**Slide 1**

Good morning or afternoon, for my data mining course project, I developed an anime recommendation system.

**Slide 2**

* As the title suggests, my goal for this project was to create a recommendation system to recommend anime shows.
* I was inspired to do this because I wanted to help those who are new to anime or interested in it, find more interesting and lesser-known shows outside of the more popular ones.
* Done by using a weighted hybrid recommendation system algorithm.
* Which uses both content-based filtering and collaborative filtering recommended systems.
* And weights the results to output a final score.
* Along with the pandas library to work with the data, sklearn library to process the data and matplotlib library to plot the data.

**Slide 3**

* I used a dataset found on Kaggle, mined from an anime review and rating site, called MyAnimeList.
* Similar to IMBD, as its a database.
* The Kaggle database contains:
* A dataset of the animes and all the relevant information about it.
* The dataset of the user profiles
* The dataset of each user’s anime list, which is the progression for each show a user followed or favourited.
* For this project, I only used the dataset of the animes and its relevant information
* Contains the title of a show, the average score, the number of users who rated the show, the number of favourites and follows, the ranking, the popularity ranking, and many more classes.

**Slide 4**

* When I imported the csv file, the first thing I did, was to filter out any shows which had less than 100 votes.
* Then using the average scores and number of votes, I calculated the weighted average using this formula.
* Which is the formula for weighted averages.
* I also had to reverse the order of the popularity class because it was a ranking instead of an actual score.
* This is so that the most popular show in the dataset now has the largest popularity score.
* Then, I normalized the weighted averages, popularity score and number of favourites between 0 and 1.
* And finally, I calculated the final scores with 33% priority between the three normalized classes.

**Slide 5**

* Once the scores are calculated, the list is sorted with descending order of final score.
* But some of the instances did not even have a final score, nor did it have any normalized values.
* And so, these instances were filtered out.
* Then I split the results into 6 output files, each corresponding to the six anime types or categories.
* Which included all the relevant information about each show.
* Which are: TV, Movies, OVAs, Specials, ONAs and Music.
* And then, each of these are split further into only the first 50 instances.
* So, in other words, the resulting output showed the top 50 recommended shows, in each category.

**Slide 6/END**

And that is the end of my presentation. Thank you.