SPRING SEMESTER 2023
IST 3015 (B): BUSSINESS DATA ANALYTICS
INSTRUCTOR: JAPHETH MURSI
DATE: 5th APRIL 2023, Venue: ICTLAB

END OF SEMESTER EXAMS

Duration: 1hr 45 Mins

Total marks (30)

Instruction

1. Show the output below each question

Question 1 (10mks)

a) Discuss any four 4 properties of data frame **(2mks)**

b) You have two Dataframes df1 and df2, both with columns col1 and col2. How can you combine the two Dataframes into a single dataframe df3 where the rows from df1 come first, followed by the rows from df2? **(2mks)**

c) The values of x and their corresponding values of y are shown in the table below

x	101	211	322	332	451	552	170	70	171
y	162	173	285	294	299	232	144	26	68

i) Find the least squares regression line y = a + bx. **(3mks)**

ii) Estimate the value of y when x = 397. **(1mks)**

iii) Calculate coefficient correlation r **(2mks)**

Question 2 (10mks)

a) Giving examples, discuss three ways Subsetting a list in R **(3mks)**

b) Using "Bank Churners" Dataset attached, Conduct Exploratory data analysis on the dataset and comment on few interesting observations **(4mks)**

c) Plot out a "Violin plot" to show relationship between "Months_on_book" and "Total_Amt_Chng_Q4_Q1" column. **(3mks)**

Question 3 (10mks)

a) Discuss process analysis workflow**(3mks)**

b) Using "Bank Churners Dataset create a scatter plot using ggplot2, where each plot shows the relationship between "Months_on_book" and "Credit_Limit" and show the differnet education levels in your plot. Use facet_wrap() to arrange the plots based on Marital status. **(3mks)**

c) Test the hypothesis whether the Credit_Limit is independent of their Total_Revolving_Bal at .05 significance level. **(2mks**

d) Using Bank Churners Dataset, create stacked barchart of column Total_Trans_Amt use Card_Category as fill. Label the plot as "Total AMount". **(2mks)**