

Problem 2882: Drop Duplicate Rows

Problem Information

Difficulty: Easy

Acceptance Rate: 0.00%

Paid Only: No

Problem Description

DataFrame customers +-----+-----+ | Column Name | Type | +-----+-----+ |
customer_id | int | | name | object | | email | object | +-----+-----+ |

There are some duplicate rows in the DataFrame based on the

email

column.

Write a solution to remove these duplicate rows and keep only the

first

occurrence.

The result format is in the following example.

Example 1:

Input:

```
+-----+-----+-----+ | customer_id | name | email |
+-----+-----+-----+ | 1 | Ella | emily@example.com | | 2 | David |
michael@example.com | | 3 | Zachary | sarah@example.com | | 4 | Alice | john@example.com
| | 5 | Finn | john@example.com | | 6 | Violet | alice@example.com |
+-----+-----+-----+
```

Output:

```
+-----+-----+-----+ | customer_id | name | email |  
+-----+-----+-----+ | 1 | Ella | emily@example.com | | 2 | David |  
michael@example.com | | 3 | Zachary | sarah@example.com | | 4 | Alice | john@example.com  
| | 6 | Violet | alice@example.com | +-----+-----+-----+
```

Explanation:

Alic (customer_id = 4) and Finn (customer_id = 5) both use john@example.com, so only the first occurrence of this email is retained.

Code Snippets

Pandas:

```
import pandas as pd  
  
def dropDuplicateEmails(customers: pd.DataFrame) -> pd.DataFrame:
```

Solutions

Pandas Solution:

```
import pandas as pd  
  
def dropDuplicateEmails(customers: pd.DataFrame) -> pd.DataFrame:
```