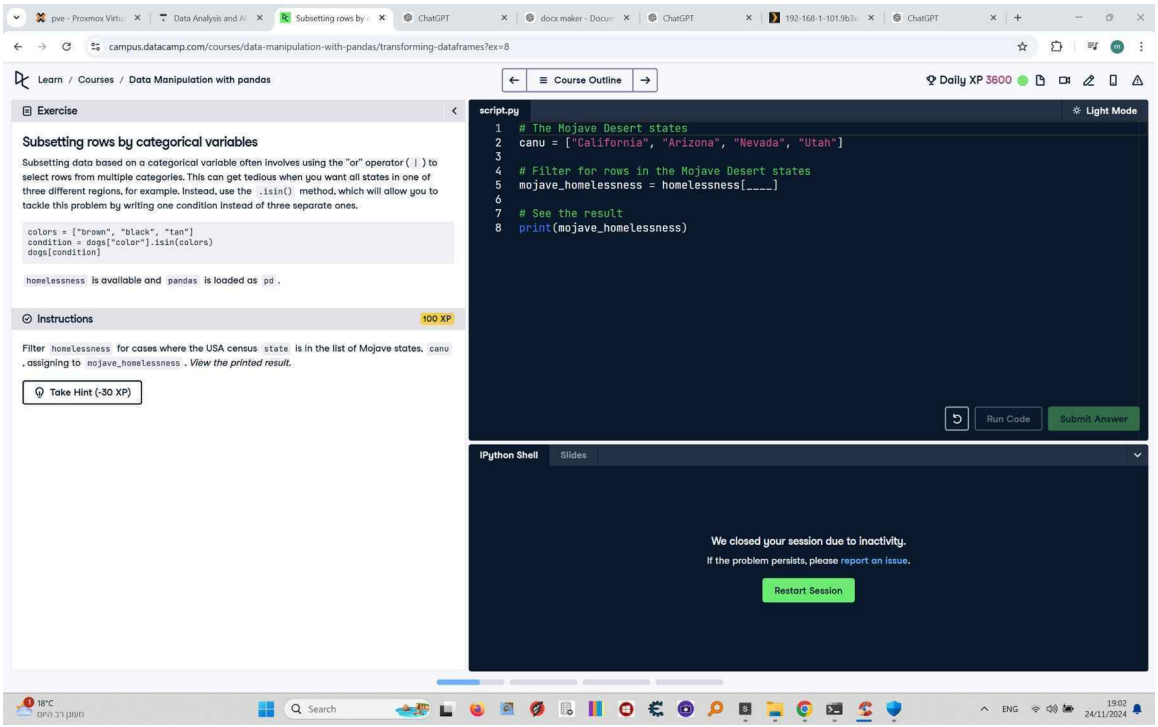
Subsetting Rows by Categorical Variables

This document includes the question, the solution, and a breakdown of the code provided in the screenshot.

Uploaded Screenshot

Below is the screenshot of the task:



Question

Filter `homelessness` for cases where the USA census `state` is in the list of Mojave states, `canu`, assigning the result to `mojave_homelessness`. View the printed result.

Answer

```
# The Mojave Desert states
canu = ["California", "Arizona", "Nevada", "Utah"]

# Filter for rows in the Mojave Desert states
mojave_homelessness = homelessness[homelessness["state"].isin(canu)]

# See the result
print(mojave_homelessness)
```

Code Explanation

Explanation of the code:

1. `canu = ["California", "Arizona", "Nevada", "Utah"]`: Creates a list of Mojave states to be used for filtering.

2. `homelessness["state"].isin(canu)`: Checks if each value in the `state` column is present in the `canu` list, creating a boolean mask.

3. `homelessness[homelessness["state"].isin(canu)]`: Filters the `homelessness` DataFrame to include only rows where the mask is `True`.

4. `mojave_homelessness`: Stores the filtered DataFrame for further use.

5. `print(mojave_homelessness)`: Prints the resulting DataFrame to verify the filter.