**Answer script**

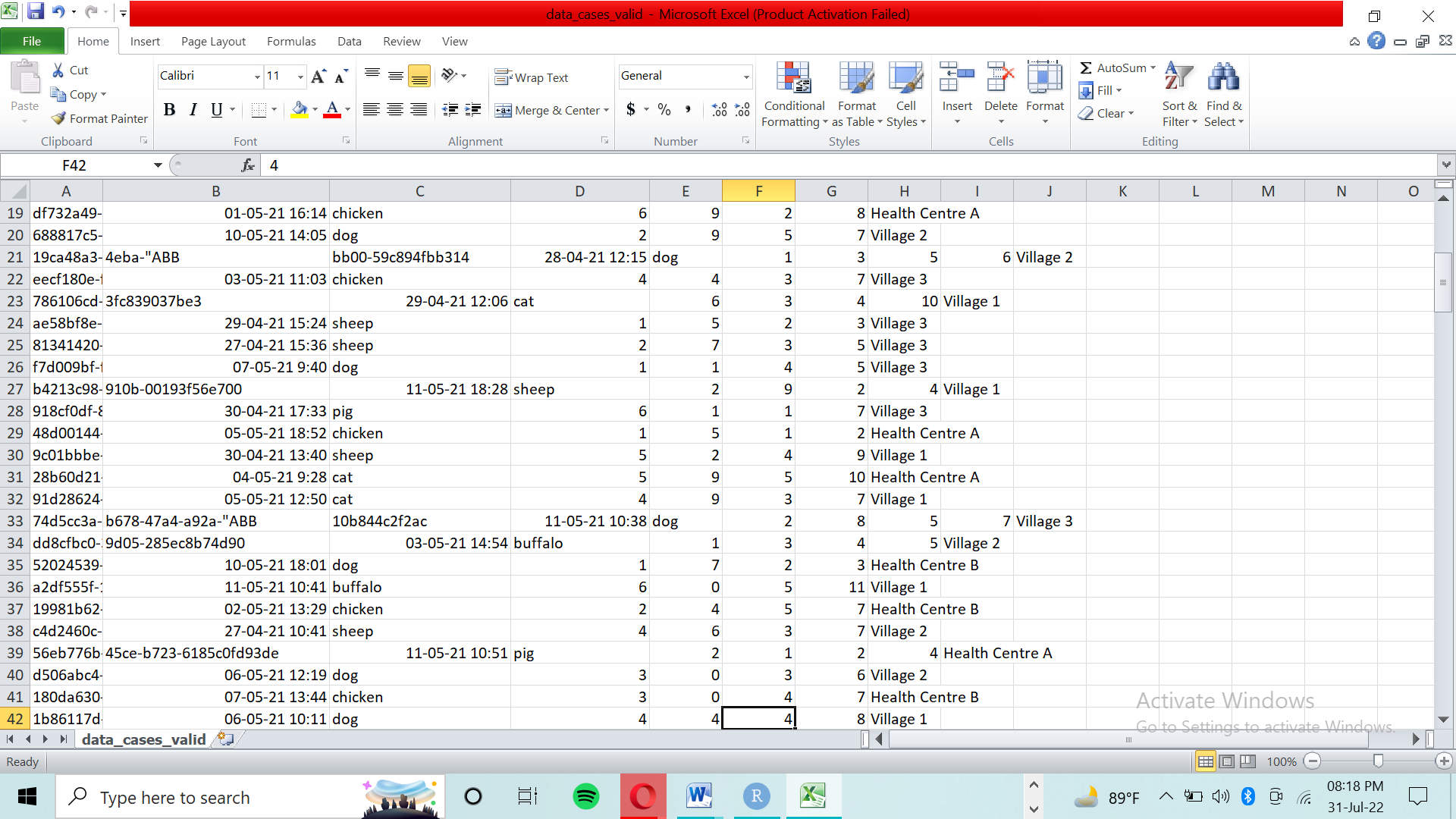
**Task [D]**

|  |  |
| --- | --- |
| **Input File** | data\_cases\_corrupted.csv  data\_cases\_valid.csv |
| **Output File** | Output\_valid.json |
| **Installed Packages** | dplyr |

**Program**

**# Program for cleaning data\_case\_corrupted.csv and Find required Output**

**Data Cleaning**



* Found some garbage’s in some preceding columns of the rows.
* Data is reordered as their column head
* Replaced data right to left cell for having cleaned data set data\_cases\_valid.csv

**#STEP1** To Know about working directory

getwd()

**#STEP2** Set working directory in folder Test\_wd

setwd(“C:/Users/HP/OneDrive/Documents/Test\_wd”)

**#STEP3** Read CSV file data\_cases\_valid and stored in data frame data

data <- read.csv(“data\_cases\_valid.csv, header = TRUE, sep =”,“ )

**#STEP4** Remove 1st Header uuid from imported CSV file

data$uuid <- NULL

**# Program for output file as given examples**

**#STEP6** Find total number of cases

tcases <- sum(data$total\_number\_cases)

sprintf(“Total number of reported cases is: %d“ , tcases)

tcase\_op <- sprintf(“Total number of reported cases is: %d“ , tcases)

**#STEP7** Find total number of death cases at each location

deathbyloc\_op <- print(“Total number of deaths reported at each location : ”)

**#STEP8** Find Sum for Group by data : location

aggregate(data$number\_mortality, list(data$location), FUN = sum)

groupdata <- aggregate(datafile$number\_mortality, list(data$location), FUN = sum)

**#STEP8** install Package for json

install.packages("rjson")

library(rjson)

**#STEP9** convert data frame to Json

myfile = toJSON(groupdata)

**#STEP10** store data in tcase\_op for showing output of Total number of reported death cases

tcase\_op <- sprintf("total number of reported cases is : %d ", tcase)

**#STEP11** data bind

Output <- rbind( tcase\_op, deathbyloc\_op)

**#STEP12** final code for showing output in output.json file

write(output, "output.json")

write(myfile, "output.json", append = TRUE)