Introduction:

I was an avid anime fan and this lead me to pursue this project of creating a recommendation system for animes a user may enjoy based on their enjoyment of similar animes and similar people. This made me try to create both a content based filtering system and a collaborative filtering system to recommend animes to the user.

using various data analysis and machine learning techniques to gain insights and make recommendations based on user ratings.

What I learnt:

- o How to replace missing or sub-optimal values in a dataset
- o Creating pairplot to compare numerical data and distributions
- o Use pandas and collections' counter to analyse frequency of values of multiple elements in strings
- o Create content based recommendation system using SVD and KNN

Included Files:

Notebooks:

- o Anime Recommendation System-final.ipynb(Notebook for cleaning the anime and ratings datasets that will be used,
- o Exploring the data,
- o containing the code for the content based recommender

Datasets:
- o anime.csv (Original anime dataset),
- o rating.csv (Original user and rating data)

Content:

Anime.csv

anime_id - myanimelist.net's unique id identifying an anime.

name - full name of anime.

genre - comma separated list of genres for this anime.

type - movie, TV, OVA, etc.

episodes - how many episodes in this show. (1 if movie).

rating - average rating out of 10 for this anime.

members - number of community members that are in this anime's "group".

rating.csv

user_id - non identifiable randomly generated user id.

anime_id - the anime that this user has rated.

rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating)

Tools:

The following Python libraries were used in this project:

Pandas: for data manipulation and analysis

NumPy: for numerical computing

Matplotlib and Seaborn: for data visualization

Tk interface: is the standard Python interface to the Tcl/Tk GUI toolkit

Scikit-learn: for machine learning models and evaluation metrics

Surprise: for implementing the SVD model.

Data Cleaning and Preprocessing:

The dataset was first cleaned and preprocessed to remove missing values, duplicate rows, and outliers.

The data was also transformed to make it suitable for machine learning models, such as one-hot encoding for categorical variables and scaling for numerical variables.

Exploratory Data Analysis:

Exploratory data analysis was performed to gain insights into the data using various data visualization techniques such as histograms, box plots, and scatter plots. The analysis focused on identifying trends and patterns in the data, such as the distribution of ratings and the relationship between ratings and genre.

Machine Learning Models:

Two machine learning models were implemented in this project for recommending anime to users based on their ratings:

1) **K-nearest neighbors (KNN) algorithm:** This model uses a KNN algorithm to find similar users based on their ratings of different anime and recommends new anime to watch based on their ratings.
2) **SVD model:** This model uses Singular Value Decomposition to recommend anime to users based on their ratings.
3) **K-medoids clustering algorithm:** This model uses a K-medoids clustering algorithm to group anime into clusters based on their ratings and recommends new anime to watch based on the clusters.

Conclusion:

This project demonstrates the use of various data analysis and machine learning techniques to gain insights and make recommendations based on user ratings of anime. The project can be extended to include other machine learning models and evaluation metrics to further improve the accuracy of the recommendations