Group By

Announcement:

1) Assignment

1) Clear on Alias, use "num" or num

      If you want to use an alias with space, use [], not quotes,

e.g. [number of players]

2) Desc, Asc, which is default?

1. Student Table

StudentID	Name	Department	Final
s10001	Victor	MIS	88
s10002	Miranda	Accounting	85
s10003	Joey	Accounting	91
s10004	Heather	Marketing	90
s10005	Tim	Management	89
s10006	Joel	Management	88
s10007	Bobby	Marketing	86
s10008	Lynn	Finance	86
s10009	Jess	Marketing	81
s10010	Kenneth	Finance	92
s10011	David	Finance	84
s10012	Martin	MIS	88
s10013	Andrew	MIS	84
s10014	Amanda	MIS	86
s10015	Lyle	Finance	89
s10016	Lillian	Accounting	85
s10017	Kevin	Management	85

StudentID	Name	Department	Final	
s10002	Miranda	Accounting	85	Group1
s10003	Joey	Accounting	91	
s10016	Lillian	Accounting	85	
s10008	Lynn	Finance	86	Group2
s10010	Kenneth	Finance	92	
s10011	David	Finance	84	
s10015	Lyle	Finance	89	
s10005	Tim	Management	89	Group3
s10006	Joel	Management	88	
s10017	Kevin	Management	85	
s10004	Heather	Marketing	90	Group4
s10007	Bobby	Marketing	86	
s10009	Jess	Marketing	81	
s10001	Victor	MIS	88	Group5
s10012	Martin	MIS	88	
s10013	Andrew	MIS	84	
s10014	Amanda	MIS	86	

Up till now, we have learned how to calculate the average, count, max, min of a specific group with aggregate function.

Practice:

Find the average final score for Finance Department.

In practice, usually we need to compare the aggregate results for all the groups.

For example, we want to know the average score for each department

department	average grade
Accounting	87
Finance	87.75
Management	87.33333333
Marketing	85.66666667
MIS	86.5

More examples:

1) Find the number of students in each department

2) Find the highest score in each department

SELECT *column_name(s)*
FROM *table_name*
GROUP BY *column_name(s);*

Question:

Select studentName, count(*) from gradebook group by department;

What is wrong here? The variable "name" is an "individual level", but your table after using "group by" becomes a "group-level" (i.e. department level) data, therefore the structure is not consistent.

department	count(*)	name
Accounting	3	???
Finance	4	???
Management	3	???
Marketing	3	???
MIS	4	???

Notice

1) When using group by, the columns you select can only be two cases:

      a. An aggregate function within a group

      b. The column that you used for group by

2) Group by actually creates a new table.

Filter after group by:

department	count(*)
Accounting	3
Finance	4
Management	3
Marketing	3
MIS	4

What if we only want departments with more than 3 students?

1) The HAVING clause was added to SQL because the WHERE keyword could not be used with aggregate functions.

StudentID	Name	Department	Final
s10001	Victor	MIS	88
s10002	Miranda	Accounting	85
s10003	Joey	Accounting	91
s10004	Heather	Marketing	90
s10005	Tim	Management	89
s10006	Joel	Management	88
s10007	Bobby	Marketing	86
s10008	Lynn	Finance	86
s10009	Jess	Marketing	81
s10010	Kenneth	Finance	92
s10011	David	Finance	84
s10012	Martin	MIS	88
s10013	Andrew	MIS	84
s10014	Amanda	MIS	86
s10015	Lyle	Finance	89
s10016	Lillian	Accounting	85
s10017	Kevin	Management	85

Select TOP 3 department, count(*) FROM gradebook WHERE final>=85 GROUP BY department HAVING count(*)>2 ORDER BY department DESC;

How does this query run? (https://stackoverflow.com/questions/1130062/what-is-the-execution-sequence-of-group-by-having-and-where-clause-in-sql-serve)

1) From, which table?

2) Where, SQL will first run the "where" condition, remove the rows which do not comply with the condition

3) Group by, generate a temporary group-level table

4) Having, remove the group-level rows which do not comply with the condition

5) Order by

6) Top

Notice:

1) Where is a filter for the original table

2) Having is a filter for the results of the aggregate function, it must be used with group by

3) You cannot use alias in order by or having condition

Group by

1 Find the average height of players from university of Connecticut

2 Find the average height of players from each university

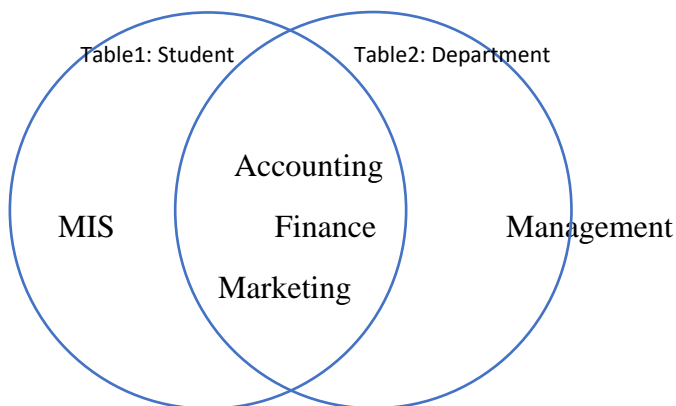>   a. Group all the players by their college

>   b. calculate the average of the height within each group
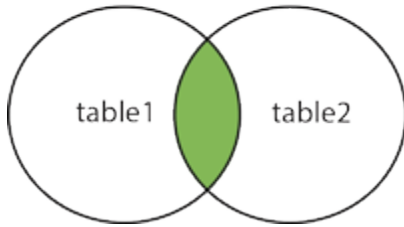
3 Find the number of players in each college

select college,count(*) as num from players group by college having count(*)>10 order by count(*) ;

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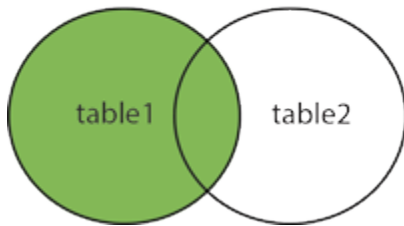ORDER BY *column_name(s);*

Join

Table1: Student   Table2: Department

Accounting

MIS   Finance   Management

Marketing

INNER JOIN

table1   table2

LEFT JOIN

table1   table2

RIGHT JOIN

table1   table2