DESCRIPTION

Table: Students

+---------------+---------+

| Column Name | Type |

+---------------+---------+

| student_id | int |

| student_name | varchar |

+---------------+---------+

student_id is the primary key (column with unique values) for this table.

Each row of this table contains the ID and the name of one student in the school.

Table: Subjects

+--------------+---------+

| Column Name | Type |

+--------------+---------+

| subject_name | varchar |

+--------------+---------+

subject_name is the primary key (column with unique values) for this table.

Each row of this table contains the name of one subject in the school.

Table: Examinations

+--------------+---------+

| Column Name | Type |

+--------------+---------+

| student_id | int |

| subject_name | varchar |

+--------------+---------+

There is no primary key (column with unique values) for this table. It may contain duplicates.

Each student from the Students table takes every course from the Subjects table.

Each row of this table indicates that a student with ID student_id attended the exam of subject_name.

Write a solution to find the number of times each student attended each exam.

Return the result table ordered by student_id and subject_name.

The result format is in the following example.

Example 1:

Input:

Students table:

+------------+--------------+

| student_id | student_name |

+------------+--------------+

| 1 | Alice |

| 2 | Bob |

| 13 | John |

| 6 | Alex |

+------------+--------------+

Subjects table:

+--------------+

| subject_name |

+--------------+

| Math |

| Physics |

| Programming |

+--------------+

Examinations table:

+------------+--------------+

| student_id | subject_name |

+------------+--------------+

| 1 | Math |

| 1 | Physics |

| 1 | Programming |

| 2 | Programming |

| 1 | Physics |

| 1 | Math |

| 13 | Math |

| 13 | Programming |

| 13 | Physics |

| 2 | Math |

| 1 | Math |

+------------+--------------+

Output:

+------------+--------------+--------------+----------------+

| student_id | student_name | subject_name | attended_exams |

+------------+--------------+--------------+----------------+

| 1 | Alice | Math | 3 |

| 1 | Alice | Physics | 2 |

| 1 | Alice | Programming | 1 |

| 2 | Bob | Math | 1 |

| 2 | Bob | Physics | 0 |

| 2 | Bob | Programming | 1 |

| 6 | Alex | Math | 0 |

| 6 | Alex | Physics | 0 |

| 6 | Alex | Programming | 0 |

| 13 | John | Math | 1 |

| 13 | John | Physics | 1 |

| 13 | John | Programming | 1 |

+------------+--------------+--------------+----------------+

Explanation:

The result table should contain all students and all subjects.

Alice attended the Math exam 3 times, the Physics exam 2 times, and the Programming exam 1 time.

Bob attended the Math exam 1 time, the Programming exam 1 time, and did not attend the Physics exam.

Alex did not attend any exams.

John attended the Math exam 1 time, the Physics exam 1 time, and the Programming exam 1 time.

SOLUTION

Option 1:

- Join students and subjects tables using merge() and 'cross' join method
- Assign a dataframe from examinations using groupby and value_counts() to get 'attended_exams' count and reset index using reset_index()
- Join df1 and exam_counts using merge() and 'left' join method and fill null with zero on 'attdended_exams' column
- Return the result in ordered by 'student_id' and 'subject_name' using sort_values()

```
import pandas as pd

def students_and_examinations(students: pd.DataFrame, subjects: pd.DataFrame,
examinations: pd.DataFrame) -> pd.DataFrame:
    df1 = pd.merge(students, subjects, how = 'cross')
    exam_counts =
pd.DataFrame(examinations.groupby('student_id').value_counts()).rename(columns={'coun
t': 'attended_exams'}).reset_index()
    df = pd.merge(df1, exam_counts, how = 'left', on = ['student_id',
'subject_name']).fillna({'attended_exams':0})
    return df.sort_values(by=['student_id', 'subject_name'])
```

- Snapshot of the same code above for readability purposes

```
import pandas as pd

def students_and_examinations(students: pd.DataFrame, subjects: pd.DataFrame, examinations: pd.DataFrame) -> pd.DataFrame:
    df1 = pd.merge(students, subjects, how = 'cross')
    exam_counts = pd.DataFrame(examinations.groupby('student_id').value_counts()).rename(columns={'count': 'attended_exams'}).reset_index()
    df = pd.merge(df1, exam_counts, how = 'left', on = ['student_id', 'subject_name']).fillna({'attended_exams':0})
    return df.sort_values(by=['student_id', 'subject_name'])
```