**Name: Saif Karborani**

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? (Saif Karborani)

Formula used: =AVERAGE(B2:B100)

Answer: 151551.4

2. What’s the most common marital status of your customers?(Saif Karborani)

Formulas used:

For Married -> =COUNTIF(customer!P:P, "Married")

For Divorced -> =COUNTIF(customer!P:P, "Divorced")

For Single -> =COUNTIF(customer!P:P, "Single")

For Together -> =COUNTIF(customer!P:P, "Together")

For Widowed -> =COUNTIF(customer!P:P, "Widowed")

**Pivot table:**

A screenshot of a graph

Description automatically generated

Answer:

Married -> 860 | Single -> 482 | Divorced -> 231 | Together 571 | Widowed 76

the most common marital status of your customers is Married -> 860

3. What’s the most common educational level? What’s the least common?(Saif Karborani)

Formulas used:

For Basic -> =COUNTIF(customer!Q:Q,"Basic")

For Graduate -> =COUNTIF(customer!Q:Q,"Graduate")

For Master -> =COUNTIF(customer!Q:Q,"Master")

For PhD -> =COUNTIF(customer!Q:Q,"PhD")

Answers:

A screenshot of a graph

Description automatically generated

Basic: 54 | Graduate:1122 | Master: 367 | PhD: 479

The most common educational level is : Graduate

The least common educational level 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?(Saif Karborani)

Formulas used:

**Oldest Age**: Use =MAX() to find the maximum value in the **Age** column.

**Youngest Age**: Use =MIN() to find the minimum value in the **Age** column.

**Average Age**: Use =AVERAGE() to calculate the average age of your customers.

Answers:

The oldest age -> 80

The Youngest age -> -64

The average age -> 51

-64: This is clearly an invalid age, as ages cannot be negative.

0: While this may indicate a missing or invalid value, an age of 0 typically doesn't make sense in this context.

5. Do customer households tend to have more kids or teenagers?(Saif Karborani)

Formulas used:

Sum of Kids -> =AVERAGE(customer!C:C)

Sum of Teenagers -> =AVERAGE(customer!D:D)

Answers:

Kids -> 0.443288

Teenagers -> 0.507883

Customer households tend to have more teenagers

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

Answer:

**The typical customer is a married individual aged 51 years, with a Graduate university education, living in a household with 0 to 2 children.**

**Products**

7. Which product generates the most sales?

Formulas: in this case, we need to SUM all Products to find what is the most sales

Teas -> =SUM(sales\_update!C:C)

Sweets -> =SUM(sales\_update!G:G)

Seafood -> =SUM(sales\_update!F:F)

Meats -> =SUM(sales\_update!E:E)

Fruits -> =SUM(sales\_update!D:D)

Answers:

Teas: 1734336 | Sweets:157502 | SeaFood: 219967 | Meats: 986946 | Fruits: 154139 |

The Teas product is the most sales product

8. Do customers buy more standard or premium products?

Formulas:

Regular -> =SUM(sales\_update!I:I)

Premium -> =SUM(sales\_update!H:H)

Answers:

Regular: 2961904 | Premium: 292422

Based on the total sales data, **customers tend to buy more standard products (Regular)** than premium products.

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

Formulas:

This formula to find the SUM of the distribution channel

Website -> =SUM(sales\_update!K:K)

Catalog ->=SUM(sales\_update!L:L)

Store -> =SUM(sales\_update!M:M)

This formula to find the percentage of distribution channel

Website-> =SUM(sales\_update!K:K) / (SUM(sales\_update!K:K) + SUM(sales\_update!L:L) + SUM(sales\_update!M:M))

Catalog-> =SUM(sales\_update!L:L) / (SUM(sales\_update!K:K) + SUM(sales\_update!L:L) + SUM(sales\_update!M:M))

Store-> =SUM(sales\_update!M:M) / (SUM(sales\_update!K:K) + SUM(sales\_update!L:L) + SUM(sales\_update!M:M))

Answers:

Website -> 9,108

Catalog -> 5,862.20

Store -> 12,905

Based on the total sales and their distribution across channels, we find that:

* **Website** generates 33% of the total sales.
* **Store** contributes 46% of total sales.
* **Catalog** represents 21% of total sales.



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

Formula:

To calculate the **average time since the last purchase**, we’ll use the **Recency** column, which represents the number of days since the last purchase for each customer.

=AVERAGE(sales\_update!B:B)

Answer: 49.02567

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:

**Key Insights:**

* **Top Product**: **Regular** products generate the most sales at **2,961,904**, making them the key revenue driver.
* **Next Best-Seller**: **Teas** with **1,734,336** in total sales are the second most popular, indicating strong customer interest.
* **Low Sales Categories**: **Sweets** and **Fruits** have the lowest sales, suggesting a need for targeted marketing or product adjustments.
* **Premium Products**: With total sales of **292,422**, **Premium** items have potential but lag behind regular products.

**Recommendations:**

1. **Focus on Regular products** to sustain revenue.
2. **Promote Teas** to capitalize on their popularity.
3. **Boost Premium product sales** through special promotions.
4. **Investigate low sales for Sweets and Fruits** and adjust strategies accordingly.

**Marketing**

12. Which campaigns generated the most and least interest?

Formulas:

we need to analyze the **marketing** table to determine which campaigns had the highest and lowest levels of customer engagement.

Answers:

First sum each column :

MC3 -> =SUM(marketing!B:B) | 165

MC4-> =SUM(marketing!C:C) |165

MC5-> =SUM(marketing!D:D) | 161

MC1-> =SUM(marketing!E:E) |143

MC2 -> =SUM(marketing!F:F) | 30

~~Complaint-> =SUM(marketing!G:G) | 20~~

~~Pilot -> =SUM(marketing!H:H) | 334~~

A graph with blue and white bars

Description automatically generated with medium confidence

Based on these results, the campaigns that generated the most interest were the MC3 and MC4 Campaigns, while the least successful campaign was Campaign MC2. This indicates that Campaign MC3 and MC4 were effective in capturing customer attention, whereas Campaign MC2 may need further evaluation and adjustments to improve its effectiveness

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

Formulas:

I copy and paste the marital status and ID from the customer.csv file to compare with ID in the marketing.csv file

=VLOOKUP(A2, 'Marital status for customers'!$A:$B, 2, FALSE)

Answer:

A screenshot of a graph

Description automatically generated

From this analysis, the most common marital status of customers participating in Campaign 4 is Married This suggests that customers in the Married category are more likely to engage with this campaign compared to other marital status groups.

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:

* 1. Copy and paste the data from the sales.csv file to a new sheet in the marketing.csv file
  2. Create a new column named Campaign 2 Participation
  3. Use the **VLOOKUP** formula to import the **MC2** data from the **Marketing** sheet,

=VLOOKUP(A2, marketing!A:H, 6, FALSE),

Drag the formula down to populate the column for all rows in the **Sales** sheet

* 1. Calculate Total Sales for Each Product Category
     + Sum the Sales for Each Product Category

=SUMIF(G:G, 1, B:B) -> Teas

=SUMIF(G:G, 1, C:C) -> Fruits

=SUMIF(G:G, 1, D:D) -> Meats

=SUMIF(G:G, 1, E:E) -> Seafood

=SUMIF(G:G,1,F:F) -> Sweets

Answer: The product category with the highest total sold is Teas with 65527

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

Formulas:

1. **Combine Data from the Sales and Customers Sheets**
2. Use **VLOOKUP** to bring the **Complaint** data from the **Marketing** sheet to the **Customer** sheet

=VLOOKUP(A2, Marketing!$A$2:$I$2221, 7, FALSE)

1. filter the data to only show rows where the **Complaint** column equals **1**.
2. Use a **Pivot Table** on the filtered data to summarize the count of education levels

Answers:

The education level with the highest count in the Pivot Table results is the Graduate level

A screenshot of a graph

Description automatically generated

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:

**Additional Data to Collect:**

1. **Campaign Cost Data: The cost associated with each campaign**
2. **Customer Feedback and Satisfaction Scores: Surveys or ratings to determine customer Satisfaction**
3. **Seasonal and Timing Data**: Information on when campaigns were run (e.g., holidays, weekends) to identify time-based trends.
4. As you know the social media for company it's very important to post the campaign on it. Measures the level of interaction (likes, shares, comments) generated by the campaign. A higher engagement rate suggests the campaign resonates well with the audience.
5. It’s important to know if the campaign led to repeat purchases and improved customer retention, which are crucial for the long-term success of the company.
6. Comparing the campaign’s performance with competitors' campaigns can help understand whether the company’s marketing is on par with industry standards or outperforming competitors.
7. Compare sales figures during the campaign period to the same period before the campaign. A significant increase in sales can indicate campaign success.

To gain deeper insights into the success or failure of the marketing campaigns, collecting more detailed data (such as campaign costs, engagement metrics, and demographic data) and applying formulas like ROI and conversion rates would be essential. Additionally, competitor benchmarking and seasonal analysis can help understand external influences on campaign performance.

Bonus Question (Optional):

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

Quantitative data is numerical and measurable, such as sales or age, allowing for statistical analysis.

Qualitative data is non-numerical and descriptive, capturing characteristics like opinions or experiences to provide deeper insights.

Quantitative shows *what* is happening, while qualitative explains *why*.

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.

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?

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?

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?