

# Assignment12: Airline Quality

2025-10-16

## Contents

1	Introduction	1
2	Assignment	1

## 1 Introduction

This assignment contains data scraped from Airline Quality [<https://www.airlinequality.com/>] about the quality ratings of three major (Dutch) airline operators:

- KLM
- Transavia
- Air France

## 2 Assignment

First we need to scrape the data from the website for three operators using “rvest” add-on package.

```
library(rvest)
klm <- read_html("https://www.airlinequality.com/airline-reviews/klm-royal-dutch-airlines/")
transavia <- read_html("https://www.airlinequality.com/airline-reviews/transavia/")
air_france <- read_html("https://www.airlinequality.com/airline-reviews/air-france/")
```

### 2.1 Number of Reviews and Overall Rating

```
review_klm <- klm %>%
  # set the key identifier for overall rating
  html_nodes(".customer-rating-total, .review-count") %>%
  html_text()
review_trans <- transavia %>%
  # set the key identifier for overall rating
  html_nodes(".customer-rating-total, .review-count") %>%
```

Table 1: Number of Reviews and Overall Ratings

Airline	Review_Count	Rating
KLM	1702	5/10
Transavia	286	4/10
Air France	1455	5/10

```

html_text()
review_airfr <- air_france %>%
  # set the key identifier for overall rating
  html_nodes(".customer-rating-total, .review-count") %>%
  html_text()
review_klm <- gsub("[\n\t]", "", review_klm)
review_trans <- gsub("[\n\t]", "", review_trans)
review_airfr <- gsub("[\n\t]", "", review_airfr)

extract_number <- function(string){
  m <- gregexpr("[0-9]+", string)
  as.numeric(unlist(regmatches(string, m)))
}
klm_review_count <- extract_number(review_klm[1])
trans_review_count <- extract_number(review_trans[1])
airfr_review_count <- extract_number(review_airfr[1])

tab1 <- data.frame(
  Airline = c("KLM", "Transavia", "Air France"),
  Review_Count = c(klm_review_count, trans_review_count, airfr_review_count),
  Rating = c(review_klm[2], review_trans[2], review_airfr[2])
)
library(knitr)
kable(tab1, format = "latex",
      caption = "Number of Reviews and Overall Ratings",
      booktabs = TRUE,
      align = c("l", "c"))

```

Table 1 presents the number of customer reviews and the corresponding overall ratings for each airline. KLM and Air France both received an average overall rating of 5 out of 10, while Transavia scored slightly lower at 4 out of 10. However, the number of reviews differs significantly across airlines — KLM and Air France have substantially more reviews (1,702 and 1,455 respectively) compared to Transavia (286).

## 2.2 Average Rating on Airline Service

Next, we will scrape the average ratings on the

1. Seat comfort
2. Inflight entertainment
3. Cabin staff
4. Food and beverage
5. Value for money

The corresponding information can be accessed with the CSS selectors `.stars`, `.fill`. You find the data on the (five) requested average ratings at the beginning of the output with the last entry before the first appearance of 12345.

Create a function `scrape.av()` that extracts the requested information and apply it to the three airline websites.

Data frame object named `average_ratings` contains the scraped information.

```
scrape.av <- function(url){
  page <- read_html(url)

  stars_raw <- page %>%
    html_nodes(".stars, .fill") %>%
    html_text()

  stars_clean <- gsub("[\n\t]", "", stars_raw)
  stars_clean <- gsub("\\s", "", stars_clean)

  # Find all "12345" positions
  idx <- which(stars_clean == "12345")

  # Get indices of the last number before each "12345"
  # (skip the first 12345 because there's nothing before it)
  last_before <- idx - 1
  last_before <- last_before[last_before > 0] # remove negatives just in case

  # Extract those values
  ratings <- as.numeric(stars_clean[last_before])

  data.frame(
    FoodandBeverage = ratings[1],
    InflightEntertainment = ratings[2],
    SeatComfort = ratings[3],
    CabinStaff = ratings[4],
    ValueforMoney = ratings[5]
  )
}

klm_service <- scrape.av("https://www.airlinequality.com/airline-reviews/klm-royal-dutch-airlin
```

```

trans_service <- scrape.av("https://www.airlinequality.com/airline-reviews/transavia/")
airfr_service <- scrape.av("https://www.airlinequality.com/airline-reviews/air-france/")
library(dplyr)
average_ratings <- bind_rows(KLM = klm_service,
  Transavia = trans_service,
  `Air France` = airfr_service,
  .id = "Airline"
)
kable(average_ratings,
  caption = "Average Service Ratings for Each Airline",
  align = "c")

```

Table 2: Average Service Ratings for Each Airline

Airline	FoodandBeverage	InflightEntertainment	SeatComfort	CabinStaff	ValueforMoney
KLM	3	3	3	4	3
Transavia	2	1	2	3	2
Air France	3	3	3	3	3

### 2.2.1 Average Rating Grouped Bar Plot

```

library(tidyr)
avg_long <- average_ratings %>%
  pivot_longer(
    cols = -Airline,
    names_to = "Category",
    values_to = "Rating"
  )

library("RColorBrewer")
brewer.pal(5, "Spectral")

library(ggplot2)
ggplot(avg_long, aes(x = Airline, y = Rating, fill = Category)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  geom_text(aes(label = Rating),
            vjust = -0.3,
            position = position_dodge(0.9),
            size = 3) +
  labs(
    title = "Average Ratings by Airline and Category",
    x = "Airline",
    y = "Average Rating (1-5)",
    fill = "Category"

```

```

) +
ylim(0, 5) +
scale_fill_brewer(palette = "Spectral") +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
  axis.text.x = element_text(angle = 0, hjust = 0.5, size = 11)
)

```

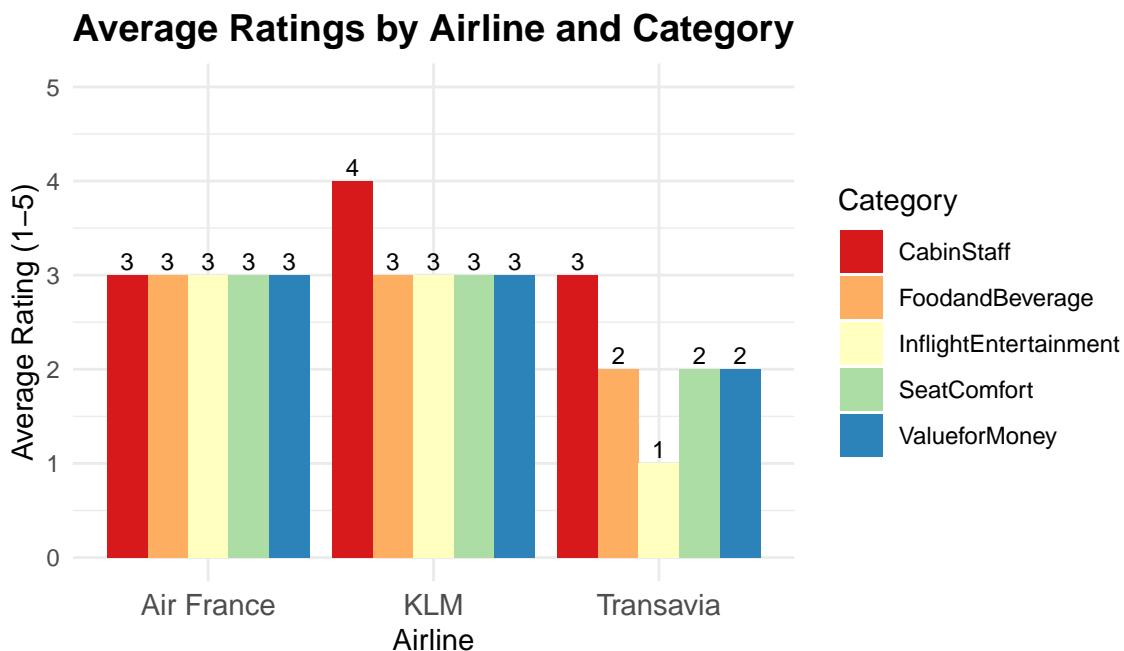


Figure 1: Average Ratings by Airline and Category

As shown in Figure 1, the grouped barplot illustrates the average service ratings across all airlines and categories. Average service ratings across five evaluation categories *Cabin Staff*, *Food and Beverage*, *Inflight Entertainment*, *Seat Comfort* and *Value for Money* - for Air France, KLM and Transavia are recorded. Overall, the barplot shows consistent ratings across most dimensions, with scores ranging from 2 to 4 (out of 5). **Air France** and **KLM** achieve the highest overall service levels, particularly in *Cabin Staff*, while **Transavia** trails slightly behind, especially in *Inflight Entertainment* and *Seat Comfort*.