

Adding New Columns

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:

The screenshot shows a web browser window displaying a DataCamp exercise page. The page title is "Adding new columns". The instructions state: "You aren't stuck with just the data you are given. Instead, you can add new columns to a DataFrame. This has many names, such as *transforming*, *mutating*, and *feature engineering*. You can create new columns from scratch, but it is also common to derive them from other columns, for example, by adding columns together or by changing their units. homelessness is a DataFrame containing estimates of homelessness in each U.S. state in 2018. The individuals column is the number of homeless individuals not part of a family with children. The family_members column is the number of homeless individuals part of a family with children. The state_pop column is the state's total population. homelessness is available and pandas is loaded as pd."

The instructions section lists two tasks:

- Add a new column to homelessness, named total, containing the sum of the individuals and family_members columns.
- Add another column to homelessness, named p_homeless, containing the proportion of the total homeless population to the total population in each state: state_pop.

Below the instructions is a "Take Hint (-30 XP)" button.

On the right side of the page is a code editor window titled "script.py". It contains the following Python code:

```
1 # Add total col as sum of individuals and family_members
2 ----
3
4 # Add p_homeless col as proportion of total homeless population to the state population
5 ----
6
7 # See the result
8 print(homelessness)
```

Below the code editor is a "Run Code" button and a "Submit Answer" button.

At the bottom of the page is a "Python Shell" window with "In [1]:" and a blank space for output.

Question

Add a new column to `homelessness`, named `total`, containing the sum of the `individuals` and `family_members` columns. Add another column to `homelessness`, named `p_homeless`, containing the proportion of the `total` homeless population to the total population in each state, `state_pop`. View the printed result.

Answer

```
# Add total column as sum of individuals and family_members
homelessness['total'] = homelessness['individuals'] +
homelessness['family_members']
```

```
# Add p_homeless column as proportion of total homeless population to
state population
homelessness['p_homeless'] = homelessness['total'] /
homelessness['state_pop']
```

```
# See the result  
print(homelessness)
```

Code Explanation

Explanation of the code:

1. `homelessness['total'] = homelessness['individuals'] + homelessness['family_members']`: Adds a new column named `total` to the `homelessness` DataFrame. This column is calculated by summing the values in the `individuals` and `family_members` columns.
2. `homelessness['p_homeless'] = homelessness['total'] / homelessness['state_pop']`: Adds a new column named `p_homeless` to the `homelessness` DataFrame. This column is calculated as the proportion of the `total` homeless population to the state's total population (`state_pop`).
3. `print(homelessness)`: Prints the updated DataFrame to verify that the new columns have been added and calculated correctly.