--select all table ds_salaries
limit 10;

123 id	123 work_year	A-Z experience_level	A-Z employment_type	A-Z job_title
0	2,020	MI	FT	Data Scientist
1	2,020	SE	FT	Machine Learning Scie
2	2,020	SE	FT	Big Data Engineer
3	2,020	MI	FT	Product Data Analyst
4	2,020	SE	FT	Machine Learning Engi
5	2,020	EN	FT	Data Analyst
6	2,020	SE	FT	Lead Data Scientist
7	2,020	MI	FT	Data Scientist
8	2,020	MI	FT	Business Data Analyst

AS: memunculkan semua kolom di tabel ds salaries , tapi 10 baris pertama

--select distinct-as, ada 50 unik data judul. distinct untuk melihat keunikan
select distinct job_title
from ds_salaries *ds* ;

A-Z job_title
Machine Learning Infrastructure Engineer
Machine Learning Engineer
Head of Machine Learning
Director of Data Science
Data Scientist
Computer Vision Engineer
Lead Data Analyst
Data Engineering Manager
Big Data Architect
Data Analytics Lead

--select left/right hanya khusus varchar.knp ga pakai tanda kutip?karna buat mengambil keseluruhan data di job title.
select left(job_title,4)
from ds_salaries *ds* ;

A-Z left
Data
Mach
Big
Prod
Mach
Data
Lead
Data
Busi
Lead

AS: ini mengambil 4 huruf awl dlm kolom job title

select right(job_title,4), job_title
from ds_salaries *ds* ;

AS: mengambil 4 angka terakhir job title

--substring
```
select substring(job_title,6,9), job_title
from ds_salaries ds ;
```


as:mengambil urutan ke 6 sampai 9 dari kolom job title

--replace
```
select replace(job_title,'Machine','data'), job_title
from ds_salaries ds ;
```


as: mengganti setiap kata machine menjadi data.

-- CTE : Berapa Total Salary dari setiap job title, dimana employee residence = US, total salary > 500.000

```
with total_salary as (select job_title, sum(salary) as totalsalary, employee_residence
                            from ds_salaries ds
                            where employee_residence = 'US'
                            group by job_title, employee_residence)
select * from total_salary --Memanggil table cte
where totalsalary > 500000;
```

	A-Z job_title	123 totalsalary	A-Z employee_residence
1	Machine Learning Scientist	995,300	US
2	Analytics Engineer	700,000	US
3	Data Science Manager	1,751,183	US
4	Applied Data Scientist	714,000	US
5	Data Analyst	7,564,865	US
6	Data Analytics Manager	889,940	US
7	Head of Data Science	501,875	US
8	ML Engineer	526,000	US
9	Financial Data Analyst	550,000	US
10	Data Architect	1,672,512	US

as: membuat tabel banyakan dengan CTE yaitu total salary. dengan tujuan ingin menampilkan jenis job title beserta total salarynya dengan kondis total salary>500.000, dan tinggal di US. yang stabilo kuning bisa ada bisa tidak, karna itu hanya menampilkan kolom employee residence aja, agar menunjukkan itu benar terfilter hanya US saja. kalau ada aggregat sum, harus ad grup by.

-- RANK : Melihat Ranking Job Title Berdasarkan Total Salary Tertinggi, namun mempertimbangkan Experience Level nya
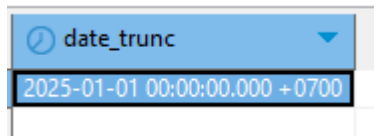
```
select sum(salary_in_usd) as total_salary, job_title, experience_level,
       rank() over (partition by experience_level order by sum(salary_in_usd) desc) as ranked
from ds_salaries ds
group by job_title , experience_level;
```

123 total_salary	A-Z job_title	A-Z experience_level	123 ranked
1,217,280	Data Scientist	EN	1
782,967	Machine Learning Engineer	EN	2
707,202	Data Engineer	EN	3
647,528	Data Analyst	EN	4
344,528	Research Scientist	EN	5
313,204	Data Science Consultant	EN	6
225,000	Machine Learning Scientist	EN	7
220,000	Computer Vision Software Engineer	EN	8
163,609	Computer Vision Engineer	EN	9
150,102	Business Data Analyst	EN	10

as: Melihat Ranking Job Title Berdasarkan Total Salary Tertinggi, namun mempertimbangkan Experience Level nya. Jadi pada level entry level (EN), ranking 1 gaji tertinggi adalah data scientist.

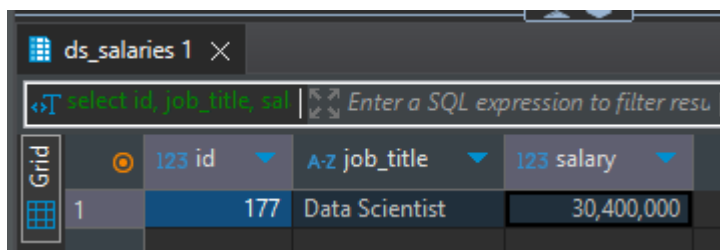--date trunc--memangkas tanggal dan waktu

```
select current_timestamp;
select date_trunc('month', current_timestamp);
```



```
date_trunc
2025-01-01 00:00:00.000 +0700
```

as: kalau dibuat month, maka menunjukkan month saat ini, tapi tanggal, jam, menit detiknya berubah jadi default.

Soal: Tampilkan id, job title, salary dari karyawan dengan gaji tertinggi.

```
select id, job_title, salary
from ds_salaries ds
where salary=(select max(salary) from ds_salaries ds2);
```


id	job_title	salary
177	Data Scientist	30,400,000

Soal: Tampilkan id, job title, salary dari karyawan dengan gaji tertinggi ke terendah masing2 job title.

```
select job_title, max (salary) as highest_salary
from ds_salaries ds
group by job_title;
```

job_title	highest_salary ↓
Data Scientist	30,400,000
BI Data Analyst	11,000,000
ML Engineer	8,500,000
Data Science Manager	7,000,000
Head of Machine Learning	6,000,000
Machine Learning Engineer	4,900,000
Data Engineer	4,450,000
Lead Data Scientist	3,000,000
Big Data Engineer	1,672,000
Lead Data Analyst	1,450,000
Business Data Analyst	1,400,000
AI Scientist	1,335,000
Principal Data Engineer	600,000
Product Data Analyst	450,000
Data Analyst	450,000
Financial Data Analyst	450,000
Research Scientist	450,000
Data Science Consultant	423,000
Applied Machine Learning Scientist	423,000
Principal Data Scientist	416.000

Soal: buat tabel job title, jumlah total employee berdasarkan job title.

```sql
select job_title, count(*) as total_employees
from ds_salaries ds
group by job_title ;
```

job_title	total_employees
Machine Learning Infrastructure Engineer	3
Machine Learning Engineer	41
Head of Machine Learning	1
Director of Data Science	7
Data Scientist	143
Computer Vision Engineer	6
Lead Data Analyst	3
Data Engineering Manager	5
Big Data Architect	1
Data Analytics Lead	1
Machine Learning Scientist	8

Soal: Hitung **total kumulatif gaji** dari semua karyawan berdasarkan urutan gaji dari yang terkecil ke terbesar.

```sql
select id, salary, sum (salary) over (order by salary) as cumulative_salary
from ds_salaries ds ;
```

Soal:

Tampilkan job_title dan avg_salary dari job title yang memiliki rata-rata gaji lebih dari 600,000.

```sql
select job_title, avg(salary) as avg_salary
from ds_salaries ds
group by job_title
having avg(salary)>600000;
```