

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



**MÔN TRỰC QUAN HÓA DỮ LIỆU**

**BÁO CÁO ĐỒ ÁN 1:  
TRỰC QUAN HÓA DỮ LIỆU VỚI PYTHON**

**GIẢNG VIÊN HƯỚNG DẪN:** thầy Lê Nhựt Nam  
**SINH VIÊN THỰC HIỆN:**

Lê Nguyễn Gia Bảo	- 22120023
Huỳnh Mẫn	- 22120201
Hồ Hoàng Nhật Minh	- 22120208
Bùi Tấn Thành Nam	- 22120216
Hồ Minh Quang	- 22120295

Thành phố Hồ Chí Minh, tháng 03 năm 2025

MỤC LỤC

**BẢNG PHÂN CÔNG CÔNG VIỆC ..... 1**

**I. GIỚI THIỆU CHỦ ĐỀ..... 2**

**II. ĐÁP ỨNG YÊU CẦU & TỰ ĐÁNH GIÁ..... 2**

**III. TÀI LIỆU THAM KHẢO ..... 3**

## BẢNG PHÂN CÔNG CÔNG VIỆC

### 1. Danh sách thành viên

Các thành viên:

STT	MSSV	Họ và tên	Email
1	22120023	Lê Nguyễn Gia Bảo	<a href="mailto:22120023@student.hcmus.edu.vn">22120023@student.hcmus.edu.vn</a>
2	22120201	Huỳnh Mẫn	<a href="mailto:22120201@student.hcmus.edu.vn">22120201@student.hcmus.edu.vn</a>
3	22120208	Hoàng Hồ Nhật Minh	<a href="mailto:22120208@student.hcmus.edu.vn">22120208@student.hcmus.edu.vn</a>
4	22120216	Bùi Tấn Thành Nam	<a href="mailto:22120216@student.hcmus.edu.vn">22120216@student.hcmus.edu.vn</a>
5	22120295	Hồ Minh Quang	<a href="mailto:22120295@student.hcmus.edu.vn">22120295@student.hcmus.edu.vn</a>

### 2. Bảng phân công công việc

STT	Người phụ trách	Công việc được giao	Mức độ hoàn thành
1	Lê Nguyễn Gia Bảo	<ul style="list-style-type: none"><li>- Tiền xử lý dữ liệu.</li><li>- Trực quan hóa dữ liệu.</li><li>- Sử dụng mô hình máy học.</li></ul>	100%
2	Huỳnh Mẫn	<ul style="list-style-type: none"><li>- Thu thập dữ liệu.</li><li>- Viết báo cáo.</li><li>- Kiểm tra tất cả nội dung trước khi nộp.</li></ul>	100%
3	Hoàng Hồ Nhật Minh	<ul style="list-style-type: none"><li>- Thu thập dữ liệu.</li><li>- Tiền xử lý dữ liệu.</li><li>- Trực quan hóa dữ liệu.</li></ul>	100%
4	Bùi Tấn Thành Nam	<ul style="list-style-type: none"><li>- Tiền xử lý dữ liệu.</li><li>- Trực quan hóa dữ liệu.</li><li>- Viết báo cáo.</li></ul>	100%
5	Hồ Minh Quang	<ul style="list-style-type: none"><li>- Trực quan hóa dữ liệu.</li><li>- Tổng hợp, trình bày kết quả.</li><li>- Sử dụng mô hình máy học.</li></ul>	100%

## I. GIỚI THIỆU CHỦ ĐỀ

**Chủ đề:** Phân tích hành vi người tiêu dùng và thói quen mua sắm.

**Nguồn dữ liệu:** [Consumer Behavior and Shopping Habits Dataset](#).

Trong thời đại mà sự phát triển của thương mại điện tử và các kênh bán lẻ trực tuyến đang bùng nổ, hành vi tiêu dùng đã trở thành một lĩnh vực nghiên cứu quan trọng của các doanh nghiệp. Nhóm chúng em đã chọn chủ đề "Hành vi tiêu dùng và thói quen mua sắm" với mong muốn khám phá rõ cách mà các yếu tố như sở thích cá nhân, thói quen mua sắm và môi trường kinh doanh tác động đến quyết định mua hàng của người tiêu dùng.

Động cơ chính của dự án xuất phát từ câu hỏi “ Những yếu tố nào đang định hình hành vi mua sắm của người tiêu dùng, và làm thế nào thông qua việc trực quan hóa dữ liệu, các doanh nghiệp có thể tận dụng những thông tin này để tối ưu hóa chiến lược tiếp thị và cải thiện trải nghiệm khách hàng? ” Sự tò mò này được thúc đẩy bởi nhu cầu hiểu rõ hơn về các xu hướng tiêu dùng hiện nay, cũng như mong muốn áp dụng các phương pháp khoa học dữ liệu để phân tích và đưa ra những nhận định dựa trên dữ liệu thực tế.

Việc trực quan hóa dữ liệu không chỉ giúp làm nổi bật các mẫu và xu hướng trong hành vi tiêu dùng mà còn hỗ trợ việc ra quyết định chiến lược cho các doanh nghiệp. Qua đó, chúng em hy vọng sẽ đóng góp một góc nhìn sâu sắc, hữu ích cho cộng đồng nghiên cứu và kinh doanh, mở ra những hướng đi mới trong việc quản lý và phát triển thị trường tiêu dùng.

## II. ĐÁP ỨNG YÊU CẦU & TỰ ĐÁNH GIÁ

### 1. Đáp ứng yêu cầu:

Trong đồ án này, nhóm chúng em sử dụng Kaggle để tìm kiếm dữ liệu công khai về chủ đề mà nhóm quan tâm. Bộ dữ liệu thỏa mãn yêu cầu *có tối thiểu 5 trường dữ liệu với 3000 dòng* và được *cập nhật trong hai năm trở lại đây*.

Giai đoạn	Yêu cầu	Đáp ứng
Thu thập dữ liệu	<ul style="list-style-type: none"><li>- Nêu ngữ cảnh và lý do chọn chủ đề của bộ dữ liệu.</li><li>- Chủ đề của bộ dữ liệu là gì?</li><li>- Phương pháp xây dựng nên bộ dữ liệu là gì? Phương pháp sử dụng ra sao?</li><li>- Dữ liệu có “hợp pháp” để sử dụng không.</li></ul>	100%

Tiền xử lý và khám phá dữ liệu	<ul style="list-style-type: none"> <li>- Mỗi dòng và mỗi cột có ý nghĩa là gì?</li> <li>- Dữ liệu phân bố như thế nào? Kiểu dữ liệu có phù hợp không?</li> <li>- Tiền xử lý dữ liệu như thế nào?</li> </ul>	100%
Khám phá mối quan hệ giữa các trường trong bộ dữ liệu	<ul style="list-style-type: none"> <li>- Chọn các trường dữ liệu để trực quan hóa và chọn loại biểu đồ thích hợp để trực quan mối quan hệ giữa các trường dữ liệu đó.</li> <li>- Đưa ra lý do chọn loại biểu đồ để trực quan.</li> <li>- Đưa ra nhận xét về mối quan hệ giữa các trường dữ liệu đã được trực quan.</li> </ul>	100%
Mô hình học máy	- Áp dụng mô hình học máy cơ bản để dự đoán thuộc tính đích.	100%

## 2. Tự đánh giá:

Tiêu chí	Tự đánh giá
Thu thập dữ liệu và tiền xử lý dữ liệu. (5%)	100%
Chọn, diễn giải và trực quan hóa các trường và mối quan hệ ẩn của chúng. (50%)	100%
Rút ra ý nghĩa logic đằng sau mỗi dữ liệu được trực quan hóa. (20%)	100%
Xem xét nhiều mối quan hệ và nhiều góc nhìn khác nhau. (10%)	100%
Báo cáo trình bày theo định dạng và bố cục hợp lý, rõ ràng. (15%)	100%
Có phân tích, trực quan hóa với biểu đồ mới lạ, thông tin hữu ích. Sử dụng các mô hình máy học cơ bản. (5%)	100%
Hiểu biết về mã nguồn đã làm. (5%)	100%

## III. TÀI LIỆU THAM KHẢO

1. [\*Persona Recommendation System\(with EDA\)\*](#)
2. [\*Create a stacked bar plot in Matplotlib\*](#)
3. [\*Vẽ Biểu Đồ Với Thư Viện Matplotlib \(Phần 2\)\*](#)
4. [\*Visualization trong python\*](#)

**HẾT.**