Assignment 1: Consider the link of Techno India University:
"https://www.technoindiauniversity.ac.in/". Scrape this webpage in R to extract the text written about each of the schools under this university.

Solution:

library(rvest)

library(stringi)

url <- "https://www.technoindiauniversity.ac.in/"

webpage <- read_html(url)

results <- webpage %>% html_nodes(".single-content")

schools <- results %>% html_nodes("p") %>% html_text(trim = TRUE)

print(schools)

Output:

```
> install.packages('rvest')
Installing package into 'C:/Users/91943/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'https://cran.icts.res.in/bin/windows/contrib/4.3/rvest_1.0.4.zip'
Content type 'application/zip' length 305071 bytes (297 KB)
downloaded 297 KB

package 'rvest' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\91943\AppData\Local\Temp\RtmpEHagWp\downloaded_packages
> library(rvest)
Warning message:
package 'rvest' was built under R version 4.3.3
> library(stringi)
>
> url <- "https://www.technoindiauniversity.ac.in/"
> webpage <- read_html(url)
>
> results <- webpage %>% html_nodes(".single-content")
>
> schools <- results %>% html_nodes("p") %>% html_text(trim = TRUE)
> print(schools)
 [1] "The Department of Engineering at Techno India University, West Bengal is a conglomeration of multiple disciplines with separate"
 [2] "The Architecture major establishes an intellectual context for the students to interpret the relation of form, space, program,"
 [3] "The School of Pharmacy was established with a view to promote center of excellence in pharmaceutical education and to prepare"
 [4] "The Department of Science at Techno India University, West Bengal is a conglomeration of multiple disciplines with separate"
 [5] "The concepts of computer science are applicable to many aspects of our lives, from investigations in the scientific and medical"
 [6] "In the era of globalization, disruptive technological change, School of Management, Techno India University, West Bengal"
 [7] "\"Today and always, there will be an obligation to pass on to the new generation the tradition of liberal"
 [8] "The work of a lawyer is broad in scope as there is no singular occupation designated to the title. A wide range of alternatives is available,allowing"
 [9] "The Commerce department at Techno India University, West Bengal aims to teach students in different domains of Commerce, Finance and"
[10] "Nursing is a basic course to introduce students to all aspects of general healthcare and unique proficiency."
```

Assignment 2: Consider the following link: "https://www.cricbuzz.com/cricket-series/5945/indian-premier-league-2023/points-table". Scrape this webpage in R, to procure the information available there.

Solution:

library(rvest)

url <- "https://www.cricbuzz.com/cricket-series/5945/indian-premier-league-2023/points-table"

webpage <- read_html(url)

team_names <- webpage %>%

  html_nodes(".cb-col-84") %>%

  html_text()

points <- webpage %>%

  html_nodes(".cb-srs-pnts-td") %>%

  html_text()

point_names <- webpage %>%

  html_nodes(".cb-srs-pnts-th") %>%

  html_text()

data <- data.frame(

  Team = team_names

)

point_matrix <- matrix(ncol = length(point_names), nrow = length(team_names))

point_matrix[, 1] <- team_names

point_matrix[, -1] <- matrix(as.numeric(points), ncol = 7, byrow = TRUE)

```
colnames(point_matrix) <- point_names


final_data <- as.data.frame(point_matrix)

html_table <- knitr::kable(final_data)


print(html_table)
```

Output:

```
> library(rvest)
>
> url <- "https://www.cricbuzz.com/cricket-series/5945/indian-premier-league-2023/points-table"
> webpage <- read_html(url)
>
> team_names <- webpage %>%
+   html_nodes(".cb-col-84") %>%
+   html_text()
>
> points <- webpage %>%
+   html_nodes(".cb-srs-pnts-td") %>%
+   html_text()
>
> point_names <- webpage %>%
+   html_nodes(".cb-srs-pnts-th") %>%
+   html_text()
>
> data <- data.frame(
+   Team = team_names
+ )
>
> point_matrix <- matrix(ncol = length(point_names), nrow = length(team_names))
> point_matrix[, 1] <- team_names
> point_matrix[, -1] <- matrix(as.numeric(points), ncol = 7, byrow = TRUE)
> colnames(point_matrix) <- point_names
>
> final_data <- as.data.frame(point_matrix)
> html_table <- knitr::kable(final_data)
>
> print(html_table)
```

Teams	Mat	Won	Lost	Tied	NR	Pts	NRR
Gujarat Titans (Q)	14	10	4	0	0	20	0.809
Chennai Super Kings (Q)	14	8	5	0	1	17	0.652
Lucknow Super Giants (Q)	14	8	5	0	1	17	0.284
Mumbai Indians (Q)	14	8	6	0	0	16	-0.044
Rajasthan Royals (E)	14	7	7	0	0	14	0.148
Royal Challengers Bengaluru (E)	14	7	7	0	0	14	0.135
Kolkata Knight Riders (E)	14	6	8	0	0	12	-0.239
Punjab Kings (E)	14	6	8	0	0	12	-0.304
Delhi Capitals (E)	14	5	9	0	0	10	-0.808