

STUDY GUIDE

USE AGGREGATE FUNCTIONS TO SUMMARIZE AND COMPARE DATA

Aggregate Data

Aggregate data refers to statistical summaries of large amounts of information, potentially from multiple sources (i.e., different columns/rows). It is especially useful when you have a large database to analyze, and/or a common recurring query to perform on said database. It can be used to:

- Compile reports.
- Identify and/or predict trends.
- Perform a statistical analysis.
- Track certain metrics over time.

Aggregate functions

Aggregate functions compute against a returned column of numeric data from your `SELECT` statement to create a summary.

- The syntax is: `SELECT AggregateFunction (Column_ name) FROM table_ name;`

- **Sum:** Adds together the value for each row of a specified column and is used to add together the values of a numeric column.
- **Average:** Returns the average of all of the values in a given column and is used to determine a value against which to compare the rest of your data set.
- **Minimum:** Returns the minimum value in a given column and is used to identify the lower limit of a data set and/or outliers.
- **Maximum:** Returns the maximum value in a given column and is used to identify the upper limit of a data set and/or outliers.
- **Count:** Counts the number of rows returned by a query.
- **Count(*):** Counts the number of null and non-null rows returned by a query.
- **Count distinct:** Counts the number of distinct rows returned by a query and is used to determine the number of unique occurrences of something.

GROUP BY

- `GROUP BY` groups all of the rows that meet certain criteria, performs the designated aggregate function on that group and returns the value.
- The syntax is: `SELECT AggregateFunction (column_ name), column_ name _ 2 FROM table_ name GROUP BY column_ name _ 2`

HAVING

`HAVING` performs the same function as `WHERE` but should only be used with aggregate functions.