

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



**BÁO CÁO ĐỒ ÁN THỰC HÀNH**

**Môn Học: Lập Trình Cho Khoa Học Dữ Liệu**

**GVHD: Thầy Trần Đại Chí – Thầy Lê Nhựt Nam**

Nhóm: 6

Danh sách sinh viên thực hiện:

- |                   |          |
|-------------------|----------|
| 1. Bùi Thanh Tùng | 20120398 |
| 2. Đỗ Tấn Tài     | 20120408 |
| 3. Trần Khắc Bình | 20120437 |
| 4. Lê Văn Hùng    | 20120485 |

Thành phố Hồ Chí Minh - 2022

**Phần 1: Bảng phân công công việc**

STT	Công việc	Người thực hiện	Ngày hoàn thành (dự kiến)	Nội dung
1	Tìm nguồn dữ liệu, đề tài.	Tùng, Tài, Bình, Hùng	13-11-2022	Chọn đề tài Olympic
2	Tiền xử lý dữ liệu.	Hùng, Tài, Bình, Tùng	1-12-2022	Thực hiện như các bước trong file hướng dẫn đồ án
3	Đặt câu hỏi 1,2	Tài	10-12-2022	Top 5 VDV tham gia nhiều năm thế vận hội nhất? Top 5 VDV đạt nhiều huy chương nhất?
4	Đặt câu hỏi 3,4	Tùng	10-12-2022	Quốc gia có thành tích tốt nhất qua các kỳ Olympic mùa hè? Thành tích của Thể thao Việt Nam trong lịch sử tham dự Olympic?
5	Đặt câu hỏi 5, 7	Bình	10-12-2022	- Trong một thế vận hội: ai là người tham gia thi nhiều nội dung nhất, ai là người giành nhiều huy chương nhất? - Mối quan hệ giữa số nội dung tham gia và số huy chương đạt được của mỗi quốc gia qua các kỳ thế vận hội.
6	Đặt câu hỏi 6	Hùng	10-12-2022	Trong những năm gần đây (từ khi Olympic mùa hè và mùa đông bắt đầu xen kẽ nhau), mỗi quốc gia đăng cai liệu có lợi thế chủ nhà hay không ?
7	Làm slide	Tùng, Tài, Bình, Hùng	14-12-2022	
8	Viết bảng báo cáo	Tùng, Tài, Bình, Hùng	14-12-2022	

Lịch trình, tiến độ công việc: <https://painted-almond-8b8.notion.site/L-p-tr-nh-Khoa-H-c-D-Li-u-e570c7d5ad0f4283a6e0249006e3b5ea>

**Phần 2: Tổng hợp lại quá trình thực hiện đồ án**

STT	Họ và tên	Khó khăn mắc phải	Bài học rút ra
1	Bùi Thanh Tùng	Việc không hiểu rõ về thông tin dữ liệu khiến em trả lời sai câu hỏi mà em đặt ra.	Tìm hiểu kỹ về dữ liệu trước khi tiến hành phân tích.
2	Đỗ Tấn Tài	- Nội dung trình bày của nhóm khá dài khiến việc làm slide mất nhiều thời gian hơn dự kiến. - Đặc điểm của dữ liệu có một số điểm đặc biệt khiến việc tiền xử lý gặp một số trục trặc nhỏ.	- Nên phân bố thời gian làm slide hợp lý để tránh việc thiếu thời gian, khiến slide chưa được hoàn thiện tốt nhất. - Nên tìm hiểu kỹ trước thông tin dữ liệu trước khi tiến hành xử lý.
3	Trần Khắc Bình	Phân tích chưa kỹ dữ liệu dẫn đến hiểu sai về dữ liệu (tính sai số huy chương của quốc gia do các môn thể thao đồng đội chỉ tính 1 huy chương cho tập thể). Dữ liệu có nhiều cột nên việc tiền xử lý hơn khó khăn.	Dành nhiều thời gian để phân tích dữ liệu.
4	Lê Văn Hùng	- Nội dung dài nên phải cắt bớt câu hỏi để đủ thời gian trình bày của nhóm và làm slide khá mất thời gian - Việc phân tích dữ liệu khá tốn thời gian vì phải tìm những bài báo nước ngoài và đã cũ để tìm kiểm đúng thông tin	- Phải ước lượng trước độ dài của đáp án định làm cũng như các câu hỏi khác để chọn lọc câu hỏi phù hợp độ dài nội dung. - Trau dồi thêm ngôn ngữ để có thể đọc hiểu báo nhanh hơn

Nếu có nhiều thời gian hơn, nhóm em sẽ:

- Cùng nhau bàn bạc, thảo luận nhiều hơn để tìm hiểu tường minh thông tin dữ liệu mà nhóm đã thu thập. Từ đó tối đa hóa việc xử lý các vấn đề phát sinh trong việc tính toán và tiền xử lý dữ liệu. Loại bỏ các dữ liệu gây nhiễu hoặc các dữ liệu không có giá trị về mặt ý nghĩa.
- Đưa ra các câu hỏi mang tính sâu sắc hơn nhằm phân tích hiệu quả giá trị mà dữ liệu đã cung cấp.
- Hoàn thiện file notebook, tuân thủ clear-coding, viết code ngắn gọn và đơn giản nhất có thể, chú thích tường minh cho từng bước xử lý.

- 
- Hoàn thiện hơn slide thuyết trình về mặt nội dung và thẩm mỹ. Giúp truyền tải có hiệu quả hơn những vấn đề mà nhóm đã đánh giá, tìm hiểu về dữ liệu và muốn truyền tải đến thầy và mọi người.

---

## TÀI LIỆU THAM KHẢO TRONG QUÁ TRÌNH LÀM ĐỒ ÁN

---

- [1]. Mark Joshua, (2018), *Why is winter Olympic Games less popular than summer Games?*, truy cập ngày 29/11/2022, từ <https://www.quora.com/Why-is-winter-Olympic-Games-less-popular-than-summer-Games>
- [2]. SAGAR CHADHA, (2018), *Olympics Data- Cleaning, Exploration, Prediction*, truy cập ngày 2/12/2022, từ <https://www.kaggle.com/code/chadalee/olympics-data-cleaning-exploration-prediction/notebook#Exploring-the-Olympics-dataset>