Step for cleaning data:

First: Based on the research questions, we reviewed all questions and checked the data in the dataset to see which columns are need for answering questions and analyzing dataset.

Second, we checked the format of the data set to see which kind of command is needed to use in R.

Third: getting the summary of data with using R to see what is the minimum, maximum etc. of each column in the dataset. With this summary, we got the overview of whole set of data.

Forth: looking for missing data and found all missing data and apply NA in the rows which has missing value.

## Fifth: try to figure out how can deal with missing value like Data prep and pattern, Deleting the observations, Deleting the variable, Prediction or Imputation with mean / median / mode.

## Sixth: We used deleting observation because there is large number of observations in the dataset, where all the classes to be predicted are sufficiently represented in the training data, then we deleted them from the dataset.

## Problem with cleaning data:

## Unfamiliarly with R and how can use comments and code was the biggest issue in cleaning data set. It took while to figure out how can use R. as an example it took time to know we need create the block for the code to run the code and also close it before running the code.

## Another issue was searching in google to find the code for finding missing data which bring different methods and trying all of them was kind of hard to get the result.