

ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN THỰC HÀNH CUỐI KÌ
BÁO CÁO PHÂN CÔNG

Nhóm: 15

Các thành viên:

Đinh Thị Hoàng Linh - 20120130

Phạm Trần Gia Phú - 20120348

Nguyễn Anh Tuấn - 20120395

Phạm Khánh Hoàng Việt - 20120626

TP HỒ CHÍ MINH - 12/2022

Mục lục

I.	CÁC THÀNH VIÊN.....	1
II.	BẢNG PHÂN CÔNG CÔNG VIỆC.....	1
III.	GHI CHÚ HỌP NHÓM	1
1.	Lần họp ngày 14/11/2022.....	1
2.	Lần họp ngày 18/11/2022.....	2
3.	Lần họp ngày 4/12/2022.....	3
4.	Lần họp ngày 11/12/2022.....	3
5.	Lần họp ngày 14/12/2022.....	4
IV.	NGUỒN THAM KHẢO	4
1.	Nguồn tham khảo kỹ thuật:	4
2.	Nguồn tham khảo khác:.....	4

I. CÁC THÀNH VIÊN

- Đinh Thị Hoàng Linh – 20120130
- Phạm Trần Gia Phú – 20120348
- Nguyễn Anh Tuấn – 20120395 (Nhóm trưởng)
- Phạm Khánh Hoàng Việt – 20120626

II. BẢNG PHÂN CÔNG CÔNG VIỆC

STT	Công việc	Thành viên phụ trách	Đã hoàn thành (X)
1	Thu thập dữ liệu	Cả nhóm	X
2	Khám phá dữ liệu	Đinh Thị Hoàng Linh	X
		Phạm Trần Gia Phú	X
		Nguyễn Anh Tuấn	X
		Phạm Khánh Hoàng Việt	X
3	Đặt và trả lời câu hỏi	Đinh Thị Hoàng Linh	X
		Phạm Trần Gia Phú	X
		Nguyễn Anh Tuấn	X
		Phạm Khánh Hoàng Việt	X
5	Tổng hợp mã nguồn	Phạm Trần Gia Phú	X
6	Làm Slide báo cáo	Cả nhóm	X
7	Viết báo cáo phân công việc	Nguyễn Anh Tuấn	X

III. GHI CHÚ HỌP NHÓM

1. Lần họp ngày 14/11/2022

- Nội dung: Tìm dữ liệu cho đồ án, thống nhất github.
- Github quản lý: <https://github.com/giaphupham/LT4KHDL-DATHCK.git>
- Phân công:

STT	Công việc	Thành viên phụ trách
1	Tìm ít nhất 1 tập dữ liệu	Cả nhóm

2	Quản lý mã nguồn trên github	Phạm Trần Gia Phú
---	------------------------------	-------------------

2. Lần họp ngày 18/11/2022

- Nội dung:
 1. Thống nhất tập dữ liệu.
 2. Phân chia công việc cho Khám phá Dữ liệu.
- Tập dữ liệu thống nhất: [Chemicals in Cosmetics: What's Really in Your? | Kaggle](#)
- Hạn hoàn thành: 2/12/2022
- Phân công:

STT	Công việc	Thành viên phụ trách
1	<ul style="list-style-type: none"> - Dữ liệu bao nhiêu dòng và bao nhiêu cột? - Mỗi dòng có ý nghĩa gì? Có vấn đề các dòng có ý nghĩa khác nhau không? - Dữ liệu có các dòng bị lặp không? 	Phạm Khánh Hoàng Việt
2	<ul style="list-style-type: none"> - Mỗi cột có ý nghĩa gì? - Mỗi cột đang có kiểu dữ liệu gì? Cột nào có kiểu dữ liệu chưa phù hợp để có thể xử lý tiếp? 	Phạm Trần Gia Phú
3	<ul style="list-style-type: none"> - Với mỗi cột có kiểu dữ liệu dạng số, các giá trị được phân bố như thế nào? <ul style="list-style-type: none"> ▪ Số lượng/tỉ lệ các giá trị thiếu? ▪ Min? Max? Có gì bất thường không? 	Đinh Thị Hoàng Linh
4	<ul style="list-style-type: none"> - Với mỗi cột có kiểu dữ liệu dạng phân loại, các giá trị được phân bố như thế? <ul style="list-style-type: none"> ▪ Số lượng/tỉ lệ các giá trị thiếu? ▪ Số lượng các giá trị khác nhau? Hiện thị một vài giá trị? Có gì bất thường không? 	Nguyễn Anh Tuấn

3. Lần họp ngày 4/12/2022

- Nội dung:
 1. Phân công tổng hợp mã nguồn.
 2. Phân chia công việc Tiền xử lý, Đặt và trả lời câu hỏi.
- Hạn hoàn thành: 11/12/2022
- Phân công:

STT	Công việc	Thành viên phụ trách
1	Tổng hợp mã nguồn Khám phá Dữ liệu	Phạm Trần Gia Phú
2	Đặt và trả lời câu hỏi 1: Top 10 loại hóa chất được sử dụng nhiều nhất?	Phạm Trần Gia Phú
3	Đặt và trả lời câu hỏi 2: Top 10 những hóa chất đã được loại bỏ nhiều nhất?	Nguyễn Anh Tuấn
4	Đặt và trả lời câu hỏi 3: Các hãng sản xuất mỹ phẩm có sản phẩm chứa nhiều hóa chất?	Đinh Thị Hoàng Linh
5	Đặt và trả lời câu hỏi 4: Top 5 nhãn hiệu có nhiều sản phẩm nhất?	Phạm Khánh Hoàng Việt

4. Lần họp ngày 11/12/2022

- Nội dung:
 1. Phân công tổng hợp mã nguồn.
 2. Phân chia công việc.
- Hạn hoàn thành: 13/12/2022
- Phân công:

STT	Công việc	Thành viên phụ trách
1	Tổng hợp mã nguồn Tiền xử lý, Đặt và trả lời câu hỏi.	Phạm Trần Gia Phú
2	Thiết kế Slide báo cáo	Đinh Thị Hoàng Linh
3	Thêm nội dung slide báo cáo	Cả nhóm
4	Viết file báo cáo phân công	Nguyễn Anh Tuấn

5. Lần họp ngày 14/12/2022

- Nội dung: Kiểm tra lần cuối các file nộp trước buổi vấn đáp.

IV. NGUỒN THAM KHẢO

1. Nguồn tham khảo kỹ thuật:

- Tài liệu thư viện Numpy: [NumPy Documentation](#)
- Tài liệu thư viện Pandas: [pandas documentation — pandas 1.5.2 documentation \(pydata.org\)](#)
- Tài liệu thư viện Matplotlib: [Matplotlib documentation — Matplotlib 3.6.2 documentation](#)

2. Nguồn tham khảo khác:

- [\(2142\) Solving real world data science tasks with Python Pandas! - YouTube](#)
- [Cocamide DEA \(Surfactant\) - Cosmetic Ingredient INCI \(specialchem.com\)](#)