

Client Update

Week 4 (July 30th, 2024)

Project Goal

To analyze, visualize and create data prediction model on movie ratings and recommendations. [Link to Github](#)

Meet the Team: Team Avengers (DS5)



Chankrisna
Chham



Michael
Fischer



Muhammad
Shamsher



Lee-Yee
Yong



Data Source

9 datasets ~ 1 GB

Metadata set

- Movie_metadata.csv
- Credit.csv
- Keywords.csv
- Ratings.csv
- Links.csv

Smaller sets

- Tmdb_5000_movies.csv
- Tmdb_5000_credits.csv
- Ratings_small.csv
- links_small.csv

Project Summary

Week 4 (July 30th, 2024)



Achievements so far:

- Project setup (roles, GitHub project and repo)
- Preliminary EDA completed
- Clarification of [data definitions](#) from source (TMDB)
- [Data Relationship mapping](#) completed
- Data Cleanup (ongoing finetune)

Focus for next two weeks:

Topics	Priority	Completed by:
Confirm movie rating math formula	High	Aug 2
Visualize data for current dataset	Medium	Aug 5
Explore text-clustering prediction model	Medium	Aug 9
Set up prediction targets and features for model	High	Aug 9

Lessons learned:

- Team organization
- Communication
- Focus on outcomes (perfection is our enemy)
- Using GitHub
- Apply learnings in EDA