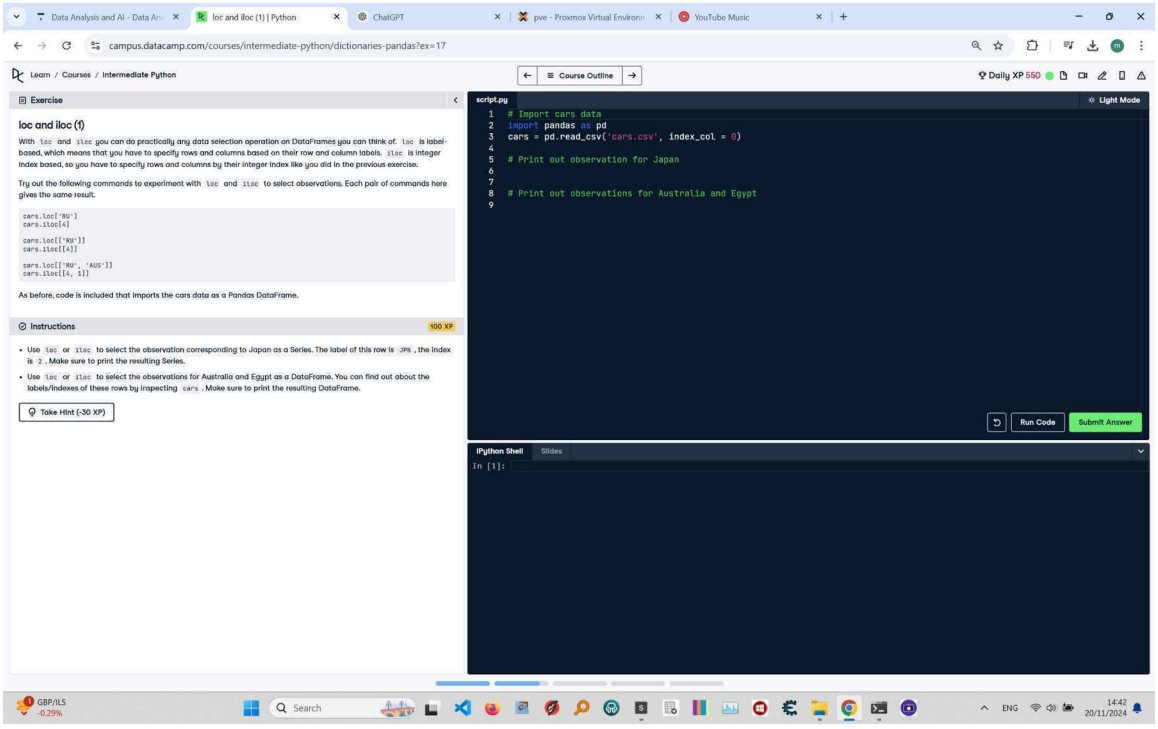
Your First Inner Join



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)
```

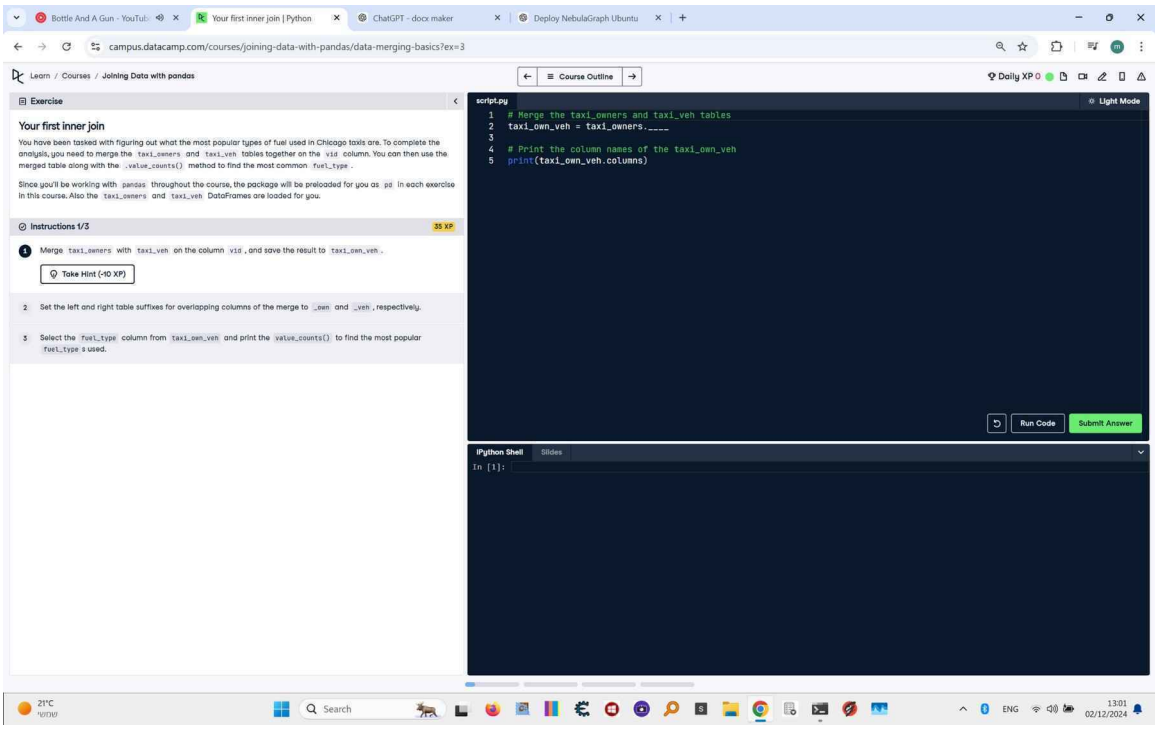
```
# Instruction 2: Merge the taxi_owners and taxi_veh tables setting a suffix
taxi_own_veh = taxi_owners.merge(taxi_veh, on='vid', suffixes=('_own',
'_veh'))

# 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.


Bottle And A Gun - YouTub
Your first inner join | Python
ChatGPT - docx maker
Deploy NebulaGraph Ubuntu
campus.datacamp.com/courses/joining-data-with-pandas/data-merging-basics?ex=3
Learn / Courses / Joining Data with pandas
Course Outline
Daily XP 0
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_owners and taxi_veh tables together on the vid 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_owners and taxi_veh DataFrames are loaded for you.
Instructions 1/3
35 XP
1 Merge taxi_owners with taxi_veh on the column vid , and save the result to taxi_own_veh .
Take Hint (-10 XP)
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 s used.
script.py
Light Mode
1 # Merge the taxi_owners and taxi_veh tables
2 taxi_own_veh = taxi_owners.____
3
4 # Print the column names of the taxi_own_veh
5 print(taxi_own_veh.columns)
Run Code
Submit Answer
IPython Shell
Slides
In [1]:
21°C
Search
ENG
13:01
02/12/2024