

Named Aggregations

Question and Screenshot:

Exercise

Named aggregations

You've seen how `groupby()` and `agg()` can be combined to show summaries across categories. Sometimes, it's helpful to name new columns when aggregating so that it's clear in the code output what aggregations are being applied and where.

Your task is to create a DataFrame called `continent_summary` which shows a row for each continent. The DataFrame columns will contain the mean unemployment rate for each continent in 2021 as well as the standard deviation of the 2021 employment rate. And of course, you'll rename the columns so that their contents are clear!

The `unemployment` DataFrame is available, and `pandas` has been imported as `pd`.

Instructions 100 XP

- Create a column called `mean_rate_2021` which shows the mean 2021 unemployment rate for each continent.
- Create a column called `std_rate_2021` which shows the standard deviation of the 2021 unemployment rate for each continent.

[Take Hint \(-30 XP\)](#)

```
script.py
1 continent_summary = unemployment.groupby("continent").agg(
2     # Create the mean_rate_2021 column
3     # Create the std_rate_2021 column
4     # Create the std_rate_2021 column
5     # Create the std_rate_2021 column
6 )
7 print(continent_summary)
```

IPython Shell Slides

In [1]:

Question Explanation:

The task requires creating a summary table using pandas that aggregates data from a DataFrame grouped by continents. The goal is to calculate the mean and standard deviation of the 2021 unemployment rate for each continent.

Corrected Code Solution:

```
import pandas as pd
```

```
# Assuming 'unemployment' DataFrame is already available
continent_summary = unemployment.groupby("continent").agg(
    mean_rate_2021=('2021', 'mean'), # Create the mean_rate_2021 column
    std_rate_2021=('2021', 'std')    # Create the std_rate_2021 column
)
```

```
# Display the summary
print(continent_summary)
```

Solution Explanation:

The error occurs because the column `'unemployment_rate_2021'` does not exist in the DataFrame. Instead, the column name is simply `'2021'`. The corrected code replaces `'unemployment_rate_2021'` with `'2021'` in the

aggregation function. This ensures that the mean and standard deviation are calculated correctly for the unemployment rate in 2021 for each continent.