Training Report Day-23

2 July 2024

Data visualization in Python is made easier with libraries like Matplotlib and Seaborn, which work well with Pandas. You can use Pandas' built-in plotting functions, or combine it with these libraries to create a wide range of visualizations directly from DataFrames.

Here's a guide on how to perform basic data visualizations using Pandas, along with some examples.

1. Basic Plotting with Pandas

Pandas provides some simple plotting capabilities through its plot() method, which is based on Matplotlib. Here's a look at some basic plot types.

Example Data

Let's start with a sample dataset:

```
import pandas as pd

import numpy as np



# Sample data

data = {

    'Category': ['A', 'B', 'C', 'D', 'E'],

    'Values1': [23, 45, 12, 67, 34],

    'Values2': [56, 34, 23, 12, 78]

}

df = pd.DataFrame(data)

print(df)
```

2. Plot Types in Pandas

2.1 Line Plot

A line plot is ideal for visualizing trends over time or sequential data.

df.plot(x='Category', y=['Values1', 'Values2'], kind='line', title='Line Plot')

3. Customizing Pandas Plots

Pandas plots can be customized in several ways using Matplotlib's parameters, allowing you to modify title, color, legend, and more.

Adding Labels and Customizing Style

import matplotlib.pyplot as plt

ax = df.plot(x='Category', y='Values1', kind='bar', color='skyblue', legend=False)

ax.set_xlabel('Categories')

ax.set_ylabel('Values')

ax.set_title('Customized Bar Plot')

plt.show()

4. Advanced Visualization Using Seaborn with Pandas

For more advanced visualizations, Seaborn offers statistical plotting options, making it easy to create heatmaps, pair plots, and other complex visualizations directly from Pandas data.

4.1 Heatmap

Heatmaps are useful for showing correlations in data.

```
import seaborn as sns

# Sample data

corr = df.corr()

sns.heatmap(corr, annot=True, cmap='coolwarm')

plt.title('Heatmap of Correlations')

plt.show()
```

5. Time Series Plotting

If you’re working with time series data, you can plot it easily using Pandas' time series plotting functionality.

```
# Sample time series data

dates = pd.date_range('20230101', periods=100)

df_time = pd.DataFrame(np.random.randn(100, 2), index=dates, columns=['A', 'B'])

df_time.plot(title='Time Series Plot')

plt.show()
```