

## Dream to Stream Stage 1 Project Proposal

1. Describe what data is stored in the database. (Where is the data from, and what attributes and information would be stored?)

The dataset contains movie/tv show information on Disney+, Netflix, Prime Video, and Hulu. Categories include Title, Id, year made, genre, length and more.

Link to data:

- Movies on Netflix, Prime Video, Hulu, and Disney Plus:  
<https://www.kaggle.com/datasets/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>
  - Netflix Movies & TV Shows:  
<https://www.kaggle.com/datasets/shivamb/netflix-shows>
  - Prime Video Movies & TV Shows:  
<https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows>
  - Hulu Movies & TV Shows:  
<https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows>
  - Disney+ Movies & TV Shows:  
<https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows>

Attributes to use:

show\_id, title, date\_added, release\_year, genre, length, type, rating, rotten\_tomatoes, duration

2. What are the basic functions of your web application? (What can users of this website do? Which simple and complex features are there?)
  1. Users will be able to create their own account
  2. With their account they can rate a Netflix/Hulu/Prime/Disney+ show/movie
  3. After ratings have been made they can view other users' ratings
  4. Complex: Users can make complex aggregate database searches to find movies with very specific filters and characteristics (genre, platform it's on, age\_rating, location). Also, combining multiple databases together to connect all the data to the streaming services will be used to help filter the data
3. What would be a good creative component (function) that can improve the functionality of your application? (What is something cool that you want to include? How are you planning to achieve it?)

Tools to use: ReactUI or Python Flask App for the frontend, GCP to have cloud experience and deploying a cloud application and hosting the backend

1. The plan is to create a web application using the tools mentioned above, that will display all the top movies/shows for all platforms (Netflix/Prime Video/Hulu/Disney+) for a given time period based on user ratings. Users can create accounts to provide their own ratings and get recommendations on previous things that users have watched and rated on before.
2. We could potentially combine multiple datasets to find more information on the media selections across several streaming platforms. One example would be adding a rotten tomatoes component that would compare entered user data to rotten tomato data. This would be done with some ETL to combine the datasets together.
3. We could potentially utilize location data of publishment to discover trends with time

4. Project Title: Dream to Stream

5. Project Summary:

Our project is a website that is used to rate different shows/movies on multiple streaming services. When you open Netflix or any other streaming service, it will show you the top ten movies in your region, but that is based on data they collect from show interaction. This may not always be a great representation of what people actually thought of the show. Therefore, we want to make a website that can show you any number of shows/movies from multiple categories and allow you to rate the media you have watched.

This website will be different from other sites like Rotten Tomatoes because it is not for movie critics to provide a whole explanation of the movie, rather just a simple rating based on their opinion. Social media generates a lot of hype around shows/movies and this site will help users determine what is actually worth the hype. Our application hopes to provide users an enjoyable experience while browsing for what their next watch is without having to worry about taking a risk on a “hyped-up” series and wasting their time watching something they might not enjoy.

6. **Description** of an application of your choice. State as clearly as possible what you want to do. What problem do you want to solve, etc.?

This application will allow users to create their own account to rate shows or movies and view other ratings of similar shows/movies. It will also allow users to filter the ratings by a variety of categories including genre, to platform, year, etc. This will allow users to see an overall rating of shows before deciding to watch and also

allows users to give current feedback for a show at any given time. Without a rating system, it is difficult to predict whether a show or movie will be good, so this application will be useful in terms of giving users a better understanding of what to expect before they decide on a particular show or movie.

On the technical side we will first start with data preprocessing, involving extracting, transforming, and loading the data into a database. We will use the pandas library in python to format the data as needed after downloading the data in csv format. The Pandas dataframe will be exported to a database hosted on google cloud. From here we can split the data into tables based on attributes like genre. We will also prepare stored procedures like aggregates for groupings that the users may want to see on the front end. Finally we will design the UI using Python Flask and React app by having a backend connecting to the database on Google Cloud and hosting a Web Page as the UI.

7. **Usefulness.** Explain as clearly as possible why your chosen application is useful. Make sure to answer the following questions: Are there any similar websites/applications out there? If so, what are they, and how is yours different?

The current problem with streaming platforms is that they are currently suggesting Top 10 Movies/Shows for the current time based on user clicks/selections - which does not indicate whether a user would enjoy watching the current selection. There are many factors that would play into this, including social media hype or a new release, that could cause it to be the most watched. Also, with so many streaming platforms, users have to open up all four individually to find new releases to watch without a centralized location to see all movies/shows and to filter on them.

Our application would help resolve this issue by allowing users to see rankings of the movies/shows in a centralized location and also be able to see user ratings as well as Rotten tomato ratings to understand better if this is going to be something worth watching. The benefit of our application is that it provides streamers one page that combines Netflix, Hulu, Prime Video, and Disney+ and save regular users hours every year of mindlessly browsing the streaming services to find what to watch, so they can focus on finding a good show/movie to watch at any given time.

8. **Realness.** Describe what your data is and where you will get it.

The data will be provided via Kaggle (all linked above in Question 1) and there is 1 big table that has all the movies on Netflix, Prime Video, Hulu, and Disney+ that will link all the streaming services together. Then we also have datasets for each of the streaming services that provides metadata on each show/movie available (genre, rating, Rotten Tomato ranking, duration, Movie/Show). Combining

these data together via a primary key (name and provider), this will link the main database with all the metadata for each show/movie.

Also, the team is planning to create a User's Sign In table that will contain all the data for user accounts and their movies/shows watched as well as the ranking for it.

9. Description of the **functionality** that your website offers. This is where you talk about what the website delivers. Talk about how a user would interact with the application (i.e., things that one could create, delete, update, or search for). Read the requirements for stages 4 and 5 to see what other functionalities you want to provide to the users. You should include:

1. **A low-fidelity UI mockup:** What do you imagine your final application's interface might look like? A PowerPoint slide or a pencil sketch on a piece of paper works!

***Dream to Stream***

Sign In

Search.... GO

Title	Movie/Show	Platform	Year	Duration	Genre
The Office	Show	Hulu	2006	8 Seasons	Comedy
<u>OuterBanks</u>	Show	Netflix	2018	3 Seasons	Adventure
Shutter Island	Movie	Netflix	2014	2 Hours	Psychological Thriller
Mandalorian	Show	Disney+	2021	2 Seasons	Sci-Fi

2. **Project work distribution:** Who would be responsible for each of the tasks or subtasks?

List of the person responsible for which exact functionalities in section 6. Explain how backend systems will be distributed across members. Be as specific as possible as this could be part of the final peer evaluation metrics.

The work has been broken up into 3 sections, Data Preprocessing, Google Cloud, and Frontend UI. As a group with even distribution amongst City Scholars and MCS students, we have decided to break up into two groups for the backend work and combine our minds together to work on the Frontend UI. Below is the list of tasks that the team has agreed upon for main areas of focus:

Data Preprocessing: (Rohan, Divya)

- Extract
- Transform
- Load

Google Cloud: (Anthony, Mohammad)

- Database Architecture (Table Design)
- DB Aggregates

UI (Rohan, Mohammed, Divya, Anthony):

- Python Flask-App
- React UI

It is worth mentioning that the team will be meeting 2x a week to provide updates on the progress of their section as well as using this time to debug problems as a team and provide each other with details on the tasks each group member has contributed to have better visibility for progress and to provide everyone with a learning experience in all areas of focus.