**Does book success guarantee its adapted movie success?**

**Abstract**

In this project, we explore if success of a book guarantees success of the movie adapted on that book. To ensure we have a feasible dataset to work on, we have restricted our search to genre of science fiction.

**Project Objective:**

Our question is whether the 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?

**Introduction**

To answer these questions, we pulled a list of science fiction books and their average reader ratings and mapped it with movies based on these books and their viewers/critics ratings and box-office revenue.

**Team Members:**

* Purnima Chandel
* Naim Panjwani
* Callan Yan
* Tigran

**Methods**

Data Sources Used:

* Websites Scrapped: Wikipedia, Goodreads
* API Used: OMDB API
* CSV: Kaggle

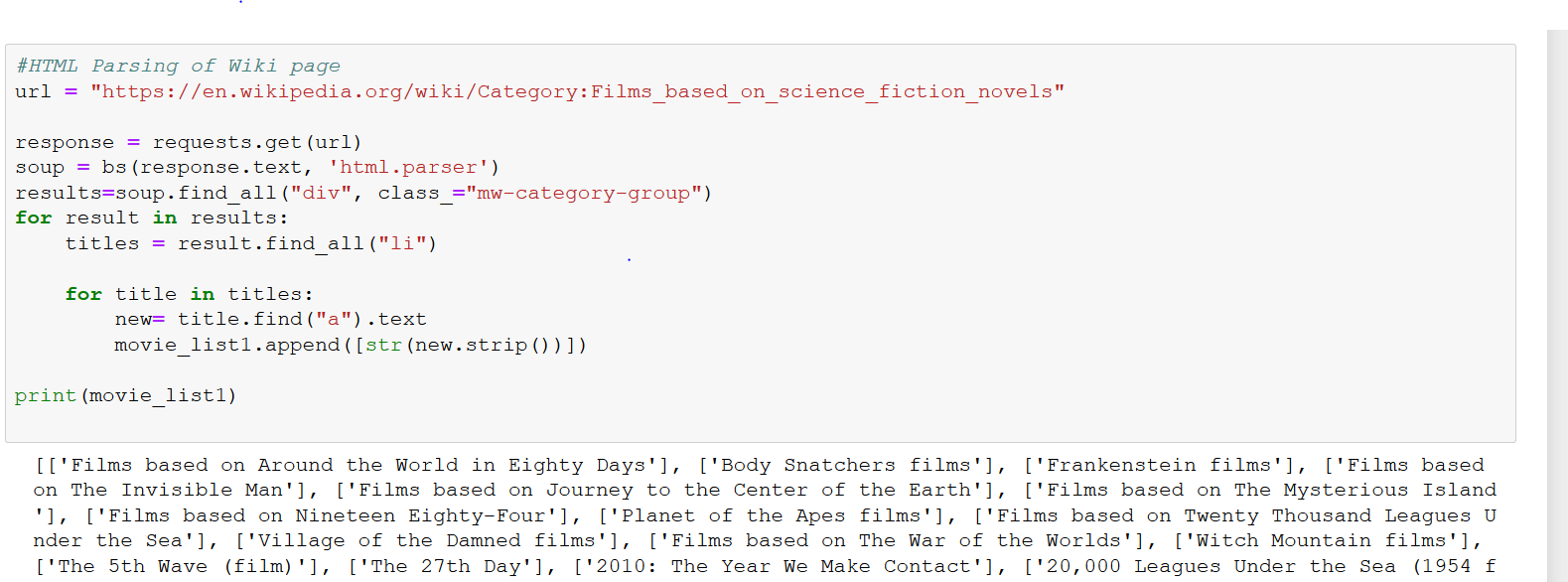
We scraped Wikipedia to get a mapping file to map films that have been adapted from books. These film titles are then queried on the OMDb API database to obtain IMDB/ rotten tomatoes ratings and box-office revenue.

For book ratings, we scraped Goodreads 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 (e.g. one in the 80s and one more recently), we focused on the most recent film.

**Extract:**

Step 1: Scrapping of Wikipedia to get the list of movies based on books:



Step 2: Querying the movie titles from Goodreads to get the corresponding books and their reader ratings:



Step 3: Querying the movies from OMDb to get the viewer ratings and box-office revenue:



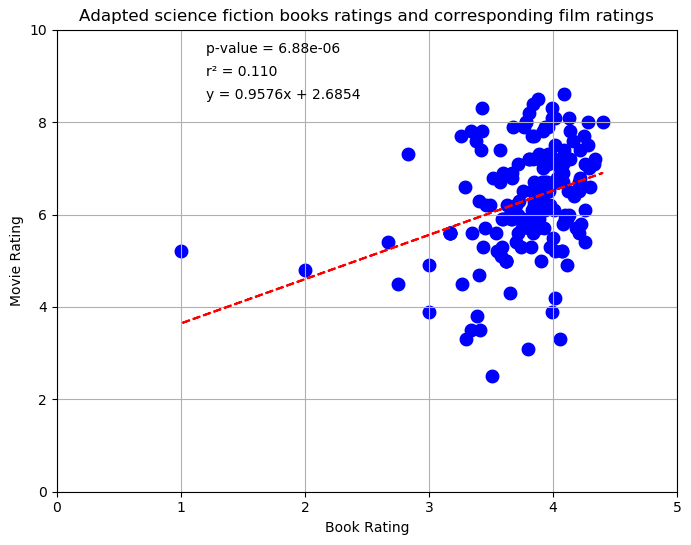
**Transform**

Step 1: Combine the data from different sources into one



Step 2: Run linear regression and plot various parameters

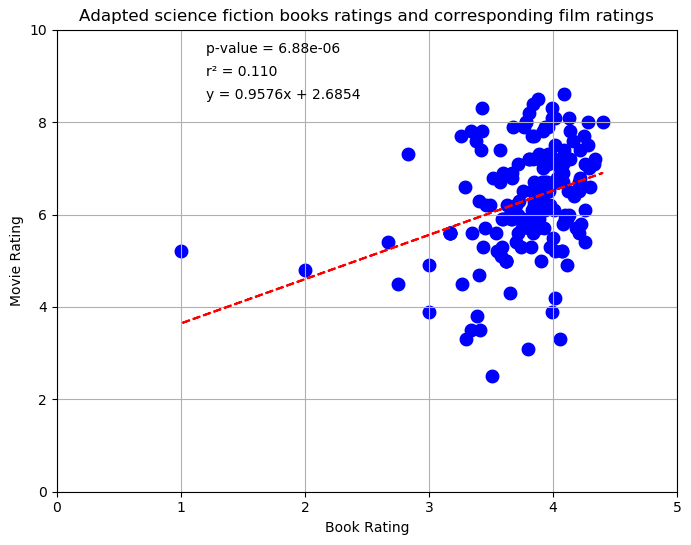




**Load**

Step 1: Load the final database into MongoDB

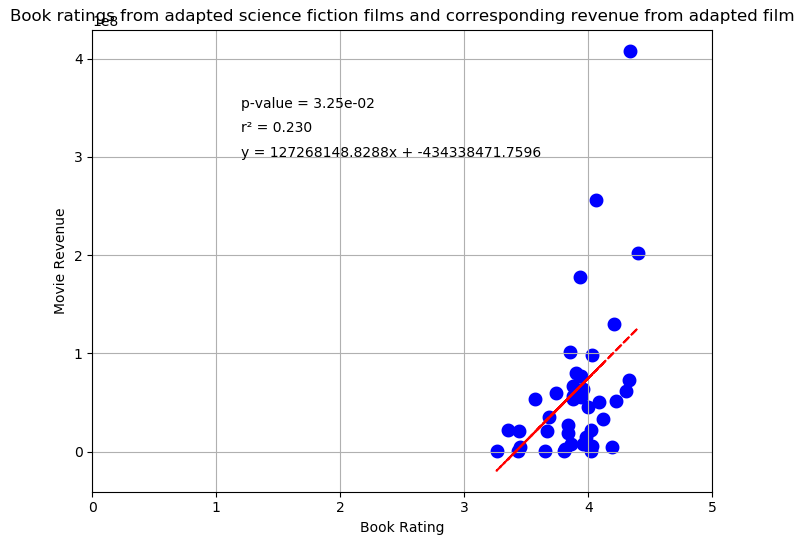


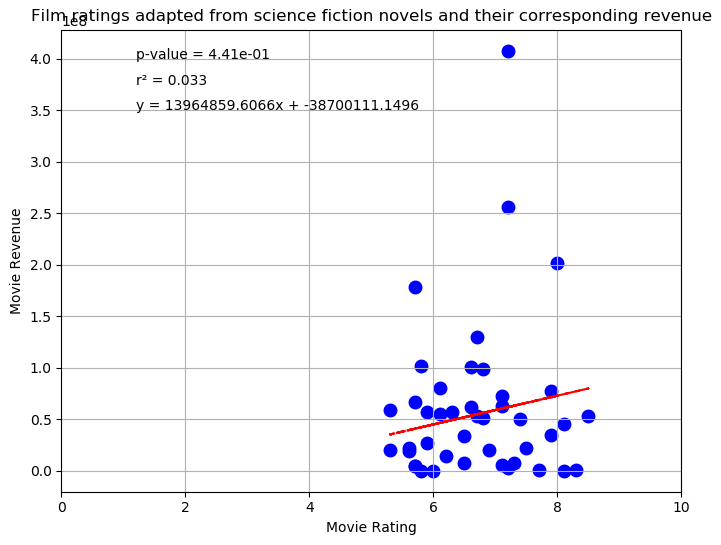
**Findings:**

***Graph 1: Book ratings vs. Movie ratings***

The p-value of the book ratings vs movie ratings is quite low, which suggests that there is some correlation between the book ratings and movie ratings.

***Graph 2: Book ratings vs Movie revenue***

The p-value for book ratings vs movie revenue is also significantly low and suggests some correlation. But due to low r-square suggests, that we cannot fully explain the variance of the data points

***Graph 3: Movie ratings vs Movie revenue***

The plot of movie ratings vs movie revenues shows the similar trends as Graph 2(book ratings vs movie revenue)

**Conclusions:**

* We did find correlation between book ratings and movie ratings, which suggests that successful book do guarantee a successful movie adaptation due to some extent
* There are various other factors which would play an important role in the success of the movie, such as production value, marketing, direction, etc.
* We did find some relation between ratings and box-office success of the movie, but any conclusion cannot be drawn due to limited data points

**Discussion:**

Some of the issues that we faced were:

* Titles of the books and movies were different at times
* Box office data for all the queried movies were not found on OMDb
* Data cleaning took a lot of time as there were additional characters/spaces in the titles
* Scanning and identifying the right sources took bit of an effort. For example: we wanted to scrape Amazon for book reviews and add it to Goodreads reviews but Amazon blocked us out (and Chapters Indigo does not authorize scrapping of the user ratings from their website).