Extracting Data and Computing Passenger Percentages

GENEVIEVE GONZALES

2025-03-10

```
library(dplyr)
library(readr)

# Loading our cleaned updated dataset (main)
df <- read_csv("updated_Air_Traffic_Passenger_cleaned.csv")
```

```
# crash info for each airline
crashes <- list(
  list(name = "Asiana Airlines", crash_year = 2013, pre_crash_month = 6,
       crash_month = 7, post_1_month = 8, post_2_months = c(8,9),
       post_5_months = c(8,9,10,11,12)),
  list(name = "Delta Air Lines", crash_year = 2006, pre_crash_month = 7,
       crash_month = 8, post_1_month = 9, post_2_months = c(9,10),
       post_4_months = c(9,10,11,12)),  # Only 4 months
  list(name = "Alaska Airlines", crash_year = 2024, pre_crash_month = 12,
       crash_month = 1, post_1_month = 2, post_2_months = c(2,3),
       post_5_months = c(2,3,4,5,6)),
  list(name = "Air France", crash_year = 2009, pre_crash_month = 5,
       crash_month = 6, post_1_month = 7, post_2_months = c(7,8),
       post_5_months = c(7,8,9,10,11)),
  list(name = "US Airways", crash_year = 2009, pre_crash_month = 12,
       crash_month = 1, post_1_month = 2, post_2_months = c(2,3),
       post_5_months = c(2,3,4,5,6))
)

# initializing an empty list to store results
all_results <- list()


for (crash in crashes) {
  airline_name <- crash$name
  crash_year <- crash$crash_year
  pre_crash_month <- crash$pre_crash_month
  crash_month <- crash$crash_month
  post_1_month <- crash$post_1_month
  post_2_months <- crash$post_2_months
  post_5_months <- crash$post_5_months %||% NULL  # handle the 5-month cases
  post_4_months <- crash$post_4_months %||% NULL  # handle the 4-month cases

  # filter airline data to the airline name
  airline_df <- df %>%
    filter(grepl(airline_name, `Operating Airline`, ignore.case = TRUE))
```

```r
results <- list()

# loop through each year (2005-2024)
for (year in 2005:2024) {

  total_passengers_year <- airline_df %>%
    filter(Year == year) %>%
    summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
    pull(Total_Passengers)

  # one of the airlines doesn't go to 2024, this breaks out of the loop
  if (total_passengers_year == 0) {
    break
  }
  # skip if pre-crash month is December and there's no prior year data
  if (pre_crash_month == 12 & year == 2005) {
    next
  }

  pre_crash_passenger_count <- airline_df %>%
    filter(Year == ifelse(pre_crash_month == 12, year - 1, year) & Month == pre_crash_month) %>%
    summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
    pull(Total_Passengers)

  crash_month_passenger_count <- airline_df %>%
    filter(Year == year & Month == crash_month) %>%
    summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
    pull(Total_Passengers)

  post_1_month_passenger_count <- airline_df %>%
    filter(Year == year & Month == post_1_month) %>%
    summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
    pull(Total_Passengers)

  post_2_months_passenger_count <- airline_df %>%
    filter(Year == year & Month %in% post_2_months) %>%
    summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
    pull(Total_Passengers)

  post_4_months_passenger_count <- if (!is.null(post_4_months)) {
    airline_df %>%
      filter(Year == year & Month %in% post_4_months) %>%
      summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
      pull(Total_Passengers)
  } else {
    NA
  }

  post_5_months_passenger_count <- if (!is.null(post_5_months)) {
    airline_df %>%
      filter(Year == year & Month %in% post_5_months) %>%
      summarise(Total_Passengers = sum(`Passenger Count`, na.rm = TRUE)) %>%
      pull(Total_Passengers)
```

```
    } else {
      NA
    }

    post_2_months_percentage <- (post_2_months_passenger_count / total_passengers_year) * 100
    post_4_months_percentage <- if (!is.null(post_4_months_passenger_count)) {
      (post_4_months_passenger_count / total_passengers_year) * 100
    } else {
      NA
    }
    post_5_months_percentage <- if (!is.null(post_5_months_passenger_count)) {
      (post_5_months_passenger_count / total_passengers_year) * 100
    } else {
      NA
    }

    # store the results
    results[[as.character(year)]] <- data.frame(
      Year = year,
      Airline = airline_name,
      Pre_Crash_Month_Passenger_Count = pre_crash_passenger_count,
      Crash_Month_Passenger_Count = crash_month_passenger_count,
      Post_1_Month_Passenger_Count = post_1_month_passenger_count,
      Post_2_Months_Passenger_Count = post_2_months_passenger_count,
      Post_4_Months_Passenger_Count = post_4_months_passenger_count,
      Post_5_Months_Passenger_Count = post_5_months_passenger_count,
      Post_2_Months_Percentage = post_2_months_percentage,
      Post_4_Months_Percentage = post_4_months_percentage,
      Post_5_Months_Percentage = post_5_months_percentage
    )
  }

  # combine results for this airline
  airline_results <- bind_rows(results)
  all_results[[airline_name]] <- airline_results
}

# combine all the airlines' results into one df
final_results <- bind_rows(all_results)

# save into csv file
write_csv(final_results, "All_Airline_Crashes_Statistics.csv")

head(final_results)
```

```
##   Year         Airline Pre_Crash_Month_Passenger_Count
## 1 2005 Asiana Airlines                           10175
## 2 2006 Asiana Airlines                            9188
## 3 2007 Asiana Airlines                            9667
## 4 2008 Asiana Airlines                            9553
## 5 2009 Asiana Airlines                            8994
## 6 2010 Asiana Airlines                            9989
##   Crash_Month_Passenger_Count Post_1_Month_Passenger_Count
```

```
## 1                        9785                         7475
## 2                        8838                         9285
## 3                        9445                        10109
## 4                        9127                        10206
## 5                        9848                         9784
## 6                        9672                         9439
##   Post_2_Months_Passenger_Count Post_4_Months_Passenger_Count
## 1                         17407                            NA
## 2                         17666                            NA
## 3                         17779                            NA
## 4                         18494                            NA
## 5                         19050                            NA
## 6                         19001                            NA
##   Post_5_Months_Passenger_Count Post_2_Months_Percentage
## 1                         46661                 15.39380
## 2                         44037                 16.60807
## 3                         42555                 16.69280
## 4                         45702                 16.96822
## 5                         45866                 17.66490
## 6                         44692                 17.19439
##   Post_4_Months_Percentage Post_5_Months_Percentage
## 1                       NA                41.26444
## 2                       NA                41.39983
## 3                       NA                39.95512
## 4                       NA                41.93152
## 5                       NA                42.53113
## 6                       NA                40.44269
```