Question 1 (The only question) – 24 points

The data file provides transaction records for a store from January 1, 2023 to April 1, 2024. Each transaction has a date, a unique customer id, demographic information (age and sex) about the customer, and a dollar amount. Using the data, perform the following:

a. Load the data and split it into two parts, one with records from 2023 and one with records from 2024 (2 points).
b. Using the 2023 data, group the data based on customer id. Create a new data frame with this grouped data with the following information (3 points):
    - Customer age
    - Customer sex
    - Number of unique purchases
    - Total value of purchases
c. Using the 2024 data, create a similar grouped data frame (1 point).
d. Join the data frames from part (b) and part (c), being careful not to drop any data points Fill any missing values with appropriate values. (3 points).
e. Fit a generalized additive model with predictor variables from the 2023 data to predict the **if** a customer makes in the first three months of 2024. Your GAM should use a smoothed function for the numeric columns in your independent variables (6 points).
f. Using the same independent variables as part (e), fit a second GAM to predict the **value** of purchases each customer makes in the first three months of 2024 (3 points).
g. Using kernel density estimators, create plots demonstrating each of the following. Your plot can include a histogram for reference but must include a smooth KDE prediction. Each bullet point should be one plot with all required distributions labelled:
    - The distributions of purchase values in each of 2023 and 2024 (3 points).
    - The predicted and actual distributions of total purchase value by customer in 2024 (3 points).