

Correlation between GDP and S&P500

The screenshot shows a web browser window with the URL `campus.datacamp.com/courses/joining-data-with-pandas/merging-ordered-and-time-series-data?ex=2`. The page is titled "Correlation between GDP and S&P500". The instructions state: "In this exercise, you want to analyze stock returns from the S&P 500. You believe there may be a relationship between the returns of the S&P 500 and the GDP of the US. Merge the different datasets together to compute the correlation. Two tables have been provided for you, named `sp500`, and `gdp`. As always, `pandas` has been imported for you as `pd`."

The code editor shows the following code:

```
1 # Use merge_ordered() to merge gdp and sp500, and forward fill missing values
2 gdp_sp500 = pd.merge_ordered(____)
3
4
5 # Print gdp_sp500
6 print(gdp_sp500)
```

The terminal output shows the following table:

	country	code	year	gdp	date	returns
0	USA	1.499e+13	2010.0			12.78
1	USA	1.354e+13	2011.0			0.00
2	USA	1.420e+13	2012.0			13.41
3	USA	1.420e+13	2012.0			13.41
4	USA	1.678e+13	2013.0			29.60
5	USA	1.732e+13	2014.0			11.39
6	USA	1.822e+13	2015.0			-0.73
7	USA	1.871e+13	2016.0			9.54
8	USA	1.949e+13	2017.0			19.42
9	USA	2.049e+13	NaN			NaN

Question:

Use the `merge_ordered()` function again, similar to before, to merge `gdp` and `sp500`, using the function's ability to fill in missing data for returns by forward-filling the missing values. Assign the resulting table to the variable `gdp_sp500`.

Answer:

```
# Import pandas as pd
import pandas as pd
```

```
# Assuming gdp and sp500 dataframes are already loaded
# Use merge_ordered() to merge gdp and sp500, forward filling missing values
gdp_sp500 = pd.merge_ordered(
    gdp, sp500,
    left_on='year',
    right_on='date',
    how='left',
    fill_method='ffill' # Forward-fill the missing values
)
```

```
# Print gdp_sp500
print(gdp_sp500)
```

Code Explanation:

1. `import pandas as pd:`

Import the pandas library, which is used for data manipulation and analysis.

2. `gdp_sp500 = pd.merge_ordered(...):`

This line merges the `gdp` and `sp500` dataframes using the `'merge_ordered'` function. The `'left_on'` and `'right_on'` parameters specify the columns to merge on (`'year'` in `gdp` and `'date'` in `sp500`). The `'how'` parameter is set to `'left'`, ensuring all rows from `gdp` are preserved. The `'fill_method='ffill'` argument is used to forward-fill missing values by propagating the last valid data point.

3. `print(gdp_sp500):`

This line displays the merged and processed dataframe.