

Feb 15 Meeting Minutes, 2025 | CSC 442 Team Project Meeting

Attendees

- Anthony Brown
- Nikhil Vasudeva
- Fisher Eskew

Notes

- The team reviewed the requirements for **Homework 5b** and identified tasks to complete prior to the deadline
- The team examined different strategies to merge the datasets including left, right, outer, and inner join.
- Based on feedback received during the dataset review session and on the results of the different merging strategies, the team decided that an inner join would be best.
- The team experimented with merging on multiple column subsets and found that merging on film titles may be insufficient because a new film with a duplicate title could be a reboot or continuation of a film franchise.
- To address the above point, the team decided to merge both on film title and release year. The reasoning behind this decision is that it is highly unlikely for two distinct films to be released in the same year.
- The team found that the merged dataset contained inconsistent release year information. For example, a movie released in January 2017 was included in either dataset as released in 2016 or 2017. This affects at most hundreds of entries in the merged dataset which contains thousands of entries. Some of these errors could be manually corrected.

Action items

Task	Deadline	Person Responsible	Comments
Merge datasets	February 17, 2025	Anthony Brown	Some progress has been done but issues related to inconsistent data may still need to be addressed

Write the data dictionary	February 21, 2025	Fisher Eskew, Nikhil Vasudeva	
Identify and address inconsistencies	February 21, 2025	All team members	