

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



BÁO CÁO LAB 1

Học phần: **TRỰC QUAN HÓA DỮ LIỆU**

Mã lớp : 20_21

Nhóm sinh viên thực hiện: Nguyễn Quang Hiền	- 20120077
Trần Xuân Hòa	- 20120085
Lê Nguyễn Thanh Hoàng	- 20120088
Lê Chí Nghĩa	- 20120144
Đào Khoa Nguyên	- 20120147

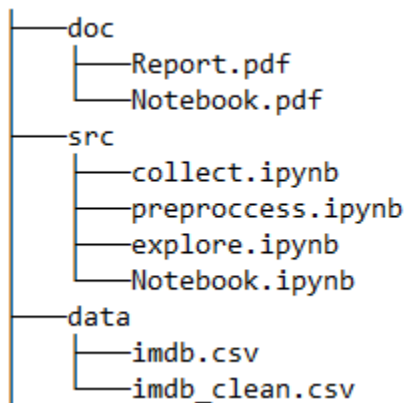
TP.HCM, tháng 4 năm 2022

1. Phân công

Công việc	Phân công
Tìm kiếm và lựa chọn dữ liệu	Hòa
Tiền xử lí dữ liệu	Hoàng
Phân tích dữ liệu trong từng cột	Hiền
Phân tích dữ liệu giữa cột Rate và các cột Votes, Dates, Duration,...	Nguyên
Phân tích dữ liệu giữa cột Genre với cột Rate và Votes	Hoàng
Phân tích dữ liệu giữa các cột Nudity, Violence, Profanity, Alcohol và Frightening	Nghĩa

2. Cấu trúc thư mục nộp bài

Thư mục nộp bài của nhóm có dạng



- Thư mục **doc** chứa:
 - File **Report.pdf**, chứa nội dung báo cáo phân công, cấu trúc thư mục nộp bài và các đường dẫn liên quan.
 - File **Notebook.pdf**, được export từ file **Notebook.ipynb** trong thư mục **src**, chứa toàn bộ code của nhóm đã được convert sang pdf
- Thư mục **src** chứa:

- Các file **collect.ipynb**, **preprocess.ipynb** và **explore.ipynb** chứa code và markdown giai đoạn thu thập, tiền xử lý và khám phá dữ liệu
- File **Notebook.ipynb**, chứa code tổng hợp của 3 file notebook trên.
- Thư mục **data** chứa: Các file **imdb.csv** và **imdb_clean.csv** là dữ liệu trước và sau khi tiền xử lý

3. Các đường dẫn liên quan

- Link của bộ dữ liệu: [tại đây](#)
- Link Github của nhóm: [tại đây](#)