School of Technology Management & Engineering | Navi Mumbai Campus
Course: B. Tech. (AI&DS)
Class: Second Year | Semester- III
Subject: Data Science Wrangling

MID TERM 1 (23rd Aug 2022) 20 Marks

Students are requested to read the instructions carefully before attempting the questions
All the 4 question are compulsory
Marks for each question is given after every question

Instructions:

- *Create a folder in the desktop of your <SAP number_MT1> and set this as your working directory.*
- *Example: If my SAP no. is 1234567 then the folder created will be 1234567_MT1*
- *Save all your files in this directory only*

Install the packages tidyr, dplyr, validate if not installed
Your R file should consists of these 3 lines at the start of the program
library(tidyr)
library(dplyr)
library(validate)

Use Case (Data): Please refer ICC.csv

Indian Cricket Team Coach has shared a dataset (ICC.csv) with you. This data is of 3 bowlers (AVK, KYV, BUVI). Develop / write a R program which covers all the syntax of the below questions:

Question 1: **(2 Marks)**
Use R application program to create a new column "Ball_Type" in the data set ICC.csv, the values of "Ball_Type" are conditional, referencing to column "Runs_Given".
Condition: If the value in the "Runs_Given" column is 0 then the status or the string in the "Ball_Type" column should be "Dot_Ball" else "Runs_Scored".
(Hint:
Read this csv file in a variable
Use mutate function)

Question 2: **(2 Marks)**
The three bowlers have bowled 3 overs in each spell so together they have bowled 36 balls. Find the percentage of "Dot_Ball" (Print the value in a statement).

Question 3: **(4 Marks)**
Replace all the "NB" values in "Wicket_Type" column by "RUNS".
In the column "Wicket_Type", wherever the value is "RUNS", replace it by "NO_Run" if the value of "Ball_Type" is "Dot_Ball".

Question 4: **(12 Marks)**

Create a new Data Frame "Analys1" and save it as Analys1.csv file with the following information:

Bowler_Name	Each_Spell	Total_Runs_Given	No_of_Wickets
AVK	1		
AVK	2		
KYV	1		
KYV	2		
BUVI	1		
BUVI	2		

Calculate the "Total_Runs_Given" i.e. conceded and "No_of_Wickets" taken by each bowler (it should be in the above table).

(Hint:
<your_variable><- sum(ICC[which(ICC['Bowler'] == "AVK" & ICC['Spell'] == 1),'Runs_Given'])
<your_variable><- length(which(ICC$Wicket_Taken == "Yes" & ICC$Bowler == "AVK" & ICC$Spell == 1))
Similarly, find the values for other bowlers.
Once all the variable values are calculated, create a Data frame and then write it to the working directory as csv file)

Finally, save your R file into the working directory.