# Report on Creating a Dataset of 300-Level Computer Science Students in Bowen University

## R Programming Language

R is a programming language and free software environment used primarily for statistical computing and graphics. It is highly extensible and provides a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and more. The comprehensive R Archive Network (CRAN) offers a large repository of packages that extend R's functionality.

## Question Analysis

The task requires the creation of a dataset for 300-level Computer Science students at Bowen University. The dataset should include fields for Matric Number, First Name, Other Name, Gender, and Date of Birth. Additionally, we need to derive the age of each student from their Date of Birth and the current date, then add this information as a new column to the dataset.

## Problem Requirment

To solve this problem, we need the following:

- An input CSV file containing the student data (First Name, Other Name, Gender, Matric Number, Date of Birth).

- R programming environment to read the input data, process it, and write the output.

- Functions to calculate the age of each student from their Date of Birth.

- Proper handling of date formats in R.

## Code

Here is the R code used to achieve the required solution:

```r

# Read the data from a csv

class\_data\_file <- read.csv("class\_attendance.csv")

# Function to calculate age from DateOfBirth

calculate\_age <- function(dob) {

dob <- as.Date(dob, format = "%d/%m/%y") # Adjusting the date format to match the provided data

current\_date <- Sys.Date()

age <- as.numeric(difftime(current\_date, dob, units = "weeks")) %/% 52.25

return(age)

}

# Apply the age calculation function to the DateOfBirth column

class\_data\_file$Age <- sapply(class\_data\_file$DateOfBirth, calculate\_age)

# Writes the data to a CSV file

write.csv(class\_data\_file, file = "output.csv", row.names = FALSE, quote = FALSE)

print(class\_data\_file)

```

## Result

The result of running the above R code is a new CSV file named `output.csv`, which includes the original student data along with an additional column for Age. The Age column is calculated based on the Date of Birth and the current date at the time of running the script. Here is an example of the output data frame:

```

FirstName OtherName Gender Matric DateOfBirth Age

1 David Sopirinye Jaja Male BU21CSC1001 2001-10-20 22

2 Jane Smith Female BU21CSC1002 2001-05-12 22

...

```

## Conclusion

In this project, we successfully created a dataset for 300-level Computer Science students at Bowen University, including Matric Number, First Name, Other Name, Gender, Date of Birth, and an automatically calculated Age. Using R's powerful data manipulation capabilities, we efficiently read, processed, and wrote the required data to a new CSV file. This process illustrates the utility of R in handling and analyzing structured data.