**Team members (Team 2) – ETL Project**

* Callan Yan
* Tigran Zohrabyan
* Purnima Chandel
* Naim Panjwani

**Proposal**

It is interesting to see novels being adapted to films. Our question is whether science fiction novels’ ratings correlate to ratings of films. Also, is there a correlation between science fiction ratings or film ratings with revenue obtained from a film?

**Methods**

We will scrape Wikipedia to get a mapping file to map book titles to film titles as they may not necessarily be the same. These film titles can then be queried on the OMDB API database to obtain IMDB and rotten tomatoes ratings (which we can scale and average out for each movie).

For book ratings, we will scrape Amazon and Chapters for the list of book titles of interest, and obtained the average user ratings for these books.

For the instances where one book was made into more than one version of a movie (eg. one in the 80s and one more recently), we will focus on the most recent film.

Summary of sources:

* OMDB API
* Amazon scraping for book titles and ratings
* Chapter scraping for book titles and ratings
* Optional: Kaggle provides a list of >80k (will need another source for getting book genre)
* Wikipedia to link book title to movie title <https://en.wikipedia.org/wiki/Category:Films_based_on_science_fiction_novels>

**Results**

* We will graph and investigate the relationship of film ratings versus book ratings
* We will graph revenue of film versus film rating
* And graph revenue of film versus average book rating (weighted average between Chapters and Amazon)