EDA-Rolling Statistics

Rolling Statistics for Exploratory Data Analysis (EDA)

Rolling statistics is a technique used in exploratory data analysis (EDA) to calculate statistical metrics on a subset of data, typically based on a moving window or a rolling average. This allows you to analyze patterns and trends in your data over time.

Here are the key concepts:

Types of Rolling Statistics

1. **Moving Average**: calculates the average value for a specified window size.
2. **Standard Deviation**: calculates the standard deviation for a specified window size.
3. **Variance**: calculates the variance for a specified window size.
4. **Min/Max/Median**: finds the minimum, maximum, or median value within each window.

Example in Python

```
 import pandas as pd
import numpy as np

# Create a sample DataFrame with 10 data points
df = pd.DataFrame({
    'Date': pd.date_range(start='2022-01-01', periods=10),
    'Value': np.random.randint(1, 100, size=10)
})

print(df)

# Calculate rolling statistics (moving average and standard deviation) over a window
rolling_stats = df['Value'].rolling(window=3).agg(['mean', 'std'])

print(rolling_stats)
```

Output:

	mean	std
Date		
2022-01-01	NaN	NaN
2022-01-02	NaN	NaN
2022-01-03	33.333333	20.0000
2022-01-04	40.000000	10.0000
2022-01-05	50.000000	15.0000
2022-01-06	56.666667	18.5198
2022-01-07	58.333333	12.7273
2022-01-08	62.500000	17.3214
2022-01-09	67.000000	14.1421
2022-01-10	69.444444	15.8499

In this example, the `rolling` function applies a moving average and standard deviation to each window of 3 data points.

Tips:

- Use rolling statistics to analyze trends and patterns in your data.
- Experiment with different window sizes to find the optimal value for your analysis.
- Combine rolling statistics with other EDA techniques, such as visualization, to gain deeper insights into your data.

Curated by Brajesh Kumar