# Các loại dữ liệu chuỗi thời gian

Dữ liệu chuỗi thời gian là một chuỗi các điểm dữ liệu được lập chỉ mục theo thứ tự thời gian. Thông thường nhất, chuỗi thời gian là một chuỗi được lấy tại các thời điểm cách đều nhau liên tiếp, làm cho nó trở thành một chuỗi dữ liệu theo thời gian rời rạc. Nó được sử dụng rộng rãi trong kinh tế, tài chính, kinh doanh, khoa học và kỹ thuật, cùng các lĩnh vực khác, để dự báo xu hướng trong tương lai dựa trên dữ liệu lịch sử. Có một số loại dữ liệu chuỗi thời gian, mỗi loại có đặc điểm và kiểu mẫu riêng. Trong bài viết này, tôi sẽ hướng dẫn bạn về các loại dữ liệu Chuỗi thời gian và thuật toán nào bạn nên sử dụng để dự báo chúng.

## Các loại dữ liệu chuỗi thời gian

Có một số loại dữ liệu chuỗi thời gian (Time Series), mỗi loại có đặc điểm và kiểu mẫu riêng. Hiểu các loại này là rất quan trọng để chọn thuật toán phân tích hoặc dự báo thích hợp. Dưới đây là các loại dữ liệu chuỗi thời gian mà bạn nên biết với tư cách là chuyên gia Khoa học dữ liệu:

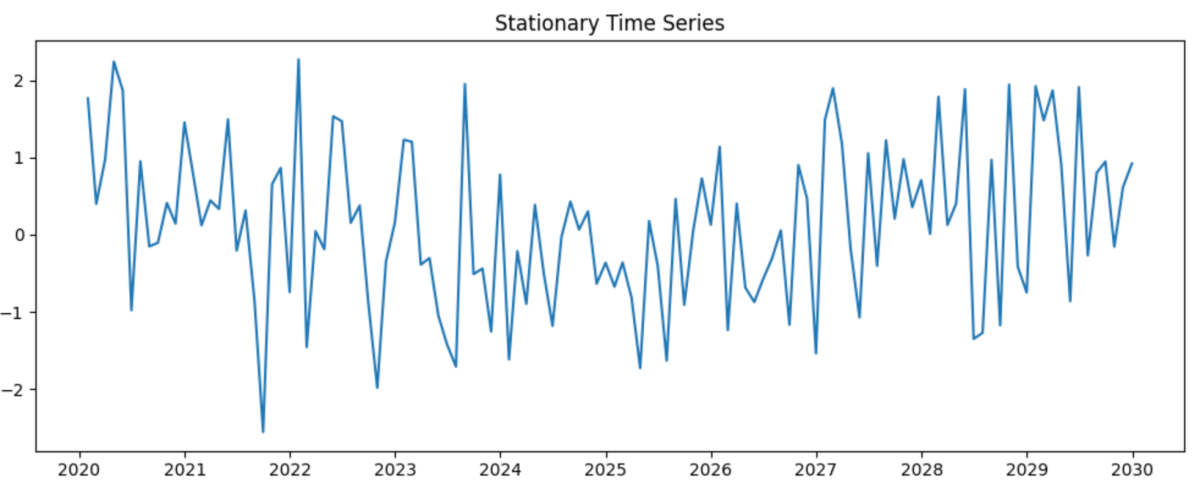
1. Chuỗi thời gian dừng (Stationary Time Series)
2. Chuỗi thời gian không dừng (Non-stationary Time Series)
3. Chuỗi thời gian theo chu kỳ (Cyclical Time Series)

Hãy hiểu tất cả các loại này và cách chọn thuật toán dự báo chuỗi thời gian cho từng loại.

## Chuỗi thời gian dừng

Chuỗi thời gian dừng có các đặc tính không phụ thuộc vào thời điểm quan sát chuỗi đó. Vì vậy, nó không có xu hướng hoặc hiệu ứng theo mùa. Các thuộc tính thống kê như giá trị trung bình, phương sai và tự tương quan không đổi theo thời gian.

Dưới đây là một ví dụ về chuỗi thời gian dừng:



Biểu đồ ở trên cho thấy một chuỗi thời gian dừng (stationary), trong đó trung bình và phương sai không đổi theo thời gian, và không có xu hướng hay tính mùa vụ rõ ràng. ARIMA (AutoRegressive Integrated Moving Average-Tự Hồi Quy Tích Hợp Trung Bình Động) và SES (Simple Exponential Smoothing-Làm Mịn Hàm Số Mũ Đơn Giản) là hai phương pháp cơ bản để phân tích dữ liệu chuỗi thời gian dừng.

ARIMA kết hợp ba thành phần chính: tự hồi quy (AR-AutoRegression), sai phân (I) để đạt được tính dừng và trung bình động (MA). Trong phần AR, mô hình dự đoán giá trị tương lai bằng cách sử dụng kết hợp tuyến tính các giá trị trong quá khứ. Thành phần I liên quan đến việc phân biệt dữ liệu một hoặc nhiều lần để loại bỏ bất kỳ xu hướng hoặc tính thời vụ nào và đạt được tính ổn định. Phần MA mô hình hóa lỗi dự đoán dưới dạng tổ hợp tuyến tính của các thuật ngữ lỗi từ các dự đoán trong quá khứ. Sự kết hợp này cho phép ARIMA nắm bắt được nhiều mẫu chuỗi thời gian trong dữ liệu dừng.

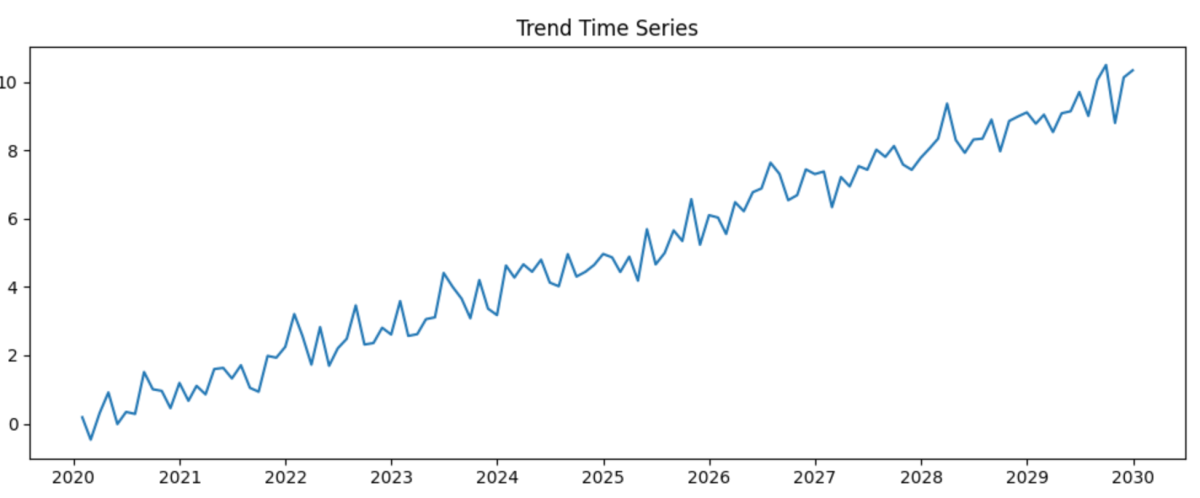
Mặt khác, SES tập trung vào việc làm mịn dữ liệu, đưa ra trọng số giảm theo cấp số nhân cho các quan sát trong quá khứ khi chúng tiến xa hơn về quá khứ. Phương pháp này đặc biệt hữu ích đối với dữ liệu không có xu hướng hoặc tính thời vụ, vì nó tạo ra một phiên bản mượt mà của chuỗi thời gian, nhấn mạnh các xu hướng gần đây nhất đồng thời làm giảm các biến động cũ một cách tự nhiên, khiến nó trở nên lý tưởng cho việc dự báo ngắn hạn trong chuỗi dừng.

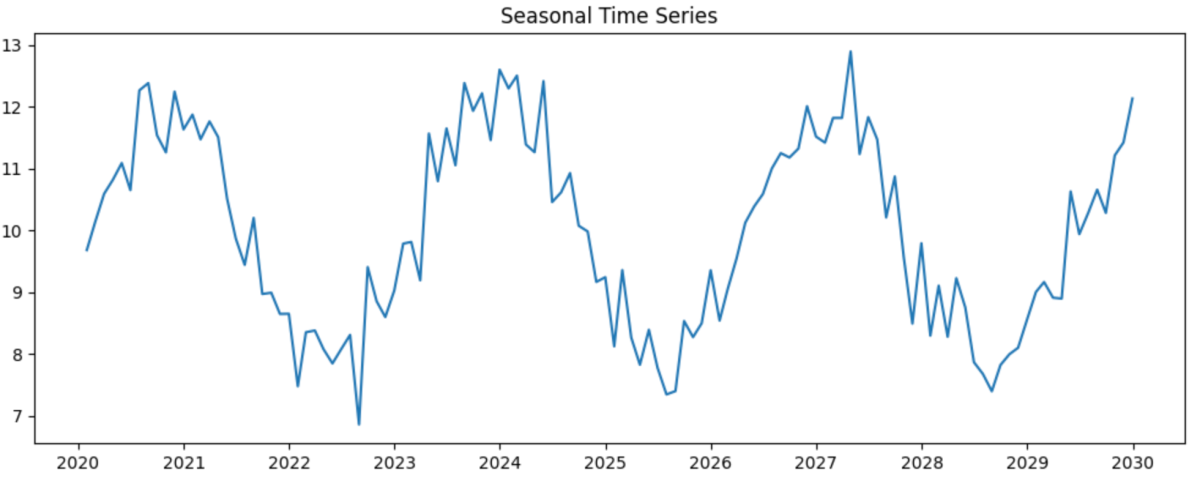
## Chuỗi thời gian không dừng

Chuỗi thời gian không dừng có các đặc tính thay đổi theo thời gian. Nó có thể bao gồm việc có xu hướng, tính thời vụ hoặc cả hai.

* **Xu hướng (Trend)**: Một xu hướng tồn tại khi dữ liệu tăng hoặc giảm trong thời gian dài. Nó không nhất thiết phải tuyến tính.
* **Tính thời vụ (Seasonality)** : Tính thời vụ xảy ra khi có các mô hình biến động thường xuyên ở những khoảng thời gian đều đặn cụ thể, chẳng hạn như hàng quý, hàng tháng hoặc hàng ngày.

Dưới đây là ví dụ về xu hướng và dữ liệu chuỗi thời gian theo mùa:





Biểu đồ đầu tiên ở trên thể hiện chuỗi thời gian có xu hướng rõ ràng, cho thấy dữ liệu tăng lên trong thời gian dài. Biểu đồ thứ hai ở trên minh họa tính thời vụ, trong đó các mô hình lặp lại đều đặn. Đối với dữ liệu không dừng thể hiện xu hướng, có thể sử dụng phép sai phân để làm cho chuỗi dữ liệu có tính dừng trước khi áp dụng ARIMA. Đối với dữ liệu có cả xu hướng và tính thời vụ, **SARIMA** hoặc **Holt-Winters** có thể đặc biệt hiệu quả.

ARIMA hoạt động trên dữ liệu không dừng có xu hướng bằng cách kết hợp bước phân biệt ban đầu (phần “Tích hợp” của ARIMA), bước này chuyển đổi chuỗi thành dạng dừng bằng cách loại bỏ xu hướng. Quá trình này cho phép mô hình áp dụng các thành phần tự hồi quy (AR) và trung bình trượt (MA) cho các dữ liệu khác nhau để nắm bắt một cách hiệu quả các mối phụ thuộc và mô hình nhiễu trong chuỗi tĩnh.

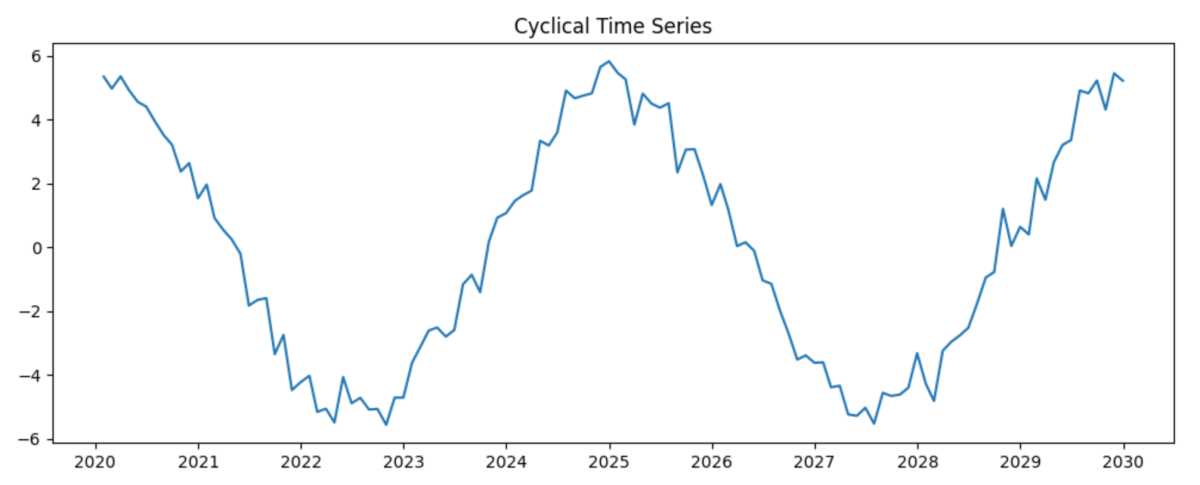
Đối với dữ liệu không dừng có cả xu hướng và tính thời vụ, SARIMA (ARIMA theo mùa) mở rộng ARIMA bằng cách thêm sự khác biệt theo mùa và các thuật ngữ AR và MA theo mùa bổ sung, đặc biệt là giải quyết và lập mô hình thành phần theo mùa của dữ liệu chuỗi thời gian. Nó khiến SARIMA đặc biệt thành thạo trong việc xử lý các mô hình phức tạp thể hiện cả xu hướng và tính thời vụ.

Mặt khác, phương pháp Holt-Winters, còn được gọi là Làm mịn hàm mũ ba, mô hình trực tiếp cả xu hướng và tính thời vụ bằng cách áp dụng ba phương trình làm mịn: một cho cấp độ, một cho xu hướng và một cho thành phần theo mùa. Cách tiếp cận này cho phép Holt-Winters điều chỉnh dự báo dựa trên cả xu hướng quan sát được và tính thời vụ trong dữ liệu, mang lại hiệu quả cao cho chuỗi thời gian với các mô hình theo mùa rõ ràng và nhất quán.

## Chuỗi thời gian theo chu kỳ

Các mô hình chu kỳ xảy ra khi có những biến động không có tần số dừng. Chúng thường được quan sát thấy trong các tập dữ liệu dài hơn và có thể được coi là các dao động dài hạn hơn, khác với tính thời vụ ở chỗ không có khoảng thời gian dừng.

Dưới đây là một ví dụ về chuỗi thời gian theo chu kỳ:



Biểu đồ trên biểu thị một chuỗi thời gian có tính chu kỳ với các chu kỳ không có tần số dừng, hiển thị các mức tăng và giảm không có độ dài dừng. Nó giống như các chu kỳ kinh tế có thể kéo dài trong các khoảng thời gian khác nhau trên một năm. [**Vector Autoregression**](https://otexts.com/fpp2/VAR.html) (VAR-Tự hồi quy vec tơ) và mô hình cấu trúc là những phương pháp phức tạp được thiết kế để phân tích và dự báo chuỗi thời gian mang tính chu kỳ, thường thấy trong kinh tế và tài chính, trong đó các chu kỳ không hoàn toàn theo mùa và có thể thay đổi theo thời gian.

VAR mô hình hóa sự phụ thuộc lẫn nhau và ảnh hưởng đồng thời giữa nhiều chuỗi thời gian bằng cách sử dụng các giá trị trong quá khứ của hệ thống để dự đoán các giá trị tương lai của nó, nắm bắt hiệu quả các động lực mang tính chu kỳ thông qua sự tương tác của các chuỗi này. Nó làm cho VAR đặc biệt phù hợp để kiểm tra xem các biến số tiến triển cùng nhau như thế nào theo thời gian, xác định mối quan hệ dẫn đầu-độ trễ giữa chúng và hiểu tác động của các cú sốc đối với một biến số đối với các biến số khác.

Mặt khác, các Mô hình cấu trúc kết hợp kiến ​​thức lý thuyết về các quy trình cơ bản tạo ra dữ liệu, bao gồm cả các hành vi có tính chu kỳ tiềm ẩn. Chúng chia chuỗi thời gian thành các thành phần: xu hướng, chu kỳ, tính thời vụ và các thành phần không đều dựa trên lý thuyết kinh tế hoặc kết quả thực nghiệm, cho phép mô hình hóa rõ ràng các mô hình chu kỳ và cách chúng bị ảnh hưởng bởi các biến ngoại sinh hoặc sự thay đổi cơ cấu trong nền kinh tế.

## Kết luận

Vì vậy, dưới đây là các loại dữ liệu chuỗi thời gian mà bạn nên biết với tư cách là chuyên gia Khoa học dữ liệu:

1. **Chuỗi thời gian dừng** : Chuỗi thời gian dừng có các đặc tính không phụ thuộc vào thời điểm chuỗi được quan sát. Vì vậy, nó không có xu hướng hoặc hiệu ứng theo mùa. Các thuộc tính thống kê như giá trị trung bình, phương sai và tự tương quan không đổi theo thời gian.
2. **Chuỗi thời gian không dừng** : Chuỗi thời gian không dừng có các đặc tính thay đổi theo thời gian. Nó có thể bao gồm việc có xu hướng, tính thời vụ hoặc cả hai.
3. **Chuỗi thời gian theo chu kỳ** : Các mô hình chu kỳ xảy ra khi có những biến động không có tần số dừng. Chúng thường được quan sát thấy trong các tập dữ liệu dài hơn và có thể được coi là các dao động dài hạn hơn, khác với tính thời vụ ở chỗ không có khoảng thời gian dừng.