

SESSIONS 1

Grouping Result Sets

Data Science Program

GROUP BY

- The GROUP BY statement groups rows that have the same values into summary rows, like "find the number of customers in each country".
- The GROUP BY statement is often used with aggregate functions (COUNT, MAX, MIN, SUM, AVG) to group the result-set by one or more columns.

```
SELECT column_name(s)
FROM table_name
WHERE condition
GROUP BY column_name(s)
ORDER BY column_name(s);
```

GROUP BY

- Show how many city name in every country code based on City ID

```
SELECT COUNT(ID), CountryCode  
FROM City  
GROUP BY CountryCode;
```

COUNT(ID)	CountryCode
1	ABW
4	AFG
5	AGO
2	AIA
1	ALB

GROUP BY & WHERE

- Show average city population in every district in Indonesia (IDN)

```
SELECT AVG(Population), District
FROM CITY
WHERE CountryCode = 'IDN'
GROUP BY District;
```

AVG(Population)	District
9604900.0000	Jakarta Raya
377868.3846	East Java
321531.4400	West Java
373713.8571	Sumatera Utara
673382.0000	Sumatera Selatan
286595.7273	Central Java
1060257.0000	Sulawesi Selatan
680332.0000	Lampung
534474.0000	Sumatera Barat
482931.0000	Kalimantan Selatan
206803.0000	Riau
435000.0000	Bali
262872.0000	Yogyakarta
409632.0000	Kalimantan Barat
368963.5000	Kalimantan Timur
385201.0000	Jambi
213173.0000	Sulawesi Utara
306600.0000	Nusa Tenggara Barat
249312.0000	Molukit
146439.0000	Bengkulu
126504.5000	Aceh
142800.0000	Sulawesi Tengah
129300.0000	Nusa Tenggara Timur
99693.0000	Kalimantan Tengah
94800.0000	Sulawesi Tenggara
94700.0000	West Irian

GROUP BY, WHERE, & AS

- Show average city population in every district in Indonesia (IDN).
- AS is Alias syntax to give new column name in result set.

```
SELECT AVG(Population) AS Rata_rata,  
District AS Provinsi  
FROM CITY  
WHERE CountryCode = 'IDN'  
GROUP BY District;
```

Rata_rata	Provinsi
9604900.0000	Jakarta Raya
377868.3846	East Java
321531.4400	West Java
373713.8571	Sumatera Utara
673382.0000	Sumatera Selatan
286595.7273	Central Java
1060257.0000	Sulawesi Selatan
680332.0000	Lampung
534474.0000	Sumatera Barat
482931.0000	Kalimantan Selatan
206803.0000	Riau
435000.0000	Bali
262872.0000	Yogyakarta
409632.0000	Kalimantan Barat
368963.5000	Kalimantan Timur
385201.0000	Jambi
213173.0000	Sulawesi Utara
306600.0000	Nusa Tenggara Barat
249312.0000	Molukit
146439.0000	Bengkulu
126504.5000	Aceh
142800.0000	Sulawesi Tengah
129300.0000	Nusa Tenggara Timur
99693.0000	Kalimantan Tengah
94800.0000	Sulawesi Tenggara
94700.0000	West Irian

GROUP BY, WHERE, AS, & HAVING

- Show average city population in every district in Indonesia (IDN), but only showing the average greater than 500.000.
- **HAVING** is similar with WHERE which is syntax to filter result set. But, HAVING is used after grouping (GROUP BY).

```
SELECT AVG(Population) AS Rata_rata,  
District AS Provinsi  
FROM CITY  
WHERE CountryCode = 'IDN'  
GROUP BY District  
HAVING Rata_rata > 500000;
```

Rata_rata	Provinsi
9604900.0000	Jakarta Raya
673382.0000	Sumatera Selatan
1060257.0000	Sulawesi Selatan
680332.0000	Lampung
534474.0000	Sumatera Barat

GROUP BY, WHERE, AS, & HAVING

- Show average city population in every district in Indonesia (IDN), but only showing the 'Provinsi' which is started with 'K'.

```
SELECT AVG(Population) AS Rata_rata,  
District AS Provinsi  
FROM CITY  
WHERE CountryCode = 'IDN'  
GROUP BY District  
HAVING Provinsi LIKE 'K%';
```

Rata_rata	Provinsi
482931.0000	Kalimantan Selatan
409632.0000	Kalimantan Barat
368963.5000	Kalimantan Timur
99693.0000	Kalimantan Tengah

Reference

- https://www.w3schools.com/sql/sql_groupby.asp