Netflix recommender project call
14th June 2021
Present: Michael Hoang (MH), Rucha Sharma (RS), Vivek Kotecha (VK)

Progress

MH: Drafted up article for 3rd stage
RS: Uploaded current progress stage 1 analysis
VK: Explored the dataset too. Set up Github repository.

Discussion

Discussion around data cleaning and prepping dataset for Netflix recommender.

Discussion about splitting actors/directors into separate columns.

Release_year = column inaccurate, as shows the year of the latest season of the tv show, not the original year.

Easy cop-out – mostly TV shows with missing directors (as per dataset definition). Shows with missing directors typically documentaries, reality etc where less likely.

What about films with multiple release years? E.g. reboots

Discuss whether to separate data or not during this stage. Decision to separate.

VK: Can run tier 2 and 3 analysis at same time.

Next steps

Timeline is delayed from original.

RS: complete missing value by around Tuesday 15th June.

So start write up – 1 week for documentation.

Next week: meeting 21st June 2021
Agenda: draft documentation for tier 1, discuss ideas for tier 2

Clean up work done by then.

Revised timetable:

28th June – 1st intro article
5/12th July – Tier 1 article
2-3 weeks gap between each step