PLOT-A-THON

STORE DATASET

Submission link: https://forms.gle/2Nkty2STt2na6Mzz7

Instructions:

- Merge the files and pre-process the database.
- Create visuals based on the following questions.
- The Visualization output should be in pdf format.
    - There should be creativity in the visuals.
    - The logical questions should be solved.
    - There should be an explanation for each visual.

Preprocessing Task:

The candidate will be required to:

1. **Merge Datasets**:
    - Join customers, sales, products, and stores datasets using keys (CustomerKey, StoreKey, ProductKey).
    - Incorporate exchange rate data by mapping Currency Code in sales to Currency in exchange rates based on the Order Date.
2. **Handle Missing Values**:
    - Fill missing values in categorical fields with "Unknown".
    - Fill numeric fields (e.g., Unit Price USD, Exchange) with the mean or median value.
3. **Standardize Data**:
    - Format dates (e.g., Order Date, Delivery Date) in a consistent format (e.g., YYYY-MM-DD).
    - Ensure consistency in column names (e.g., snake_case format).
4. **Create Derived Metrics**:
    - Calculate the **total revenue in USD** (Total Revenue = Quantity × Unit Price in USD).
    - Compute the **delivery time** as the difference between Delivery Date and Order Date in days.

- Determine the **customer age** using the Birthday field.

Visualization Questions (5 Questions):

1. **Customer gender distribution**:
   - Create a **pie chart** to visualize the gender distribution of customers.
2. **Product category distribution**:
   - Use a **tree map** to display the revenue share of each product category.
3. **Delivery time analysis by state**:
   - Create a **box plot** showing the delivery times across different states.
4. **Top 10 products by revenue**:
   - Use a **bar chart** to display the products contributing the most to total revenue.
5. **Store performance by square meters**:
   - Create a **scatter plot** with Square Meters on the x-axis, Total Revenue on the y-axis, and the bubble size representing the number of orders.