196. Delete Duplicate Emails

Easy | Topics | Companies

SQL Schema > Pandas Schema >

Table: `Person`

```
+-------------+---------+
| Column Name | Type    |
+-------------+---------+
| id          | int     |
| email       | varchar |
+-------------+---------+
id is the primary key (column with unique values) for this table.
Each row of this table contains an email. The emails will not contain
uppercase letters.
```

Write a solution to **delete** all duplicate emails, keeping only one unique email with the smallest `id`.

For SQL users, please note that you are supposed to write a `DELETE` statement and not a `SELECT` one.

For Pandas users, please note that you are supposed to modify `Person` in place.

After running your script, the answer shown is the `Person` table. The driver will first compile and run your piece of code and then show the `Person` table. The final order of the `Person` table **does not matter**.

```
The result format is in the following example.


Example 1:

  Input:
  Person table:
  +----+------------------+
  | id | email            |
  +----+------------------+
  | 1  | john@example.com |
  | 2  | bob@example.com  |
  | 3  | john@example.com |
  +----+------------------+
  Output:
  +----+------------------+
  | id | email            |
  +----+------------------+
  | 1  | john@example.com |
  | 2  | bob@example.com  |
  +----+------------------+
  Explanation: john@example.com is repeated two times. We keep the row with
  the smallest Id = 1.
```

MySQL:

```sql
# Write your MySQL query statement below
DELETE p1
FROM Person p1
JOIN Person p2
ON p1.email = p2.email
AND p1.id > p2.id;
```

Pandas:

```python
import pandas as pd

def delete_duplicate_emails(person: pd.DataFrame) -> None:
    # Sort by id so that the smallest id for each email comes first
    person.sort_values(by="id", inplace=True)
    # Drop duplicates based on email, keeping the first (smallest id)
    person.drop_duplicates(subset="email", keep="first", inplace=True)
```

PostgreSQL:

```
-- Write your PostgreSQL query statement below
DELETE FROM Person
WHERE id NOT IN (
    SELECT MIN(id)
    FROM Person
    GROUP BY email
);
```