileging the extensive use of bars). Simplicity in charts types but also simplicity in colours, sizes and visualization. We have kept the same colour pallet for comparison of the same values and also chose colours schemes accounting for colour blindness . All titles and notes share same fonts and size. Dashboard size has been put to automatic and all worksheets have been inserted inside containers as directed by best practices.

All charts/worksheet titles have been edited towards something understandable for audience that will accessing our dashboard without having heard our presentation and same care has been also given to modify aliases and headers in order to present an aesthetic visual result.

The dashboards presented aim to focus on individual comparisons between categorized groups from our dataset. Besides replying to specific questions in our analysis via the worksheets produces, we have also tried to apply the same filters in most of our worksheets. That way we pass the opportunity to the shareholders to instantly "draw" a different comparison path than the one we present. Idea behind keeping the same filters in most of our worksheet, is to provide a relative short range of possible criteria for shareholders of groups they should focus on in order to achieve the wanted changes.

Point 4

Patterns, trends, and insights: Clearly articulate any patterns, trends, or insights you discovered. Ensure they relate to the business scenario. Include recommendations on any areas for further exploration.

Via our analysis we tried to reply to a series of points identified from our worksheets as follows:

1) Despite its global presence 2Market has its 50% of its clientele originating from Spain and then sharing 15% and 13% accordingly between South Africa and Canada and 5-6% for each of Australia, India, Germany and US customer base. (*WorkSheet: %Nationalities*)
2) We have created a Buyer Persona trying to form the average buyer visiting both online and physical stores. Our globally average client is 52 years old, with about 607 USD expenditure per month, earning about 52,000 USD, been about 9 years as customer with 2Market and doing on average 5 online and 6 in-store purchases per month. (*WorkSheet: DemoPersona*)
3) As expected in most cases, younger ages belong to lower income groups, and in our dataset the highest income groups gather the highest average age group percentages (please see note in Point 2 ref age and income groups). Around 60 % of our clients come with almost equal percentages from 4 income groups ranging between 30,000 and 80,000 USD income per year. Interesting insight , and also request of the shareholders, is that Spain holds the biggest % of clients with annual income between 90 and 100k per year. Also the youngest age group (25-34) (despite its low population) have an average income below 30k per year (*WorkSheet: IncomeGroup%)*
4) The following worksheet continues along the expected trend that high earners are usually high spenders and in our case, the 35% of clients who earn between 70,000 and 120,000 USD per year spend on average about 1,500 USD monthly, which is almost double from the overall average with 95% Confidence level. Again, targeted income group of 90-100k , manages an average monthly expenditure of about 1800 USD per month hence focus and attention would be recommended towards this type of clientele that 2Market would not want to disappoint. (interesting insight is that earners of above 100k per year spend close to the global average of about 700 USD per month) (*WorkSheet: TotalSpentIncome)*
5) Lastly in our demographics examination, the age group 25-34 and groups 55-64 and over 65 years old, keep the same elevated average for Total monthly expenditure in comparison to age groups 35-44 and 45-54 that mark below average. Maybe 2Market should attempt in targeting specific age groups to further increase monthly profits. (*WorkSheet: TotalSpentAge)*
6) Analysis of the products against various parameters as Age, Income, Country or Marital Status is constantly showing a clear winner. Alcoholic products sell the better and with a big difference from the second in place. Interesting points to reflect on considering only Alcohol related sales are the following:
    a. Age groups 55-64 and Over 65 top the chart with a significant difference from the other groups (and as seen from demographics examination these are the groups with the highest earners and also the big spenders) (*WorkSheet: ProductPerAge)*

   b. To verify the above point, for the income group 80-90k the average monthly expenditure in Alcohol equals the expenditure in all other categories together. And for the income group 90-100k the sales of alcohol just blows the average expenditure to the already mentioned astonishing average of 1800 USD per month (ie an average of 850 USD for Alcohol products and 950 USD for the rest of all the categories together) (*WorkSheet: ProductPerIncome)*
   c. Most countries average about 300 USD per month in Alcoholic sales expect India where the monthly average is 250 USD. (*WorkSheet: ProductPerCountry)*
   d. While all other categories of products share same average expenditure between people in couples and singles (created small group Singles & Couple including all declared Marital Statuses for sake of simplifying certain patterns), Alchohol shows higher sale levels from customers being single than customers in couples (*WorkSheet: ProductPerIncome)*
7) The only category which seems to be quite competing to Alcohol (always in terms of monthly sales) is the Meats products category. While in all charts always come in second place behind Alcohol Sales, we can see that in relation to households with no kids or Teens, meat products expenditure is considerably higher than households with at least 1 kid or teen. (*WorkSheets: ProductPerKid, ProductPerTeen*)
8) View that the business operates with physical stores but also with online purchases, we compare the later and witness that in-store sales significantly surpass online sales in all age and income groups . More importantly if we focus in Spain, where we have 50% of our clientele, the difference in online vs in-store sales for the age group 25-34 is almost double which clearly indicates that progress needs to be done in that segment. (*WorkSheet: OnlineVsOffline)*
9) The argument that work needs to be done in improving 2Markets online sales is also being supported from the results of the marketing campaign launched. For the big market of Spain, the successful lead conversions were due to responses from their Bulk Mail campaign instead of the campaigns launched in the rest of the 3 social media (Facebook, Instagram and Twitter). Remarkable insight is that Canada does not share the same pattern with Spain or the other countries, as in Canada the Twitter campaign returned almost same results as the Bulk Mail campaign (*WorkSheet: CountryCampaign)*
10) Further examining the campaign in between the 3 social media (Facebook, Instagram, Twitter) we can see that from the successful leads that replied back from the said campaign Twitter is winning the age groups 45-54, 55-64 and Over 65 while Instagram is the chosen medium of age groups 25-34 and 35-44. Also in total percentages, Twitter seems to be the media platform the most effective per country