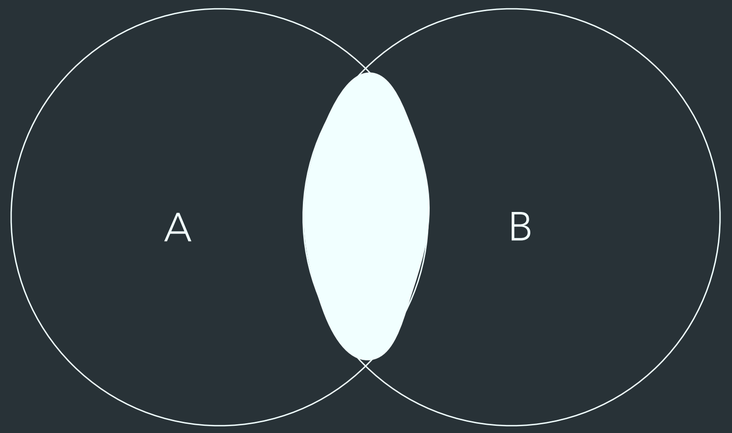
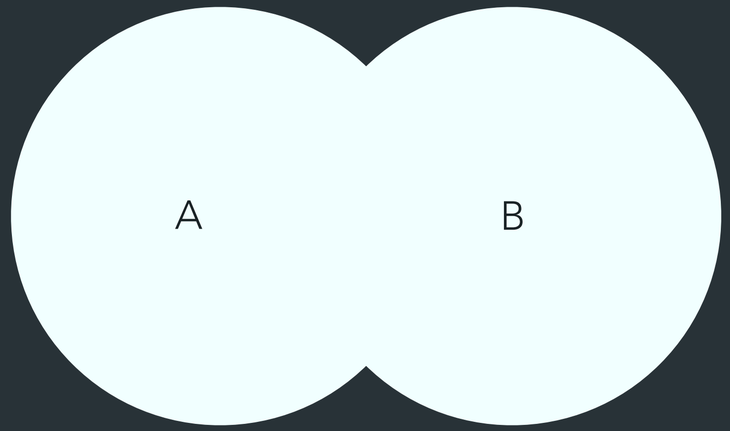
kokchun giang

working with **sets** in sql to manipulate and compare multiple queries

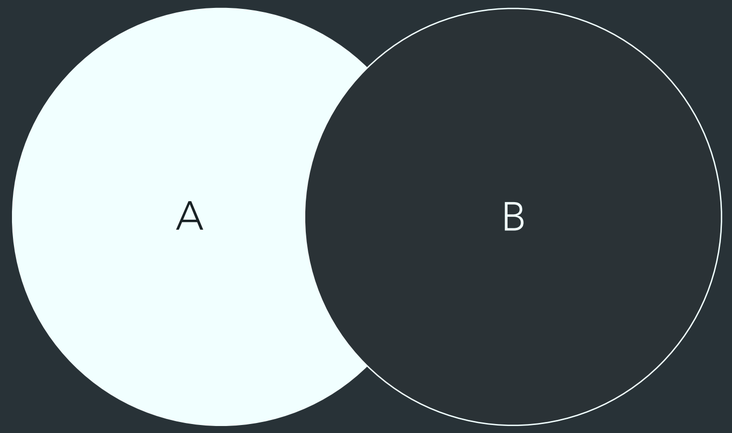
set theory explained with **venn diagrams**



A intersect B



A union B



A except B

A and B corresponds to
result sets from queries

conditions for set operations:

1. same number of columns
2. corresponding data types

combine rows from two result sets and removing duplicates with **union**

```
SELECT * FROM synthetic.sales_jan
UNION
SELECT * FROM synthetic.sales_feb;
```

id int32	product_name varchar	amount decimal(10,2)	sale_date date
3	Monitor	300.00	2024-02-10
4	Keyboard	100.00	2024-02-11
5	Mouse	50.00	2024-02-18

sales_feb

id int32	product_name varchar	amount decimal(10,2)	sale_date date
1	Laptop	1200.50	2024-01-05
2	Headphones	200.00	2024-01-12
3	Monitor	300.00	2024-01-15

sales_jan

UNION

id int32	product_name varchar	amount decimal(10,2)	sale_date date
4	Keyboard	100.00	2024-02-11
2	Headphones	200.00	2024-01-12
3	Monitor	300.00	2024-01-15
5	Mouse	50.00	2024-02-18
1	Laptop	1200.50	2024-01-05
3	Monitor	300.00	2024-02-10

combine rows from two result sets without removing duplicates with **union all**

```sql
SELECT product_name, amount FROM synthetic.sales_jan
UNION ALL
SELECT product_name, amount FROM synthetic.sales_feb;
```

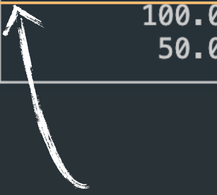
id int32	product_name varchar	amount decimal(10,2)	sale_date date
3	Monitor	300.00	2024-02-10
4	Keyboard	100.00	2024-02-11
5	Mouse	50.00	2024-02-18

sales_feb

id int32	product_name varchar	amount decimal(10,2)	sale_date date
1	Laptop	1200.50	2024-01-05
2	Headphones	200.00	2024-01-12
3	Monitor	300.00	2024-01-15

sales_jan

product_name varchar	amount decimal(10,2)
Laptop	1200.50
Headphones	200.00
Monitor	300.00
Monitor	300.00
Keyboard	100.00
Mouse	50.00



gets common rows from two result sets with **intersection**

```
SELECT product_name, amount FROM synthetic.sales_jan
INTERSECT
SELECT product_name, amount FROM synthetic.sales_feb;
```

id int32	product_name varchar	amount decimal(10,2)	sale_date date
3	Monitor	300.00	2024-02-10
4	Keyboard	100.00	2024-02-11
5	Mouse	50.00	2024-02-18

sales_feb

id int32	product_name varchar	amount decimal(10,2)	sale_date date
1	Laptop	1200.50	2024-01-05
2	Headphones	200.00	2024-01-12
3	Monitor	300.00	2024-01-15

sales_jan

product_name varchar	amount decimal(10,2)
Monitor	300.00

removing rows that are in the other result set from the first one using **except**

```
SELECT product_name, amount FROM synthetic.sales_jan
EXCEPT
SELECT product_name, amount FROM synthetic.sales_feb;
```

id int32	product_name varchar	amount decimal(10,2)	sale_date date
3	Monitor	300.00	2024-02-10
4	Keyboard	100.00	2024-02-11
5	Mouse	50.00	2024-02-18

sales_feb

id int32	product_name varchar	amount decimal(10,2)	sale_date date
1	Laptop	1200.50	2024-01-05
2	Headphones	200.00	2024-01-12
3	Monitor	300.00	2024-01-15

sales_jan

product_name varchar	amount decimal(10,2)
Headphones	200.00
Laptop	1200.50