

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**



**BÁO CÁO ĐỒ ÁN CUỐI KỲ**  
**THỰC HÀNH LẬP TRÌNH CHO KHOA HỌC DỮ LIỆU**

**Nhóm 13:**

20120530: Trần Thông Lực

20120571: Phùng Hữu Tài

20120588: Lê Quang Thọ

20120614: Nguyễn Anh Tuấn

TP Hồ Chí Minh 2022

**Mục lục**

I.	Phân công công việc và mức độ hoàn thành .....	2
II.	Đánh giá công việc .....	2
1.	20120530 – Trần Thông Lực .....	2
2.	20120571 – Phùng Hữu Tài .....	3
3.	20120588 – Lê Quang Thọ .....	3
4.	20120614 – Nguyễn Anh Tuấn.....	4
5.	Nhóm.....	4

## I. Phân công công việc và mức độ hoàn thành

Công việc	Chi tiết	Phụ trách	Hoàn thành
<b>Thu thập dữ liệu</b>	<ul style="list-style-type: none"> <li>- Tìm chủ đề, nguồn dữ liệu</li> <li>- Trình bày tổng quan về dữ liệu</li> <li>- Trình bày markdown ở notebook</li> </ul>	Tuấn	100%
<b>Khám phá dữ liệu</b>	<ul style="list-style-type: none"> <li>- Đặt câu hỏi khám phá dữ liệu</li> <li>- Thực hiện khám phá dữ liệu theo câu hỏi đã được đặt ra</li> <li>- Tiền xử lý dữ liệu</li> <li>- Trình bày notebook</li> </ul>	Tuấn	100%
<b>Đặt và trả lời câu hỏi</b>	- Dựa vào dữ liệu, tìm hiểu và đặt câu hỏi.	Lực, Tài, Thọ	100%
	- Trả lời câu 1	Tài	100%
	- Trả lời câu 2	Lực	100%
	- Trả lời câu 3	Thọ	100%
	- Trả lời câu 4	Tuấn	100%
<b>Đánh giá cá nhân</b>	- Trình bày khó khăn, những thứ học được trong quá trình làm đồ án, nếu có thêm thời gian sẽ làm gì?	Lực, Tài, Thọ, Tuấn	100%
<b>Làm slide</b>	- Mỗi thành viên tự điền phần thô nội dung mà mình đảm nhận	Lực, Tài, Thọ, Tuấn	100%
	- Tổng hợp, định dạng và trình bày slide	Tài	100%
<b>Làm bảng phân công công việc</b>		Tuấn	100%

## II. Đánh giá công việc

### 1. 20120530 – Trần Thông Lực

- Khó khăn khi làm đồ án:
  - o Đa số cách xử lý đã được học trước trong các lab thực hành nên ít gặp phải khó khăn khi xử lý dữ liệu.
  - o Dữ liệu hơi ít cột để nghĩ ra câu hỏi sáng tạo.

- Đối với loại dữ liệu số thì đơn vị quá lớn và có 1 số giá trị less than và more theo em thì xử lý vẫn chưa phù hợp vì sai số đơn vị vẫn còn lớn
- Những điều học được khi làm đồ án:
  - Sử dụng git nhiều nhánh
  - Thay thế đa số for = các hàm có thể sử dụng trong thư viện
  - Được ôn tập lại kiến thức về các thư viện đã học
  - Học được cách phân tích đơn giản 1 bộ dữ liệu..

## **2. 20120571 – Phùng Hữu Tài**

- Khó khăn:
  - Lúc mới bắt đầu làm thì chưa biết xử lý dữ liệu thô như thế nào.
  - Chưa thể khai thác tốt dữ liệu để đưa ra các câu hỏi hay
  - Còn khó khăn trong việc sử dụng các thư viện để trực quan hóa dữ liệu
- Học được:
  - Cách phối hợp làm việc nhóm với các bạn
  - Cơ bản xử lý được dữ liệu thô
  - Cách xây dựng 1 file notebook logic như trong các lab đã làm trong thực hành
  - Kết hợp các thư viện xử lý dữ liệu với các thư viện trực quan dữ liệu

## **3. 20120588 – Lê Quang Thọ**

- Khó khăn:
  - Khó đặt được câu hỏi để đào sâu dữ liệu đã thu thập
  - Việc xử lý dữ liệu gặp nhiều khó khăn do dữ liệu thu thập không theo một form cố định. Đặc biệt là Job Title
  - Khó thống nhất trong việc tiền xử lý dữ liệu
- Học được:
  - Học được thêm nhiều câu lệnh để xử lý, vẽ biểu đồ trong các thư viện của python

- Đặt câu hỏi để tìm hiểu lợi ích của việc thu thập và phân tích dữ liệu từ đó đưa ra những thông tin mà ta có thể suy ra được từ dữ liệu

#### **4. 20120614 – Nguyễn Anh Tuấn**

- Khó khăn:
  - Khó khăn trong việc tìm câu hỏi
  - Tìm hướng đi của câu hỏi
  - Tốn khá nhiều thời gian cho việc trực quan hóa dữ liệu
  - Gặp nhiều khó khăn khi lần đầu sử dụng github
  - Khá lúng túng khi không biết chọn biểu đồ nào cho phù hợp
- Học được:
  - Cách làm việc nhóm, quản lý công việc với Trello, quản lý source code với github
  - Biết thêm nhiều dạng biểu đồ mới, nhiều cách vẽ mới

#### **5. Nhóm**

Nếu có nhiều thời gian hơn, nhóm sẽ:

- Tìm cách xử lý phù hợp hơn đối với 1 số kiểu dữ liệu của cột.
- Tìm những nguồn dữ liệu hay hơn và sẽ tìm cách để tự cào dữ liệu về
- Tìm hiểu kĩ hơn về dữ liệu để đặt thêm nhiều câu hỏi hay và có ý nghĩa hơn
- Tìm hiểu thêm về các loại biểu đồ và ứng dụng của chúng, cũng như là tìm hiểu thêm cách để làm biểu đồ trông đẹp mắt và truyền đạt nhiều thông tin hơn
- Tìm hiểu thêm về cơ chế pull request của github