

# Final Project

Group 4

2023-12-22

```
library(rvest)
library(httr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(polite)
```

```
url <- "https://www.airlinequality.com/airline-reviews/emirates/page/1/?sortby=post_date%3ADesc&pagesize=10"
session <- bow(url, user_agent = "Educational")
session
```

```
## <polite session> https://www.airlinequality.com/airline-reviews/emirates/page/1/?sortby=post_date%3ADesc&pagesize=10
##   User-agent: Educational
##   robots.txt: 2 rules are defined for 1 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
webpage <- read_html(url)
```

```
ReviewTextP1 <- webpage %>%
  html_nodes("div.text_content") %>%
  html_text()
```

```
url <- "https://www.airlinequality.com/airline-reviews/emirates/page/2/?sortby=post_date%3ADesc&pagesize=10"
session <- bow(url, user_agent = "Educational")
session
```

```
## <polite session> https://www.airlinequality.com/airline-reviews/emirates/page/2/?sortby=post_date%3ADesc&pagesize=10
##   User-agent: Educational
##   robots.txt: 2 rules are defined for 1 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
webpage <- read_html(url)
```

```
ReviewTextP2 <- webpage %>%
```

```

html_nodes("div.text_content") %>%
html_text()

url <- "https://www.airlinequality.com/airline-reviews/emirates/page/3/?sortby=post_date%3ADesc&pagesize=30"
session <- bow(url, user_agent = "Educational")
session

## <polite session> https://www.airlinequality.com/airline-reviews/emirates/page/3/?sortby=post_date%3ADesc&pagesize=30
##   User-agent: Educational
##   robots.txt: 2 rules are defined for 1 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent

webpage <- read_html(url)

ReviewTextP3<- webpage %>%
  html_nodes("div.text_content") %>%
  html_text()

R1 <- as.data.frame(ReviewTextP1)
R2 <- as.data.frame(ReviewTextP2)
R3 <- as.data.frame(ReviewTextP3)

names(R1) <- names(R2) <- names(R3) <- "Emirates Costumer Reviews"

Reviews_300 <- rbind(R1, R2, R3)

```