



UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Công nghệ dữ liệu lớn

Báo cáo đồ án

Hệ Thống Gợi Ý Anime Dựa
Trên Nội Dung Sử Dụng
PySpark và Kafka

Nhóm thực hiện: nhóm 8

- 21520565 - Lê Quốc Anh
- 21522207 - Nguyễn Đoàn Nhật Khánh

GVHD: TS. Đỗ Trọng Hợp



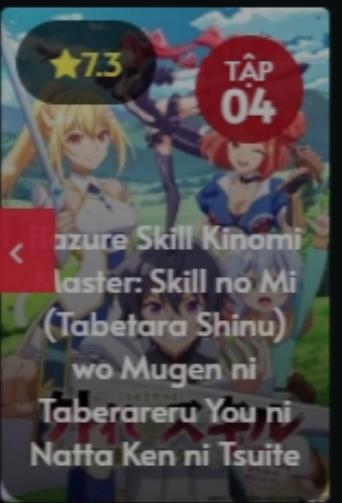
Mục lục

- ▶ Mục tiêu 01
- ▶ Bộ dữ liệu 02
- ▶ Sơ đồ luồng dữ liệu - DFD 03
- ▶ Code 04
- ▶ Demo 05
- ▶ Kết luận 06



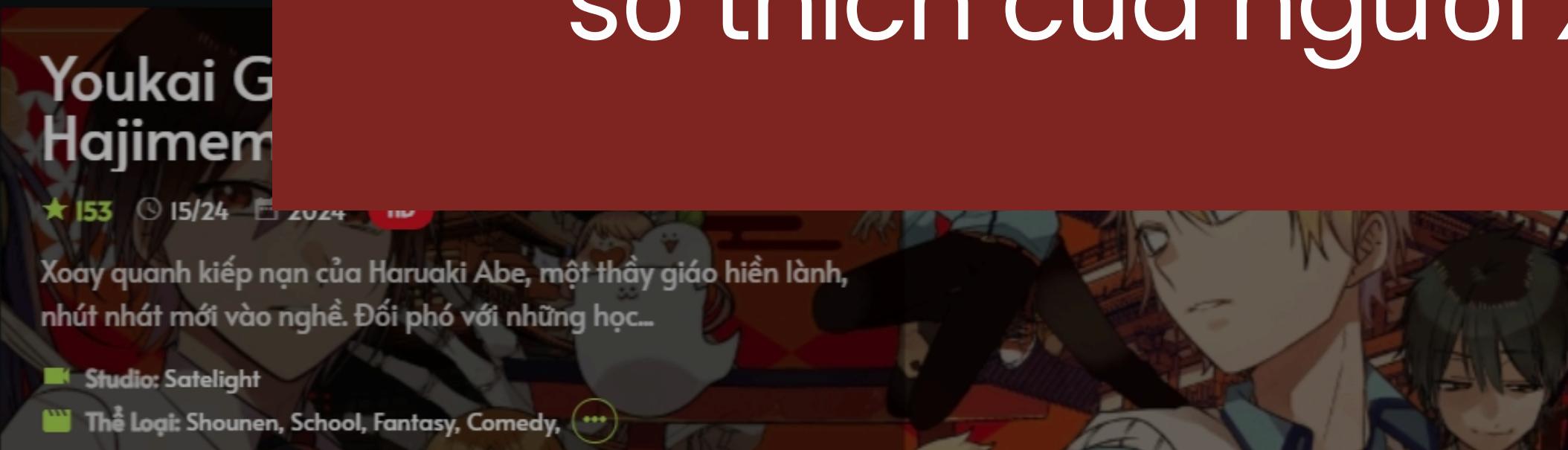


- Lưu hoặc nhớ ngay link rút gọn bit.ly/animevietsubtv để truy cập khi nhà mạng chặn!
- Mời bạn tham gia Group [tại đây!](#) hoặc tham gia Discord [tại đây!](#) để ủng hộ AnimeVietsub
- Lấy dữ liệu



Mục tiêu

Gợi ý những bộ anime phù hợp
sở thích của người xem



★ 153 ⏸ 15/24 ⏰ 2024

Xoay quanh kiếp nạn của Haruaki Abe, một thầy giáo hiền lành, nhút nhát mới vào nghề. Đối phó với những học...

Studio: Satelight

Thể Loại: Shounen, School, Fantasy, Comedy,

▶ Xem Anime Ngẫu Nhiên

Hỏi/đáp anime

Bộ dữ liệu

Before cleaning:

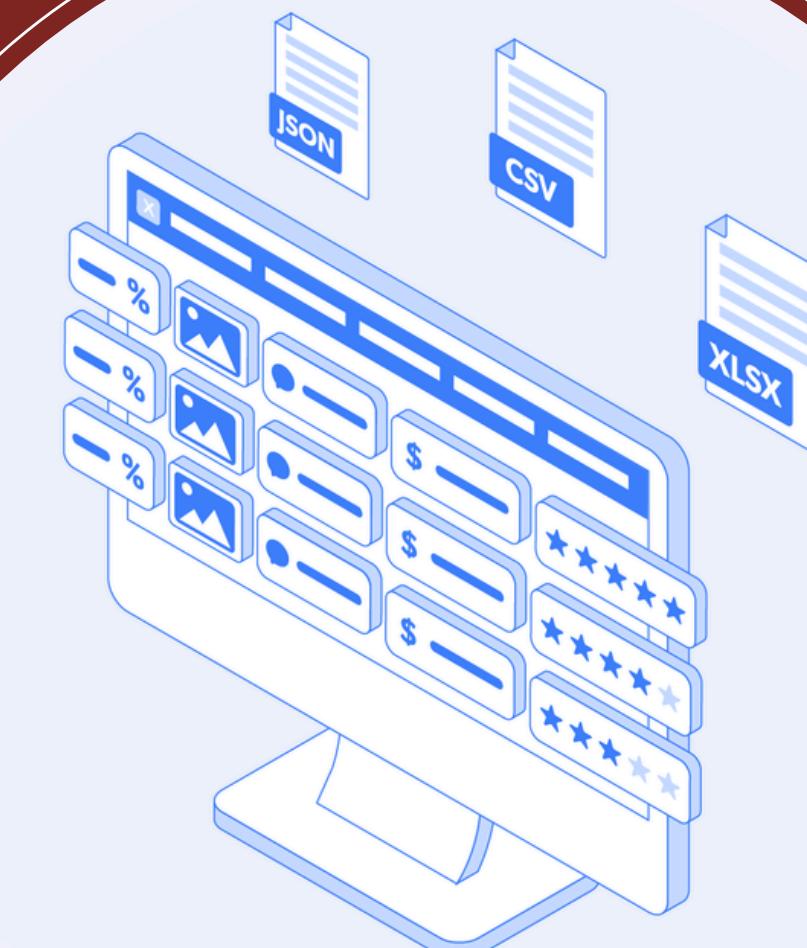
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5389 entries, 0 to 5388
Data columns (total 12 columns):
 #   Column      Non-Null Count Dtype  
--- 
 0   title        5389 non-null   object  
 1   link         5389 non-null   object  
 2   image        5389 non-null   object  
 3   episode       4146 non-null   object  
 4   rating        5389 non-null   object  
 5   views         5389 non-null   object  
 6   quality       5347 non-null   object  
 7   year          5378 non-null   object  
 8   genres        5359 non-null   object  
 9   description    5351 non-null   object  
 10  season        5389 non-null   object  
 11  type          5389 non-null   object  
dtypes: object(12)
memory usage: 505.3+ KB
```



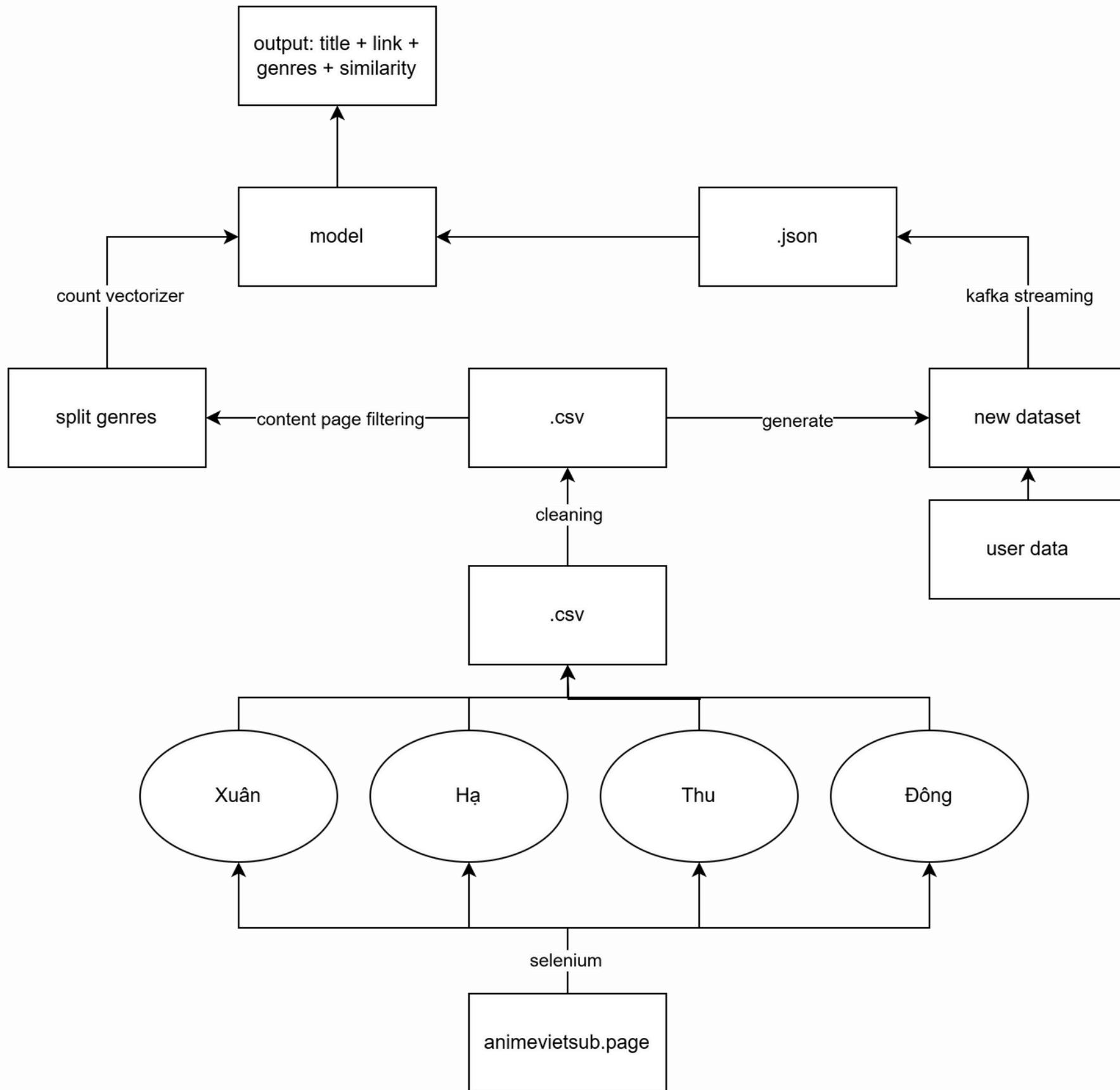
Bộ dữ liệu

After cleaning:

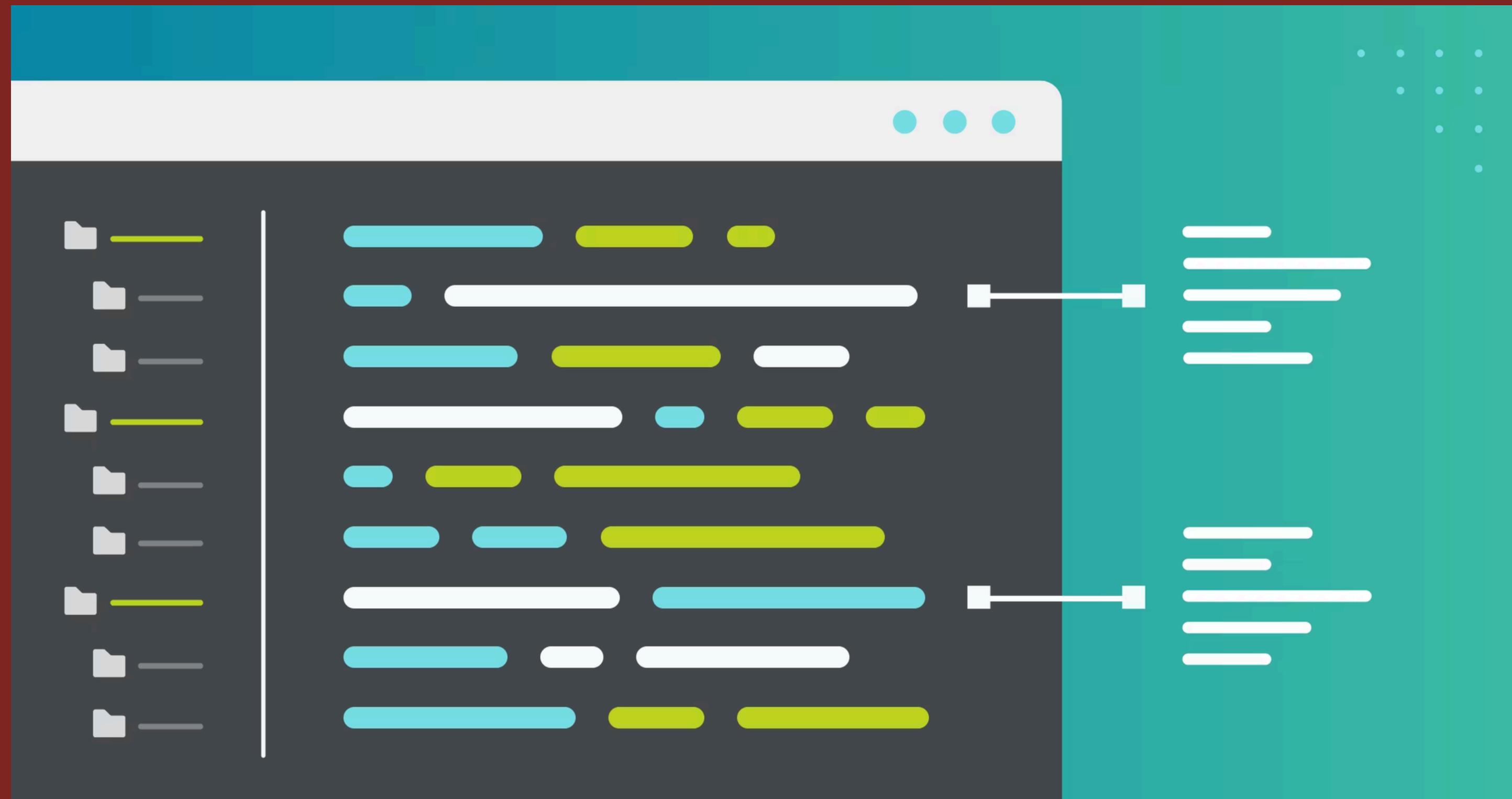
```
... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 5389 entries, 0 to 5388
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   title       5389 non-null    object  
 1   link        5389 non-null    object  
 2   image       5389 non-null    object  
 3   episode     5389 non-null    object  
 4   rating      5389 non-null    object  
 5   views       5389 non-null    int64  
 6   quality     5389 non-null    object  
 7   year        5389 non-null    object  
 8   genres      5389 non-null    object  
 9   description 5389 non-null    object  
 10  season      5389 non-null    object  
 11  type        5389 non-null    object  
 12  rate        5389 non-null    float64 
 13  nums_of_vote 5389 non-null    int64  
dtypes: float64(1), int64(2), object(11)
memory usage: 589.5+ KB
```



Sơ đồ luồng dữ liệu - DFD



Code



Demo



Kết luận

Qua những gì đã làm, nhóm xin đưa ra các nhận định về đồ án và quá trình thực hiện đồ án dưới đây.

01

Những điều đã làm được

02

Những điều chưa làm được

03

Hướng phát triển



UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

THANK YOU

