SPRING SEMESTER 2023
IST 3015 (A): BUSSINESS DATA ANALYTICS
INSTRUCTOR: JAPHETH MURSI
DATE: 23rd March 2023, Venue: Online

QUIZZ 2

Duration: 1hr 45 Mins

Total marks (20)

Instruction

1. Show the output below each question

Question 1 (8mks)

a) How would you apply plotly function to plots in R package? **(2mks)**
b) Elucidate four data structures in R? **(2mks)**
c) Create a vector with 8 of your friend's names. Display the 2nd and 3rd friends **(2mk)**
d) Sort your friends by names using two sorting methods **(2mks)**

Question 2 (12mks)

In the CRB_Score _dataset attached, Subset the dataset and create two data frames for JobID 1 and 2.

a) Merge the two data frames and sort by "Housing". **(2mks)**

b) Using the merged dataframe in (a) above, create a boxplot for *Credit Amount and Duration*. Facet the plot by the column *'Jobid'*. **(2mks)**

c) Test the hypothesis whether the customers default score is independent of their credit amount at .05 significance level. **(2mks)**

d) Using the whole dataset, create a scatterplot for *Age* and Duration column. Use *color and shape functions* in your plot. Label the plot as "Duration distribution". Connect the plot points using geomline. **(4mks)**

e) Is there a relationship between the age of the customer and the Duration of payment? **(2mks)**