

# WEB SCRAPPING at TIKET.COM USING SQL

---

MARCH 2

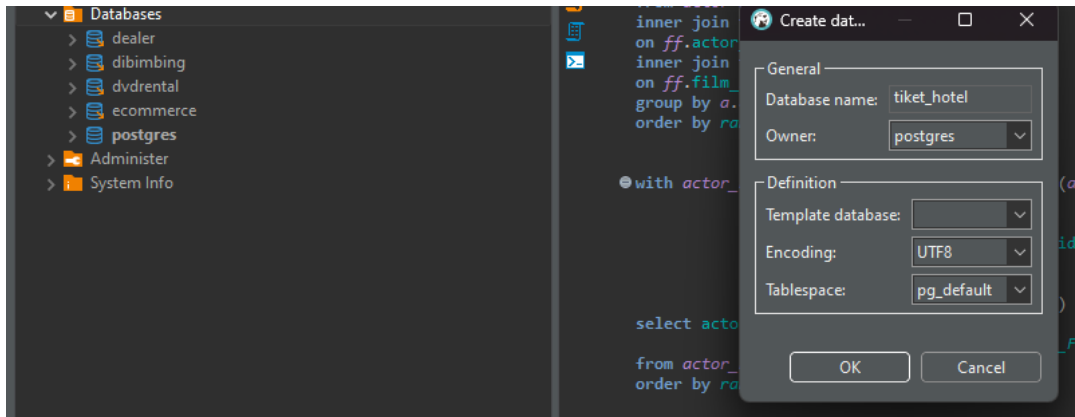
---

Authored by: FERDYANSYAH PERMANA PUTRA

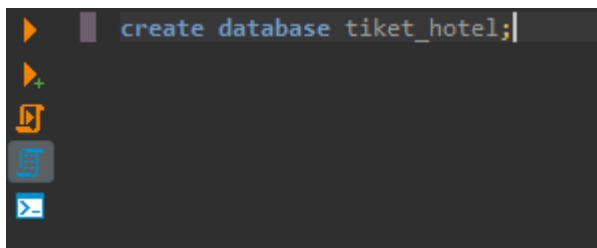


Setelah melakukan proses scrapping menggunakan **Python** kemudian melakukan **pembersihan data** setelah data scrapping. Selanjutnya, akan melakukan proses SQL untuk menyimpan data hasil scrapping dan melakukan **query sederhana** untuk **menarik data** dari PostgreSQL.

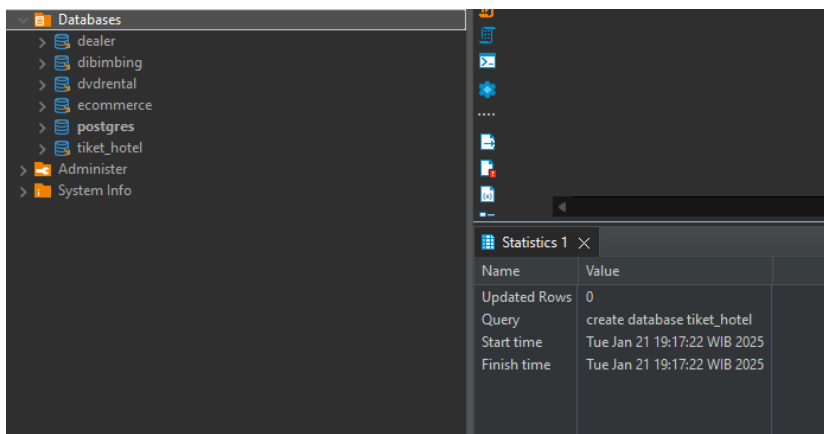
1. Buatlah database di PostgreSQL dan tabel untuk menyimpan data hasil scraping.



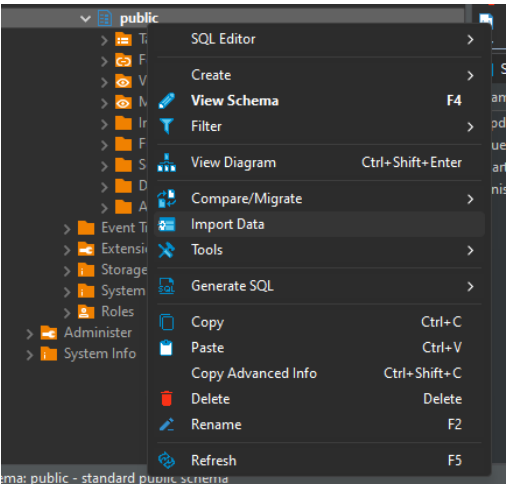
atau menggunakan query



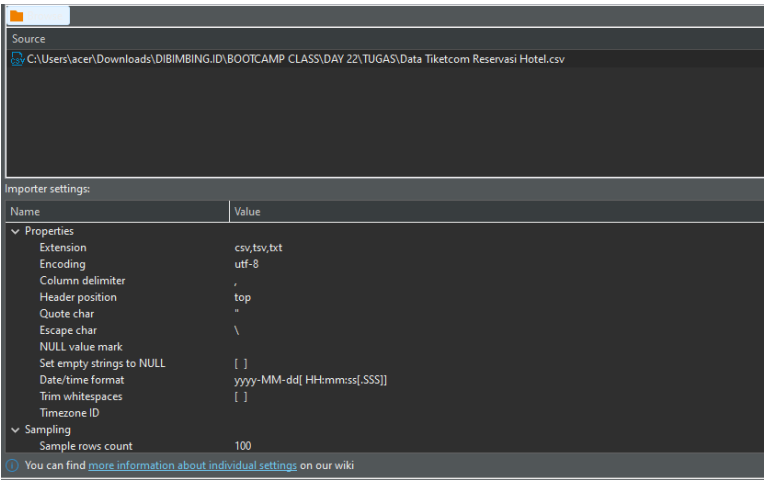
Output:



Kemudian melakukan import data dari hasil data scrapping dalam bentuk csv ke SQL.



Kemudian melakukan column delimiter dengan “,”

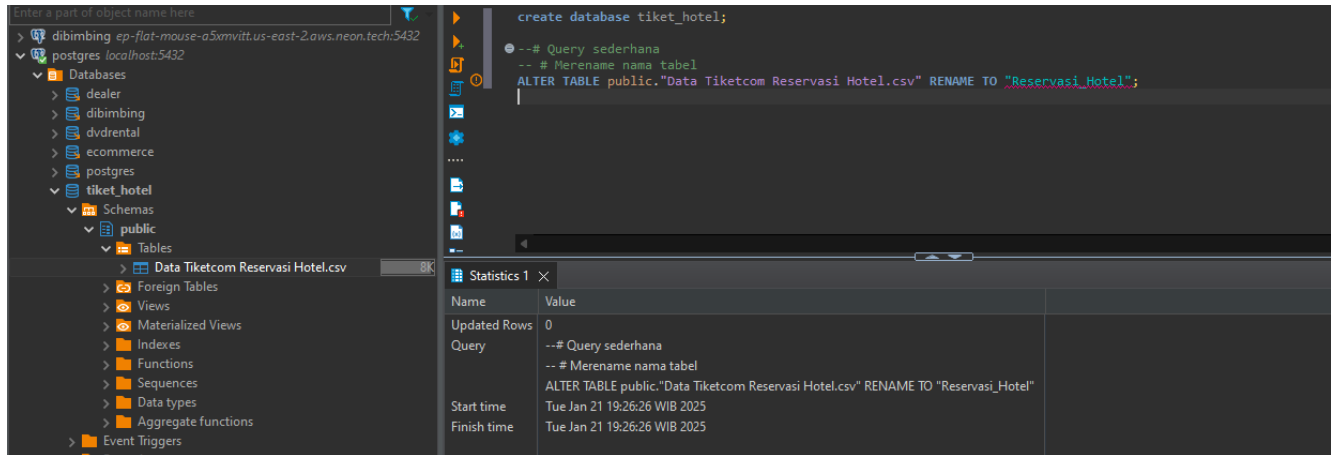


Dengan konfigurasinya →

Columns mapping						
Table properties						
Target DDL						
Source container: Data Tiketcom Reservasi Hotel.csv						
Source entity: Data Tiketcom Reservasi Hotel.csv						
Target container: postgres						
Target entity: "Data Tiketcom Reservasi Hotel.csv"						
Source Column	Source Type	Target Column	Target Type	Mapping	Transform	
123 "No"	INTEGER	No	integer	new		
A-Z "Nama Hot...	VARCHAR(64)	Nama_Hotel	varchar(64)	new		
A-Z Alamat	VARCHAR(50)	Alamat	varchar(50)	new		
123 "Angka Rati...	REAL	Angka_Rating	real	new		
123 "Total Revi...	REAL	Total_Review...	real	new		
123 "Harga Asli ...	INTEGER	Harga_Asli_Fix	integer	new		
123 "Harga Sesa...	INTEGER	Harga_Sesuda...	integer	new		
A-Z "Kategori H...	VARCHAR(50)	Kategori_Harga	varchar(50)	new		
123 "Diskon (%)"	REAL	Diskon	real	new		

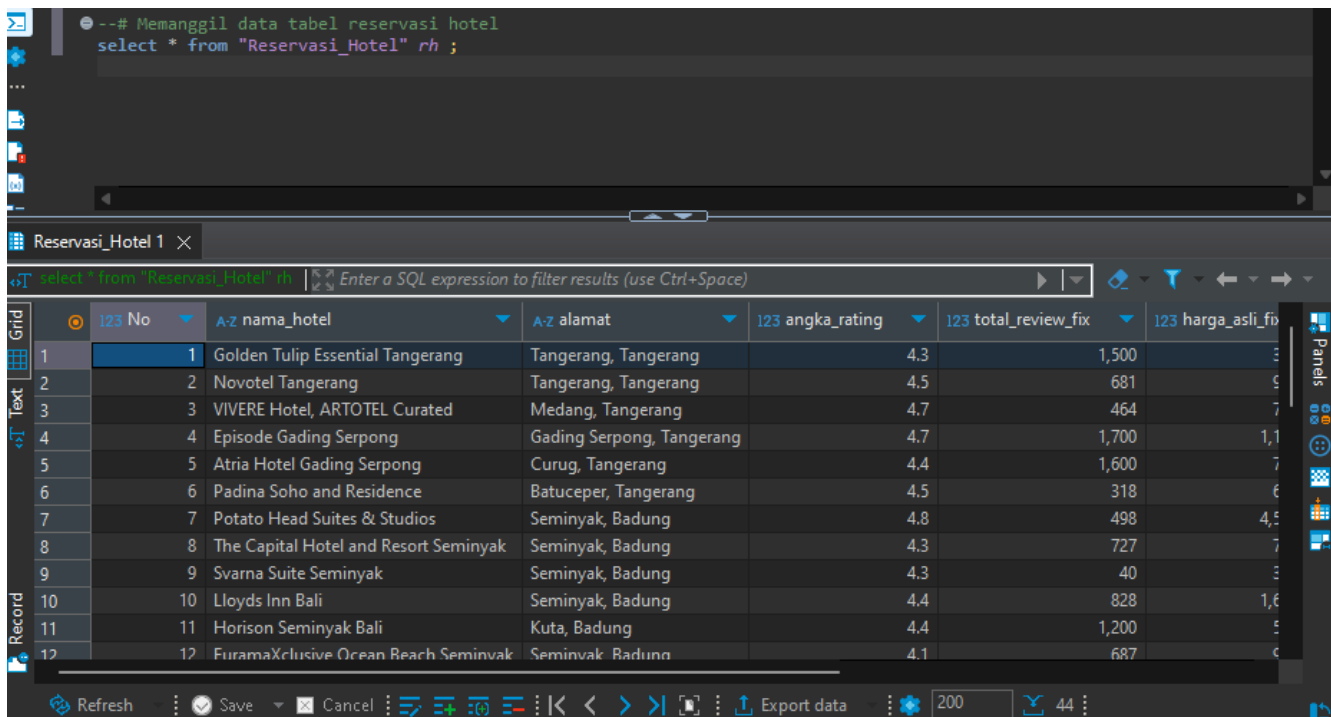
## 2. Melakukan rename tabel dengan fungsi Alter table

`ALTER TABLE public."Data Tiketcom Reservasi Hotel.csv" RENAME TO "Reservasi_Hotel";`



## 3. Memanggil Kolom Reservasi\_Hotel

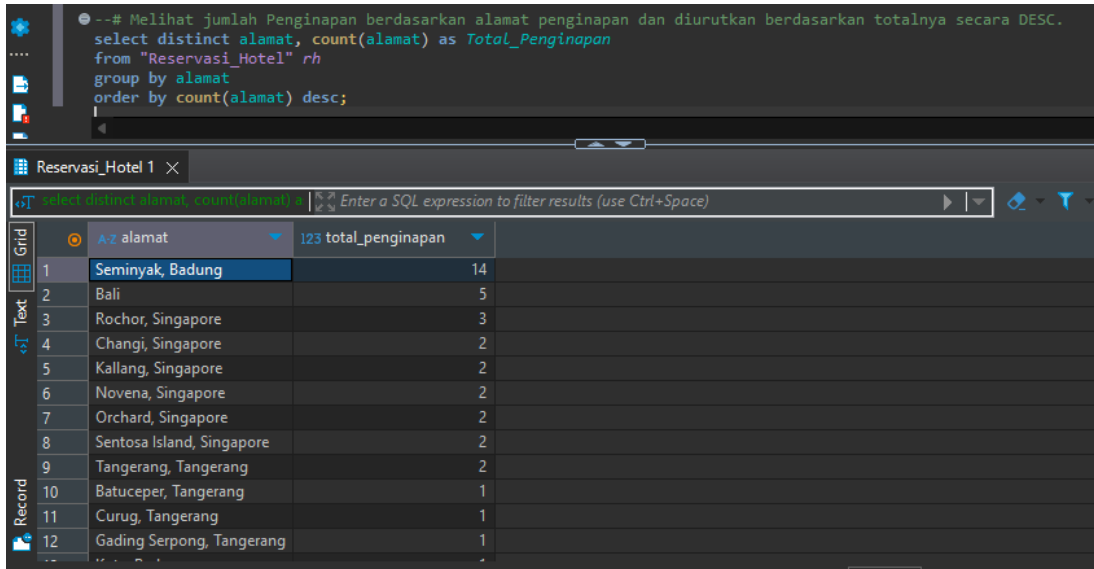
`select * from "Reservasi_Hotel" rh;`



## 4. Melihat jumlah penginapan berdasarkan alamat penginapan dan diurutkan berdasarkan totalnya secara DESC.

--# Melihat jumlah Penginapan berdasarkan alamat penginapan dan diurutkan berdasarkan totalnya secara DESC.

```
select distinct alamat, count(alamat) as Total_Penginapan
from "Reservasi_Hotel" rh
group by alamat
order by count(alamat) desc;
```

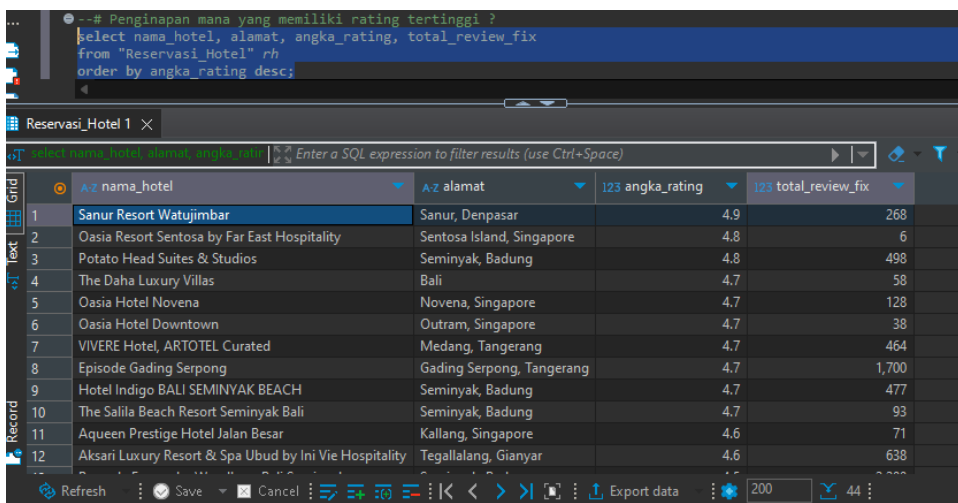


	A:2 alamat	123 total_penginapan
1	Seminyak, Badung	14
2	Bali	5
3	Rochor, Singapore	3
4	Changi, Singapore	2
5	Kallang, Singapore	2
6	Novena, Singapore	2
7	Orchard, Singapore	2
8	Sentosa Island, Singapore	2
9	Tangerang, Tangerang	2
10	Batuceper, Tangerang	1
11	Curug, Tangerang	1
12	Gading Serpong, Tangerang	1

Interpretasi → Ternyata total penginapan yang paling banyak adalah dengan alamat Seminyak, Badung sebesar 14 total penginapan yang tersedia.

## 5. Penginapan mana yang memiliki rating tertinggi?

```
select nama_hotel, alamat, angka_rating, total_review_fix
from "Reservasi_Hotel" rh
order by angka_rating desc;
```

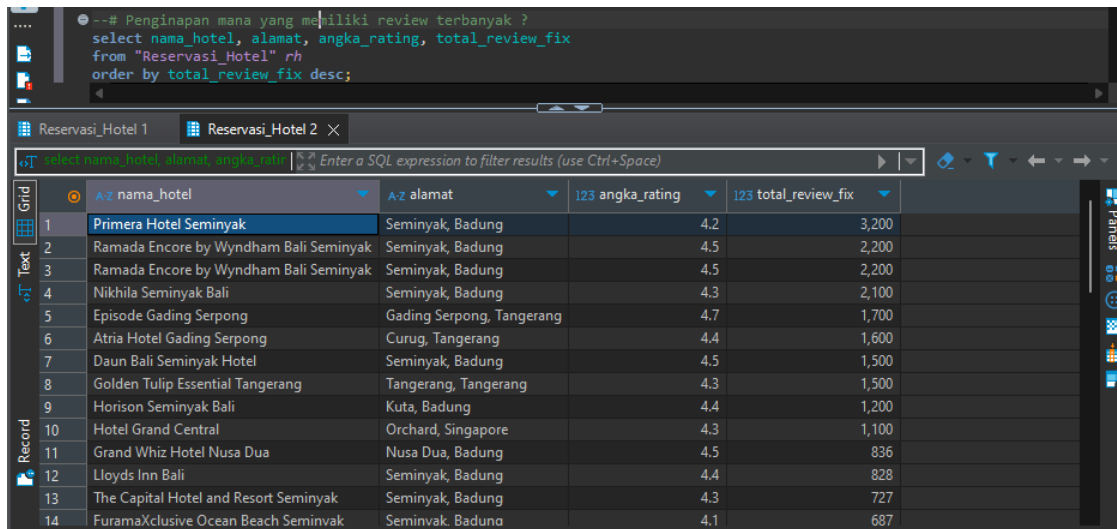


	A:2 nama_hotel	A:2 alamat	123 angka_rating	123 total_review_fix
1	Sanur Resort Watujimbar	Sanur, Denpasar	4.9	268
2	Oasia Resort Sentosa by Far East Hospitality	Sentosa Island, Singapore	4.8	6
3	Potato Head Suites & Studios	Seminyak, Badung	4.8	498
4	The Daha Luxury Villas	Bali	4.7	58
5	Oasia Hotel Novena	Novena, Singapore	4.7	128
6	Oasia Hotel Downtown	Outram, Singapore	4.7	38
7	VIVERE Hotel, ARTOTEL Curated	Medang, Tangerang	4.7	464
8	Episode Gading Serpong	Gading Serpong, Tangerang	4.7	1,700
9	Hotel Indigo BALI SEMINYAK BEACH	Seminyak, Badung	4.7	477
10	The Salila Beach Resort Seminyak Bali	Seminyak, Badung	4.7	93
11	Aqueen Prestige Hotel Jalan Besar	Kallang, Singapore	4.6	71
12	Aksari Luxury Resort & Spa Ubud by Ini Vie Hospitality	Tegallalang, Gianyar	4.6	638

Interpretasi → Ternyata penginapan dengan Hotel Sanur Resort Watujimbar yang memiliki angka rating terbanyak sebesar 4,9 dari 5 dengan total review sebanyak 268 oleh konsumen dimana beralamat pada Sanur, Denpasar.

## 6. Penginapan mana yang memiliki total reviews tertinggi?

```
select nama_hotel, alamat, angka_rating, total_review_fix
from "Reservasi_Hotel" rh
order by total_review_fix desc;
```

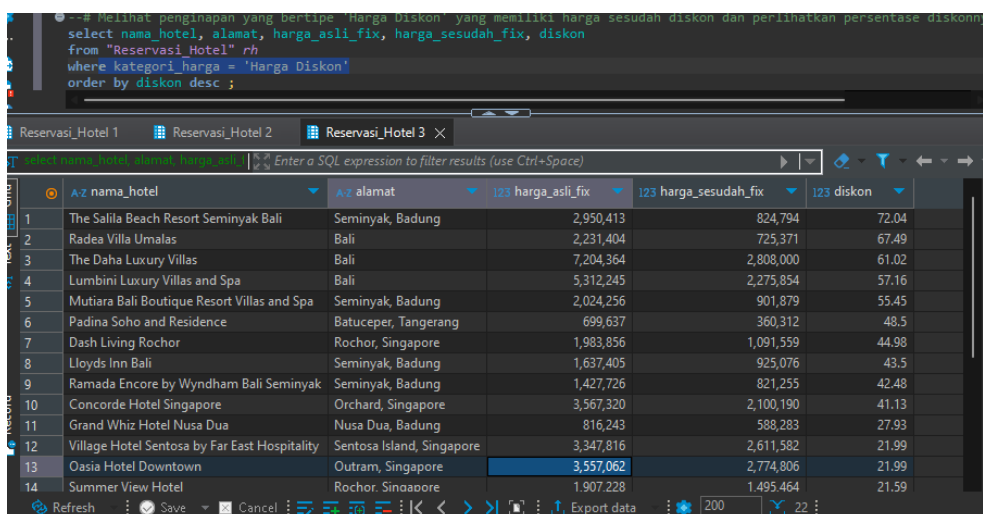


	A-Z nama_hotel	A-Z alamat	123 angka_rating	123 total_review_fix
1	Primera Hotel Seminyak	Seminyak, Badung	4.2	3,200
2	Ramada Encore by Wyndham Bali Seminyak	Seminyak, Badung	4.5	2,200
3	Ramada Encore by Wyndham Bali Seminyak	Seminyak, Badung	4.5	2,200
4	Nikhila Seminyak Bali	Seminyak, Badung	4.3	2,100
5	Episode Gading Serpong	Gading Serpong, Tangerang	4.7	1,700
6	Atria Hotel Gading Serpong	Curug, Tangerang	4.4	1,600
7	Daun Bali Seminyak Hotel	Seminyak, Badung	4.5	1,500
8	Golden Tulip Essential Tangerang	Tangerang, Tangerang	4.3	1,500
9	Horison Seminyak Bali	Kuta, Badung	4.4	1,200
10	Hotel Grand Central	Orchard, Singapore	4.3	1,100
11	Grand Whiz Hotel Nusa Dua	Nusa Dua, Badung	4.5	836
12	Lloyds Inn Bali	Seminyak, Badung	4.4	828
13	The Capital Hotel and Resort Seminyak	Seminyak, Badung	4.3	727
14	FuramaXclusive Ocean Beach Seminyak	Seminyak, Badung	4.1	687

Interpretasi → Total review terbanyak untuk tempat penginapan yaitu pada Primera Hotel Seminyak sebanyak 3200 reviews.

## 7. Melihat penginapan yang bertipe 'Harga Diskon' yang memiliki harga sesudah diskon dan perlihatkan persentase diskonnya yang paling tinggi

```
select nama_hotel, alamat, harga_asli_fix, harga_sesudah_fix, diskon
from "Reservasi_Hotel" rh
where kategori_harga = 'Harga Diskon'
order by diskon desc ;
```

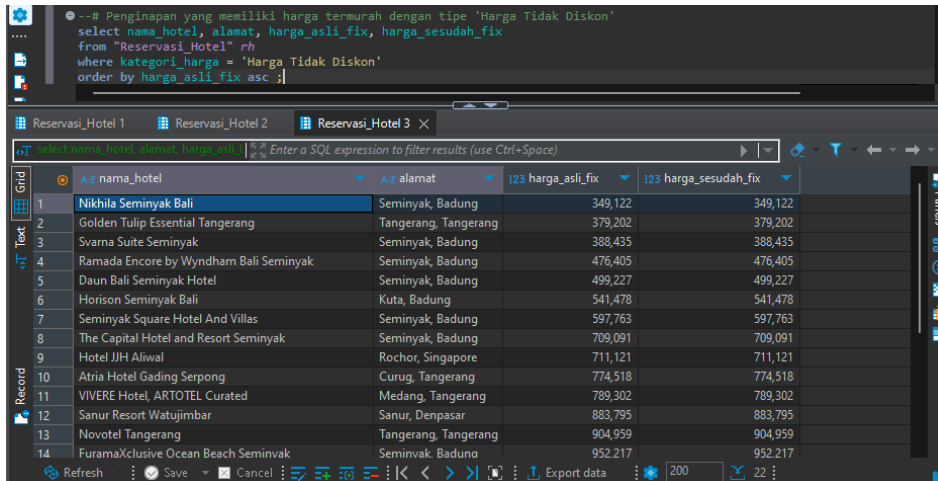


	A-Z nama_hotel	A-Z alamat	123 harga_asli_fix	123 harga_sesudah_fix	123 diskon
1	The Salila Beach Resort Seminyak Bali	Seminyak, Badung	2,950,413	824,794	72.04
2	Radea Villa Umalas	Bali	2,231,404	725,371	67.49
3	The Daha Luxury Villas	Bali	7,204,364	2,808,000	61.02
4	Lumbini Luxury Villas and Spa	Bali	5,312,245	2,275,854	57.16
5	Mutiara Bali Boutique Resort Villas and Spa	Seminyak, Badung	2,024,256	901,879	55.45
6	Padina Soho and Residence	Batucape, Tangerang	699,637	360,312	48.5
7	Dash Living Rochor	Rochor, Singapore	1,983,856	1,091,559	44.98
8	Lloyds Inn Bali	Seminyak, Badung	1,637,405	925,076	43.5
9	Ramada Encore by Wyndham Bali Seminyak	Seminyak, Badung	1,427,726	821,255	42.48
10	Concorde Hotel Singapore	Orchard, Singapore	3,567,320	2,100,190	41.13
11	Grand Whiz Hotel Nusa Dua	Nusa Dua, Badung	816,243	588,283	27.93
12	Village Hotel Sentosa by Far East Hospitality	Sentosa Island, Singapore	3,347,816	2,611,582	21.99
13	Oasia Hotel Downtown	Outram, Singapore	3,557,062	2,774,806	21.99
14	Summer View Hotel	Rochor, Singapore	1,907,228	1,495,464	21.99

Interpretasi → Penginapan yang memiliki diskon tertinggi adalah The Salila Beach Resort Seminyak Bali sebesar 72,04% dari harga aslinya Rp. 2.950.413 menjadi Rp. 824.794 per malamnya yang beralamat di Seminyak, Badung.

## 8. Melihat penginapan yang bertipe 'Harga Tidak Diskon' yang memiliki harga termurah

```
select nama_hotel, alamat, harga_asli_fix, harga_sesudah_fix
from "Reservasi_Hotel" rh
where kategori_harga = 'Harga Tidak Diskon'
order by harga_asli_fix asc ;
```



SQL Query:

```
--# Penginapan yang memiliki harga termurah dengan tipe 'Harga Tidak Diskon'
select nama_hotel, alamat, harga_asli_fix, harga_sesudah_fix
from "Reservasi_Hotel" rh
where kategori_harga = 'Harga Tidak Diskon'
order by harga_asli_fix asc ;
```

	A-Z nama_hotel	A-Z alamat	123 harga_asli_fix	123 harga_sesudah_fix
1	Nikhila Seminyak Bali	Seminyak, Badung	349,122	349,122
2	Golden Tulip Essential Tangerang	Tangerang, Tangerang	379,202	379,202
3	Svarna Suite Seminyak	Seminyak, Badung	388,435	388,435
4	Ramada Encore by Wyndham Bali Seminyak	Seminyak, Badung	476,405	476,405
5	Daun Bali Seminyak Hotel	Seminyak, Badung	499,227	499,227
6	Horison Seminyak Bali	Kuta, Badung	541,478	541,478
7	Seminyak Square Hotel And Villas	Seminyak, Badung	597,763	597,763
8	The Capital Hotel and Resort Seminyak	Seminyak, Badung	709,091	709,091
9	Hotel JIH Aliwal	Rochor, Singapore	711,121	711,121
10	Atria Hotel Gading Serpong	Curug, Tangerang	774,518	774,518
11	VIVERE Hotel, ARTOTEL Curated	Medang, Tangerang	789,302	789,302
12	Sanur Resort Watujimbar	Sanur, Denpasar	883,795	883,795
13	Novotel Tangerang	Tangerang, Tangerang	904,959	904,959
14	FuramaKlusive Ocean Beach Seminyak	Seminyak, Badung	952,217	952,217

Interpretasi → Penginapan yang termurah dengan kategori 'Harga Tidak Diskon' adalah Nikhila Seminyak Bali di Seminyak, Badung dengan harga Rp. 349.122 permalamnya.

## 9. Memilih penginapan yang harganya di atas rata-rata harga sesudahnya dan di atas rating rata-rata

```
with filter_hotel as (select nama_hotel, alamat, angka_rating, harga_sesudah_fix
                      from "Reservasi_Hotel" rh
                      where angka_rating > (select avg(angka_rating) from "Reservasi_Hotel"
                      rh) and
                      harga_sesudah_fix > (select avg(harga_sesudah_fix) from
                      "Reservasi_Hotel" rh )
                      group by nama_hotel, alamat, angka_rating, harga_sesudah_fix
                      )
select nama_hotel, alamat, angka_rating, harga_sesudah_fix
from filter_hotel
group by nama_hotel, alamat, angka_rating, harga_sesudah_fix;
```



```
--# Memilih penginapan yang harganya di atas rata-rata harga sesudahnya dan di atas rating rata-rata
with filter_hotel as (select nama_hotel, alamat, angka_rating, harga_sesudah_fix
from "Reservasi_Hotel" rh
where angka_rating > (select avg(angka_rating) from "Reservasi_Hotel" rh) and
harga_sesudah_fix > (select avg(harga_sesudah_fix) from "Reservasi_Hotel" rh )
group by nama_hotel, alamat, angka_rating, harga_sesudah_fix
)
select nama_hotel, alamat, angka_rating, harga_sesudah_fix
from filter_hotel
group by nama_hotel, alamat, angka_rating, harga_sesudah_fix;
```

	A-Z nama_hotel	A-Z alamat	123 angka_rating	123 harga_sesudah_fix
1	Aksari Luxury Resort & Spa Ubud by Ini Vie Hospitality	Tegallalang, Gianyar	4.6	2,561,275
2	Aqueen Prestige Hotel Jalan Besar	Kallang, Singapore	4.6	1,513,664
3	Hotel Indigo BALI SEMINYAK BEACH	Seminyak, Badung	4.7	3,429,540
4	Oasia Hotel Downtown	Outram, Singapore	4.7	2,774,806
5	Oasia Hotel Novena	Novena, Singapore	4.7	2,133,089
6	Oasia Resort Sentosa by Far East Hospitality	Sentosa Island, Singapore	4.8	2,543,492
7	Potato Head Suites & Studios	Seminyak, Badung	4.8	4,041,163
8	The Daha Luxury Villas	Bali	4.7	2,808,000

Interpretasi → Terdapat 8 penginapan yang memiliki angka rating dan harga sesudahnya di atas rata-rata

## 10. Melakukan rank pada angka rating berdasarkan kategori harganya di seluruh penginapan-penginapan

```
select nama_hotel, alamat, kategori_harga , angka_rating, harga_sesudah_fix,
dense_rank() over (partition by kategori_harga order by angka_rating desc) as Ranked
from "Reservasi_Hotel" rh;
```

```
--# Melakukan rank pada angka rating berdasarkan kategori harganya di seluruh penginapan
select nama_hotel, alamat, kategori_harga , angka_rating, harga_sesudah_fix,
dense_rank() over (partition by kategori_harga order by angka_rating desc) as Ranked
from "Reservasi_Hotel" rh;
```

	A-Z nama_hotel	A-Z alamat	A-Z kategori	123 angka_rating	123 harga_sesudah_fix	123 ranked
1	Potato Head Suites & Studios	Seminyak, Badung	Harga Diskon	4.8	4,041,163	1
2	Oasia Resort Sentosa by Far E	Sentosa Island, Singapore	Harga Diskon	4.8	2,543,492	1
3	Oasia Hotel Downtown	Outram, Singapore	Harga Diskon	4.7	2,774,806	2
4	The Daha Luxury Villas	Bali	Harga Diskon	4.7	2,808,000	2
5	The Salila Beach Resort Semir	Seminyak, Badung	Harga Diskon	4.7	824,794	2
6	Episode Gading Serpong	Gading Serpong, Tangerang	Harga Diskon	4.7	1,136,446	2
7	Ramada Encore by Wyndham	Seminyak, Badung	Harga Diskon	4.5	821,255	3
8	Grand Whiz Hotel Nusa Dua	Nusa Dua, Badung	Harga Diskon	4.5	588,283	3
9	Padina Soho and Residence	Batuceper, Tangerang	Harga Diskon	4.5	360,312	3
10	Mutiara Bali Boutique Resort	Seminyak, Badung	Harga Diskon	4.4	901,879	4
11	Concorde Hotel Singapore	Orchard, Singapore	Harga Diskon	4.4	2,100,190	4
12	Village Hotel Sentosa by Far E	Sentosa Island, Singapore	Harga Diskon	4.4	2,611,582	4

Interpretasi → Untuk rating tertinggi pada 4,8 dengan kategori Harga Diskon pada penginapan Potato Head Suites & Studios di Seminyak, Badung memiliki harga Rp. 4.041.163 dengan peringkatnya urut 1. Sedangkan, yang kategori



---

Tidak Harga Diskon untuk rating tertinggi 4,9 pada penginapan Sanur Resort Watujimbar di Sanur, Denpasar memiliki harga Rp. 883.795 dengan peringkatnya urutan 1.

## CONCLUSION

Berdasarkan beberapa query sederhana yang telah dilakukan di atas untuk data reservasi hotel hasil scrapping website Tiket.com dapat kita temukan atau telaah informasi penting mengenai preferensi penginapan-penginapan mana yang paling murah berdasarkan harga, atau berdasarkan rating yang paling bagus, atau memiliki diskon yang tertinggi, atau berdasarkan total reviews terbanyak yang menunjukkan bahwa penginapan tersebut sering sekali diboeking oleh banyak konsumen atau menjadi top rekomendasi untuk booking/reservasi hotel.

Selain itu, data-data ini dapat digunakan analisis lebih lanjut seperti halnya melakukan pemetaan geografis dimana menampilkan distribusi hotel/villa berdasarkan lokasi menggunakan visualisasi geografis. Analisis terkait hubungan harga dengan rating atau membuat analisis terkait Menghitung rata-rata diskon berdasarkan kategori harga dan lokasi. Kemudian, juga bisa sistem membuat rekomendasi bisa berkaitan dengan harga, lokasi, atau rating, atau hal lainnya.

Harapannya, semakin banyak data yang discrapping agar mendapatkan informasi sebanyak-banyaknya. Hal ini, berguna untuk sebagai preferensi yang penting untuk rekomendasi atau analisis lanjutan yang bisa berguna untuk user ataupun perusahaan dalam research market, atau segmentasi pasarnya yang sesuai dengan reservasi hotel. Untuk konsumen, dapat mempermudah dalam memilih reservasi yang tepat dan sesuai dengan minatnya.