

Bài tập lý thuyết #02

PHÂN TÍCH DỮ LIỆU CHUỖI THỜI GIAN

GÓC NHÌN LÝ THUYẾT & KINH DOANH

Bối cảnh

Công ty A là một nhà bán lẻ trực tuyến không có cửa hàng (non-store online retailer) được đăng ký và hoạt động tại Vương quốc Anh (the UK). Công ty chủ yếu bán các sản phẩm quà tặng độc đáo cho mọi dịp. Nhiều khách hàng của công ty là các nhà bán sỉ.

Dữ liệu giao dịch của công ty từ ngày 01/12/2010 đến 09/12/2011 sẽ được cung cấp để đánh giá hiệu suất hoạt động và đưa ra các khuyến nghị cho doanh nghiệp.

Nếu bạn có bất kỳ thắc mắc nào về bộ dữ liệu hoặc muốn tìm cảm hứng cho phần phân tích, bạn có thể truy cập trang Kaggle tại đây:

<https://www.kaggle.com/datasets/carrie1/ecommerce-data>.

Mục tiêu

- Hiểu rõ sự khác biệt giữa cross-sectional, panel và time series data
- Có khả năng tái cấu trúc dữ liệu giao dịch thành chuỗi thời gian hợp lệ
- Phân tích và lập luận học thuật về:
 - Trend
 - Seasonality
 - Cyclical fluctuations
 - Irregularity
- Đánh giá hạn chế của dữ liệu và phương pháp
- Kết nối phân tích chuỗi thời gian với ra quyết định chiến lược.

Các yêu cầu.

1. Yêu cầu 1- Đánh giá bản chất chuỗi thời gian của dataset

Nhiệm vụ

Phân tích và trả lời:

- Dataset gốc có phải time series data không?

2) Nếu không, thì đang thuộc loại dữ liệu nào?

3) Đề xuất chiến lược chuyển đổi dữ liệu để phục vụ phân tích chuỗi thời gian

Yêu cầu học thuật:

- Sử dụng đúng thuật ngữ

- Trình bày giả định (assumptions) khi xây dựng chuỗi thời gian (không được hiểu ngầm và phải viết rõ lập luận)

2. **Yêu cầu 2-Thiết kế chuỗi thời gian & lựa chọn độ phân giải thời gian**

Nhiệm vụ: Xây dựng ít nhất 2 chuỗi thời gian:

- 1) Doanh thu theo thời gian

- 2) Số lượng giao dịch hoặc số khách hàng theo thời gian

- 3) Thử nghiệm ít nhất 2 mức độ thời gian (ví dụ: daily vs monthly)

Yêu cầu cụ thể, so sánh:

- Tính ổn định

- Mức độ nhiễu

- Khả năng diễn giải kinh doanh

- Giải thích vì sao một độ phân giải thời gian phù hợp hơn độ còn lại

3. **Yêu cầu 3 – Phân tích xu hướng (Trend)**

Nhiệm vụ

- Trực quan hóa chuỗi thời gian

- Áp dụng moving average (ít nhất 2 window size khác nhau)

Kiểm tra:

- Xu hướng quan sát được có bền vững không?

- Có thể bị “đánh lừa” bởi seasonal effect hay không?

4. **Yêu cầu 4 – Phân tích seasonality**

Nhiệm vụ

- Phân tích mùa vụ theo:

- o Tháng trong năm
 - o Ngày trong tuần

- So sánh seasonality ở các cấp độ thời gian khác nhau

Yêu cầu nâng cao

- Liên hệ seasonality với:

- Đặc thù ngành bán lẻ online
- Bối cảnh UK
- Hành vi mua sỉ (wholesale customers)

📌 *Phải chứng minh seasonality là “business-driven”, không chỉ là pattern hình học*

Yêu cầu 5 – So sánh Time Series vs Snapshot Analysis

Nhiệm vụ

- Chọn một insight từ Milestone 1
- Chứng minh:
 - Insight đó thay đổi như thế nào khi nhìn dưới góc độ chuỗi thời gian?

Ví dụ: “Sản phẩm bán chạy nhất” có ổn định theo thời gian hay chỉ là hiện tượng ngắn hạn?

SẢN PHẨM NỘP:

- 1) **Báo cáo Milestone 2 (10–15 trang):**
 - a. Có section riêng: *Time Series Methodology*
 - b. Có lập luận, phản biện, hạn chế
- 2) **Hệ thống biểu đồ chuẩn học thuật**
- 3) **Mã nguồn phân tích dữ liệu (file Jupyter Notebook)**