

SQL – DBeaver Assignment

1. What is the mean temperature for flights from each airport?

The screenshot shows the DBeaver 24.2.2 interface with the SQL Editor displaying the command `DESCRIBE;`. The Results panel shows the following table structure:

Grid	Az database	Az schema	Az name	column_names	column_types	temporary
1	flights	main	airlines	> carrier [+1]	> VARCHAR [+1]	[]
2	flights	main	airports	> faa [+7]	> VARCHAR [+7]	[]
3	flights	main	flights	> year [+18]	> INTEGER [+18]	[]
4	flights	main	planes	> tailnum [+8]	> VARCHAR [+8]	[]
5	flights	main	weather	> origin [+14]	> VARCHAR [+14]	[]

5 row(s) fetched - 0.071s, on 2024-10-19 at 22:19:38

The screenshot shows the DBeaver 24.2.2 interface with the SQL Editor displaying the following query:

```
CREATE TEMP TABLE combined_table AS
SELECT
  f.*, w.temp
FROM
  flights f
JOIN
  weather w
ON f.origin = w.origin
AND f.year = w.year
AND f.month = w.month
AND f.day = w.day
AND f.hour = w.hour;
```

The Statistics panel shows the following information:

Name	Value
f.*, w.temp	
FROM	
flights f	
JOIN	
weather w	
ON f.origin = w.origin	
AND f.year = w.year	
AND f.month = w.month	
AND f.day = w.day	
AND f.hour = w.hour	
Start time	Sat Oct 19 22:27:52 EDT 2024
Finish time	Sat Oct 19 22:27:52 EDT 2024

EST en Writable Smart Insert 1:33:32 Set: 0 | 0

DBeaver 24.2.2 - <flights.duckdb> Script-1

File Edit Navigate Search SQL Editor Database Window Help

SQL Commit Rollback Auto flights.duckdb flights

Results 1 x

Enter a part

SQL: `SELECT f.*, w.temp, w.dewp, w.humid, w.wind_speed`

	air_time	distance	hour	minute	time_hour	temp	dewp	humid	wind_speed
1	100	762	6	0	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
2	[NULL]	719	6	0	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
3	63	416	6	0	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
4	45	212	6	0	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
5	78	502	6	0	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
6	316	2,454	6	0	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
7	164	733	6	0	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
8	148	1,096	6	5	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
9	128	963	6	5	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
10	309	2,475	6	0	2013-06-26 10:00:00.000	75.92	64.94	68.78	12.6585
11	75	541	6	15	2013-06-26 10:00:00.000	75.92	64.94	68.78	12.6585
12	128	944	6	5	2013-06-26 10:00:00.000	75.92	64.94	68.78	12.6585
13	[NULL]	711	6	5	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
14	79	502	6	10	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
15	104	762	6	5	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
16	350	2,586	6	10	2013-06-26 10:00:00.000	75.92	64.94	68.78	12.6585
17	147	1,068	6	7	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
18	182	1,411	6	0	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554
19	142	1,020	6	10	2013-06-26 10:00:00.000	77	62.96	61.91	9.2062
20	145	1,085	6	15	2013-06-26 10:00:00.000	73.94	64.94	73.49	8.0554

Refresh Save Cancel Export data 200 200+

200 row(s) fetched - 0.113s (0.013s fetch), on 2024-10-19 at 22:21:11

DBeaver 24.2.2 - <flights.duckdb> Script-1

File Edit Navigate Search SQL Editor Database Window Help

SQL Commit Rollback Auto flights.duckdb flights

Results 1 x

Enter a part

SQL: `SELECT origin AS airport, AVG(temp) AS mean_temp FROM combined_table GROUP BY origin;`

	airport	mean_temp
1	LGA	57.3684382611
2	EWB	57.4185864066
3	JFK	56.1880268755

Refresh Save Cancel Export data 200 3

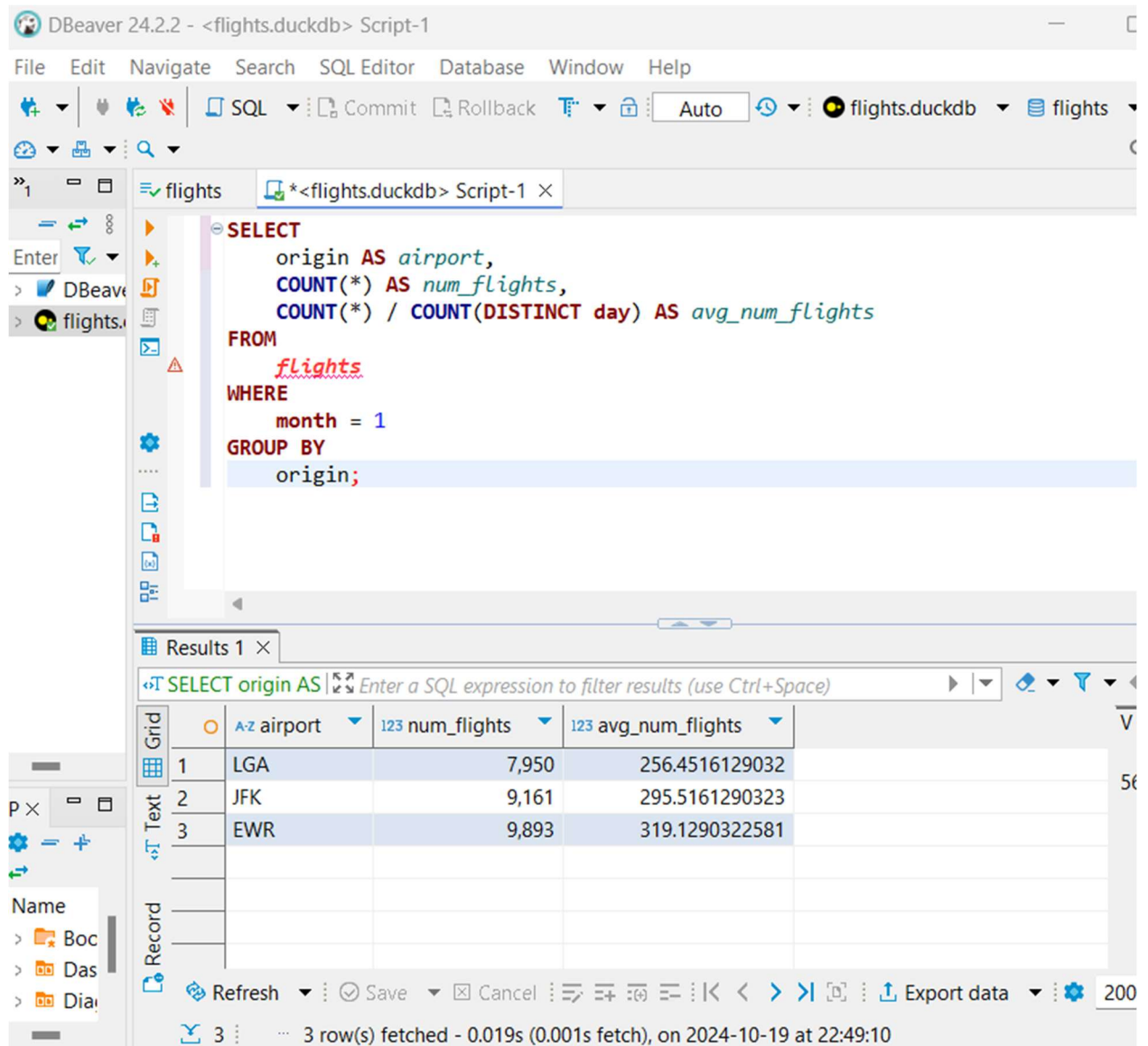
3 row(s) fetched - 0.015s, on 2024-10-19 at 22:29:54

To analyze the mean temperatures for flights, we used the 'DESCRIBE' command to view the different schemas in the database, which helped me identify the available columns in each table. Then we joined the 'weather' table with the 'flights' table based on matching attributes like origin, year, month, day, and hour, as a temporary table. This allowed us to query the mean temperatures for flights departing from the three major airports in the New York City area: LaGuardia Airport (LGA), Newark Liberty International Airport (EWB), and John F. Kennedy International Airport (JFK).

The mean temperatures are:

- LaGuardia Airport (LGA): 57.37°F
- Newark Liberty International Airport (EWR): 57.42°F
- John F. Kennedy International Airport (JFK): 56.19°F

2. What is the average number of flights from each airport per day in January?



The screenshot shows the DBeaver 24.2.2 interface with a SQL query executed against the 'flights.duckdb' database. The query calculates the total number of flights and the average number of flights per day for each airport in January. The results are displayed in a table with three columns: 'airport', 'num_flights', and 'avg_num_flights'.

```
SELECT
    origin AS airport,
    COUNT(*) AS num_flights,
    COUNT(*) / COUNT(DISTINCT day) AS avg_num_flights
FROM
    flights
WHERE
    month = 1
GROUP BY
    origin;
```

Grid	A-Z airport	123 num_flights	123 avg_num_flights
1	LGA	7,950	256.4516129032
2	JFK	9,161	295.5161290323
3	EWR	9,893	319.1290322581

3 row(s) fetched - 0.019s (0.001s fetch), on 2024-10-19 at 22:49:10

The average number of flights from each airport in January is as follows:

- LaGuardia Airport (LGA): 7,950 total flights, averaging 256.45 flights per day.
- John F. Kennedy International Airport (JFK): 9,161 total flights, averaging 295.52 flights per day.
- Newark Liberty International Airport (EWR): 9,893 total flights, averaging 319.13 flights per day.

This indicates that EWR had the highest average number of flights per day, while LGA had the lowest.

3. What are the top destinations for each airport?

The screenshot shows the DBeaver SQL Editor interface. The SQL Editor pane contains a query to find the top destinations for each airport. The query is as follows:

```
SELECT
  origin,
  dest,
  COUNT(*) AS num_flights
FROM
  flights AS f
GROUP BY
  origin, dest
HAVING
  COUNT(*) = (
    SELECT
      MAX(flight_count)
    FROM (
      SELECT
        COUNT(*) AS flight_count
      FROM
        flights
      WHERE
        origin = f.origin
      GROUP BY
        dest
    ) AS counts
  )
ORDER BY
  origin
```

The Results pane shows the results of the query, displaying a table with 3 rows and 4 columns: A-Z origin, A-Z dest, num_flights, and a total count of 10263. The table is as follows:

	A-Z origin	A-Z dest	num_flights
1	EWR	ORD	6,100
2	JFK	LAX	11,262
3	LGA	ATL	10,263

The bottom status bar indicates that 3 row(s) were fetched in 0.027s on 2024-10-19 at 23:24:26. The interface also shows a sidebar with a tree view containing 'Bookmarks', 'Dashboards', 'Diagrams', and 'Scripts'.

- EWR (Newark) has its top destination as ORD (Chicago O'Hare) with 6,100 flights.
- JFK (John F. Kennedy) has LAX (Los Angeles) as its busiest route with 11,262 flights.
- LGA (LaGuardia) has ATL (Atlanta) as its top destination, with 10,263 flights.

GitHub Link: https://github.com/jk7830a/HW_kimbrough.git

*Please be advised, all group work will be in folder called *Group*.*

