Quantium Virtual Internship - Retail Strategy and Analytics - Task 1

.Certainly! Here's a comprehensive R code example for a typical data preparation and customer analytics task. This example will guide you through:

1.Data Preparation: Loading data, cleaning it, and creating relevant subsets.

2.Customer Analytics: Analyzing customer segments, summarizing data, and visualizing insights.

```
# Load necessary libraries

library(dplyr)

library(ggplot2)

#### Point the filePath to where you have downloaded the datasets to and

#### assign the data files to data.tables

filepath <- "C:/Users/indup/AppData/Local/R/win-library/4.4"

transactionData <- fread(paste0(filePath,"QVI_transaction_data.csv"))

customerData <- fread(paste0(filePath,"QVI_purchase_behaviour.csv"))
```

Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process that involves examining and summarizing the main characteristics of a dataset.

Examining Transaction Data:

Examining transaction data can be quite insightful for various purposes, such as understanding customer behavior, assessing financial performance, or identifying trends.

```
# Simulate a large dataset

set.seed(123)

n <- 100000  # Large number of rows

data <- data.frame(

  CustomerID = 1:n,

  LIFESTAGE = sample(c("Pre-School", "Primary", "Secondary", "No Children"), n, replace = TRUE),

  PREMIUM_CUSTOMER = sample(c("High", "Medium", "Low"), n, replace = TRUE),
```

```
  PurchaseAmount = runif(n, 5, 500)  # Random purchase amounts between $5 and $500
)

# Create a sample dataset
sample_size <- 1000  # Size of the sample
sample_data <- data %>% sample_n(sample_size)

# Function to create pie charts
create_pie_chart <- function(data, group_var, value_var, title) {
  data %>%
    group_by_at(group_var) %>%
    summarise(TotalAmount = sum(get(value_var))) %>%
    mutate(Percentage = TotalAmount / sum(TotalAmount) * 100,
           Label = paste0(round(Percentage, 1), "%")) %>%
    ggplot(aes(x = "", y = Percentage, fill = .data[[group_var]], label = Label)) +
    geom_bar(width = 1, stat = "identity") +
    coord_polar(theta = "y") +
    labs(title = title, x = NULL, y = NULL) +
    theme_void() +
    theme(legend.title = element_blank())
}

# Create pie charts for sample data
create_pie_chart(sample_data, "PREMIUM_CUSTOMER", "PurchaseAmount", "Sample Data: Purchase Amount by PREMIUM_CUSTOMER")
create_pie_chart(sample_data, "LIFESTAGE", "PurchaseAmount", "Sample Data: Purchase Amount by LIFESTAGE")
```

 # Create pie charts for full dataset

create_pie_chart(data, "PREMIUM_CUSTOMER", "PurchaseAmount", "Full Data: Purchase Amount by PREMIUM_CUSTOMER")

create_pie_chart(data, "LIFESTAGE", "PurchaseAmount", "Full Data: Purchase Amount by LIFESTAGE")