**Name:**

**Team members: Amneh Ehab, Ayat Dwekat, Bushra Al-Nabrawi, Eyad Alqinaa, Yazan Alabed, Yazeed Mshayekh**

Write your answers below and paste the formulas you used to solve the problems, if applicable.

**Customers**

1. What’s the average amount of your customers’ yearly income?

**Formula used:**

=AVERAGE(B2:B2221), =average select all the cells in the income column.

**Answer:**

**151551.3527 $/Year**

2. What’s the most common marital status of your customers?

**Formulas used:**

pivot table (sum values) which is the most efficient, or =INDEX(F$1:J$1, MATCH(MAX(Q2:Q6), Q2:Q6, 0)

**Note:**

=COUNTIFS(F2:F2221, 1) used to calculate Q2,

=COUNTIFS(G2:G2221, 1) used for Q3,

=COUNTIFS(I2:I2221, 1) used for Q4,

=COUNTIFS(H2:H2221, 1) used for Q5,

=COUNTIFS(J2:J2221, 1) used for Q6

**Answer:**

**Married**

3. What’s the most common educational level? What’s the least common?

**Formulas used:**

pivot table (sum values) which is the most efficient, or =INDEX(K1:N1, MATCH(MAX(R2:R5), R2:R5, 0)) and =INDEX(K1:N1, MATCH(MIN(R2:R5), R2:R5, 0))

**Note:**

=COUNTIFS(K2:K2221, 1) used to calculate R2,

=COUNTIFS(L2:L2221, 1) used for R3,

=COUNTIFS(M2:M2221, 1) used for R4,

=COUNTIFS(N2:N2221, 1) used for R5

**Answers:**

The most common education level is **“Graduate”**, and the least common is **“Basic”**.

4. What’s the oldest age of your customers? What is the youngest? How about the average? Are there any concerns with the data?

**Formulas used:**

=MAX(E2:E2221), =MIN(E3:E2222), AVERAGE(E2:E2221)

**Answers:**

Oldest customer **“80”**, Youngest customer **“0”**, Average age **“51”**  
There is a concern in the data which is in cell **E551 “-64”,** this is not possible for the   
age.

5. Do customer households tend to have more kids or teenagers?

**Formulas used:**

pivot table or =INDEX(C1:D1, MATCH(MAX(S2:S3), S2:S3, 0))

**Note:**

=COUNTIFS(C2:C2221, 1) used for S2,

=COUNTIFS(D2:D2221, 1) used for S3

**Answers:**

**Teens**

6. Based on your understanding of the customer data, how would you describe the typical customer for your company?

**Answer**:

The typical customer is a **middle-aged individual around 50** with an **average annual income of $151,551**, likely **married**, **single**, or **living with a partner**, and often **well-educated**, with a **significant portion holding university or advanced degrees**. Many customers have **children** in the household, suggesting a **family-oriented demographic**. The customer base is geographically diverse across the U.S**.,** with **similar representation from states** like California, Florida, Illinois, New York, Texas, and others. This profile indicates a **financially stable, educated, and family-focused audience**.

**Products**

7. Which product generates the most sales?

**Formulas:**

Sum the total sales for each product category:

Total Sales = Sum (Teas, Fruits, Meats, Seafood, Sweets)

**Answers:**

**Teas** generate the most sales, with **1,734,336** units sold.

8. Do customers buy more standard or premium products?

**Formulas:**

=INDEX(H1:I1, MATCH(MAX(P2:P3), P2:P3, 0))

Compare the sum of "Premium" and "Regular" product sales1:

Total Premium = Sum(Premium)  
Total Regular = Sum(Regular)

**Answers:**

Customers buy more **“Regular” (standard) products**, with **2,961,904** units compared to **292,422** units for **“Premium” products**.

9. What is the distribution of sales by distribution channel?

**Formulas:**

Sum the total sales for each channel:

Channel Sales = Sum(Website, Catalog, Store)

=SUM(K:K) / SUM(K:M) \* 100,

=SUM(L:L) / SUM(K:M) \* 100,

=SUM(M:M) / SUM(K:M) \* 100

**Answers:**

* Website: **9,108 units.**
* Catalog: **5,862.2 units**
* Store: **12,905 units**

10. What's the average time since the last purchase?

**Formula:**

Calculate the average of the "**Recency**" column:

Average Recency = Mean(Recency)

=AVERAGE(B:B)

**Answer:**

The average time since the last purchase is approximately **49 days**.

11. Based on your understanding of the sales data, do you have an idea of what products the company may want to focus on?

**Answer:**

The company should sell more Teas because people buy them the most. Also, since people buy regular teas more than premium teas, the company should make the premium teas more attractive to increase sales and make more money

**Marketing**

12. Which campaigns generated the most and least interest?

**Formulas:**

=SUM(marketing!B2:B2221),

=SUM(marketing!C1:C2221),

=SUM(marketing!D1:D2221),

=SUM(marketing!E1:E2221),

=SUM(marketing!F9:F2221)

**Answers:**

The most generated interest is **MC3 and MC4**, and the least interest is **MC2**.

13. For campaign 4, what is the typical marital status of a customer?

**Formulas:**

=VLOOKUP($A3,'C:\Users\Hp\Downloads\[customer.csv]customer'!$A$2:$H$2221,8,FALSE)&&=SUM(N3:N167). (for each status)

**Answer:**

The typical marital status of a customer is **married (62 married).**

14. For campaign 2, what product category sold the most? *(Note: Premium & Regular are product quality categories and not separate product categories. Do not directly use in your calculations for this question.)*

**Formulas:**

=VLOOKUP($A2,'C:\Users\Hp\Downloads\[sales\_update.csv]sales\_update'!$A$2:$C$2221,3,FALSE)&&=SUM(J2:J31). (for each product)

**Answer:**

The product category sold the most is **teas (65527 sales from teas).**

15. What level of education do customers have who typically complain?

**Formulas:**

=VLOOKUP($A2,'C:\Users\Hp\Downloads\[customer(1).csv]customer(1)'!$A$2:$N$2221,11,FALSE)&&=SUM(J2:J21). (for each education level)

**Answers:**

The level of education of customers who typically complain is **graduate (14 graduate).**

16. What additional information would you need to better understand the success or failure of the company's marketing campaigns? (*This could be formulas, or suggestions for other types of data to collect.)*

**Answers:**

**Campaign Costs:** The campaign costs can be calculated by adding up all the expenses related to the campaign, such as advertising costs, discounts offered, and any other execution costs.  
**Campaign Type:** You can categorize the campaign by its type, such as "Advertisements" if it involves paid ads, or "Discounts and Offers" if it provides discounts or special deals on products or services.  
**Campaign Display Location:** You should identify the channels used to display the campaign, such as the internet if the campaign ran on platforms like Facebook or Google, or television if the ads were shown on TV channels.

Bonus Question (Optional):

A. Write a small paragraph that explains the difference between quantitative data and qualitative data in your own words.

**Answer:**

Quantitative data refers to information that contains numerical values, obtained using measurement techniques and analyzed using statistical methods. In the other hand, qualitative data focuses on the characteristics or qualities of the information, often expressed in descriptive terms or language. This type of data is collected through observations and analyzed by organizing it into various categories.

B. Using the website www.kaggle.com, find a business problem that is interesting to you and review the quantitative data provided to be analyzed. Can you differentiate the quantitative data from the qualitative data the business may have also used? Be inspired by MP1 questions and try to do the same thing with your new dataset.

**Answer:**

**Quantitative Data:**

* **Date (numeric):** the date when each transaction was generated.
* **Price (numeric):** the price of each product per unit in pound sterling (£).
* **Quantity (numeric):** the quantity of each product per transaction. Negative values related to cancelled transactions.

**Qualitative Data:**

* **CustomerNo (categorical):** a five-digit unique number that defines each customer.
* **Country (categorical):** name of the country where the customer resides.
* **ProductNo (categorical):** a five or six-digit unique character used to identify a specific product.
* **Product (categorical):** product/item name.
* **TransactionNo (categorical):** a six-digit unique number that defines each transaction. The letter “C” in the code indicates a cancellation.

I’ve chosen this [dataset](https://www.kaggle.com/datasets/gabrielramos87/an-online-shop-business) from Kaggle called “**E-commerce Business Transaction**”, This is a sales transaction data set of UK-based e-commerce (online retail) for one year.

**Given the tables you have been provided, what type of analysis do you foresee performing that will help you inform management as to the next steps to improve business results?**

**Answer:**

**1. Sales Analysis:**

Tracking sales trends over time to pinpoint peak periods and seasonal patterns. This information would be valuable in scenarios such as determining the optimal timing for discounts on specific products or deciding when and on which products to offer promotions to customers.

**2. Customer Analysis:**

Group customers based on purchasing habits, frequency, and monetary value to tailor marketing efforts.

**3. Market Analysis:**

Utilize analysis to identify product combinations that are commonly purchased together and uncover recurring patterns.

C. In addition to the data you have been provided in this case, is there additional information you believe would be helpful in your analysis that has not been provided to you?

**Answer:**

* Customer segmentation data, such as age, gender, and income level, would be beneficial.
* Customer feedback, including ratings, would assist in making improvements to products and services.
* Competitor data would help identify the business's strengths and weaknesses.

D. As a data analyst, your job is to be the subject matter expert in data analytics; however, you also need to understand how your business performs and the objectives of management. If you were part of the management team asking a business analyst to solve this business problem, what are the questions you would ask of your data scientists, and what deliverables would you expect from them?

**Answer:**

**Questions:**

* How does our performance compare to competitors?
* What are the characteristics of our most valuable customers?
* Are there any patterns in customer purchasing behavior?
* Which products generate the highest and lowest revenue, and why?
* How can we optimize inventory management?

**Expected Deliverables:**

* Forecasts for future sales, revenue.
* Inventory management strategies to enhance efficiency.
* Insights into market trends and competitor strategies.
* Identify the high-value customers.
* Recommendations for optimizing product offerings.