

# Aggregate Queries

## TABLE OF CONTENTS

1. Aggregate Functions
2. Group By
3. Having
4. Having v/s Where



Notes

15<sup>th</sup> Hard day challenge :

1. Assignments + Revision
2. Backlog (Assignments of prev. session)
3. Additional Questions



**So far we have discussed queries like :**

1. Find students who...
2. Find batches who...
3. Find name of every student...

**But now we will be answering queries like :**

1. What is avg salary of the employees?
2. Count of students in each batch.
3. What is maximum psp of students?



# Aggregate Functions

To find answer for aggregate functions we need to combine multiple row data.

1. Count

2. Max

3. Min

4. Avg

5. Sum

**NOTE** : Distinct isn't an aggregate function.



# Count

## Students

id	name	b_id	psp
1	John	Null	80
2	Jane	1	90
3	Jim	2	85
4	Jenny	3	95
5	Jack	2	78

- Only takes Not Null values.

**< Question > :** Find count of b\_id in students table.

- Count function takes count of only Not Null values

```
select count(b_id)
from students ;
```

Count of b\_id



&lt;/&gt; Pseudo-code

```
table = [...]  
  
count = 0  
  
for row in table:  
    if row[b-id]:  
        count ++  
  
print(count)
```

&lt; Question &gt; : Will

1. Avg (psp) and
2. sum (psp) / count ( \* ) yield same result?

**NOTE :** Nesting isn't allowed for aggregates.



# Group By

So far we wrote queries for questions like -

**Find :**

1. Avg (psp) of students.
2. Min (psp) of students.

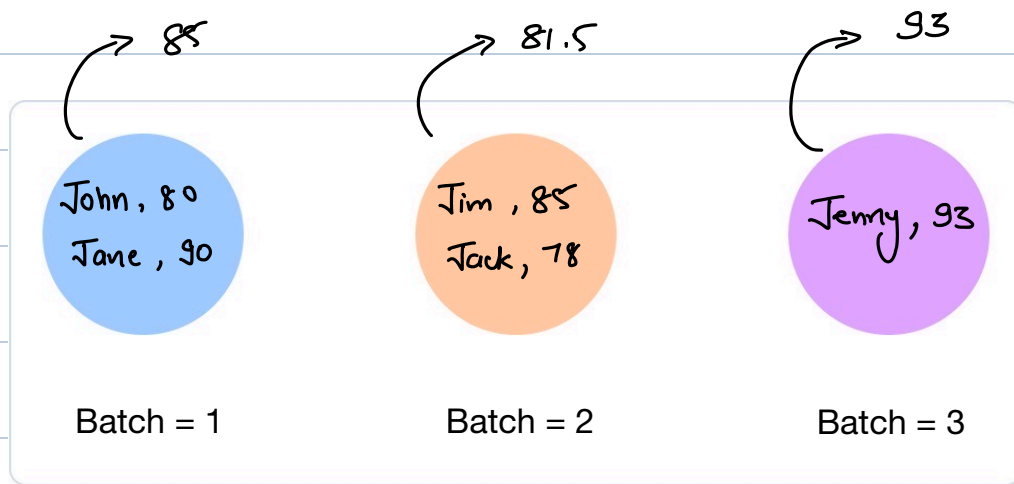
**What if I want to find avg (psp) of every batch?**

**Students**

id	name	b_id	psp
1	John	1	80
2	Jane	1	90
3	Jim	2	85
4	Jenny	3	95
5	Jack	2	78

**Batches**

b_id	name
1	A
2	B
3	C
4	D



```
select avg(psp)
from students
group by batch-id;
```

\* We should use only those col<sup>m</sup> in select clause which are common to a group.



**< Question > :** Get avg (psp) of every batch along with their batch names.

• **How to approach these kind of questions?**

1. From which table do we need data.
2. Then apply joins accordingly if required.
3. Filtering, grouping data.





# Having

- Get avg (psp) of every batch where avg (psp) > 85.

**'Where'** - On rows (row by row filtering)

Can't filter groups.

**'Where'** - Can't be applied once groups are formed.

**'Having'** - Will be used to filter groups.



# Sequence

1. Get data using from and joins from table.
2. Then we can group our data.
3. Filter data based on 'Having Clause' condition.
4. Order By.
5. Print our data.

*group by*  
`from / joins → where → group by → having → select`