

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**  
**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO BÀI TẬP NHÓM 2**

**Xử lý và trực quan hóa dữ liệu chuỗi thời gian**

**MÔN: TRỰC QUAN HÓA DỮ LIỆU**

**Lớp: 22\_21**

Giảng viên hướng dẫn: **Thầy Lê Nhật Nam**

Nhóm thực hiện: **Nhóm 01**

Thông tin thành viên:

22120095 Nguyễn Xuân Hạng

22120110 Đặng Minh Hoàng

22120424 Phạm Ngọc Bảo Uyên

22120449 Lê Nguyễn Huyền Vy

**THÀNH PHỐ HỒ CHÍ MINH, THÁNG 12/ 2024**

## Mục lục

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Yêu cầu bài tập</b>  | <b>2</b> |
| 1.1      | Thu thập dữ liệu . . . . .  | 2        |
| 1.2      | Khám phá dữ liệu . . . . .  | 2        |
| 1.3      | Khám phá các mối quan hệ ẩn trong tập dữ liệu . . . . .               | 2        |
| <b>2</b> | <b>Lựa chọn chủ đề</b>  | <b>2</b> |
| <b>3</b> | <b>Một số câu hỏi trong Khám phá các mối quan hệ ẩn trong dữ liệu</b> | <b>3</b> |
| 3.1      | Tổng quan về thời gian . . . . .                                      | 3        |
| 3.2      | Phân tích chi tiết . . . . .  | 3        |
| 3.3      | Phân tích xu hướng mùa vụ . . . . .                                   | 3        |
| 3.4      | Phân tích hiệu quả hoạt động theo cửa hàng . . . . .                  | 3        |
| 3.5      | Phân tích hiệu suất sản phẩm . . . . .                                | 3        |
| <b>4</b> | <b>Phân công công việc và mức độ hoàn thành</b>                       | <b>4</b> |
| <b>5</b> | <b>Bảng tự đánh giá</b>   | <b>4</b> |

# 1 Yêu cầu bài tập

Sử dụng [Kaggle](#) để tìm kiếm dữ liệu công khai có sẵn liên quan đến chuỗi thời gian về một chủ đề sinh viên quan tâm. Tập dữ liệu phải được tổ chức theo một bảng có ít nhất 5 cột dữ liệu và 1000 dòng.

## 1.1 Thu thập dữ liệu

- Bối cảnh hoặc câu chuyện chính có ý nghĩa khiến nhóm của bạn chọn chủ đề này để tìm dữ liệu phù hợp cho dự án này?
- Chủ đề chính của tập dữ liệu của bạn là nguồn để xây dựng nó?
- Mọi người xây dựng dữ liệu này như thế nào? Phương pháp là gì?
- Chúng ta sử dụng tập dữ liệu này như thế nào? Điều này có hợp pháp để sử dụng trong giáo dục?

## 1.2 Khám phá dữ liệu

- Mỗi dòng có ý nghĩa gì? Có vấn đề gì nếu các dòng có ý nghĩa khác nhau?
- Mỗi cột có ý nghĩa gì?
- Mỗi cột hiện có kiểu dữ liệu gì? Có cột nào có kiểu dữ liệu không phù hợp để xử lý tiếp?
- Đối với mỗi cột, các giá trị (số, phân loại) được phân bố như thế nào?
- Có cần phải xử lý trước dữ liệu không và nếu có thì bạn cần thực hiện như thế nào?

## 1.3 Khám phá các mối quan hệ ẩn trong tập dữ liệu

- Xác định các mối quan hệ bằng cách đặt ra những câu hỏi có ý nghĩa giữa các thuộc tính trong tập dữ liệu.
- Thể hiện các mối quan hệ bằng cách trực quan hóa với các biểu đồ phù hợp.
- Đưa ra các đánh giá, nhận xét về các mối quan hệ dựa trên biểu đồ đã trực quan.

# 2 Lựa chọn chủ đề

**Chủ đề:** Phân tích dữ liệu về các thông tin giao dịch của một cửa hàng cà phê (từ tháng 1/2023 đến tháng 6/2023)

**Nguồn dữ liệu:** [Kaggle - Coffee Shop Sales](#)

### 3 Một số câu hỏi trong Khám phá các mối quan hệ ẩn trong dữ liệu

Nhóm đã thực hiện chia ra các vấn đề theo từng mục và đặt ra các câu hỏi để giải quyết các vấn đề đặt ra này.

#### 3.1 Tổng quan về thời gian

1. Doanh thu tổng cộng theo tháng-năm là bao nhiêu?
2. Lượng giao dịch có xu hướng tăng hay giảm theo thời gian?
3. Liệu rằng cuối tuần sẽ có doanh thu cao hơn các ngày trong tuần?

#### 3.2 Phân tích chi tiết

1. Cửa hàng nào tạo ra doanh thu cao nhất?
2. Sản phẩm nào bán chạy theo khung giờ trong ngày?
3. Giá sản phẩm có ảnh hưởng đến số lượng bán ra không?
4. Khung giờ nào trong ngày sẽ có lượng giao dịch cao nhất?
5. Liệu có sự khác nhau về số lượng giao dịch cao nhất theo khung giờ giữa ngày trong tuần và ngày cuối tuần?

#### 3.3 Phân tích xu hướng mùa vụ

1. Có sự biến động doanh thu nào vào các tháng đầu năm không?
2. Doanh số trong các ngày lễ đặc biệt so với các ngày còn lại?

#### 3.4 Phân tích hiệu quả hoạt động theo cửa hàng

1. Hiệu quả hoạt động trong từng buổi trong ngày của từng chi nhánh cửa hàng cà phê khác nhau như thế nào?
2. Cửa hàng nào hoạt động hiệu quả nhất vào giờ cao điểm?
3. Doanh thu trung bình theo mỗi loại sản phẩm tại mỗi cửa hàng là bao nhiêu?

#### 3.5 Phân tích hiệu suất sản phẩm

1. Danh mục sản phẩm nào đóng góp lớn nhất vào tổng doanh thu theo tháng?
2. Sản phẩm nào bán chạy nhất theo tháng? Liệu sản phẩm đó có thuộc danh mục sản phẩm mang lại doanh thu lớn nhất theo tháng?
3. Liệu một sản phẩm được bán ra nhiều nhất thì có phải là sản phẩm đạt được doanh thu cao nhất?

4 Phân công công việc và mức độ hoàn thành

| STT | Người phụ trách    | Công việc được giao   | Mức độ hoàn thành |
|-----|--------------------|---|-------------------|
| 1   | Nguyễn Xuân Hạnh   | Thu thập dữ liệu, Khám phá các mối quan hệ ẩn trong dữ liệu | 100%              |
| 2   | Đặng Minh Hoàng    | Thu thập dữ liệu, Khám phá các mối quan hệ ẩn trong dữ liệu | 100%              |
| 3   | Phạm Ngọc Bảo Uyên | Khám phá dữ liệu, Khám phá các mối quan hệ ẩn trong dữ liệu | 100%              |
| 4   | Lê Nguyễn Huyền Vy | Khám phá dữ liệu, Khám phá các mối quan hệ ẩn trong dữ liệu | 100%              |

Bảng 1: Bảng phân công công việc và mức độ hoàn thành

5 Bảng tự đánh giá

| Tiêu chí   | Mức độ hoàn thành |
|--|-------------------|
| Thu thập và tiền xử lý dữ liệu (5%)  | 5%                |
| Chọn lựa, giải thích, trực quan các trường và các mối quan hệ giữa chúng (50%)   | 50%               |
| Rút ra ý nghĩa hợp lý sau mỗi dữ liệu được trực quan (20%)   | 20%               |
| Xem xét trên nhiều quan hệ, nhiều góc nhìn khác nhau (10%)   | 10%               |
| Báo cáo trình bày bố cục và định dạng hợp lý, rõ ràng (15%)  | 15%               |
| Báo cáo trình bày bố cục và định dạng hợp lý, rõ ràng (15%)  | 15%               |
| Có những phân tích, trực quan hóa bằng những biểu đồ mới lạ và rút ra những thông tin hữu ích. Sử dụng mô hình học máy cơ bản (5%) | 5%                |
| Hiểu và giải thích được tổng quan mã nguồn (5%)  | 5%                |
| Tổng (110%)  | 110%              |

Bảng 2: Các yêu cầu và mức độ hoàn thành của bài tập