Introduction to Data Science

Instructor: Daniel D. Gutierrez

MODULE 7 QUIZ

Question 1

The following are ways you can use the features of the R statistical environment in support of the numeric exploratory data analysis phase of a data science project EXCEPT which one?

(a) Applying general statistical methods

(b) Calculating numeric summaries and aggregations

(c) Creating a new variable based on one or more other variables

(d) Reviewing all levels of factor variables

Question 2

The `summary()` function is useful for numeric EDA. Which of one the following best describes this function?

(a) Calculates summary statistics for each numeric variable: Min, Max, Meadian, first quartile, and third quartile

(b) Shows the count of the most frequently occurring values for factor variables

(c) Processes all the observations in a data frame

(d) All of the above

Question 3

Which one of the following best characterizes the `cor()` function in R?

(a) Computes a statistical correlation

(b) Used to indicate how widely individuals in a group vary

(c) Returns a vector containing both the minimum and maximum value

(d) None of the above

Question 4

Which one of the following best describes a feature of the `table()` function in R?

(a) Displays the unique levels in a factor variable

(b) Tool for viewing the distribution of a numeric variable

(c) Calculates the cumulative sum of a numeric vector

(d) Provides the ability to produce a "contingency table"

Question 5

Which one of the following describes ways in R to perform counts?

(a) Use the `unique()` function

(b) Use the `sqldf` package and perform the count using SQL

(c) Use the `length()` function along with subsetting of a vector

(d) All of the above