# N Kaso SkyTrax Seats **Draft**

N Kaso

21/09/2020

## Description

**A scraped dataset from all user reviews on Skytrax (www.airlinequality.com). This set contains traveller ratings of the airline seat over a number of categories.**

**TBC...**

## Questions

**Which are the most frequently rated airlines?**

**How does the average rating distribution compares across the top ten most often rated?**

**How does overall rating vary between Boeing and Airbus?**

**Which aircraft models show the best rating performance?**

### Read the data

```
seats <- read_csv("https://raw.githubusercontent.com/quankiquanki/skytrax-reviews-dataset/master/data/s
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   date = col_date(format = ""),
##   overall_rating = col_double(),
##   seat_legroom_rating = col_double(),
##   seat_recline_rating = col_double(),
##   seat_width_rating = col_double(),
##   aisle_space_rating = col_double(),
##   viewing_tv_rating = col_double(),
##   power_supply_rating = col_double(),
##   seat_storage_rating = col_double(),
##   recommended = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
seats
```

```
## # A tibble: 1,258 x 21
##    airline_name link  title author author_country date       content aircraft
##    <chr>        <chr> <chr> <chr>  <chr>          <date>     <chr>   <chr>
##  1 aegean-airl~ /sea~ Aege~ Jay S~ United Kingdom 2015-07-20 LHR to~ A320-200
##  2 aegean-airl~ /sea~ Aege~ Paul ~ United Kingdom 2013-01-21 For a ~ AIRBUS ~
##  3 aer-lingus   /sea~ Aer ~ L Pul~ United States  2015-07-07 The se~ A330
##  4 aer-lingus   /sea~ Aer ~ D Bro~ United States  2010-10-22 Appear~ Airbus ~
##  5 aeroflot-ru~ /sea~ Aero~ Konst~ Greece         2015-08-02 Boeing~ Boeing ~
##  6 aeroflot-ru~ /sea~ Aero~ Dan K~ United States  2015-06-17 Almost~ Boeing ~
##  7 aeroflot-ru~ /sea~ Aero~ Josef~ Israel         2014-07-25 There ~ AIRBUS ~
##  8 aeroflot-ru~ /sea~ Aero~ Laure~ Thailand       2011-04-05 Its no~ Boeing ~
##  9 aeroflot-ru~ /sea~ Aero~ Boris~ Slovakia       2009-10-06 I did ~ Airbus ~
## 10 aeroflot-ru~ /sea~ Aero~ Boris~ Slovakia       2009-10-06 Leg ro~ IL96
## # ... with 1,248 more rows, and 13 more variables: seat_layout <chr>,
## #   date_flown <chr>, cabin_flown <chr>, type_traveller <chr>,
## #   overall_rating <dbl>, seat_legroom_rating <dbl>, seat_recline_rating <dbl>,
## #   seat_width_rating <dbl>, aisle_space_rating <dbl>, viewing_tv_rating <dbl>,
## #   power_supply_rating <dbl>, seat_storage_rating <dbl>, recommended <dbl>
```

### Tidy and wrangle the data

```
airbus_mods <- c("A300", "A310-300","A319","A320","A321", "A320-200","A330","A330-200","A330-300","A340

boeing_mods <- c("B737-900", "B767", "B767-300", "B777", "B777-200", "B777-200LR", "B777-200ER", "B777-

seats_tidy <- seats %>%
              select(-link, -title, -author, -content, -power_supply_rating, -seat_storage_rating) %>%
              separate(col = "aircraft", c("aircraft_make", "aircraft_model"), sep = " ") %>%
              mutate(aircraft_model = ifelse(is.na(aircraft_model), aircraft_make, aircraft_model)) %>%
              mutate(aircraft_make = ifelse(aircraft_make %in% airbus_mods, "AIRBUS", aircraft_make)) %
              mutate(aircraft_make = ifelse(aircraft_make %in% boeing_mods, "BOEING", aircraft_make)) %
              mutate(aircraft_make = ifelse(aircraft_make == "Embraer", "EMBRAER", aircraft_make)) %>%
              filter(aircraft_make %in% c("AIRBUS", "BOEING", "EMBRAER"))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 25 rows [22, 46, 101,
## 102, 119, 178, 199, 238, 342, 356, 357, 386, 395, 466, 469, 474, 530, 565, 573,
## 688, ...].
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 393 rows [1, 3,
## 10, 11, 17, 18, 23, 34, 37, 38, 40, 41, 42, 43, 47, 48, 49, 87, 96, 99, ...].
```

```
seats_tidy$airline_name <- str_replace_all(seats_tidy$airline_name,"-"," ")
seats_tidy$airline_name <- toupper(seats_tidy$airline_name)

seats_tidy$aircraft_model <- str_replace_all(seats_tidy$aircraft_model,"B","")
seats_tidy$aircraft_model <- str_replace_all(seats_tidy$aircraft_model,"ER","")
seats_tidy$aircraft_model <- str_replace_all(seats_tidy$aircraft_model,"LR","")
### Code to remove part of aircraft_model string after the '-' below, however it is retains the last pa
```

```
### seats_tidy$aircraft_model <- str_remove(seats_tidy$aircraft_model, ".+?(?=-)")

seats_tidy
```

```
## # A tibble: 1,232 x 16
##    airline_name author_country date       aircraft_make aircraft_model
##    <chr>        <chr>          <date>     <chr>         <chr>
##  1 AEGEAN AIRL~ United Kingdom 2015-07-20 AIRBUS        A320-200
##  2 AEGEAN AIRL~ United Kingdom 2013-01-21 AIRBUS        A320
##  3 AER LINGUS   United States  2015-07-07 AIRBUS        A330
##  4 AER LINGUS   United States  2010-10-22 AIRBUS        A330
##  5 AEROFLOT RU~ Greece         2015-08-02 BOEING        737-800
##  6 AEROFLOT RU~ United States  2015-06-17 BOEING        777-300
##  7 AEROFLOT RU~ Israel         2014-07-25 AIRBUS        A330-300
##  8 AEROFLOT RU~ Thailand       2011-04-05 BOEING        767
##  9 AEROFLOT RU~ Slovakia       2009-10-06 AIRBUS        A330
## 10 AEROFLOT RU~ Germany        2009-01-13 AIRBUS        A319
## # ... with 1,222 more rows, and 11 more variables: seat_layout <chr>,
## #   date_flown <chr>, cabin_flown <chr>, type_traveller <chr>,
## #   overall_rating <dbl>, seat_legroom_rating <dbl>, seat_recline_rating <dbl>,
## #   seat_width_rating <dbl>, aisle_space_rating <dbl>, viewing_tv_rating <dbl>,
## #   recommended <dbl>
```

## Calculate

```
### seats_tidy <- seats_tidy %>% mutate(ave_rating = ave(11:15)) ###unsure how to calculate an average
seats_tidy
```

```
## # A tibble: 1,232 x 16
##    airline_name author_country date       aircraft_make aircraft_model
##    <chr>        <chr>          <date>     <chr>         <chr>
##  1 AEGEAN AIRL~ United Kingdom 2015-07-20 AIRBUS        A320-200
##  2 AEGEAN AIRL~ United Kingdom 2013-01-21 AIRBUS        A320
##  3 AER LINGUS   United States  2015-07-07 AIRBUS        A330
##  4 AER LINGUS   United States  2010-10-22 AIRBUS        A330
##  5 AEROFLOT RU~ Greece         2015-08-02 BOEING        737-800
##  6 AEROFLOT RU~ United States  2015-06-17 BOEING        777-300
##  7 AEROFLOT RU~ Israel         2014-07-25 AIRBUS        A330-300
##  8 AEROFLOT RU~ Thailand       2011-04-05 BOEING        767
##  9 AEROFLOT RU~ Slovakia       2009-10-06 AIRBUS        A330
## 10 AEROFLOT RU~ Germany        2009-01-13 AIRBUS        A319
## # ... with 1,222 more rows, and 11 more variables: seat_layout <chr>,
## #   date_flown <chr>, cabin_flown <chr>, type_traveller <chr>,
## #   overall_rating <dbl>, seat_legroom_rating <dbl>, seat_recline_rating <dbl>,
## #   seat_width_rating <dbl>, aisle_space_rating <dbl>, viewing_tv_rating <dbl>,
## #   recommended <dbl>
```

```
by_airline <- seats_tidy %>% group_by(airline_name)
most_reviewed <- by_airline %>% summarise(no_of_reviews = n()) %>%
                                filter(no_of_reviews >= 45) %>%
                                arrange(desc(no_of_reviews))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
most_reviewed
```

```
## # A tibble: 10 x 2
##    airline_name            no_of_reviews
##    <chr>                           <int>
##  1 BRITISH AIRWAYS                    86
##  2 EMIRATES                           83
##  3 CATHAY PACIFIC AIRWAYS             75
##  4 VIRGIN ATLANTIC AIRWAYS            75
##  5 AIR FRANCE                         64
##  6 LUFTHANSA                          64
##  7 QANTAS AIRWAYS                     64
##  8 SINGAPORE AIRLINES                 54
##  9 ETIHAD AIRWAYS                     46
## 10 QATAR AIRWAYS                      45
```