Practice SQL

Gloria Baidoo

2024-10-12

```
## Loading required package: DBI
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
con <- dbConnect(duckdb(dbdir = "../data/flights.duckdb", read_only = TRUE))
```

```
SHOW TABLES;
```

Table 1: 5 records

name
airlines
airports
flights
planes
weather

```
DESCRIBE;
```

Table 2: 5 records

database	schema	name	column_names	column_types	temporary
flights	main	airlines	carrier, name	VARCHAR, VARCHAR	FALSE
flights	main	airports	faa, name , lat , lon , alt , tz , dst , tzone	VARCHAR, VARCHAR, DOUBLE , DOUBLE , DOUBLE , DOUBLE , VARCHAR, VARCHAR	FALSE

database	schema	name	column_names	column_types	temporary
flights	main	flights	year , month , day , dep_time , sched_dep_time, dep_delay , arr_time , sched_arr_time, arr_delay , carrier , flight , tailnum , origin , dest , air_time , distance , hour , minute , time_hour	INTEGER , INTEGER , INTEGER , INTEGER , INTEGER , DOUBLE , INTEGER , INTEGER , DOUBLE , VARCHAR , INTEGER , VARCHAR , VARCHAR , VARCHAR , DOUBLE , DOUBLE , DOUBLE , DOUBLE , TIMESTAMP	FALSE
flights	main	planes	tailnum , year , type , manufacturer, model , engines , seats , speed , engine	VARCHAR, INTEGER, VARCHAR, VARCHAR, VARCHAR, INTEGER, INTEGER, INTEGER, VARCHAR	FALSE
flights	main	weather	origin , year , month , day , hour , temp , dewp , humid , wind_dir , wind_speed, wind_gust , precip , pressure , visib , time_hour	VARCHAR , INTEGER , INTEGER , INTEGER , INTEGER , DOUBLE , DOUBLE , DOUBLE , DOUBLE , DOUBLE , DOUBLE , DOUBLE , DOUBLE , DOUBLE , TIMESTAMP	FALSE

Practice 1: Practice 1: Select Flights from a Specific Carrier

```sql
SELECT *
FROM flights
WHERE carrier = 'UA';
```

Table 3: Displaying records 1 - 10

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	1	1	517	515	2	830	819	11	UA	1545	N14228	EWR	IAH	227	1400	5	15	2013-01-01 10:00:00
2013	1	1	533	529	4	850	830	20	UA	1714	N24211	LGA	IAH	227	1416	5	29	2013-01-01 10:00:00
2013	1	1	554	558	-4	740	728	12	UA	1696	N39463	EWR	ORD	150	719	5	58	2013-01-01 10:00:00
2013	1	1	558	600	-2	924	917	7	UA	194	N29129	JFK	LAX	345	2475	6	0	2013-01-01 11:00:00
2013	1	1	558	600	-2	923	937	-14	UA	1124	N53441	EWR	SFO	361	2565	6	0	2013-01-01 11:00:00
2013	1	1	559	600	-1	854	902	-8	UA	1187	N76515	EWR	LAS	337	2227	6	0	2013-01-01 11:00:00

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	1	1	607	607	0	858	915	-17	UA	1077	N53442	EWR	MIA	157	1085	6	7	2013-01-01 11:00:00
2013	1	1	611	600	11	945	931	14	UA	303	N532UA	JFK	SFO	366	2586	6	0	2013-01-01 11:00:00
2013	1	1	623	627	-4	933	932	1	UA	496	N459UA	LGA	IAH	229	1416	6	27	2013-01-01 11:00:00
2013	1	1	628	630	-2	1016	947	29	UA	1665	N33289	EWR	LAX	366	2454	6	30	2013-01-01 11:00:00

Practice 2: Count the Number of Flights for Each Carrier

```
SELECT carrier, COUNT(*) AS flight_count
FROM flights
GROUP BY carrier;
```

Table 4: Displaying records 1 - 10

carrier	flight_count
OO	32
DL	48110
B6	54635
FL	3260
HA	342
EV	54173
AA	32729
US	20536
9E	18460
YV	601

Practice 3: Find the Flights with the Longest Distance

```
SELECT *
FROM flights
ORDER BY distance DESC
LIMIT 10;
```

Table 5: Displaying records 1 - 10

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	6	26	954	1000	-6	1421	1435	-14	HA	51	N384HA	JFK	HNL	601	4983	10	0	2013-06-26 14:00:00
2013	6	27	957	1000	-3	1411	1435	-24	HA	51	N386HA	JFK	HNL	589	4983	10	0	2013-06-27 14:00:00
2013	6	28	955	1000	-5	1426	1435	-9	HA	51	N392HA	JFK	HNL	614	4983	10	0	2013-06-28 14:00:00
2013	6	29	953	1000	-7	1409	1435	-26	HA	51	N390HA	JFK	HNL	595	4983	10	0	2013-06-29 14:00:00
2013	6	30	955	1000	-5	1415	1435	-20	HA	51	N381HA	JFK	HNL	601	4983	10	0	2013-06-30 14:00:00
2013	7	1	1005	1000	5	1527	1430	57	HA	51	N384HA	JFK	HNL	588	4983	10	0	2013-07-01 14:00:00
2013	7	2	952	1000	-8	1402	1430	-28	HA	51	N383HA	JFK	HNL	590	4983	10	0	2013-07-02 14:00:00
2013	7	3	957	1000	-3	1410	1430	-20	HA	51	N388HA	JFK	HNL	585	4983	10	0	2013-07-03 14:00:00
2013	7	4	950	1000	-10	1359	1430	-31	HA	51	N380HA	JFK	HNL	590	4983	10	0	2013-07-04 14:00:00
2013	7	5	950	1000	-10	1423	1430	-7	HA	51	N383HA	JFK	HNL	600	4983	10	0	2013-07-05 14:00:00

Practice 4: Calculate Average Arrival Delay for Each Carrier

```sql
SELECT carrier, AVG(arr_delay) AS avg_arrival_delay
FROM flights
WHERE arr_delay IS NOT NULL
GROUP BY carrier;
```

Table 6: Displaying records 1 - 10

carrier	avg_arrival_delay
EV	15.7964311
AA	0.3642909
US	2.1295951
9E	7.3796692
YV	15.5569853

carrier	avg_arrival_delay
FL	20.1159055
HA	-6.9152047
WN	9.6491199
AS	-9.9308886
F9	21.9207048

Practice 5: Find Flights Departing from JFK to LAX

```
SELECT *
FROM flights
WHERE origin = 'JFK' AND dest = 'LAX';
```

Table 7: Displaying records 1 - 10

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	1	1	558	600	-2	924	917	7	UA	194	N29129	JFK	LAX	345	2475	6	0	2013-01-01 11:00:00
2013	1	1	658	700	-2	1027	1025	2	VX	399	N627VA	JFK	LAX	361	2475	7	0	2013-01-01 12:00:00
2013	1	1	702	700	2	1058	1014	44	B6	671	N779JB	JFK	LAX	381	2475	7	0	2013-01-01 12:00:00
2013	1	1	743	730	13	1107	1100	7	AA	33	N338AA	JFK	LAX	358	2475	7	30	2013-01-01 12:00:00
2013	1	1	829	830	-1	1152	1200	-8	UA	443	N554UA	JFK	LAX	360	2475	8	30	2013-01-01 13:00:00
2013	1	1	856	900	-4	1226	1220	6	AA	1	N324AA	JFK	LAX	358	2475	9	0	2013-01-01 14:00:00
2013	1	1	859	900	-1	1223	1225	-2	VX	407	N846VA	JFK	LAX	359	2475	9	0	2013-01-01 14:00:00
2013	1	1	921	900	21	1237	1227	10	DL	120	N713TW	JFK	LAX	333	2475	9	0	2013-01-01 14:00:00
2013	1	1	941	945	-4	1300	1258	2	B6	679	N806JB	JFK	LAX	352	2475	9	45	2013-01-01 14:00:00
2013	1	1	1026	1030	-4	1351	1340	11	AA	19	N328AA	JFK	LAX	356	2475	10	30	2013-01-01 15:00:00

Practice 6: Calculate the Total Number of Flights Each Month

```
SELECT month, COUNT(*) AS total_flights
FROM flights
GROUP BY month
ORDER BY month;
```

Table 8: Displaying records 1 - 10

month	total_flights
1	27004
2	24951
3	28834
4	28330
5	28796
6	28243
7	29425
8	29327
9	27574
10	28889

Practice 7: Find Flights with Departure Delays Greater than 2 Hours

```
SELECT *
FROM flights
WHERE dep_delay > 120;
```

Table 9: Displaying records 1 - 10

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	1	1	848	1835	853	1001	1950	851	MQ	3944	N942MQ	JFK	BWI	41	184	18	35	2013-01-01 23:00:00
2013	1	1	957	733	144	1056	853	123	UA	856	N534UA	EWR	BOS	37	200	7	33	2013-01-01 12:00:00
2013	1	1	1114	900	134	1447	1222	145	UA	1086	N76502	LGA	IAH	248	1416	9	0	2013-01-01 14:00:00
2013	1	1	1540	1338	122	2020	1825	115	B6	705	N570JB	JFK	SJU	193	1598	13	38	2013-01-01 18:00:00
2013	1	1	1815	1325	290	2120	1542	338	EV	4417	N17185	EWR	OMA	213	1134	13	25	2013-01-01 18:00:00

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	1	1	1842	1422	260	1958	1535	263	EV	4633	N18120	EWR	BTV	46	266	14	22	2013-01-01 19:00:00
2013	1	1	1856	1645	131	2212	2005	127	AA	181	N323AA	JFK	LAX	336	2475	16	45	2013-01-01 21:00:00
2013	1	1	1934	1725	129	2126	1855	151	MQ	4255	N909MQ	JFK	BNA	154	765	17	25	2013-01-01 22:00:00
2013	1	1	1938	1703	155	2109	1823	166	EV	4300	N18557	EWR	RIC	68	277	17	3	2013-01-01 22:00:00
2013	1	1	1942	1705	157	2124	1830	174	MQ	4410	N835MQ	JFK	DCA	60	213	17	5	2013-01-01 22:00:00

Practice 8: Find the Number of Flights per Day

```
SELECT year, month, day, COUNT(*) AS total_flights
FROM flights
GROUP BY year, month, day
ORDER BY year, month, day;
```

Table 10: Displaying records 1 - 10

year	month	day	total_flights
2013	1	1	842
2013	1	2	943
2013	1	3	914
2013	1	4	915
2013	1	5	720
2013	1	6	832
2013	1	7	933
2013	1	8	899
2013	1	9	902
2013	1	10	932

Practice 9: Find Flights That Arrived Early

```
SELECT *
FROM flights
WHERE arr_delay < 0;
```

Table 11: Displaying records 1 - 10

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance	hour	minute	time_hour
2013	1	1	544	545	-1	1004	1022	-18	B6	725	N804JB	JFK	BQN	183	1576	5	45	2013-01-01 10:00:00
2013	1	1	554	600	-6	812	837	-25	DL	461	N668DN	LGA	ATL	116	762	6	0	2013-01-01 11:00:00
2013	1	1	557	600	-3	709	723	-14	EV	5708	N829AS	LGA	IAD	53	229	6	0	2013-01-01 11:00:00
2013	1	1	557	600	-3	838	846	-8	B6	79	N593JB	JFK	MCO	140	944	6	0	2013-01-01 11:00:00
2013	1	1	558	600	-2	849	851	-2	B6	49	N793JB	JFK	PBI	149	1028	6	0	2013-01-01 11:00:00
2013	1	1	558	600	-2	853	856	-3	B6	71	N657JB	JFK	TPA	158	1005	6	0	2013-01-01 11:00:00
2013	1	1	558	600	-2	923	937	-14	UA	1124	N53441	EWR	SFO	361	2565	6	0	2013-01-01 11:00:00
2013	1	1	559	559	0	702	706	-4	B6	1806	N708JB	JFK	BOS	44	187	5	59	2013-01-01 10:00:00
2013	1	1	559	600	-1	854	902	-8	UA	1187	N76515	EWR	LAS	337	2227	6	0	2013-01-01 11:00:00
2013	1	1	600	600	0	851	858	-7	B6	371	N595JB	LGA	FLL	152	1076	6	0	2013-01-01 11:00:00

Practice 10: Find the Average Air Time per Carrier

```
SELECT carrier, AVG(air_time) AS avg_air_time
FROM flights
WHERE air_time IS NOT NULL
GROUP BY carrier;
```

Table 12: Displaying records 1 - 10

carrier	avg_air_time
EV	90.07619
AA	188.82230
US	88.57380
9E	86.78160
YV	65.74081

carrier	avg_air_time
FL	101.14394
HA	623.08772
OO	83.48276
UA	211.79135
MQ	91.18025