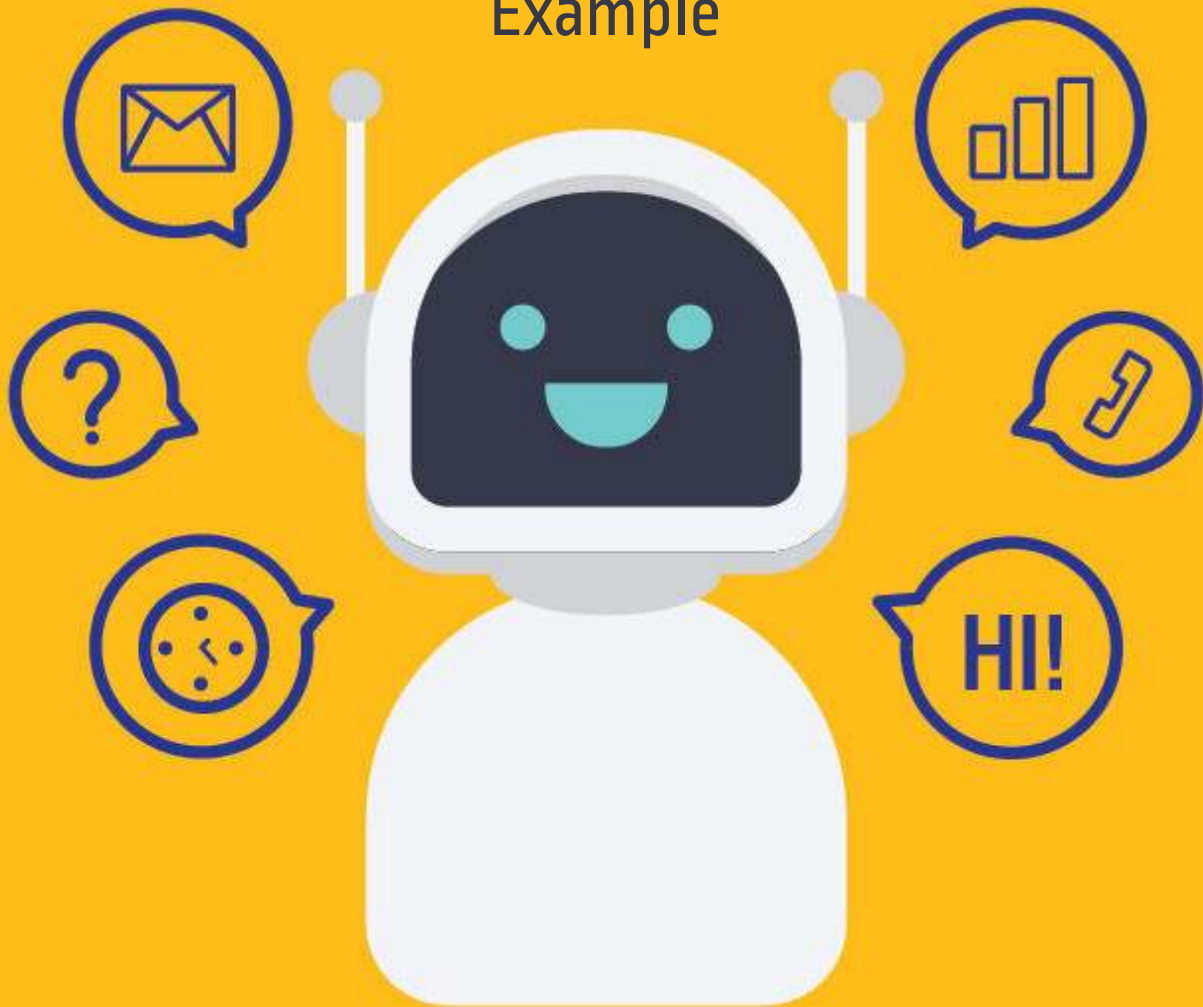


SQL for Citizen Data Scientists

SQL SELECT DISTINCT Statement - Learn By Example



SETScholars & WACAMLDS

As you know SELECT returns all matched rows. But what if you don't want every occurrence of every value? The solution is to use the DISTINCT keyword.

The DISTINCT keyword, as its name implies, instructs the database to only return distinct values.

Syntax

```
SELECT DISTINCT column_name(s)
FROM table_name;
```

Sample Table

To help you better understand the examples, and enable you to follow along with the tutorial, we are going to use the following sample table.

This table is part of an 'Employee Management System' that contains basic information about employees.

ID	Name	Age	City	Job	Salary
1	Bob	28	New York	Manager	60000
2	Eve	24	New York	Developer	32000
3	Max	26	New York	Janitor	9000
4	Kim	25	Chicago	Manager	55000
5	Joe	23	Chicago	Developer	30000
6	Sam	27	Chicago	Janitor	10000

SELECT DISTINCT Values

In some cases, a query might return duplicate rows of data. For example, if you were to retrieve the list of jobs, you would see the following:

```
SELECT Job  
FROM Employees;
```

Job

Manager

Developer

Janitor

Manager

Developer

Janitor

Here the SELECT statement returned 6 rows (even though there are only three unique jobs) because there are 6 employees listed. What you probably want in this case is the distinct job titles.

To achieve this, add the DISTINCT keyword just after the SELECT keyword.

```
SELECT DISTINCT Job  
FROM Employees;
```

Job
Developer
Janitor
Manager

DISTINCT on Multiple Columns

You can also get distinct results for more than one column. For example, if you need the distinct 'City' and 'Job', just include both of those columns in the SELECT statement.

When you specify multiple columns in the SELECT list, the DISTINCT clause displays all rows where the combination of columns is distinct.

```
SELECT DISTINCT City, Job  
FROM Employees;
```

City	Job
Chicago	Developer
Chicago	Janitor
Chicago	Manager
New York	Developer
New York	Janitor
New York	Manager

Note that the DISTINCT can be used only once in a SELECT list, and it must precede all column names in that list. Otherwise SQL will raise a syntax error.

```
SELECT City, DISTINCT Job  
FROM Employees;  
--Incorrect syntax near the keyword 'DISTINCT'.
```

DISTINCT in Aggregate Functions

You can also use the DISTINCT keyword to eliminate duplicate values in aggregate function calculations.

The following example uses the `COUNT()` function to return the count of unique jobs. Here the DISTINCT keyword makes sure that the COUNT only takes into account unique jobs.

```
SELECT COUNT(DISTINCT Job)  
FROM Employees;
```