****

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

A picture containing text

Description automatically generated

**CSC10108 – TRỰC QUAN HOÁ DỮ LIỆU**

LAB-01: TRỰC QUAN HOÁ DỮ LIỆU VỚI PYTHON

Icon

Description automatically generated with medium confidence

**Nhóm thực hiện: ID-16**

*TP. Hồ Chí Minh, ngày 01/04/2023*

**MỤC LỤC**

[I. THÔNG TIN NHÓM 2](#_Toc131538131)

[II. MỨC ĐỘ HOÀN THÀNH MỖI YÊU CẦU 2](#_Toc131538132)

[III. TÀI LIỆU THAM KHẢO 3](#_Toc131538133)

# I. THÔNG TIN NHÓM

**Nhóm 16:**

Link dataset:

<https://www.kaggle.com/datasets/rakkesharv/spotify-top-10000-streamed-songs?fbclid=IwAR0I5mLsdM0TxUdUDcqc5WV_9w5dTSpb7lkBGhMev45ywErhkgbY9hX24vs>

Link Github: <https://github.com/PhuongAnh57/LAB1-TQHDL.git>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | HỌ VÀ TÊN | MSSV | VAI TRÒ | MỨC ĐỘ HOÀN THÀNH |
| 1 | Trần Huỳnh Hương | 20120099 | Thành viên | 100% |
| 2 | Nguyễn Dương Tuấn Phương | 20120166 | Thành viên | 100% |
| 3 | Trần Hoài Thiện | 20120195 | Thành viên | 100% |
| 4 | Trần Thị Mỹ Trinh | 20120224 | Thành viên | 100% |
| 5 | Nguyễn Phương Anh | 20120247 | Nhóm trưởng | 100% |

# II. MỨC ĐỘ HOÀN THÀNH MỖI YÊU CẦU

|  |  |  |  |
| --- | --- | --- | --- |
| STT | YÊU CẦU | PHÂN CÔNG | MỨC ĐỘ HOÀN THÀNH |
| 1. Thu thập dữ liệu | | | |
| 1 | Ngữ cảnh, câu chuyện gì khiến nhóm sinh viên thực hiện tìm kiếm dữ liệu? | Huỳnh Hương | 100% |
| 2 | Dữ liệu mà nhóm sinh viên là về chủ đề gì và được lấy từ nguồn nào? | Huỳnh Hương | 100% |
| 3 | Người ta có cho phép sử dụng dữ liệu như thế này không? | Tuấn Phương | 100% |
| 4 | Người ta đã thu thập dữ liệu như thế nào? Phương pháp thực hiện? | Tuấn Phương | 100% |
| 1. Khám quá dữ liệu | | | |
| 1 | Mỗi dòng có ý nghĩa gì? Có vấn đề các dòng có ý nghĩa khác nhau không? | Mỹ Trinh | 100% |
| 2 | Mỗi cột có ý nghĩa gì? | Mỹ Trinh | 100% |
| 3 | Mỗi cột hiện đang có kiểu dữ liệu gì? Có cột nào có kiểu dữ liệu chưa phù hợp để có thể xử lý tiếp hay không? | Phương Anh | 100% |
| 4 | Với mỗi cột, các giá trị (dạng số, dạng phân loại) được phân bố như thế nào? | Phương Anh | 100% |
| 5 | Có cần phải tiền xử lý dữ liệu hay không và nếu có thì nhóm sinh viên cần phải xử lý như thế nào? | Hoài Thiện | 100% |
| 1. Khám quá mối quan hệ trong dữ liệu | | | |
| 1 | Mối quan hệ giữa Days - Total Streams | Phương Anh | 100% |
| 2 | Mối quan hệ giữa Days - Top 10 (xTimes) - Total Streams | Phương Anh | 100% |
| 3 | Các mối quan hệ trong Artist Name | Tuấn Phương | 100% |
| 4 | Mối quan hệ giữa Top 10 (xTimes) – Peak Streams | Tuấn Phương | 100% |
| 5 | Mối quan hệ giữa các trường Artist Name, Song Name, Peak Streams và Total Streams | Huỳnh Hương | 100% |
| 6 | Mối quan hệ giữa Song Name – Total Streams | Huỳnh Hương | 100% |
| 7 | Mối quan hệ trong Artist Name | Mỹ Trinh | 100% |
| 8 | Mối quan hệ giữa Song Name, Peak Position, Peak Position (xTimes) | Mỹ Trinh | 100% |
| 9 | Mỗi quan hệ giữa tất cả các trường | Mỹ Trinh | 100% |
| 10 |  | Hoài Thiện | 100% |
| 11 |  | Hoài Thiện | 100% |

# III. TÀI LIỆU THAM KHẢO

--- HẾT ---