

SQL Homework

Assessment

Scenario:

Kamu bekerja sebagai data analyst di perusahaan penerbangan komersial. Divisi strategi korporat tengah mengkaji pricing strategi dalam membuat lapisan harga dan membutuhkan informasi preferensi harga tiket yg dipilih oleh penumpang.

Oleh divisi strategi korporat, kamu diminta untuk:

- Membuat daftar harga berdasarkan keberangkatan dan tujuan. Dari setiap harga yang ada, buat ranking (harga paling mahal = ranking pertama, kedua paling mahal ranking kedua, dst) berdasarkan lokasi keberangkatan dan tujuan.
- Dapatkan ranking harga dari setiap transaksi penerbangan. Hitung jumlah penerbangan pada masing-masing ranking harga untuk masing-masing penumpang.
- Dapatkan ranking yang paling banyak ditransaksikan oleh masing-masing penumpang.
- Lengkapi data user dengan ranking yang harga yang paling banyak ditransaksikan oleh penumpang tersebut.

Dataset:

Kamu akan menggunakan tabel flight_dataset dan user_dataset untuk latihan ini. Jika kamu belum memiliki datanya, kamu bisa kembali ke tugas pertama di Lesson 6.

Yang harus kamu lakukan:

Catatan: jika kamu akan mengambil kolom “from” dan “to” dari tabel flight_dataset, selalu gunakan tanda kutip dua ("). Contoh:

```
SELECT "from", "to" FROM flight_dataset
```

- 1 Buat CTE pertama untuk mendapatkan nilai unik (tanpa duplikat) atas kombinasi “from”, “to”, dan price.

Hint: Gunakan **DISTINCT**.

- 2 Buat CTE kedua, dengan menggunakan CTE pertama, buat ranking harga berdasarkan keberangkatan dan tujuan. Harga paling tinggi = ranking satu.

Hint: Gunakan **ROW_NUMBER() OVER(PARTITION BY “from”, “to” ORDER BY ...)**

- 3 Buat CTE ketiga. Di CTE ini, dapatkan data dari flight_dataset yang dijoin dengan CTE kedua seperti berikut

```
FROM flight_dataset AS fd
LEFT JOIN [nama CTE kedua] AS cte_dua
  ON cte_dua."from" = fd."from"
  AND cte_dua."to" = fd."to"
  AND cte_dua.price = fd.price
```

Dengan **COUNT**, hitung banyaknya travel_id yang dikelompokkan berdasarkan user_id dan ranking harga.

- 4 Buat CTE keempat. Dengan menggunakan CTE ketiga, dapatkan maksimum banyaknya travel_id untuk setiap user menggunakan window function:

```
MAX( [kolom total travel_id] ) OVER(PARTITION BY user_id) AS max_transaction
```

Dengan demikian, ranking harga yang paling sering dipilih oleh user adalah ranking harga dengan [kolom total travel_id] = max_transaction.

- 5 Join user_dataset dengan CTE keempat, kondisikan supaya setiap user hanya mendapatkan ranking harga yang paling banyak ditransaksikan.
- 6 Output yang dihasilkan diharapkan seperti ini:

	user_id	company	name	gender	age	price_rank
	integer	text	text	text	integer	bigint
1	1	4You	Roy Braun	male	21	1
2	2	4You	Joseph Holsten	male	37	4
3	2	4You	Joseph Holsten	male	37	5
4	3	4You	Wilma Mcinnis	female	48	5
5	4	4You	Paula Daniel	female	23	1
6	4	4You	Paula Daniel	female	23	3
7	4	4You	Paula Daniel	female	23	4
8	4	4You	Paula Daniel	female	23	7
9	5	4You	Patricia Carson	female	44	5
10	6	4You	Trina Thomas	none	47	4
11	7	4You	Jesse Decelle	male	46	1
12	8	4You	Gregoria Gil	female	21	5
13	9	4You	Jack Sabo	none	41	5
14	10	4You	Debbie Helms	none	35	1
15	11	4You	Malinda Smith	female	26	1