

## DESCRIPTION

Table: Project

+-----+-----+	
Column Name   Type	
+-----+-----+	
project_id   int	
employee_id   int	
+-----+-----+	

(project\_id, employee\_id) is the primary key of this table.

employee\_id is a foreign key to Employee table.

Each row of this table indicates that the employee with employee\_id is working on the project with project\_id.

Table: Employee

+-----+-----+	
Column Name   Type	
+-----+-----+	
employee_id   int	
name   varchar	
experience_years   int	
+-----+-----+	

employee\_id is the primary key of this table. It's guaranteed that experience\_years is not NULL.

Each row of this table contains information about one employee.

Write an SQL query that reports the **average** experience years of all the employees for each project, **rounded to 2 digits**.

Return the result table in **any order**.

The query result format is in the following example.

**Example 1:****Input:**

Project table:

+-----+-----+		
project_id	employee_id	
+-----+-----+		
1	1	
1	2	
1	3	
2	1	
2	4	
+-----+-----+		

Employee table:

+-----+-----+-----+			
employee_id	name	experience_years	
+-----+-----+-----+			
1	Khaled	3	
2	Ali	2	
3	John	1	
4	Doe	2	
+-----+-----+-----+			

**Output:**

+-----+-----+		
project_id	average_years	
+-----+-----+		
1	2.00	
2	2.50	
+-----+-----+		

**Explanation:** The average experience years for the first project is  $(3 + 2 + 1) / 3 = 2.00$  and for the second project is  $(3 + 2) / 2 = 2.50$

## SOLUTION

Option 1:

- Join dataframes
- Groupby the dataframe with 'project\_id' and compute 'average\_years' using mean()
- Round to 2 digits using round()

```
def project_employees_i(project: pd.DataFrame, employee: pd.DataFrame) -> pd.DataFrame:
    df = project.merge(employee, how='inner', on='employee_id')
    df = df.groupby('project_id')['experience_years'].mean().reset_index(name='average_years').round(2)
    return df
```

- Snapshot of the same code above for readability purposes

```
import pandas as pd
```

```
def project_employees_i(project: pd.DataFrame, employee: pd.DataFrame) -> pd.DataFrame:
    df = project.merge(employee, how='inner', on='employee_id')
    df = df.groupby('project_id')['experience_years'].mean().reset_index(name='average_years').round(2)
    return df
```

Option 2:

- Join dataframes
- Groupby the dataframe with 'project\_id' and compute 'average\_years' using agg()
- Round to 2 digits using round()

```
def project_employees_i(project: pd.DataFrame, employee: pd.DataFrame) -> pd.DataFrame:
    df = project.merge(employee, how='inner', on='employee_id')
    df = df.groupby('project_id').agg(average_years = ('experience_years', 'mean')).round(2).reset_index()
    return df
```

- Snapshot of the same code above for readability purposes

```
import pandas as pd
```

```
def project_employees_i(project: pd.DataFrame, employee: pd.DataFrame) -> pd.DataFrame:
    df = project.merge(employee, how='inner', on='employee_id')
    df = df.groupby('project_id').agg(average_years = ('experience_years', 'mean')).round(2).reset_index()
    return df
```