I

**SPRING SEMESTER 2023**

**IST 3015: BUSSINESS DATA ANALYTICS**

**INSTRUCTOR: JAPHETH MURSI**

**DATE: 20TH March 2022**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Exercises**

# Question 1 (13Mks)

1. Using survey dataset (attached) generate summary statistic of the dataset (2mks)
2. Create four vectors a, b, c, d with integers and each vector has 4 elements. Combine the three vectors to become a 4×4 matrix and name it M1. Where each column represents a vector. Change the row names to Vec1, Vec2, Vec3, Vec4 and show the output

c) Subset the survey dataset and create a dataframe for the month of November only

(2mks)

1. Create a second data frame of the same format as above for september (1mks)
2. Merge the two dataframe and sort by plot\_id (2mks)
3. Using Survey dataset attached, write a code to display boxplot of *Gender (sex*) and *Weight*
4. Write a program to display boxplot of Gender (sex) and Weight (3mks)
5. Create a stacked Barchat for the column “plot\_id” and categorize based on species\_id
6. Write a R program to create a matrix from a list of given vectors and show a line plot of Month vs sales and label the plot(4mks)

Month <-c(Jan,Feb,Mar,Apr, May, June, July, Aug, Sept, Oct, Nov, Dec)

Sales<-c(50000,35000,90000,150000,25700,34000,56000,45000,78000,98000,88000,92000)