

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



**BÁO CÁO LAB 01:**  
**TRỰC QUAN HÓA DỮ LIỆU VỚI PYTHON**

**Môn Học: Trực Quan Hóa Dữ Liệu**

**GVHD: Thầy Bùi Tiến Lên, Cô Nguyễn Thị Thu Hằng, Thầy Nguyễn Bảo Long,**  
**Thầy Lê Nhựt Nam**

Nhóm: 5

Danh sách sinh viên thực hiện:

- |                       |          |
|-----------------------|----------|
| 1. Đinh Thị Thùy Linh | 20120130 |
| 2. Lương Vĩnh Phú     | 20120347 |
| 3. Bùi Thanh Tùng     | 20120398 |
| 4. Đỗ Tấn Tài         | 20120408 |
| 5. Trần Khắc Bình     | 20120437 |

Thành phố Hồ Chí Minh - 2023

---

## TÓM TẮT

---

Chủ đề mà nhóm 5 chọn để làm lab01 là về **Shop Customer Data**. Đối với một doanh nghiệp, một trong những yếu tố cốt lõi để cấu thành nên sự thành công đó chính là khách hàng. Chính vì thế, việc tìm hiểu và phân tích nhu cầu khách hàng chính là một trong những việc làm quan trọng, một bài toán toán khó cho các doanh nghiệp hiện nay. Thấy được tầm quan trọng của việc phân tích khách hàng, nhóm 5 lựa chọn bộ dữ liệu **Shop Customer Data** cho lab 01.

Bộ dữ liệu được chia sẻ miễn phí trên nền tảng Kaggle. (Link dataset: <https://www.kaggle.com/datasets/datascientistanna/customers-dataset>)

Các bước thực hiện bài lab01:

- Tìm dữ liệu.
- Tiền xử lý dữ liệu: xử lý, làm sạch dữ liệu (các bước trình bày chi tiết được ghi trong file **lab01.ipynb**)
- Trực quan hóa mối quan hệ giữa các trường dữ liệu và rút ra các nhận xét (các bước trình bày chi tiết được ghi trong file **lab01.ipynb**)
- Xây dựng mô hình K-Means để phân nhóm khách hàng. (các bước trình bày chi tiết được ghi trong file **lab01.ipynb**)

**Mức độ hoàn thành lab01**

STT	Yêu cầu	Công việc	Mức độ hoàn thành
1	Thu thập dữ liệu	Thực hiện theo các yêu cầu phần 2A trong file CSC10108-Lab01.pdf	100%
2	Khám phá dữ liệu	Thực hiện theo các yêu cầu phần 2B trong file CSC10108-Lab01.pdf	100%
3	Khám phá mối quan hệ trong dữ liệu	Trực quan hóa mối quan hệ giữa các trường dữ liệu và rút ra các nhận xét	100%
4	Sử dụng mô hình học máy cơ bản.	Dùng mô hình k-mean để phân nhóm khách hàng	100%

**Bảng phân công công việc**

STT	Công việc	Người thực hiện	Nội dung
1	Tìm nguồn dữ liệu, đề tài.	Tùng, Tài, Bình, Linh, Phú	Chọn bộ dữ liệu: Shop Customer Data
2	Tiền xử lý dữ liệu.	Tùng, Tài	Thực hiện như các bước theo file hướng dẫn đồ án, trình bày trong file <b>lab01.ipynb</b>
3	Khám phá mối quan hệ giữa các biến dữ liệu	Tùng, Tài, Bình, Linh, Phú	Trực quan hóa mối quan hệ giữa các trường dữ liệu và rút ra các nhận xét
4	Xây dựng mô hình học máy	Tùng	Dùng mô hình k-means để phân nhóm khách hàng.
5	Tổng hợp, chỉnh sửa	Tùng, Tài, Bình, Linh, Phú	Tổng hợp, chỉnh sửa file <b>lab01.ipynb</b>
6	Viết báo cáo	Tùng, Tài, Bình, Linh, Phú	

Link github: [https://github.com/tungbtt/Lab01\\_DV](https://github.com/tungbtt/Lab01_DV)

---

## TÀI LIỆU THAM KHẢO

---

- [1]. <https://simerp.io/blog/tim-hieu-nhu-cau-khach-hang/>
- [2]. <https://vitalflux.com/pca-explained-variance-concept-python-example/>
- [3]. <https://www.datacamp.com/tutorial/decision-tree-classification-python>
- [4]. [https://www.kaggle.com/code/utkarshsaxenadn/shop-customer-clustering?fbclid=IwAR29tq6\\_MlfU848Svyr\\_Rghg9U3nxiIYaZa3kHgURPZvRh1NwN8Lo8wrkp0](https://www.kaggle.com/code/utkarshsaxenadn/shop-customer-clustering?fbclid=IwAR29tq6_MlfU848Svyr_Rghg9U3nxiIYaZa3kHgURPZvRh1NwN8Lo8wrkp0)