

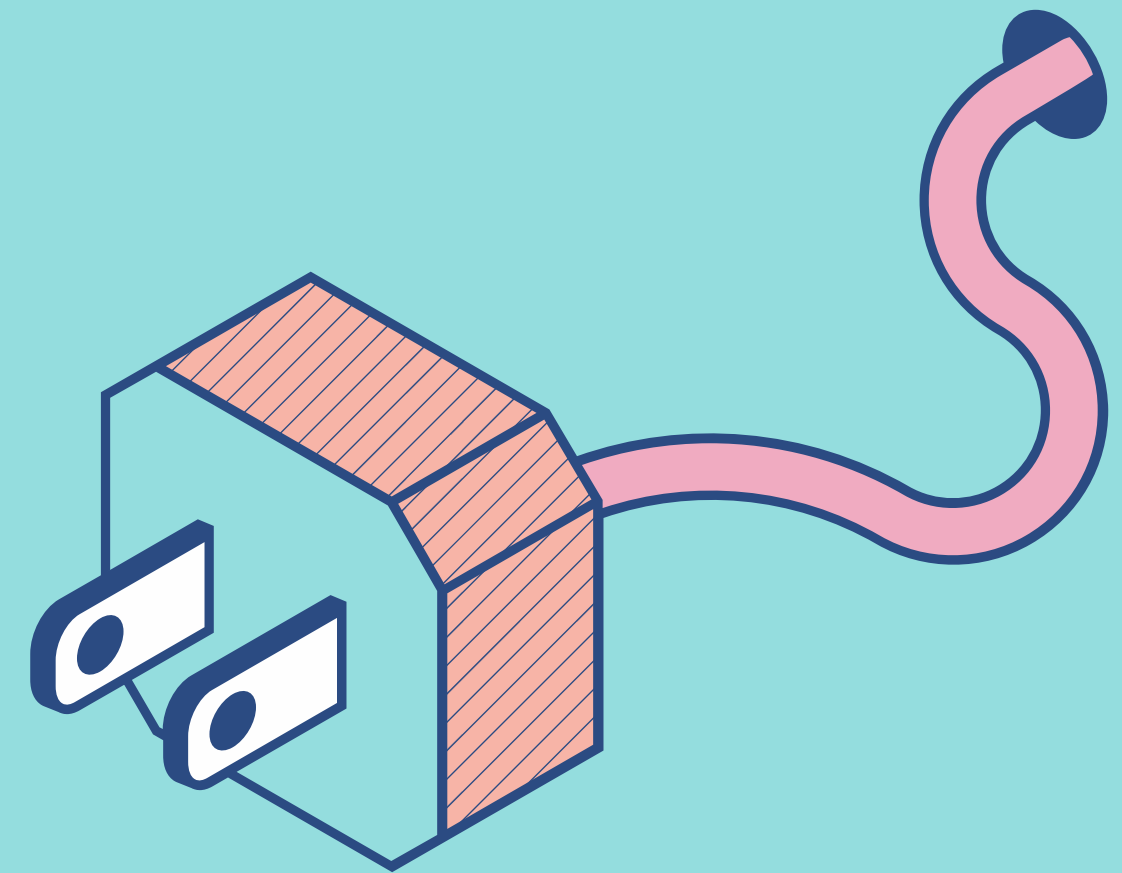


Time Series

- Tổng quan về dữ liệu chuỗi thời gian.
- Các thành phần của chuỗi thời gian: xu hướng, tính thời vụ và độ nhiễu.
- Phân rã chuỗi thời gian.
- Các khái niệm cơ bản trong mô hình chuỗi thời gian.
- Python và Pandas cho dữ liệu chuỗi thời gian.

Dữ liệu chuỗi thời gian là gì?

- Dữ liệu chuỗi thời gian là tập hợp các quan sát thu được thông qua các phép đo lặp lại theo thời gian. Về. các điểm biểu diễn trên đồ và một trong các trục sẽ luôn là thời gian.





Dữ liệu chuỗi thời gian ứng dụng vào đâu?

- Với dữ liệu chuỗi thời gian chúng ta có thể thấy rõ rệt các ứng dụng của nó ở ngay trong đời sống của chúng ta ví dụ như việc đo hoạt động điện trong não, đo lượng mưa, giá cổ phiếu... Với những ứng dụng của dữ liệu chuỗi thời gian có thể cho chúng ta thấy được tiềm năng phát triển của nó trong nhiều lĩnh vực khác nhau.

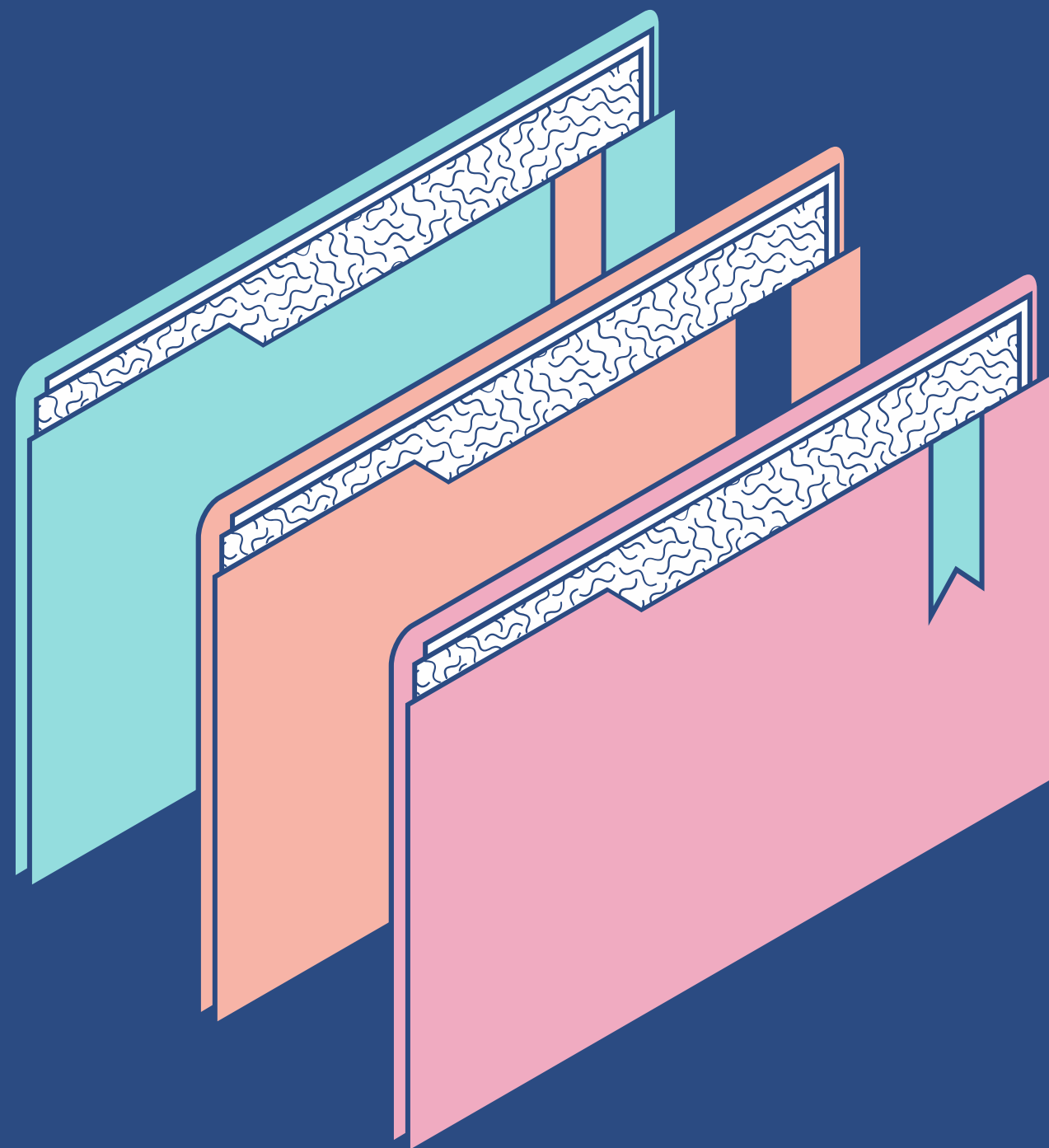
Dữ liệu chuỗi thời gian ảnh hưởng đến chứng khoán như thế nào?

- Với việc dùng dữ liệu chuỗi thời gian và các thông kế mà chứng khoán cho trước ta có thể hoàn toàn dự đoán gần đúng với giá trị của chứng khoán trong tương lai.
- Việc dữ liệu chuỗi thời hoàn toàn có thể được nhằm đến mục đích mở rộng các hướng đầu tư trong trứng khoán.

Thành phần của chuỗi thời gian



Mô hình chuỗi thời gian

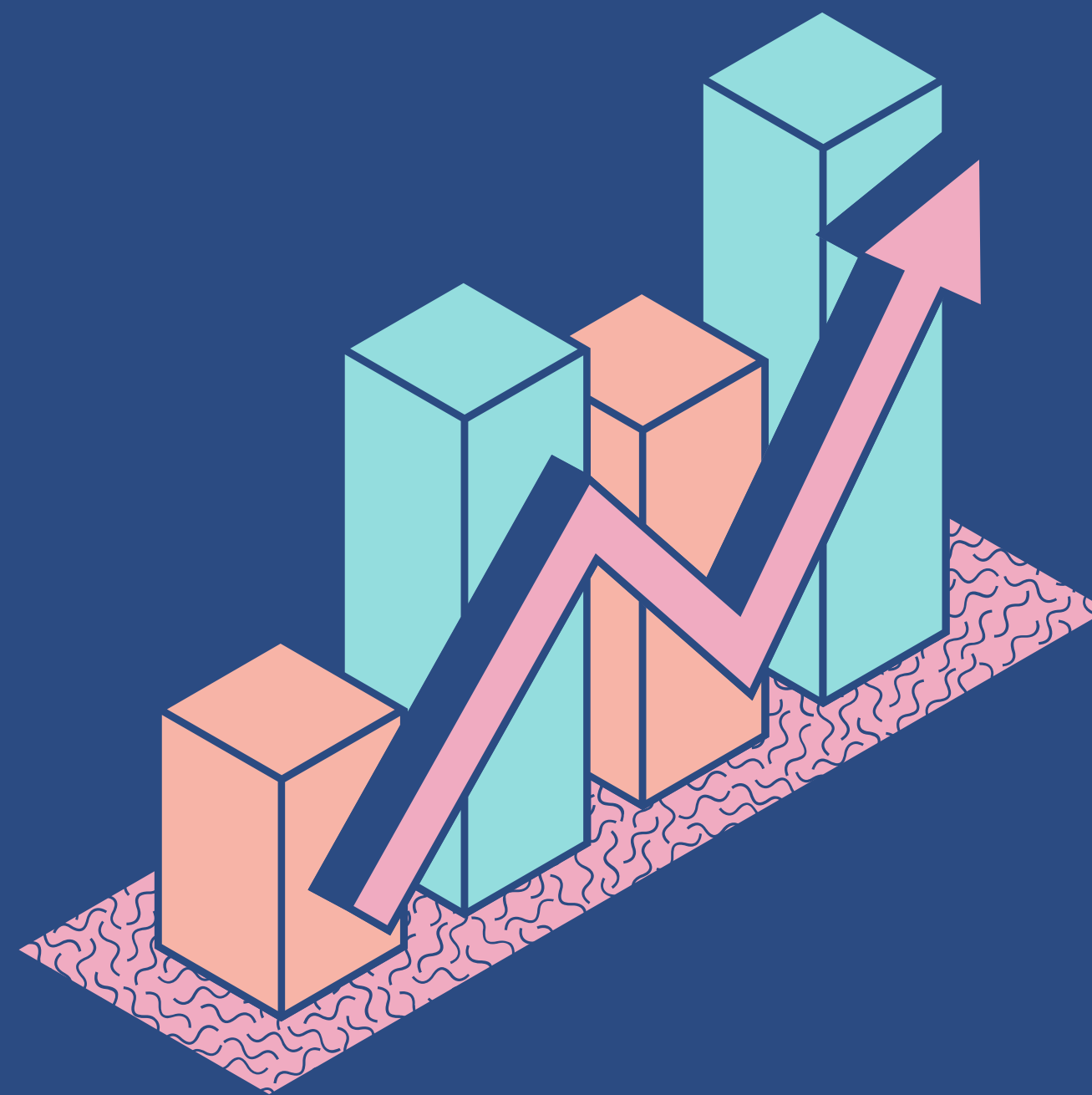


Một mô hình chuỗi thời gian được sử dụng vì rất nhiều lý do. Tuy nhiên, các kỹ thuật được sử dụng trong mô hình chuỗi thời gian phải tính đến mối tương quan chuỗi thời gian.

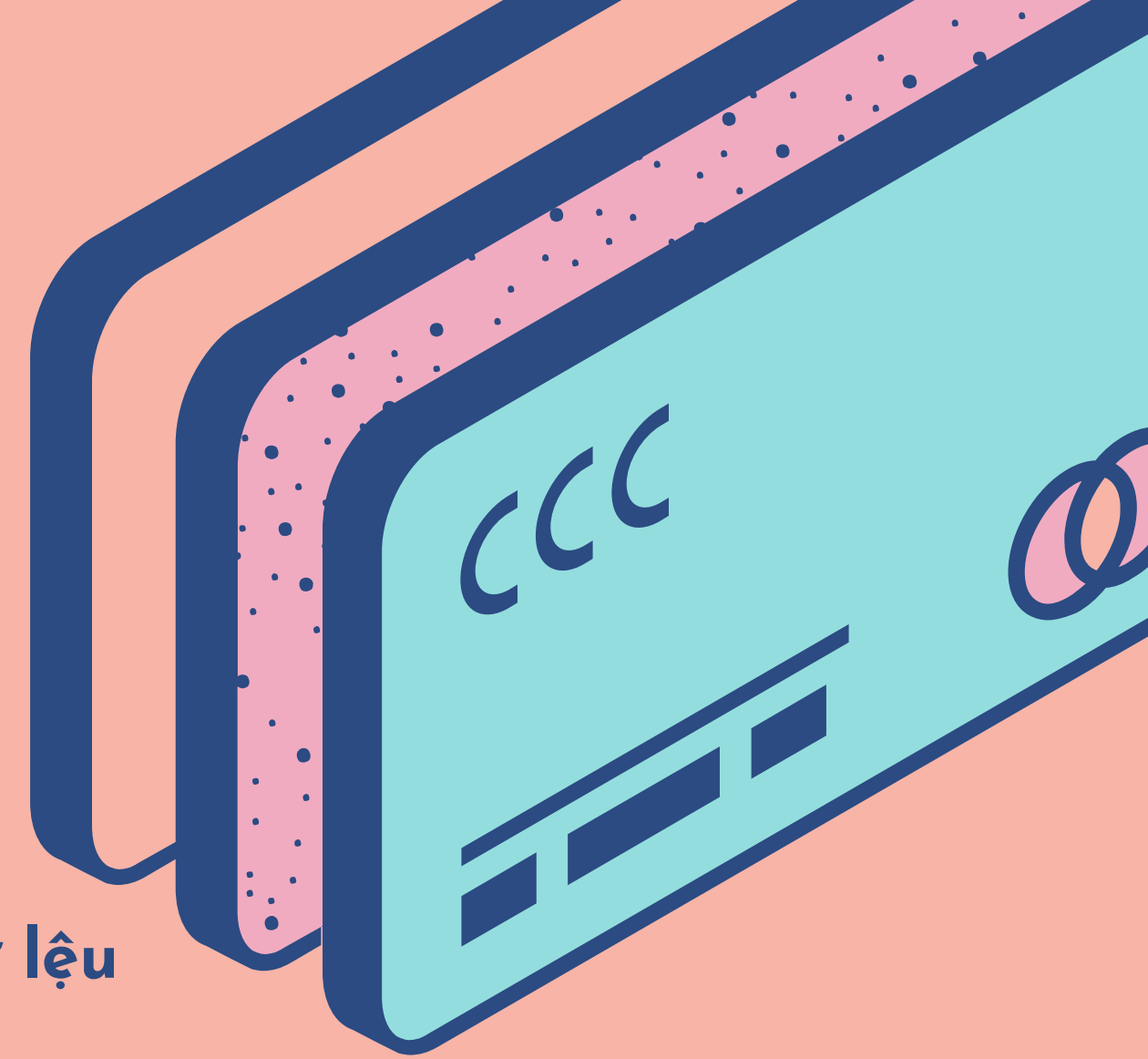
Trong mô hình chuỗi thời gian cơ bản sau sẽ giúp. ta hình dung rõ ràng hơn về một mô hình chuỗi thời gian hoàn chỉnh:

- Mô hình miền thời gian và miền tần số.
- Mô hình chuỗi thời gian đơn biến và đa biến.
- Mô hình chuỗi thời gian tuyến tính và phi tuyến tính.

Python và Pandas cho dữ liệu chuỗi thời gian



Pandas sử dụng trong dữ liệu chuỗi thời gian



Pandas chứa các khả năng và tính năng mở rộng để làm việc với dữ liệu chuỗi thời gian cho tất cả các miền. Bằng cách sử dụng Numpy `datetime64` và `timedelta64` dtypes, pandas đã hợp nhất một số lượng lớn các tính năng từ các thư viện Python khác cũng như `scikit.timeseries` tạo ra một lượng lớn chức năng để thao tác với dữ liệu chuỗi thời gian.

Mục tiêu của phân tích chuỗi thời gian với Pandas:

- Tạo chuỗi ngày.
- Làm việc với dấu thời gian dữ liệu.
- Chuyển đổi dữ liệu chuỗi thời gian thành dấu thời gian.
- Cắt dữ liệu chuỗi thời gian bằng dấu thời gian.
- Lấy mẫu lại chuỗi thời gian của bạn cho các tổng hợp/thống kê tóm tắt trong khoảng thời gian khác nhau.


```
import pandas as pd
from datetime import datetime
import numpy as np
range_date = pd.date_range(start = '1//1/2019',
                             end = '1/08/2019',
                             freq='Min')
range_date
```

```
DateTimeIndex( ['2019-01-01 00:00:00', '2019-01-01 00:01:00',
                '2019-01-01 00:02:00', '2019-01-01 00:03:00',
                '2019-01-01 00:04:00', '2019-01-01 00:05:00',
                '2019-01-01 00:06:00', '2019-01-01 00:07:00',
                '2019-01-01 00:08:00', '2019-01-01 00:09:00',
                ...,
                '2019-01-07 23:51:00', '2019-01-07 23:52:00',
                '2019-01-07 23:53:00', '2019-01-07 23:54:00',
                '2019-01-07 23:55:00', '2019-01-07 23:56:00',
                '2019-01-07 23:57:00', '2019-01-07 23:58:00',
                '2019-01-07 23:59:00', '2019-01-08 00:00:00'],
                dtype='datetime64[ns]', length=10081, freq='T')
```

Giải thích:

Với code trên, đã tạo ra dấu thời gian dựa trên số phút cho các phạm vi từ 1/1/2019-1/08/2019. Chúng ta có thể thay đổi tần số theo giờ, phút hoặc giây. Chức năng này sẽ giúp bạn theo dõi bản ghi dữ liệu được lưu trữ mỗi phút. Như chúng ta có thể thấy ở đầu ra, độ dài của tem datetime là 10081. Hãy nhớ rằng cấu trúc sử dụng kiểu datetime64[ns].


```

import pandas as pd
from datetime import datetime
import numpy as np
range_date = pd.date_range(start = '1//1/2019',
                             end = '1/08/2019',
                             freq='Min')
df=pd.DataFrame(range_date,columns = ['date'])
df['data']=np.random.randint(0, 100,
                             size = len(range_date))
print(df.head(10))

```

	date	data
0	2019-01-01 00:00:00	91
1	2019-01-01 00:01:00	21
2	2019-01-01 00:02:00	47
3	2019-01-01 00:03:00	66
4	2019-01-01 00:04:00	78
5	2019-01-01 00:05:00	87
6	2019-01-01 00:06:00	89
7	2019-01-01 00:07:00	60

Giải thích:

Trước tiên, chúng ta sẽ tạo ra chuỗi thời gian sau đó chuyển đổi dữ liệu này thành khung dữ liệu và sử dụng hàm ngẫu nhiên để tạo dữ liệu ngẫu nhiên và ánh xạ trên khung dữ liệu. Sau đó để kiểm tra kết quả chúng ta sử dụng hàm print.

Để thực hiện các thao tác chuỗi thời gian, chúng ta cần có chỉ mục ngày giờ để khung dữ liệu được lập chỉ mục trên dấu thời gian. Ở đây, chúng ta đang thêm một cột mới nữa khung dữ liệu và cấu trúc.

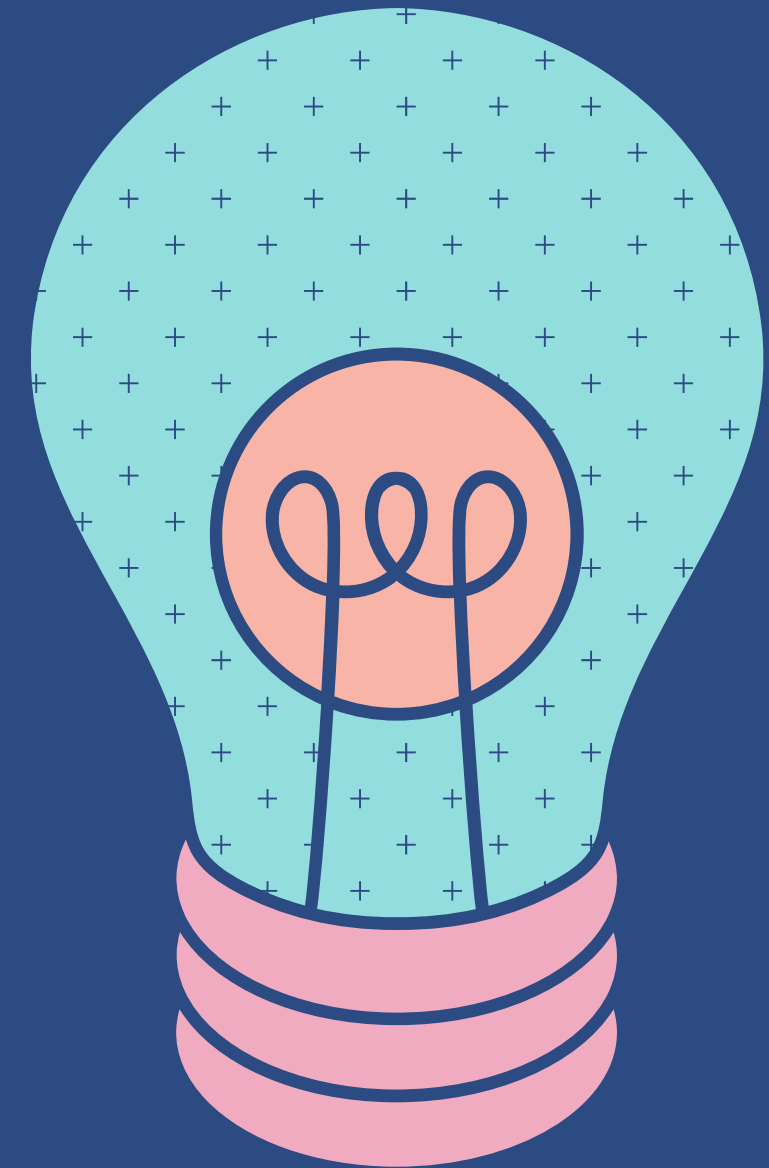


Với các ví dụ trên ta có thể tưởng tượng được một phần nào về việc mà pandas làm việc với việc phân tích chuỗi thời gian.

TÓM TẮT:

Sau khi tìm hiểu về dữ liệu chuỗi(time series) ta đã rút ra được cách làm việc và các ứng dụng về dữ liệu chuỗi thời với việc phân tích chuỗi dữ liệu chuỗi thời gian chúng ta có thể thấy được việc dự đoán trước tương lai là hoàn toàn có thể xảy ra. Riêng với việc áp dụng việc phân tích dữ liệu chuỗi thời gian cho chứng khoán chúng ta cũng có thể dự đoán được giá cổ phiếu một cách có thể gần đúng nhất

Python và Pandas rất có ích trong việc phân tích dữ liệu chuỗi thời gian với việc có thể xử lý các file như csv là rất hữu ích để làm việc và phân tích một chuỗi dữ liệu.



**Bạn có câu hỏi
nào không?**

