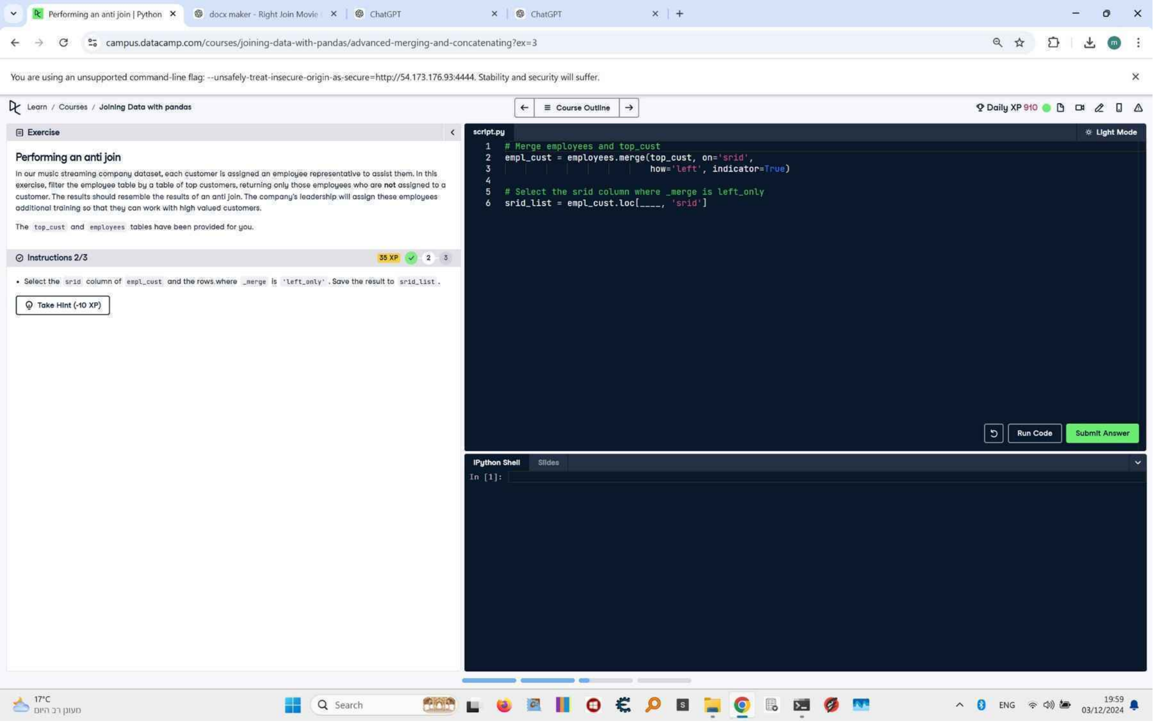
Filtering Rows in an Anti Join in Pandas


Performing an anti join | Python
docx maker - Right Join Movie
ChatGPT
ChatGPT
campus.datacamp.com/courses/joining-data-with-pandas/advanced-merging-and-concatenating?ex=3
Learn / Courses / Joining Data with pandas
Course Outline
Daily XP 910
Exercise
script.py
Light Mode
Run Code
Submit Answer
IPython Shell
Slides
In [1]:
Search
ENG
19:59
03/12/2024

You are using an unsupported command-line flag: --unsafely-treat-insecure-origin-as-secure=http://54.173.176.93:4444. Stability and security will suffer.

Performing an anti join

In our music streaming company dataset, each customer is assigned an employee representative to assist them. In this exercise, filter the employee table by a table of top customers, returning only those employees who are **not** assigned to a customer. The results should resemble the results of an anti join. The company's leadership will assign these employees additional training so that they can work with high valued customers.

The `top_cust` and `employees` tables have been provided for you.

Instructions 2/3 35 XP

- Select the `srid` column of `empl_cust` and the rows where `_merge` is `'left_only'`. Save the result to `srid_list`.

Take Hint (-10 XP)

```
# Merge employees and top_cust
empl_cust = employees.merge(top_cust, on='srid',
                            how='left', indicator=True)

# Select the srid column where _merge is left_only
srid_list = empl_cust.loc[____, 'srid']
```

Screenshot showing the exercise context for filtering rows in an anti join in pandas.

Code Answer:

Merge employees and top_cust with a left join, setting indicator to True
empl_cust = employees.merge(top_cust, on='srid', how='left', indicator=True)

Select the srid column where _merge is left_only
srid_list = empl_cust.loc[empl_cust['_merge'] == 'left_only', 'srid']

Explanation:

1. The `merge` function combines the 'employees' table and 'top_cust' table based on the 'srid' column. The `how='left'` parameter ensures that all rows from the 'employees' table are included, and the `indicator=True` parameter creates a new column, '_merge', indicating the source of each row.

2. The `loc` method is used to filter rows where the '_merge' column has the value 'left_only', which means these rows exist only in the 'employees' table and not in the 'top_cust' table.

3. The 'srid' column from these filtered rows is selected, resulting in a list of 'srid' values representing employees who are not assigned to any top customer.