DBeaver Assignment

AUTHOR
Jin Sook Song

DBeaver

```
library(DBI)
library(duckdb)
```

```
# Establish a DB connection
con <- dbConnect(duckdb(), "flights.duckdb", read_only = TRUE)

# print(con)
# dbIsValid(con)
```

```
dbListTables(con)
```

```
[1] "airlines" "airports" "flights"  "planes"   "weather"
```

Question 1.

```
# Query to get the top few rows from the weather table
query <- "SELECT * FROM weather LIMIT 5;"

# Execute query and fetch results
weather_data <- dbGetQuery(con, query)

# Display results
# print(weather_data)
```

```
# Query to compute mean temperature for each airport using the weather table
query <- "
SELECT origin, AVG(temp) AS mean_temperature
FROM weather
GROUP BY origin;
"
# Execute query and fetch results
mean_temperature <- dbGetQuery(con, query)

# Display results
print(mean_temperature)
```

```
  origin mean_temperature
1    JFK         54.47215
2    LGA         55.76261
3    EWR         55.54655
```

Question 2.

```
# Query to get the top few rows from the weather table
query <- "SELECT * FROM flights LIMIT 5;"

# Execute query and fetch results
weather_data <- dbGetQuery(con, query)
```

```
        # Display results
        # print(weather_data)
```

```
        # Query to calculate the average number of flights per day from each airport in January
        query2 <- "
        SELECT origin,
               COUNT(*) / COUNT(DISTINCT day) AS avg_flights_per_day
        FROM flights
        WHERE month = 1
        GROUP BY origin;
        "

        # Execute query and fetch results
        avg_flights_per_day <- dbGetQuery(con, query2)

        # Display results
        print(avg_flights_per_day)
```

```
  origin avg_flights_per_day
1    LGA            256.4516
2    JFK            295.5161
3    EWR            319.1290
```

Question 3.

```
        # 3. Top destinations for each airport
        query3 <- "
        WITH RankedDestinations AS (
            SELECT origin, dest, COUNT(*) AS flight_count,
                   RANK() OVER (PARTITION BY origin ORDER BY COUNT(*) DESC) AS rnk
            FROM flights
            GROUP BY origin, dest
        )
        SELECT origin, dest, flight_count
        FROM RankedDestinations
        WHERE rnk = 1;
        "

        # Execute queries and fetch results
        top_destinations <- dbGetQuery(con, query3)

        # Display results
        print(top_destinations)
```

```
  origin dest flight_count
1    JFK  LAX        11262
2    EWR  ORD         6100
3    LGA  ATL        10263
```

```
        # Disconnect from database
        dbDisconnect(con, shutdown = TRUE)
```