**Task 7:**

**Data Merge and Integration Process**

Before data could be integrated, it was cleaned. Gender and Age values were the only columns in customer demographic data that had missing values.

The mode of the gender column was used to fill in missing values while for age, the median age was used because gender is a categorical column while the median age is not affected by outliers.

Data integrated is split between online and in-store sales data and customer data.

Online and in-store data were similar except in-store data has an extra shopping mall column. Therefore, the two datasets were vertically concatenated with online sales data having NA value for the shopping mall column.

The customer data was then merged with the total sales data. An inner join was used and the joining column was customer ID. This was because the values of customer ID in total sales and customer demographic data were similar.