MDS 5102 Python Programming

Course Project on **Analysis of video platforms**

Due Date: 23:59 pm, 28 December 2021

1. Introduction

Media and Video streaming platforms are getting more and more popular in recent years and many companies are following this trend. In this project, we give you 4 tabular datasets consisting of information of movies or tv shows available on 4 top-tier video streaming platforms, including Amazon Prim, Disney+, Hulu, Netflix. You are supposed to analysis these four platforms using given datasets or compare them with each other.

2. Video Streaming Platforms

Hulu is an online movie and tv shows streaming platform owned by The Walt Disney Company, launched on 2007. Currently it is only available in the United States. As of the fourth quarter of 2021, Hulu has 43.8 billion subscribers.

Disney+ is an American subscription video on-demand over-the-top streaming service launched on 2019. It is available in America, Europe, India and Australia, and owns 118.1 million global subscribers as of October 2, 2021.

Amazon Prime is a paid subscription program from Amazon which is available in various countries and gives users access to additional services otherwise unavailable or available at a premium to other Amazon customers. In April 2021, Amazon reported that Prime had more than 200 million subscribers worldwide.

Netflix is an American subscription streaming service and production company. Launched on August 29, 1997, it offers a library of films and television series through distribution deals as well as its own productions, known as Netflix Originals. As of October 2021, it has over 214 million subscribers worldwide.

3. Dataset Introduction

The given datasets are contained in four CSV files, named in format of "{company}_titles.csv". Except the first row, each row has information of one movie / TV show. Every CSV file has 12 columns separated by comma, whose name and details are shown below. 2 consecutive commas mean one empty cell in the middle.

Column	Explanation
show_id	A unique id for each item
type	Type of this item, Movie or TV show
title	Name of this item
director	Names of directors of the item
casts	Names of main actors of the item

country	Country of production
date_added	Date added on the platform
release_year	Original release year of the item
rating	Rating of the item
duration	Total duration of the movie or tv show
listed_in	Which type is the content of the item belongs to
description	A brief description of each item

* All datasets are scraped from Internet. Note that not all cells are filled and some data may be noisy.

* Meaning of different rating are explain at the end of this document.

4. Project Requirements

You are supposed to use given datasets to conduct some analysis for the streaming platforms and/or compare among all platforms. Your analysis can include but not limited to following aspects.

(1) characteristics of shows in every platform
(2) operating strategies of each platform
(3) difference of contents available in different countries
(4) trend of movies and TV shows in online platforms

You are encouraged to discover more features from the datasets. The analysis should be conducted based on proper data visualization and processing. And you should give conclusions and show evidences to support your ideas.

Not all data is required to be used. Choose column or data if you find it useful for your analysis.

5. Assessment

This project is an individual project and owns 20% of total points in this course.

You are required to submit a zip file containing a well-documented **project report** and **relevant codes**, named as "proj_SID.zip". Your performance in this project will be assessed based on various aspects such as data processing methods, visualization result, conclusions you find and presentation skills of your report. **Any Plagiarism will NOT be tolerated.**

The due date of this project is **23:59 pm, 28 December 2021** (Tuesday). No late submission is allowed.

Your report should follow the format of a NeurIPS2021 submission (camera-ready) and we recommend you to finish this report using LaTex. The template PDF, LaTex code and style file are attached, and please read carefully for more details.

To edit LaTex file, you can use the free online editor **overleaf** by uploading your LaTex code, style file and images to your online project. Follow the checklist before submitting your work.

References

Television content rating system - Wikipedia
(https://en.wikipedia.org/wiki/Television_content_rating_system)
Motion Picture Association film rating system - Wikipedia
(https://en.wikipedia.org/wiki/Motion_Picture_Association_film_rating_system)

Appendix

Overleaf
(https://www.overleaf.com)
Latex Guidance on Overleaf
(https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes)

Rating	Meaning
TV-Y	Aimed at very young ages including 2-6
TV-Y7	Designed for children age 7 and above
TV-Y7-FV	"FV" for Fantasy violence that may be more intense or combative than other TY-Y7 programs.
TV-G	Most parents would find it suitable for all ages
TV-PG	Parental guidance is recommended; maybe unsuitable for younger children
TV-14	Many parents would find it unsuitable for children under 14
TV-MA	Intended to be viewed by adults; unsuitable for children under 17
G	All ages admitted
PG	Parental guidance suggested; maybe not suitable for children
PG-13	Parents Strongly Cautioned; maybe improper for children under 13
R	Restricted; Under 17 requires accompanying parent or adult guardian
NC-17	Adults Only; No one 17 and under admitted.
NR / Not Rated	Not Rated; has not been submitted for rating
UR / Unrated	Unrated; an uncut version of a production that was rated
ALL / ALL_AGES	All ages admitted
18+ / AGES_18_	Not recommended for viewers under the age of 18
16+	Not recommended for viewers under the age of 16
13+	Not recommended for viewers under the age of 13
7+	Not recommended for viewers under the age of 7