SVKM's NMIMS
MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING

Programme: M.Tech Data Science (Business Analytics) Year: I Semester: I

Academic Year: 2019-20

Subject: R/R Studio and Tableau Programming

Date: 18 November 2019

Marks: 100
Time: 2.00 pm - 5.00 pm
Durations: 3 (Hrs)
No. of Pages: 05

Final Examination (2019-20)

Instructions: Candidates should read carefully the instructions printed on the question paper and on the cover of the Answer Book, which is provided for their use.

1) Question No. _1___ is compulsory.
2) Out of remaining questions, attempt any __4___ questions.
3) **In all ___5_ questions to be attempted.**
4) All questions carry equal marks.
5) **Answer to each new question to be started on a fresh page.**
6) **Figures in brackets on the right hand side indicate full marks.**
7) **Assume suitable data if necessary.**

1.

A) Create a function calc() to give options of addition or multiplication for any number of inputs. (10)

B)

i. Download and read a csv file using the link to your system https://www.census.gov/2010census/csv/pop_change.csv by correctly picking the header and setting index as STATE_OR_REGION column. (1)

ii. Write a code to check if there is any missing data. (2)

iii. Select the columns containing the string "Population" (2)

iv. Create a new column calculating the percent change of the population during 2010 with respect to 2000. And filter out the states having negative percent changes. (3)

1/5

iv. Create new rows calculating the sum and mean of the columns. (2)

2. A) Use the dataset ChickWeight

i. What is the average weight of all chick during the same time? (2)

ii. Get difference of the weight. Plot a chart of these differences with the timeline and comment on which period do the chick gain the most weight. (3)

iii. Among the four diets, which diet resulted in highest and lowest gain in weight. State with reasons. (3)

iv. Create unique list of chicks who crossed the weight of 300.(2)

B) Install / import package / csv file for hflights.

i. Filter for flights originating from IAH airport

ii. Count total flights and delayed flights by each carrier

iii. Calculate the percent delay for each carrier

iv. And sort the result in descending order

3.

A) Create the following charts using iris data set using Sepal.Length as x and Sepal.Width as y:

i. Create a line chart. (2)

ii. Create a new chart showing the smooth mean for the data. (2)

iii. Add a layer of scatter plot on the previous chart. (2)

iv. Plot a new chart with color representation for the species and size for the length. (2)

v. Edit the labels and add title, subtitle and captions to the chart. (1)

vi. Save the chart in the working folder. (1)

B) Use the data iris and do the following

i. Gather the data reducing the column and groupby with the new column and calculate the mean of it. (2)
ii. In iris data which Species has the highest mean Sepal.Width (2)
iii. Filter the Iris data with Sepal.Length greater or equal to 4.6 and Petal.Width greater or equal to 0.5. and select all the columns starting with P. (2)
iv. Arrange the data in descending order with respect to Petal.Length and display the top 10 Petal.Length data. (2)
v. Fetch the last two rows in the last two columns of Iris data without explicitly specifying the location. (2)

4.

A)

i. Create the following dataframe (2)
ii. Merge them together (2)
iii. Chart the scores for each sem individually (2)
iv. Chart them together as a comparison of the subject (2)
v. Chart the subjects stacked (2)

Variable	sem_1	
Semester	Exam	Marks
Sem 1	R	70
Sem 1	Finance	63
Sem 1	Marketing	82
Sem 1	Machine Learning	63
Sem 1	Python	78

Variable	sem_2	
Semester	Exam	Marks
Sem 2	R	82
Sem 2	Finance	74
Sem 2	Marketing	64
Sem 2	Machine Learning	70
Sem 2	Python	65

3/5

B)

i. Create a function to calculate the list of prime number from 0 to nth. (10)

5.

A)

i. Get the stock data for Microsoft (MSFT), Google(GOOG), Apple(APPL) and Tesla(TSLA) from the period Jan 2018 till now. Export the google data in csv format. (1)

ii. Plot a chart of Apple and Tesla and add appropriate labels. (2)

iii. Calculate the correlation of the three stock and state which two stocks are highly correlated. (2)

iv. Calculate the daily stock's closing price movement/difference for Tesla and plot a chart of the price movement. (3)

v. Calculate mean, max and min using apply-family and function. (2)

B)

i. Write a program which checks for the user input password for login. User can only input 3 times the password. If the password match with the stored one then print 'You have successfully logged in' else print the number of failed login. (2 Failed login. 1 more attempt left). (10)

6.

A) Import the dataset airquality

i. Check if any NA in the data and the total count of NAs present in the data. (2)

ii. Clean the data by removing the NA rows. (2)

iii. Clean the data by substituting the NA with the mean of the whole column. (2)

iv. Plot a line chart on Ozone with filtered data of Month as 5. (4)

B) In the same dataset airquality

i. Clean the data by taking mean of the two adjacent rows. In case the next row is also NA then take previous row as the value. (8)

ii. Which month had the highest ozone level? (2)

7. Create a tableau data importing the google stock data.
   A) Chart the closing price. (2)
   B) Add the volume chart. (2)
   C) Color indicate the variation in volume. (2)
   D) Create a field to check the range of open and close of the day and chart it accordingly. (3)
   E) Export the file to image. (1)
   F) Perform simple linear regression on the data using R. (10)

5/5