****

**INTERNAL EXAMINATION: FEB – MAR 2020**

**MBA Batch 2019 – 21: Semester - II**

**Subject Code: 19JBBA207 Subject Name: Programming with R**

**Duration: 2 Hours Marks: 50**

**SECTION –A**

1. Answer any **TWO** of the following Three Questions: 2X8=16

**a.**

1. There are instances where we may need to selective replace occurrence(s) of a text within a string. Show with an example how the sub() and the gsub() functions perform replacement of a substring. Highlight the difference between the two functions. [6]
2. Transforming strings from lower case to upper case and vice versa is a common operation because the data are rarely set up exactly the way you need it for doing the analysis. Convert the characters in your above example to upper and lower case? [2]

**b.**

1. The Khel Utsav of CMS had three sports listed in the outdoor category – Hockey, Football and Cricket. Create a factor vector with six elements from the mentioned categories and label it as *sportfact*. How will you check the data type of *sportfact* to reconfirm whether a factor vector was created? What do the levels indicate when you print the *sportfact*? [4]
2. Create a matrix *matA* of size 2X3 using a sequence of numbers 1:6 and *matB* of size 3X2 using the sequence 6:1. Perform matrix multiplication to create *matAB* using the two matrices A and B. How will you check the size of *matAB*? Write the code to extract the element in the first row and second column position. Label the rows and columns of *matAB* suitably. [4]

**c.**

1. Around 100 datasets are supplied with base R distribution in the package datasets. How will you examine the first few rows of the in-built mtcars dataset? Which function will help you get a quick look at the data and its structure? [2]
2. Export the in-built mtcars dataset to your working directory as a .csv file. Import the file from the working directory into the environment under the name *mymtcars*. [4]
3. Change the data type of the variable cyl in the *mymtcars* data frame to a factor. [2]

**SECTION-B**

2. Answer any **TWO** of the following Three questions: 2X12=24

**a.**

1. Student data is a very crucial information for an educational institution. Maintaining the database in an appropriate tabular structure helps in accessing the information easily by the stakeholders. You are required to create a data frame *studentdf* for four students from three vectors which show their Name, UG Degree and UG Percentage. (Use imaginary names and values to create the three vectors) [6]
2. Which function will you use to see the last two rows of *studentdf*? How will you access only the name and percentage columns of all the students from *studentdf?* [2]
3. Create a new data frame consisting of student name and fee for the same four students. Merge this data frame with *studentdf* to create a *studentdf\_full* data frame. [4]

**b.** Answer the below questions using the Placement data file provided to you. Use the functions of dplyr only.

1. How will you filter students from Science and Technology (Sci&Tech) degree having degree percentage (degree\_p) ≥ 75. Create a dataset named *pldplyr* using these observations. How many observations are there?
2. Reorder (ascending) the rows in *pldplyr* based on degree\_p. What is the sl.no of the student at the top and bottom?
3. How will you select columns that starts with h? Which columns got selected?
4. Create a new variable (avg\_p) which shows the average percentage of ssc\_p and hsc\_p. What is the avg\_p of the first student in the output?

**c.** Answer the below questions using the Placement data file provided to you. Use the ggplot2 package only.

1. Draw a scatter plot between variables ssc\_p and degree\_p. Add an additional dimension of status using the color attribute.
2. Draw a bar chart to show levels under gender. Convert the bars into a stack based on status.
3. Draw a histogram for degree\_p. Use appropriate bins/bin width and fill the bars using a color of your choice.

**SECTION-C**

3. Case study - **Compulsory**: 1X10=10

Create a .csv file in your working directory for the below information and plot a line chart for nifty. Weightage will be given for beautifying the plot.

|  |  |  |  |
| --- | --- | --- | --- |
| **month** | **nifty** | **month** | **nifty** |
| 1 | 10830 | 7 | 11120 |
| 2 | 10790 | 8 | 11020 |
| 3 | 11620 | 9 | 11470 |
| 4 | 11750 | 10 | 11880 |
| 5 | 11920 | 11 | 12060 |
| 6 | 11790 | 12 | 12170 |