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

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**Logo, icon

Description automatically generated**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**MÔN HỌC: MẠNG XÃ HỘI - IS353.N11.HTCL**

**ĐỀ TÀI:**

**KHẢO SÁT MẠNG CẠNH TRANH VỀ THỂ LOẠI TRÒ CHƠI ĐIỆN TỬ GIỮA NHỮNG NHÀ PHÁT HÀNH**

**Giảng viên hướng dẫn:**

**NGUYỄN THỊ KIM PHỤNG**

**Sinh viên thực hiện:**

**NGUYỄN TRÍ MINH - 19521847**

**Thành phố Hồ Chí Minh, 10/2022**

MỤC LỤC

[MỤC LỤC 3](#_Toc117713021)

[NHẬN XÉT CỦA GIẢNG VIÊN 4](#_Toc117713022)

[QUẢN LÝ TÀI LIỆU 5](#_Toc117713023)

[phân chia công việc 6](#_Toc117713024)

[PHẦN A. XỬ LÝ BẰNG phương pháp LẬP TRÌNH 7](#_Toc117713025)

[CHƯƠNG 1. mở đầu 7](#_Toc117713026)

[1.1. Giới thiệu 7](#_Toc117713027)

[1.2. Tiếp cận bài toán 7](#_Toc117713028)

[1.3. Dữ liệu 7](#_Toc117713029)

[PHẦN b. XỬ LÝ BẰNG phương pháp THỦ CÔNG 9](#_Toc117713030)

NHẬN XÉT CỦA GIẢNG VIÊN

*……., ngày……... tháng…… năm 2022*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên)*

QUẢN LÝ TÀI LIỆU

|  |  |  |  |
| --- | --- | --- | --- |
| **Thời gian** | **Phiên bản** | **Người sửa đổi** | **Chi tiết** |
| 26/10/2022 | v1.0 | Nguyễn Trí Minh | Khởi tạo tài liệu |
| 27/10/2022 | v1.1 | Nguyễn Trí Minh | Thêm nội dung và chỉnh sửa |
| 31/10/2022 | v1.2 | Nguyễn Trí Minh | Thêm nội dung và chỉnh sửa |

phân chia công việc

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Nội dung công việc** | **Tỉ lệ đóng góp (%)** |
| 1 | Nguyễn Trí Minh | 19521847 | Thực hiện đồ án | 100 |

PHẦN A. XỬ LÝ BẰNG phương pháp LẬP TRÌNH

Giới thiệu về phương pháp lập trình:

* Ngôn ngữ sử dụng: Python.
* IDE: Visual Studio Code.
* Extension: Jupiter và Jupiter Keymap, cho phép thực thi Jupiter Notebook trực tiếp trên Visual Studio Code.

# mở đầu

## Giới thiệu

Ngày nay, nhu cầu giải trí đa phương tiện là nhu cầu không thể thiếu đối với mỗi chúng ta. Từ lâu, việc chơi các trò chơi điện tử (hay video games), vốn là một thành phần quan trọng trong nhiều thể loại giải trí đa phương tiện, luôn là một phương thức hiệu quả để giúp chúng ta giải trí sau những giờ làm việc hay học tập căng thẳng. Thị trường trò chơi điện tử luôn có nhiều biến động bởi nhiều tựa game hay, kinh điển và hấp dẫn tương ứng với nhiều thể loại game khác nhau (như Action, Sports, Platform,…) được phát triển và phát hành bởi các nhà phát hành game nổi tiếng trên khắp thế giới như Electronic Arts, Nintendo, Sony,... Với việc các tựa game nổi tiếng đó luôn được săn đón bởi các người chơi hay game thủ trên toàn cầu với thứ hạng theo doanh số bán dẫn đến các nhà phát hành game luôn phải cạnh tranh nhau trên các thể loại game nói trên thông qua các sản phẩm game mà họ phát hành. Chính vì thế, môn học *Mạng xã hội* trong phạm vi của môn học có thể giúp cho chúng ta phân tích **mạng cạnh tranh thể loại trên thị trường game** giữa các nhà phát hành thông qua các tựa game của họ bằng các độ đo (dựa trên đường đi, phổ, khoảng cách,...) và khai thác dữ liệu trên mạng này (bằng phương pháp gom cụm) một cách có hiệu quả.

## Tiếp cận bài toán

* Input: Tập dữ liệu ban đầu trên nguồn dữ liệu Kaggle đã qua tiền xử lý dữ liệu.
* Output: Đưa ra độ đo, đưa ra cộng đồng, phục vụ cho việc phân tích mạng xã hội - “bảng xếp hạng các tựa game theo doanh số bán”.

## Dữ liệu

* + 1. **Giới thiệu nguồn dữ liệu**
* Nguồn dataset: *Video Game Sales*, <https://www.kaggle.com/datasets/gregorut/videogamesales>.
* Dữ liệu “Video Game Sales” được cung cấp trên nền tảng Kaggle chứa dữ liệu các tựa game được sắp xếp theo thứ hạng doanh số bán qua nhiều năm. Dữ liệu gồm 16598 dòng với 11 thuộc tính bao gồm:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| 1 | Rank | int | Thứ hạng của game |
| 2 | Name | char | Tên game |
| 3 | Platform | char | Hệ máy |
| 4 | Year | int | Năm phát hành |
| 5 | Genre | char | Thể loại game |
| 6 | Publisher | char | Nhà phát hành |
| 7 | NA\_Sales | float | Doanh số khu vực Bắc Mĩ |
| 8 | EU\_Sales | float | Doanh số khu vực Châu Âu |
| 9 | JP\_Sales | float | Doanh số ở Nhật Bản |
| 10 | Other\_Sales | float | Doanh số khu vực khác |
| 11 | Global\_Sales | float | Doanh số toàn cầu |

* + 1. **Xử lý và phân tích dữ liệu**
* Làm sạch dữ liệu: Tiến hành loại bỏ dữ liệu bị thiếu và dữ liệu trùng:



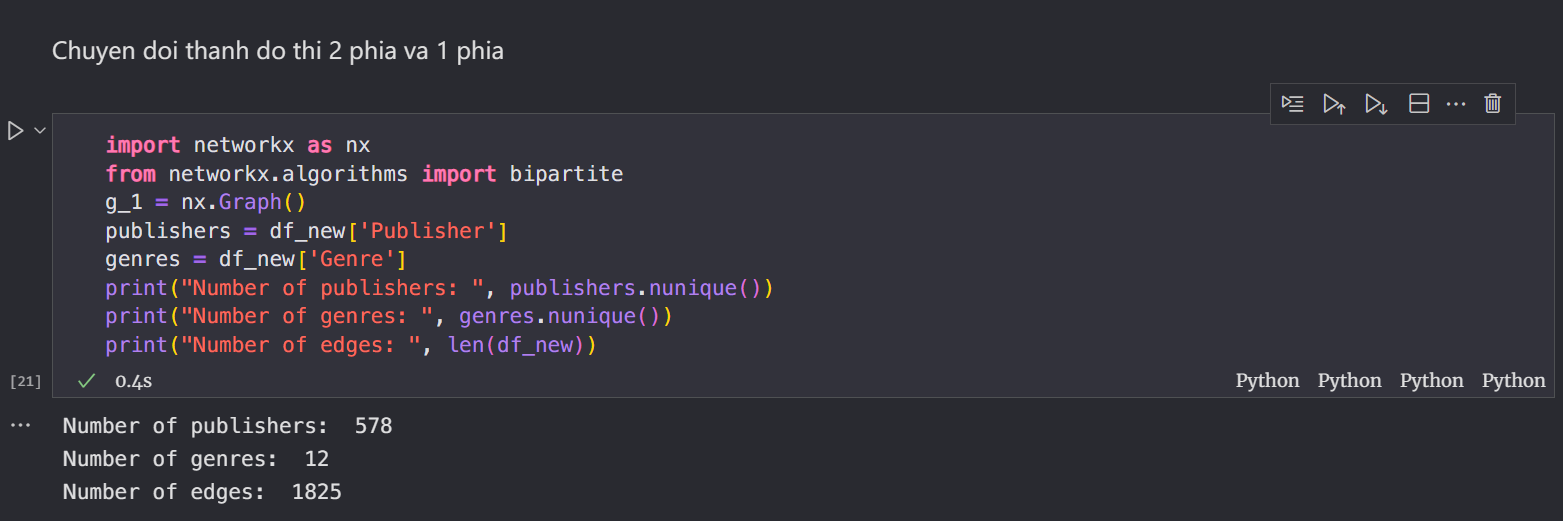
Graphical user interface, application

Description automatically generated

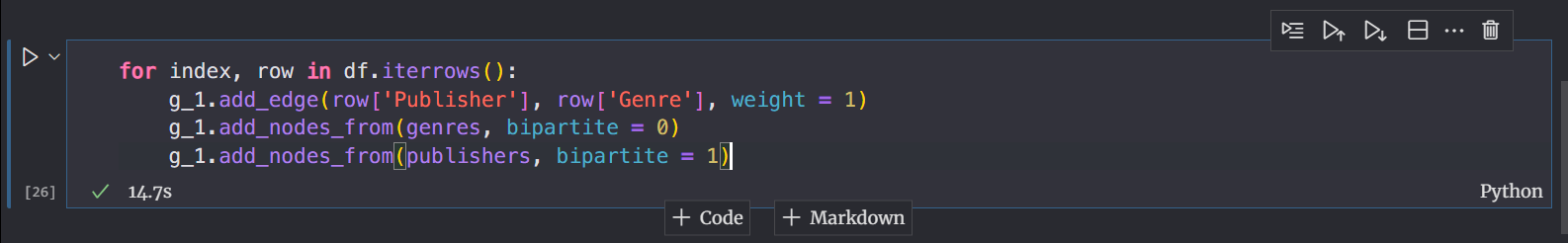
Dữ liệu được đọc từ file “vgsales.csv” đưa vào dataframe. Sau khi xoá bỏ dữ liệu trùng và chỉ lấy 2 cột “Genre” và “Publisher” thì được dataframe mới có 1825 dòng và 2 cột.

Chuyển đổi dataframe thành đồ thị:

* Đồ thị 2 phía:
* Node: Nhà phát hành và thể loại, edge: mối quan hệ giữa việc nhà phát hành cho ra mắt tựa game thuộc thể loại tương ứng nào đó.
* In ra số nhà phát hành, số thể loại và số cạnh:



* Đưa dữ liệu vào đồ thị 2 phía:



* Code hiển thị đồ thị 2 phía:

A picture containing chart

Description automatically generated

* Hiển thị đồ thị 2 phía:
* Nhận xét: Nhìn vào đồ thị có thể thấy 1 ca sĩ hay nghệ sĩ dường như chỉ phát hành nhiều bài hát ứng với 1 thể loại xuyên suốt cố định nhưng 1 thể loại nhạc có thể thuộc về nhiều bài hát từ nhiều ca sĩ khác nhau.
* Đồ thị 1 phía:

PHẦN b. XỬ LÝ BẰNG phương pháp THỦ CÔNG