

Aggregation in SQL

Aggregate Operators:

- COUNT
- SUM
- AVG
- MAX
- MIN

```
SELECT    AVG(age)
FROM      Person
```

```
SELECT    SUM(age)
FROM      Person
```

```
SELECT    MIN(age)
FROM      Person
```

Note: Find the youngest one.

```
SELECT    MAX(age)
FROM      Person
WHERE     City = "Memphis"
```

Note: You always select tuple first, and then do aggregation.

Find the ssn of the youngest one.

```
SELECT    MIN(age), ssn
FROM      Person
```

Note:

- In the SELECT statement, MIN(age) is an aggregate value, and ssn is a none aggregate value, some systems will give error message.
- There is no WHERE statement, so, all tuples will be selected.

```
SELECT    ssn
FROM      Person
WHERE     age = MIN(age) ← Error
```

You must use nested query to solve it:

```
SELECT    ssn
FROM      Person
WHERE     age IN ( SELECT    MIN(age)
                   FROM      Person
                   )
or
SELECT    ssn
FROM      Person
WHERE     age = ( SELECT    MIN(age)
                  FROM      Person
                  )
or
SELECT    ssn
FROM      Person
WHERE     age <= ALL ( SELECT    age
                       FROM      Person
                       )
```

Think about the meaning of the following query:

```
SELECT    ssn
FROM      Person
WHERE     age < ANY ( SELECT    age
                       FROM      Person
                       )
```

We will get anyone but the oldest one.

“GROUP BY” statement:

Find the average age of each city.

```
SELECT    City, AVG(age)
FROM      Person
GROUP BY  City
```

Note:

- If you have GROUP BY statement, you can SELECT on two different kinds of attributes. For the above example, it will print name of each city with average age.

- You can “group by” more than 1 attribute.

```
SELECT      City, State, SUM(income)
FROM        Person
WHERE       age > 30
GROUP BY   City, State
```

Note: The order of operations is:

- Take the tuple first. (i.e. execute WHERE statement first.)
- Group them. (i.e. execute GROUP BY statement next.)
- Do projection on attributes. (i.e. execute SELECT statement last.)

“HAVING” statement:

(It is for **picking up groups** instead of tuple.)

Find total income of people who live in the same city of the state with at least 20000 population.

```
SELECT      City, State, SUM(income)
FROM        Person P
WHERE       age > 30
GROUP BY   City, State
HAVING      20000 < (SELECT COUNT(*)
                     FROM    Person P1
                     WHERE   P1.State = P.State)
```

Note:

The sub query

```
SELECT COUNT(*)
FROM    Person P1
WHERE   P1.State = P.State
```

Note: **It counts everyone.**

will be evaluated by group and all age people are counted since this sub query is independent from the WHERE statement of original query. However, **the sum of income will be calculated by all age >30.**

The order of operations:

- Do nested part first.
- Pick up tuples.
- Group them.
- Project on attributes.