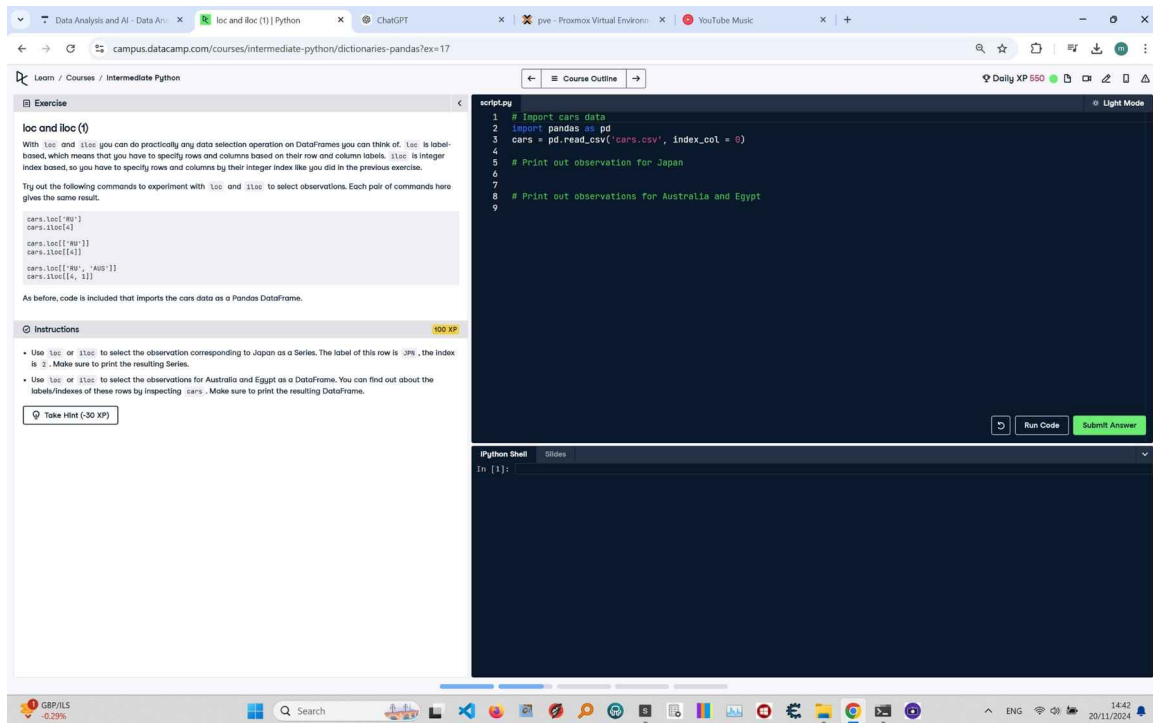


Your First Inner Join



The screenshot shows a web browser with multiple tabs. The active tab is 'loc and iloc (1) | Python' from 'campus.datacamp.com'. The page displays an exercise titled 'loc and iloc (1)' with instructions on how to use these methods for data selection. Below the instructions is a code editor with a Python script that reads a CSV file, prints out observations for Japan, and prints out observations for Australia and Egypt. The script is as follows:

```
1 # Import cars data
2 import pandas as pd
3 cars = pd.read_csv('cars.csv', index_col = 0)
4
5 # Print out observation for Japan
6
7
8 # Print out observations for Australia and Egypt
9
```

The page also includes a 'Python Shell' section with a prompt 'In [1]:'.

Question:

You have been tasked with figuring out what the most popular types of fuel used in Chicago taxis are. To complete the analysis, you need to merge the `taxi_owners` and `taxi_veh` tables on the `vid` column. You can then use the merged table along with the `value_counts()` method to find the most common `fuel_type`.

Instructions:

1. Merge `taxi_owners` with `taxi_veh` on the column `vid`, and save the result to `taxi_own_veh`.
2. Set the left and right table suffixes for overlapping columns of the merge to `_own` and `_veh`, respectively.
3. Select the `fuel_type` column from `taxi_own_veh` and print the `value_counts()` to find the most popular `fuel_type` used.

Answer:

```
# Instruction 1: Merge the taxi_owners and taxi_veh tables
# Example of merging two DataFrames on a common column
taxi_own_veh = taxi_owners.merge(taxi_veh, on='vid')
```

```
# Check the result by printing the column names
print(taxi_own_veh.columns)
```

```
# Instruction 2: Merge the taxi_owners and taxi_veh tables setting a suffix
taxi_own_veh = taxi_owners.merge(taxi_veh, on='vid', suffixes=('_own',
'_veh'))
```

```
# Print the column names of taxi_own_veh
print(taxi_own_veh.columns)
```

```
# Instruction 3: Find the most popular fuel_type
# Ensure taxi_own_veh is defined before accessing the fuel_type column
taxi_own_veh = taxi_owners.merge(taxi_veh, on='vid', suffixes=('_own',
'_veh'))
```

```
# Select the fuel_type column and compute value counts
fuel_counts = taxi_own_veh['fuel_type'].value_counts()
```

```
# Print the value counts to determine the most popular fuel type
print(fuel_counts)
```

Explanation of the Code:

1. ``taxi_own_veh = taxi_owners.merge(taxi_veh, on='vid')``: This line merges the ``taxi_owners`` and ``taxi_veh`` DataFrames on the common column ``vid``. It creates a new DataFrame containing combined data.

2. ``taxi_own_veh = taxi_owners.merge(taxi_veh, on='vid', suffixes=('_own', '_veh'))``: This line merges the tables with suffixes ``_own`` and ``_veh`` to distinguish overlapping columns in the two DataFrames. For example, if both tables had a ``name`` column, the suffixes would rename them as ``name_own`` and ``name_veh``. The columns of the resulting DataFrame are printed to verify the structure.

3. ``taxi_own_veh['fuel_type'].value_counts()``: This ensures that the ``taxi_own_veh`` variable is properly defined by merging the DataFrames before trying to access the ``fuel_type`` column. It calculates the frequency of each unique value using ``value_counts()``. It prints the frequency of each fuel type, helping identify the most popular one.

Learn / Courses / Joining Data with pandas

Exercise

Your first inner join

You have been tasked with figuring out what the most popular types of fuel used in Chicago taxis are. To complete the analysis, you need to merge the `taxi_sensors` and `taxi_veh` tables together on the `veh_id` column. You can then use the merged table along with the `value_counts()` method to find the most common `fuel_type`.

Since you'll be working with `pandas` throughout the course, the package will be preloaded for you as `pd` in each exercise in this course. Also the `taxi_sensors` and `taxi_veh` DataFrames are loaded for you.

Instructions 1/3

35 XP

1

Merge `taxi_sensors` with `taxi_veh` on the column `veh_id`, and save the result to `taxi_veh`.

Take Hint (-10 XP)

2

Set the left and right table suffixes for overlapping columns of the merge to `_sens` and `_veh`, respectively.

3

Select the `fuel_type` column from `taxi_veh` and print the `value_counts()` to find the most popular `fuel_type`'s used.

script.py

Light Mode

```
1 # Merge the taxi_sensors and taxi_veh tables
2 taxi_veh = taxi_sensors._____
3
4 # Print the column names of the taxi_veh
5 print(taxi_veh.columns)
```

Run Code Submit Answer

Python Shell

Slides

In []:

21°C
94°F

Search

ENG 13:01 02/12/2024