Lab Guide Test – Set 3

10 Marks

ABOUT DATA: The data attached is a two-year sales data of a pharma company which talks about sales in 2015 and 2016 across various regions and time frames.
Account Id : Customer ID

Account Name : Customer Name

Tier : Customer Segment

Sales 2015 : Sales for the year 2015

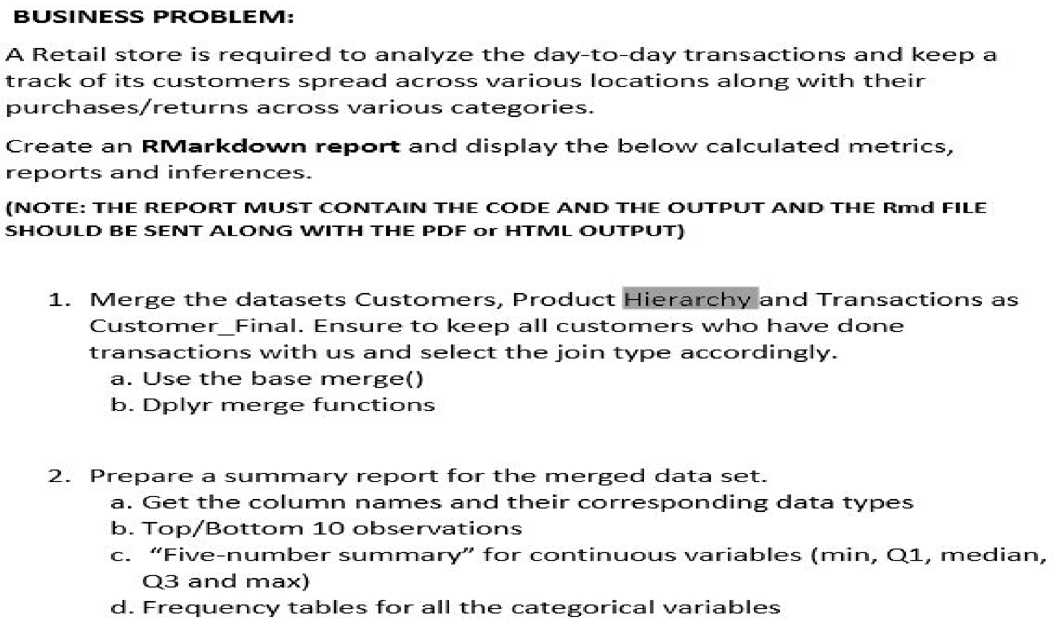
Sales 2016 : Sales for the year 2015

Units 2015 : No of Units sold for 2015

Units 2016 : No of Units sold for 2016

1. Compare Sales by region for 2016 with 2015 using bar chart
2. What are the contributing factors to the sales for each region in 2016. Visualize it using a Pie Chart.
3. Compare the total sales of 2015 and 2016 with respect to Region and Tiers
4. In East region, which state registered a decline in 2016 as compared to 2015?
5. In all the High tier, which Division saw a decline in number of units sold in 2016 compared to 2015?

15 Marks



BUSINESS PROBLEM:

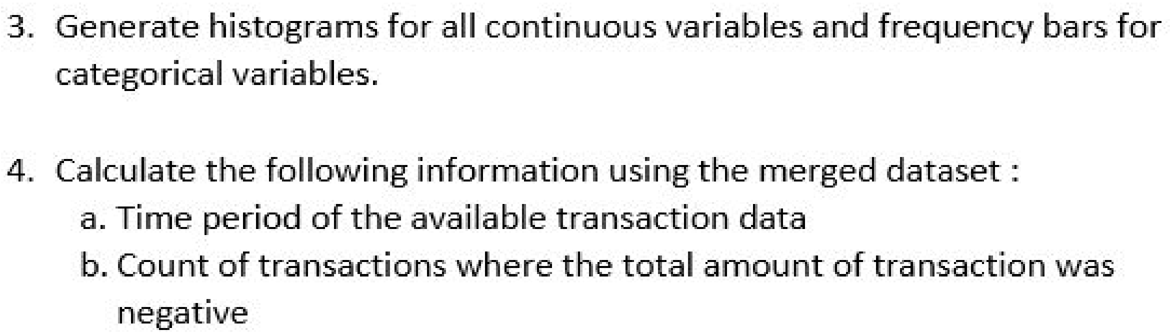
A Retail store is required to analyze the day-to-day transactions and keep a track of its customers spread across various locations along with their purchases/returns across various categories.

Create an **RMarkdown report** and display the below calculated metrics, reports and inferences.

(NOTE: THE REPORT MUST CONTAIN THE CODE AND THE OUTPUT AND THE Rmd FILE SHOULD BE SENT ALONG WITH THE PDF or HTML OUTPUT)

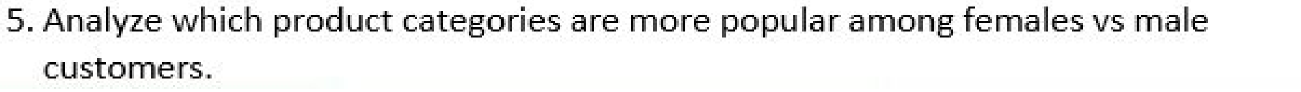
1. Merge the datasets Customers, Product Hierarchy and Transactions as Customer_Final. Ensure to keep all customers who have done transactions with us and select the join type accordingly.
   a. Use the base merge()
   b. Dplyr merge functions

2. Prepare a summary report for the merged data set.
   a. Get the column names and their corresponding data types
   b. Top/Bottom 10 observations
   c. "Five-number summary" for continuous variables (min, Q1, median, Q3 and max)
   d. Frequency tables for all the categorical variables



3. Generate histograms for all continuous variables and frequency bars for categorical variables.

4. Calculate the following information using the merged dataset :
   a. Time period of the available transaction data
   b. Count of transactions where the total amount of transaction was negative



5. Analyze which product categories are more popular among females vs male customers.