QUERY sudah setelah STAR SKEMA

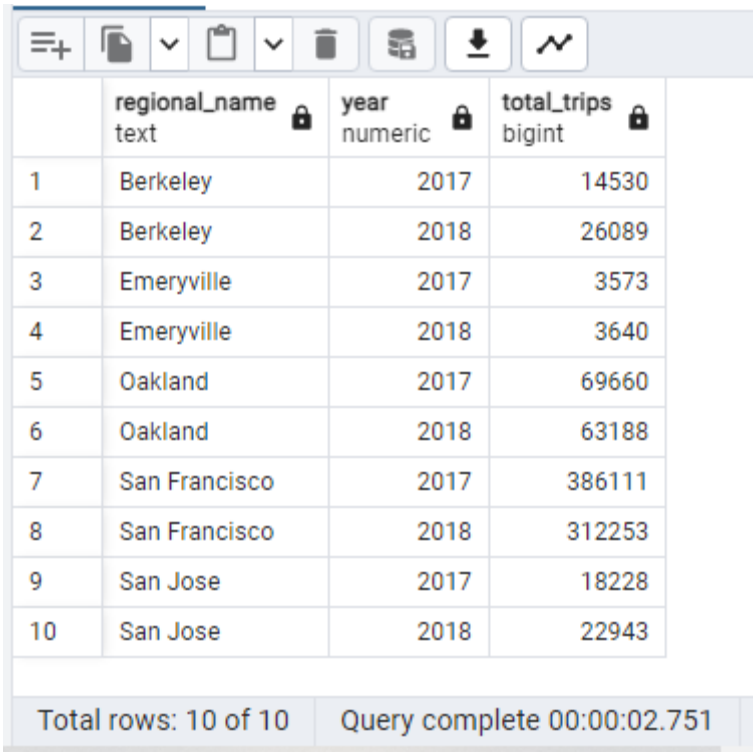
1. **Tren Penggunaan Sewa Sepeda: Jumlah Trip per Wilayah dan Tahun (2017-2018) :**

   Menampilkan jumlah trip sewa sepeda di setiap wilayah untuk tahun 2017 dan 2018. Data ini memberikan wawasan tentang tren penggunaan layanan, memungkinkan analisis pertumbuhan atau penurunan aktivitas sewa sepeda di masing-masing wilayah. Informasi ini dapat digunakan untuk evaluasi kinerja regional, perencanaan kapasitas, dan penyesuaian strategi bisnis.

   Query :

```
-- 1. Jumlah Pemakaian Sewa Sepeda per tahun per wilayah
SELECT
        dr.region_name AS regional_name,
        EXTRACT(YEAR FROM ft.start_date) AS year,
        COUNT(*) AS Total_Trips
 FROM fact_trips ft
JOIN dims_regions_station_info dr ON ft.start_station_id = dr.station_id
WHERE EXTRACT(YEAR FROM ft.start_date) BETWEEN 2017 AND 2018
GROUP BY dr.region_name, EXTRACT(YEAR FROM ft.start_date)
ORDER BY regional_name
```

   Hasil query :



	regional_name text	year numeric	total_trips bigint
1	Berkeley	2017	14530
2	Berkeley	2018	26089
3	Emeryville	2017	3573
4	Emeryville	2018	3640
5	Oakland	2017	69660
6	Oakland	2018	63188
7	San Francisco	2017	386111
8	San Francisco	2018	312253
9	San Jose	2017	18228
10	San Jose	2018	22943

Total rows: 10 of 10 Query complete 00:00:02.751

2. Rata-rata Durasi Perjalanan per Wilayah, Tahun, dan Tipe Keanggotaan

Menampilkan durasi rata-rata perjalanan (dalam menit) berdasarkan wilayah, tahun (2017-2018), dan tipe keanggotaan (subscriber/customer). Data ini memberikan wawasan untuk optimalisasi layanan, pengembangan strategi pemasaran yang ditargetkan, dan perencanaan infrastruktur berdasarkan pola penggunaan di berbagai wilayah dan segmen pelanggan.

Query :

```
-- 2. Rata - Rata Durasi Perjalanan Berdasarkan Wilayah, Tahun, dan Jenis Member (subscriber/customer):
SELECT
  dr.region_name AS Region,
  EXTRACT(YEAR FROM ft.start_date) AS Year,
  ROUND(AVG(ft.duration_sec) / 60,2) AS Average_Duration_Minutes,
  dt.subscriber_type AS Member_type
FROM fact_trips ft
JOIN dims_regions_station_info dr ON ft.start_station_id = dr.station_id
JOIN dims_trips_info dt ON ft.trip_id = dt.trip_id
WHERE EXTRACT(YEAR FROM ft.start_date) IN (2017, 2018)
GROUP BY dr.region_name, EXTRACT(YEAR FROM ft.start_date), Member_type, dt.subscriber_type
ORDER BY Region DESC, Member_type
```

Hasil query :

	region text	year numeric	average_duration_minutes numeric	member_type text
1	San Jose	2017	48.40	Customer
2	San Jose	2018	40.36	Customer
3	San Jose	2017	10.76	Subscriber
4	San Jose	2018	10.07	Subscriber
5	San Francisco	2017	42.41	Customer
6	San Francisco	2018	37.43	Customer
7	San Francisco	2017	11.67	Subscriber
8	San Francisco	2018	11.05	Subscriber
9	Oakland	2017	40.97	Customer
10	Oakland	2018	35.89	Customer
11	Oakland	2017	12.41	Subscriber
12	Oakland	2018	10.01	Subscriber
13	Emeryville	2017	53.90	Customer
14	Emeryville	2018	45.26	Customer
15	Emeryville	2017	15.86	Subscriber
16	Emeryville	2018	13.42	Subscriber
17	Berkeley	2017	47.69	Customer
18	Berkeley	2018	33.90	Customer
19	Berkeley	2017	13.09	Subscriber
20	Berkeley	2018	10.18	Subscriber

Total rows: 20 of 20 Query complete 00:00:07.704

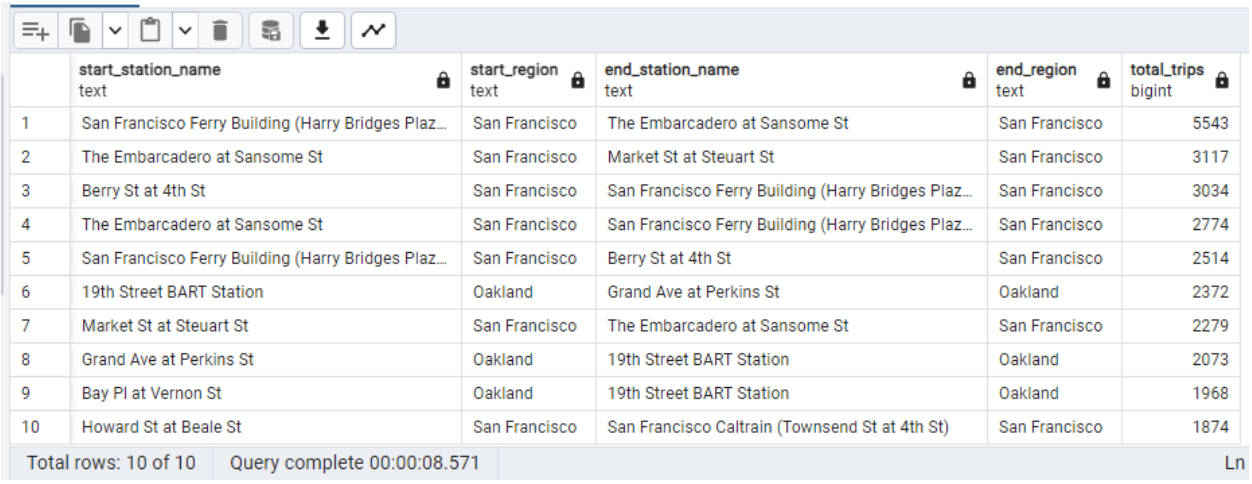
3. Top 10 Rute Perjalanan: Stasiun Awal-Akhir, Region, dan Frekuensi Perjalanan:

Menampilkan 10 rute perjalanan terpopuler, termasuk nama stasiun awal dan akhir, region masing-masing stasiun, serta jumlah total perjalanan. Mengidentifikasi rute-rute dengan frekuensi perjalanan tertinggi untuk optimalisasi layanan dan perencanaan.

Query :

```
-- 3. Nama Stasiun Awal dan Stasiun Akhir dengan Jumlah Perjalanan Terbanyak

-- dan berada di wilayah region mana stasiun tersebut

SELECT

        dr1.station_name AS start_station_name,

        dr1.region_name AS start_region,

        dr2.station_name AS end_station_name,

        dr2.region_name AS end_region,

        COUNT(ft.trip_id) AS total_trips

FROM fact_trips ft

JOIN dims_regions_station_info dr1 ON ft.start_station_id = dr1.station_id

JOIN dims_regions_station_info dr2 ON ft.end_station_id = dr2.station_id

JOIN dims_trips_info dt ON ft.trip_id = dt.trip_id

GROUP BY dr1.region_name,dr2.region_name,start_station_name,end_station_name

ORDER BY total_trips desc

Limit 10
```

Hasil query :



	start_station_name text	start_region text	end_station_name text	end_region text	total_trips bigint
1	San Francisco Ferry Building (Harry Bridges Plaz...	San Francisco	The Embarcadero at Sansome St	San Francisco	5543
2	The Embarcadero at Sansome St	San Francisco	Market St at Steuart St	San Francisco	3117
3	Berry St at 4th St	San Francisco	San Francisco Ferry Building (Harry Bridges Plaz...	San Francisco	3034
4	The Embarcadero at Sansome St	San Francisco	San Francisco Ferry Building (Harry Bridges Plaz...	San Francisco	2774
5	San Francisco Ferry Building (Harry Bridges Plaz...	San Francisco	Berry St at 4th St	San Francisco	2514
6	19th Street BART Station	Oakland	Grand Ave at Perkins St	Oakland	2372
7	Market St at Steuart St	San Francisco	The Embarcadero at Sansome St	San Francisco	2279
8	Grand Ave at Perkins St	Oakland	19th Street BART Station	Oakland	2073
9	Bay Pl at Vernon St	Oakland	19th Street BART Station	Oakland	1968
10	Howard St at Beale St	San Francisco	San Francisco Caltrain (Townsend St at 4th St)	San Francisco	1874

Total rows: 10 of 10 | Query complete 00:00:08.571 | Ln

4. Analisis Pemanfaatan Stasiun Sepeda per Regional :

menampilkan kategori utilisasi stasiun sepeda per region berdasarkan rata-rata perjalanan per sepeda per minggu, serta distribusi stasiun dalam setiap kategori. Kategori utilisasi dibagi menjadi: Underutilized: 0-6 perjalanan/sepeda/minggu, Ideal: 7-14 perjalanan/sepeda/minggu, Overutilized: >14 perjalanan/sepeda/minggu. Menampilkan total stasiun, perjalanan, kapasitas, rata-rata perjalanan, dan persentase stasiun per kategori. Analisis ini membantu mengidentifikasi stasiun kurang/over-dimanfaatkan, mengoptimalkan distribusi sepeda, merencanakan redistribusi armada, menentukan lokasi ekspansi/pengurangan kapasitas, dan mengidentifikasi tren penggunaan. Hasilnya meningkatkan efisiensi operasional, kepuasan pelanggan, dan profitabilitas di setiap wilayah

noted tambahan :

- *1 minggu = 7 hari*
- *1 bulan = 30 hari*

Kategori utilisasi:

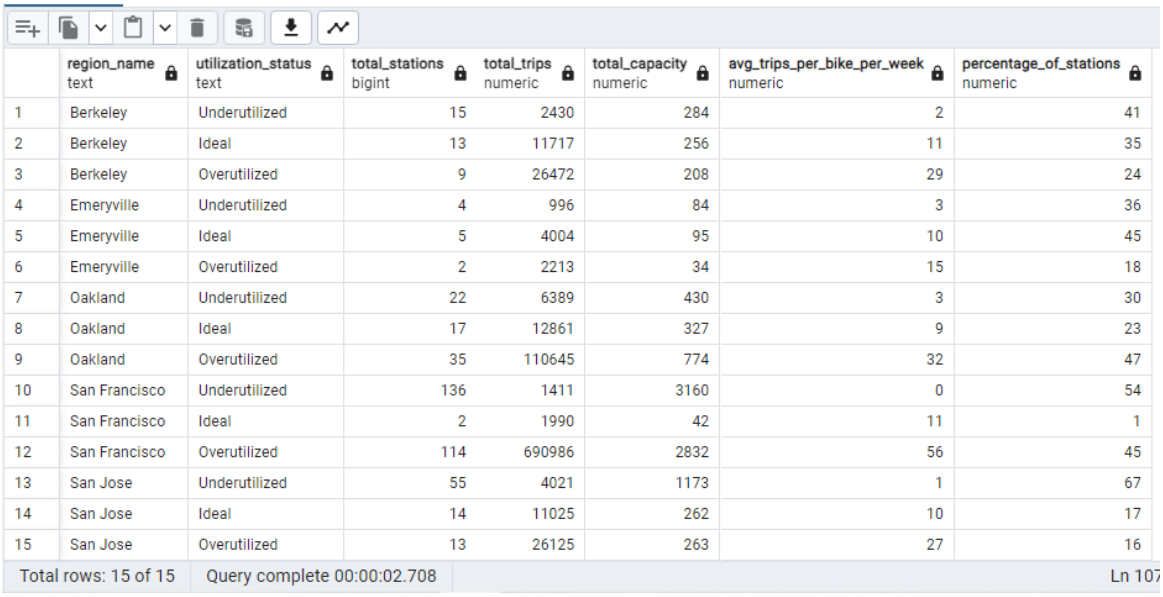
- *Underutilized: 0-6 perjalanan/sepeda/minggu >> 1 sepeda 1 minggu dipakai kurang dari 7 kali. (tidak setiap hari 1 sepeda dipakai)*
- *Ideal: 7-14 perjalanan/sepeda/minggu >> 1 sepeda 1 minggu dipakai 7-14x, dengan kata lain sehari 1 sepeda maksimal 2x dipakai.*
- *Overutilized: >14 perjalanan/sepeda/minggu >> sehari sepeda dipakai lebih dari 2x*

Query :

```
-- 4. Analisis Pemanfaatan Stasiun Sepeda per Regional

WITH station_stats AS (
  SELECT
    dr.region_name,
    dr.station_id,
    dr.station_name,
    dr.capacity,
    COUNT(ft.trip_id) as total_trips,
    30 as days_in_period, -- Asumsikan periode 30 hari
    CASE
      WHEN dr.capacity > 0 THEN
        (COUNT(ft.trip_id)::float / dr.capacity) / (30::float / 7)
      ELSE 0
    END AS trips_per_bike_per_week,
    CASE
      WHEN dr.capacity > 0 THEN
        CASE
          WHEN (COUNT(ft.trip_id)::float / dr.capacity) / (30::float / 7) < 7 THEN 'Underutilized'
          WHEN (COUNT(ft.trip_id)::float / dr.capacity) / (30::float / 7) <= 14 THEN 'Ideal'
          ELSE 'Overutilized'
        END
      ELSE 'No data'
    END AS utilization_status
```

```
  FROM
    dims_regions_station_info dr
  LEFT JOIN
    fact_trips ft ON dr.station_id = ft.start_station_id
  WHERE
    dr.capacity > 0
  GROUP BY
    dr.region_name, dr.station_id, dr.station_name, dr.capacity
),
region_totals AS (
  SELECT
    region_name,
    COUNT(station_id) as total_regional_stations
  FROM
    station_stats
  GROUP BY
    region_name
)
SELECT
  ss.region_name,
  ss.utilization_status,
  COUNT(ss.station_id) as total_stations,
  SUM(ss.total_trips) as total_trips,
  SUM(ss.capacity) as total_capacity,
  ROUND(AVG(ss.trips_per_bike_per_week)::numeric, 0) as avg_trips_per_bike_per_week,
  ROUND((COUNT(ss.station_id)::float / rt.total_regional_stations * 100)::numeric, 0) as
percentage_of_stations
FROM
  station_stats ss
JOIN
  region_totals rt ON ss.region_name = rt.region_name
GROUP BY
  ss.region_name, ss.utilization_status, rt.total_regional_stations
ORDER BY
  ss.region_name,
  CASE ss.utilization_status
    WHEN 'Underutilized' THEN 1
    WHEN 'Ideal' THEN 2
    WHEN 'Overutilized' THEN 3
    ELSE 4
  END
```

hasil Query :



	region_name text	utilization_status text	total_stations bigint	total_trips numeric	total_capacity numeric	avg_trips_per_bike_per_week numeric	percentage_of_stations numeric
1	Berkeley	Underutilized	15	2430	284	2	41
2	Berkeley	Ideal	13	11717	256	11	35
3	Berkeley	Overutilized	9	26472	208	29	24
4	Emeryville	Underutilized	4	996	84	3	36
5	Emeryville	Ideal	5	4004	95	10	45
6	Emeryville	Overutilized	2	2213	34	15	18
7	Oakland	Underutilized	22	6389	430	3	30
8	Oakland	Ideal	17	12861	327	9	23
9	Oakland	Overutilized	35	110645	774	32	47
10	San Francisco	Underutilized	136	1411	3160	0	54
11	San Francisco	Ideal	2	1990	42	11	1
12	San Francisco	Overutilized	114	690986	2832	56	45
13	San Jose	Underutilized	55	4021	1173	1	67
14	San Jose	Ideal	14	11025	262	10	17
15	San Jose	Overutilized	13	26125	263	27	16

Total rows: 15 of 15 | Query complete 00:00:02.708 | Ln 107

5. Distribusi Trip Berdasarkan Wilayah, Tahun, Gender, dan Kelompok Usia:

Menampilkan jumlah perjalanan sepeda berdasarkan wilayah, tahun, dan demografi (gender dan kelompok usia), memberikan wawasan tentang pola penggunaan layanan. Membantu dalam perencanaan promosi dan alokasi sumber daya sesuai karakteristik demografis pengguna

Query :

```
-- 5.Distribusi Trip Berdasarkan Wilayah, Tahun, Gender, dan Kelompok Usia
SELECT
  dr.region_name AS Region,
  EXTRACT(YEAR FROM ft.start_date) AS Year,
  COUNT(*) AS Total_Trips,
  SUM(CASE WHEN dt.member_gender = 'Male' THEN 1 ELSE 0 END) AS Male_Trips,
  SUM(CASE WHEN dt.member_gender = 'Female' THEN 1 ELSE 0 END) AS Female_Trips,
  CASE
    WHEN EXTRACT(YEAR FROM CURRENT_DATE) - dt.member_birth_year BETWEEN 18 AND 39 THEN
'Young Adults'
    WHEN EXTRACT(YEAR FROM CURRENT_DATE) - dt.member_birth_year BETWEEN 40 AND 64 THEN
'Middle-aged Adults'
    WHEN EXTRACT(YEAR FROM CURRENT_DATE) - dt.member_birth_year >= 65 THEN 'Senior Adults'
    ELSE 'Unidentified'
  END AS Age_Group
FROM fact_trips ft
JOIN dims_regions_station_info dr ON ft.start_station_id = dr.station_id
JOIN dims_trips_info dt ON ft.trip_id = dt.trip_id
WHERE EXTRACT(YEAR FROM ft.start_date) IN (2017, 2018)
  AND dt.member_gender IN ('Male', 'Female')  -- Filter for valid genders
  AND dt.member_birth_year IS NOT NULL        -- Filter for birth year information
GROUP BY dr.region_name, EXTRACT(YEAR FROM ft.start_date), Age_Group
ORDER BY Region DESC, Age_Group
```

Hasil query:

	region text	year numeric	total_trips bigint	male_trips bigint	female_trips bigint	age_group text
1	San Jose	2017	6127	4945	1182	Middle-aged Adults
2	San Jose	2018	5402	4286	1116	Middle-aged Adults
3	San Jose	2017	488	317	171	Senior Adults
4	San Jose	2018	527	292	235	Senior Adults
5	San Jose	2017	8681	6775	1906	Young Adults
6	San Jose	2018	15271	11522	3749	Young Adults
7	San Francisco	2017	181718	147896	33822	Middle-aged Adults
8	San Francisco	2018	141272	112290	28982	Middle-aged Adults
9	San Francisco	2017	14271	12628	1643	Senior Adults
10	San Francisco	2018	11154	9585	1569	Senior Adults
11	San Francisco	2017	135203	101572	33631	Young Adults
12	San Francisco	2018	130613	94387	36226	Young Adults
13	Oakland	2017	31073	22928	8145	Middle-aged Adults
14	Oakland	2018	29084	20970	8114	Middle-aged Adults
15	Oakland	2017	2568	2049	519	Senior Adults
16	Oakland	2018	2360	1903	457	Senior Adults
17	Oakland	2017	26329	17712	8617	Young Adults
18	Oakland	2018	26436	16784	9652	Young Adults
19	Emeryville	2017	1479	992	487	Middle-aged Adults
20	Emeryville	2018	1629	1292	337	Middle-aged Adults
21	Emeryville	2017	131	73	58	Senior Adults
22	Emeryville	2018	95	65	30	Senior Adults
23	Emeryville	2017	1239	913	326	Young Adults
24	Emeryville	2018	1477	1019	458	Young Adults
25	Berkeley	2017	4358	3218	1140	Middle-aged Adults
26	Berkeley	2018	7812	5879	1933	Middle-aged Adults
27	Berkeley	2017	363	276	87	Senior Adults
28	Berkeley	2018	780	645	135	Senior Adults
29	Berkeley	2017	7718	5707	2011	Young Adults
30	Berkeley	2018	14937	10374	4563	Young Adults

Total rows: 30 of 30 Query complete 00:00:07.519

6. Hari dan Jam Tersibuk per Wilayah dan Tahun (2017-2018):

Menampilkan hari dan jam tersibuk untuk setiap wilayah pada tahun 2017 dan 2018. Data ini penting untuk optimalisasi layanan, alokasi sumber daya, dan perencanaan operasional berdasarkan pola penggunaan puncak di setiap wilayah, memungkinkan strategi bisnis yang lebih efektif dan responsif.

Query :

```
-- 6.Hari dan Jam Tersibuk per Wilayah dan Tahun (2017-2018)

WITH regional_peaks AS (
  SELECT
    dr.region_name,
    EXTRACT(YEAR FROM ft.start_date) AS year,
    CASE EXTRACT(DOW FROM ft.start_date)
      WHEN 0 THEN 'Sunday'
      WHEN 1 THEN 'Monday'
      WHEN 2 THEN 'Tuesday'
      WHEN 3 THEN 'Wednesday'
      WHEN 4 THEN 'Thursday'
      WHEN 5 THEN 'Friday'
      WHEN 6 THEN 'Saturday'
    END AS day_of_week,
    EXTRACT(HOUR FROM ft.start_date) AS hour_of_day,
    COUNT(*) AS total_trips,
    ROW_NUMBER() OVER (PARTITION BY dr.region_name, EXTRACT(YEAR FROM ft.start_date) ORDER
BY COUNT(*) DESC) AS rank
  FROM fact_trips ft
  JOIN dims_regions_station_info dr ON ft.start_station_id = dr.station_id
  WHERE EXTRACT(YEAR FROM ft.start_date) IN (2017, 2018)
  GROUP BY dr.region_name, year, day_of_week, hour_of_day
)
SELECT
  region_name,
  year,
  day_of_week,
  hour_of_day,
  total_trips
FROM regional_peaks
WHERE rank = 1
ORDER BY region_name, year, total_trips DESC
```

Hasil query :

	region_name text	year numeric	day_of_week text	hour_of_day numeric	total_trips bigint
1	Berkeley	2017	Thursday	1	272
2	Berkeley	2018	Wednesday	1	491
3	Emeryville	2017	Thursday	0	73
4	Emeryville	2018	Wednesday	0	81
5	Oakland	2017	Wednesday	15	1307
6	Oakland	2018	Wednesday	15	1362
7	San Francisco	2017	Tuesday	15	9049
8	San Francisco	2018	Wednesday	15	8284
9	San Jose	2017	Tuesday	0	331
10	San Jose	2018	Tuesday	0	427

Total rows: 10 of 10 Query complete 00:00:05.432