Lab Guide Test – Set 4

10 Marks

1. Import the customer data into R using read.csv, read.table etc.

2. Understand the data using different functions like View, head, tail, str, names, nrow, ncol, summary, duplicates, describe etc.
3. What is the percentage of missing values for a customer Value variable?

4. Create two subsets with unique and duplicate values.

5.Create data set with list of customers whose customer value greater than 10000.
6. In customer table, create a new variable called “customer value segment” using customer value as follows. - High Value Segment - > 25000 - Medium Value Segment – Between 10000 and 25000 - Low Value Segment – less than or equal to 10000
7. Create variables “average revenue per trip” and “balance points” in the 10000.
8. How many days between last purchase date and today?

9. Calculate percentage of sales by each last city, state and region.

10. What is the count of customers, average number of purchases and average purchase transaction value by last state and city

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.