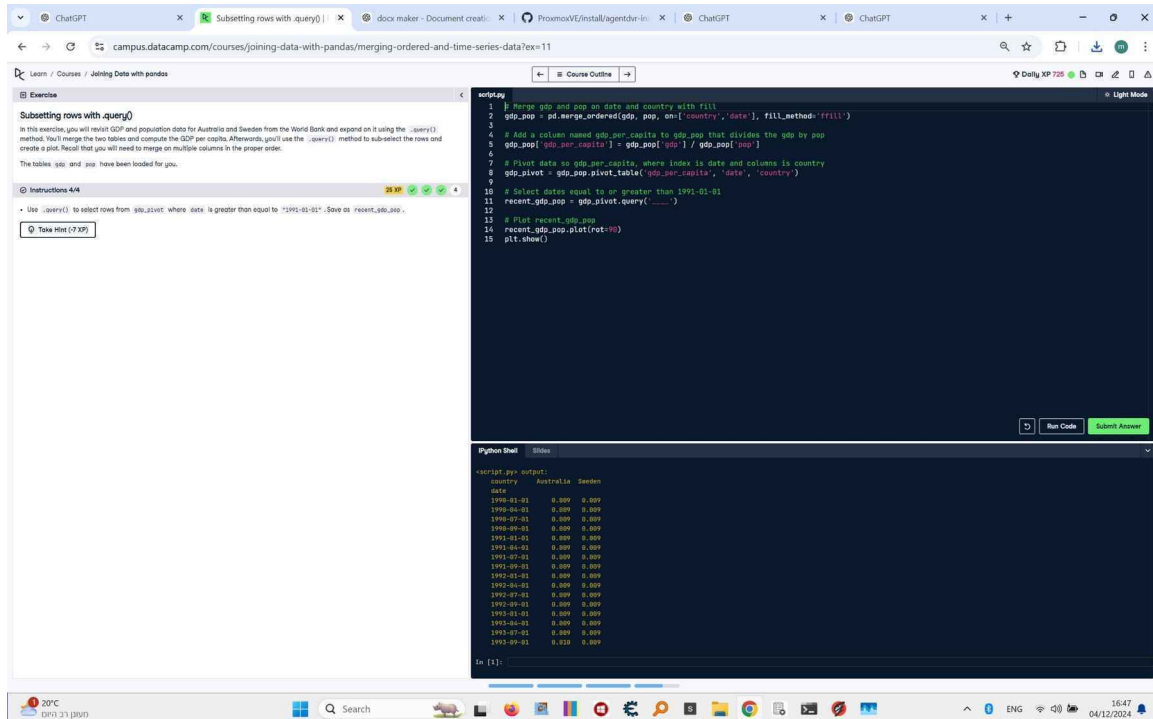


Subsetting rows with .query() - Step 4



The screenshot shows a Jupyter Notebook interface with a dark theme. The notebook is titled "Subsetting rows with .query()". The instructions section says: "In this exercise, you will merge GDP and population data for Australia and Sweden from the World Bank and expand on it using the .query() method. You'll merge the two tables and compute the GDP per capita. Afterwards, you'll use the .query() method to subset the rows and create a plot. Recall that you will need to merge on multiple columns in the proper order. The tables, gdp, and pop, have been loaded for you."

The code section contains the following Python code:

```
1 # Merge gdp and pop on date and country with fill
2 gdp_pop = pd.merge_ordered(gdp, pop, on=['country', 'date'], fill_method='ffill')
3
4 # Add a column named gdp_per_capita to gdp_pop that divides the gdp by pop
5 gdp_pop['gdp_per_capita'] = gdp_pop['gdp'] / gdp_pop['pop']
6
7 # Pivot data so gdp_per_capita, where index is date and columns is country
8 gdp_pivot = gdp_pop.pivot_table('gdp_per_capita', 'date', 'country')
9
10 # Select dates equal to or greater than 1991-01-01
11 recent_gdp_pop = gdp_pivot.query('date >= 1991-01-01')
12
13 # Plot recent_gdp_pop
14 recent_gdp_pop.plot(rot=90)
15 plt.show()
```

The output section shows the result of the .query() method, which is a DataFrame with columns 'date' and 'country'. The data is filtered for dates on or after 1991-01-01.

date	Australia	Sweden
1990-01-01	0.009	0.009
1990-04-01	0.009	0.009
1990-07-01	0.009	0.009
1990-10-01	0.009	0.009
1991-01-01	0.009	0.009
1991-04-01	0.009	0.009
1991-07-01	0.009	0.009
1991-10-01	0.009	0.009
1992-01-01	0.009	0.009
1992-04-01	0.009	0.009
1992-07-01	0.009	0.009
1992-10-01	0.009	0.009
1993-01-01	0.009	0.009
1993-04-01	0.009	0.009
1993-07-01	0.009	0.009
1993-10-01	0.011	0.009

Question:

Use .query() to select rows from gdp_pivot where date is greater than or equal to '1991-01-01'. Save as recent_gdp_pop and plot.

Answer:

```
# Merge gdp and pop on date and country with fill
gdp_pop = pd.merge_ordered(
    gdp, pop,
    on=['country', 'date'],
    fill_method='ffill'
)
```

```
# Add a column named gdp_per_capita to gdp_pop that divides the gdp by
pop
gdp_pop['gdp_per_capita'] = gdp_pop['gdp'] / gdp_pop['pop']
```

```
# Pivot table of gdp_per_capita, where index is date and columns is country
gdp_pivot = gdp_pop.pivot_table(
    values='gdp_per_capita',
    index='date',
    columns='country'
)
```

```
# Use .query() to select rows where date >= '1991-01-01'  
recent_gdp_pop = gdp_pivot.query('date >= "1991-01-01"')
```

```
# Plot recent_gdp_pop  
recent_gdp_pop.plot(  
    rot=90  
)  
plt.show()
```