Movie Recommendation Application

Project Proposal by: Himani Kaushik

- **Question/need:**
  - What is the question behind your analysis or model and what practical impact will your work have?
    - The purpose of the model is to recommend similar movies based on the content. This model will apply the text transformation and modeling techniques to create a movie recommender system and then deploy the entire system using a web application.
  - Who is your client and how will that client benefits from exploring this question or building this model/system?
    - A client wants to launch a new web-based application for recommending movies to its subscribers. The client wants to utilize the features of the movies to build the recommendation system. This project will ensure an effective and streamlined process to recommend movies based on the content of the movies.

- **Data Description:**
  - What dataset(s) do you plan to use, and how will you obtain the data?
    - The data will be obtained from the kaggle website - https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata .
    - There are two csv files - tmdb_5000_movies.csv and tmdb_5000_credits.csv. Both the files contain details for 5000 movies. The movies database has 20 columns for 5000 movies and credits database has 4 columns for 5000 movies.
  - What is an individual sample/unit of analysis in this project?
    - The individual sample for movies database has 20 fields, including these:
      - budget
      - genres
      - id
      - original_language
      - release_date
      - revenue, etc.
    - The individual sample for credits database has 4 fields:
      - movie_id
      - title
      - cast
      - crew

- **Tools:**
  - How do you intend to meet the tools requirement of the project?
    - Following Python tools would be used:
      - Pandas and Numpy: For cleaning data and preprocessing.
      - SQLite: For storing the data
      - NLTK, scikit-learn: For preprocessing and modeling data.

- Visualization tools: For creating the web application. Still deciding the best tool to use.

- **MVP Goal:**
  - What would a minimum viable product (MVP) look like for this project?
    - MVP for the project would include results and data visualizations of the initial runs of the recommendation system.

- **Data Pipeline:**