Project Proposal: Movie Popularity and Rating Trends Analysis

Objective: Analyze movie rating data to understand trends in movie popularity and user ratings over time.

Data Source:

https://www.kaggle.com/datasets/aaditya1gautam/tmdb-top-rated-9000-movies-dataset

Suggested Tasks:

1. **Data Cleaning**: Preprocess the data to handle missing values, outliers, and duplicates.
2. **Movie Rating Analysis**: Analyze the average ratings (vote_average) and identify the highest and lowest rated movies.
3. **Popularity Analysis**: Study the relationship between movie popularity (popularity) and ratings, as well as the number of votes (vote_count).
4. **Genre Analysis**: Use genre data (genre_ids) to analyze the ratings and popularity of different movie genres.
5. **Temporal Analysis**: Explore trends in movie ratings and popularity across different release years.
6. **Data Visualization**: Visualize the distribution, ratings, and popularity of movies using charts and graphs.
7. **Recommendation System Basics**: Attempt to recommend movies to users using a simple algorithm.

Requirements:

- Use Python for data processing and analysis.
- Employ the Pandas library for data manipulation.
- Use Matplotlib, Seaborn, or Plotly for data visualization.
- Document the analysis process and results in a Jupyter Notebook.
- Submit a report that includes code, analysis results, and visualizations.

Outcomes:

- Students will learn to clean and analyze data using Python.
- Students will understand the characteristics of movie data and how to conduct popularity and rating analyses.
- Students will learn to use data visualization tools to present analysis results.
- Students will gain basic knowledge of recommendation systems and how to build simple recommendation algorithms.