

Functions for Initial Exploration of Unemployment Data

The screenshot displays a Jupyter Notebook interface. On the left, a sidebar shows the 'Exercise' tab selected. The main content area is titled 'Functions for initial exploration' and contains the following text: 'You are researching unemployment rates worldwide and have been given a new dataset to work with. The data has been saved and loaded for you as a pandas DataFrame called `unemployment`. You've never seen the data before, so your first task is to use a few pandas functions to learn about this new data. `pandas` has been imported for you as `pd`.'

Below the text, there are three instructions for the exercise:

1. Use a pandas function to print the first five rows of the `unemployment` DataFrame. (10 XP)
2. Use a pandas function to print a summary of column non-missing values and data types from the `unemployment` DataFrame.
3. Print the summary statistics (count, mean, standard deviation, min, max, and quartile values) of each numerical column in `unemployment`.

On the right, the code editor shows the following code:

```
1 # Print the first five rows of unemployment
2 print(_____)
```

At the bottom, the IPython Shell is visible, showing the prompt 'In [1]:'.

Question:

You are researching unemployment rates worldwide and have been given a new dataset to work with. The data has been opened and loaded for you as a pandas DataFrame called `unemployment`. Your first task is to use a few pandas functions to learn about this new data.

1. Print the first five rows of the `unemployment` DataFrame.
2. Use a pandas function to print a summary of column non-missing values

and data types from the `unemployment` DataFrame.

3. Print the summary statistics (count, mean, standard deviation, min, max, and quartile values) of each numerical column in `unemployment`.

Explanation of the Question:

This exercise focuses on performing initial exploratory data analysis using pandas functions. These steps help in understanding the structure and statistical distribution of the dataset.

Answer:

```
# Import pandas library
import pandas as pd

# Print the first five rows of unemployment DataFrame
print(unemployment.head())

# Print a summary of column non-missing values and data types
print(unemployment.info())

# Print summary statistics of numerical columns
print(unemployment.describe())
```

Explanation of the Answer:

1. The `head()` function displays the first five rows of the DataFrame to give an overview of the data structure.
2. The `info()` function provides details on the data types, non-missing values, and memory usage.
3. The `describe()` function calculates and displays key summary statistics for each numerical column.