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

**Abstract**

For decades, readers repeated the same phrase when Hollywood adapted a beloved novel for the screen: “The book is better than the movie.” But use of that phrase has gradually faded, replaced by enthusiastic shouts on social media when Hollywood grabs the rights to a classic work of science fiction or a modern twist on fantasy.

Books often lead to movie adaptations, and so much the better for sparking a fire toward both industries -- if a movie gets more people to read, great; if avid readers get to see their beloved characters on the big-screen, that's a boon too.

In this project, we are trying to 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 tried to map 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:

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

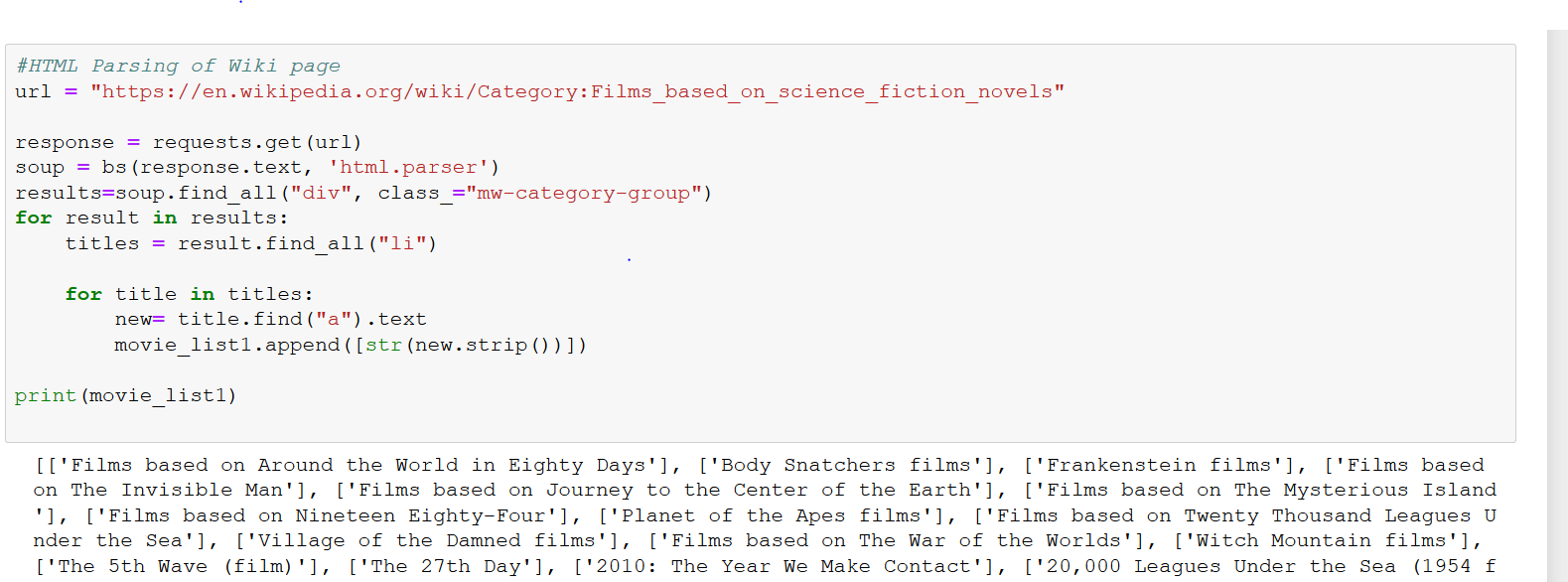
We scraped Wikipedia to get a mapping file to map films that have been adapted from books. These film titles are then queried on the OMDb API database to obtain IMDB/ rotten tomatoes ratings and box-office revenue.

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:

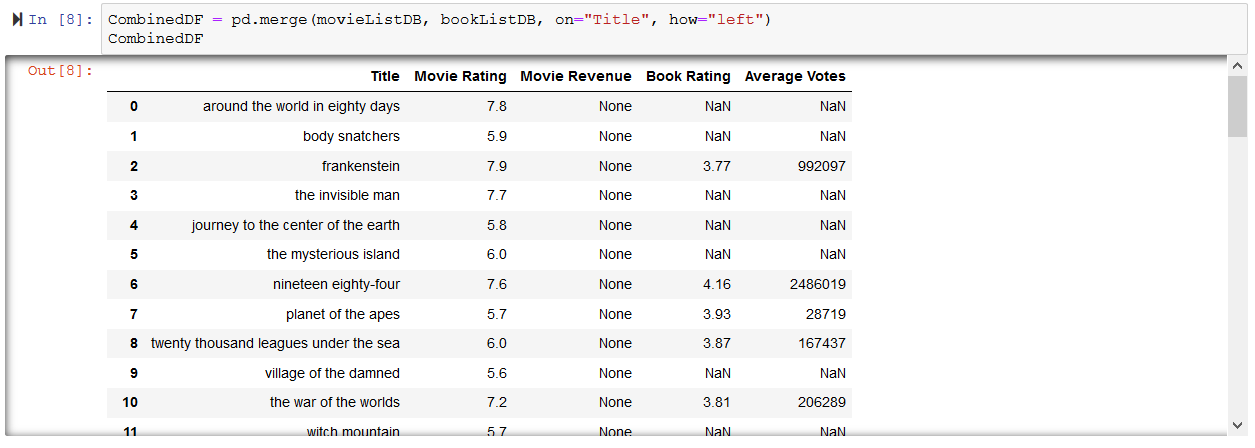


Step 3: Querying the movies form OMDB to get the viewer ratings and box-office revenue:

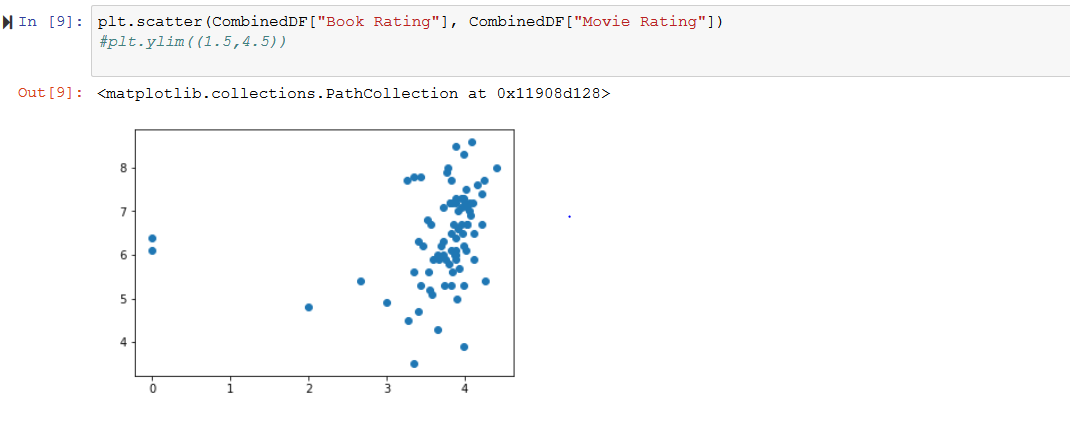


**Transform**

Step 1: Combine the data from different sources into one



Step 2: Plot the books rating with movie rating



Step 3: Run a linear regression to find the correlation between ratings of books vs movies

**Load**

Step 1: Load the final database into SQL

**Findings:**

Plot your key graphs and outline the findings in each of the plots.

**Conclusions:**

What is the answer to your hypothesis? Do your findings support this strongly?

**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 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