1. Kaggle CSV File
2. Source: <https://www.kaggle.com/promptcloud/all-marvel-cinematic-universe-movies-on-imdb>
3. Import to Pandas

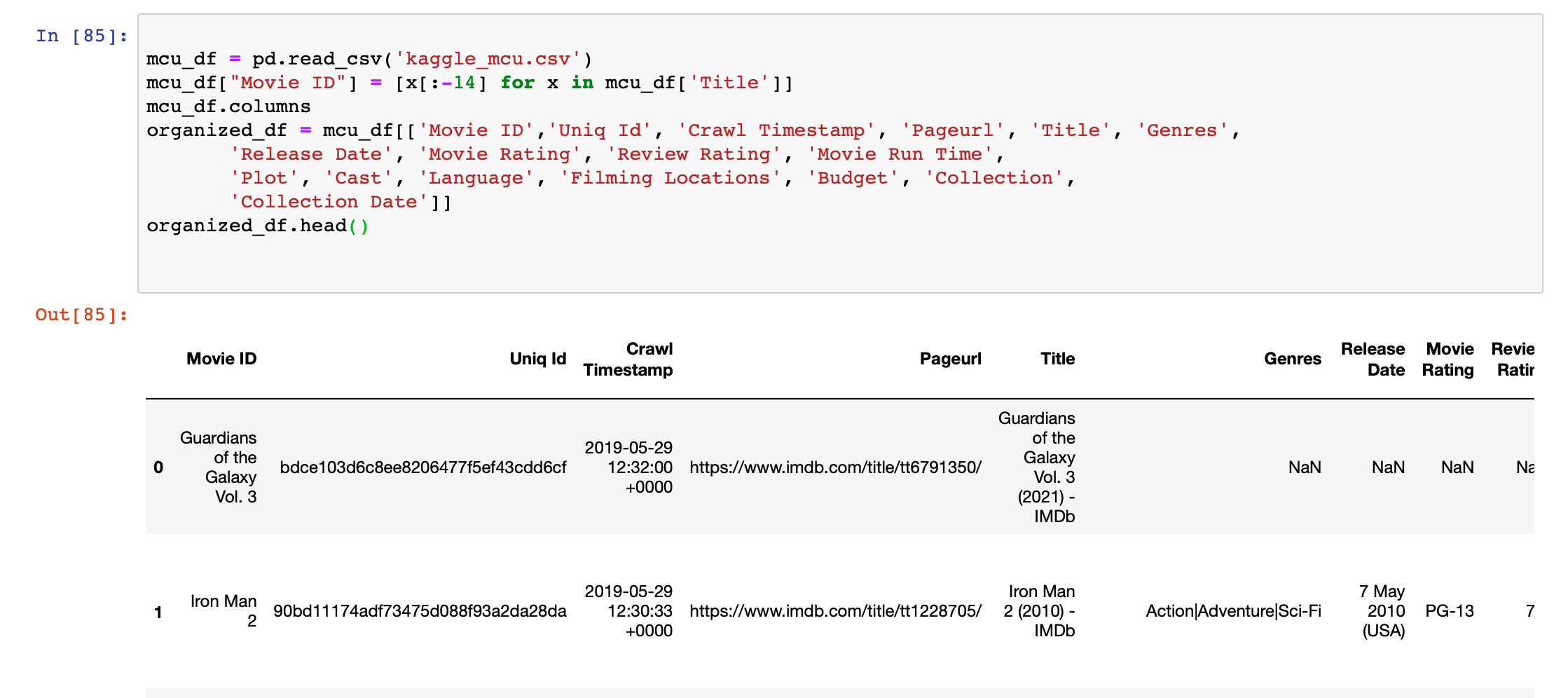


Fig. 1: Import to Pandas and Organize Data

1. Notes:

* Created a “Movie ID” column for merging
* Information available for each movie from this csv.
* 'Movie ID','Uniq Id', 'Crawl Timestamp', 'Pageurl', 'Title', 'Genres', 'Release Date', 'Movie Rating', 'Review Rating', 'Movie Run Time', 'Plot', 'Cast', 'Language', 'Filming Locations', 'Budget', 'Collection', 'Collection Date'

1. Scrape Fandom to get categorization/groups for movie analysis
2. Main URL: <https://marvelcinematicuniverse.fandom.com/wiki/Marvel_Cinematic_Universe_Wiki>
3. Scrape Phases

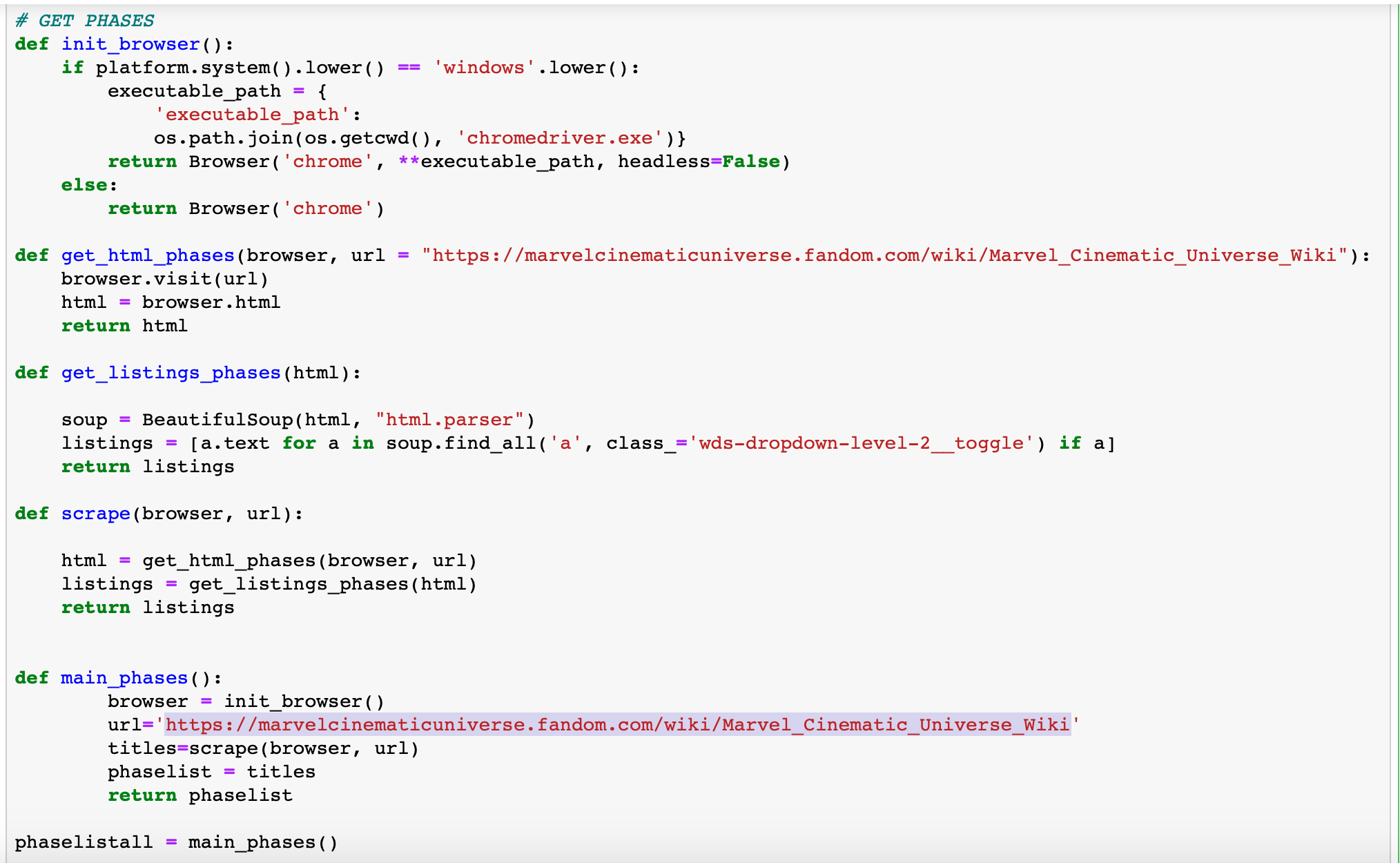


Fig. 2: Scrape Main URL to get Phases of the MCU

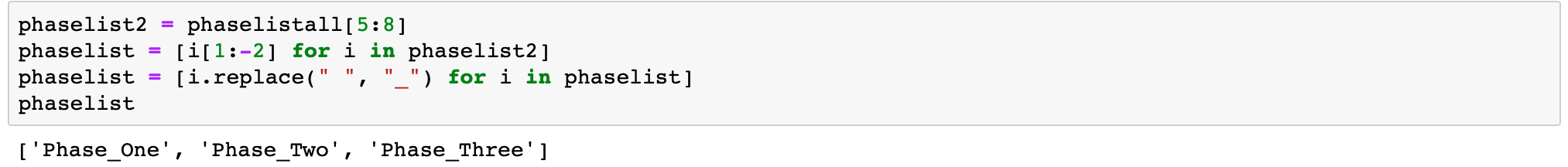


Fig.3: Use ‘phaselistall’ output to get just the required phases, then manipulate the string so it can be added to a base URL

1. Scrape Movie Links

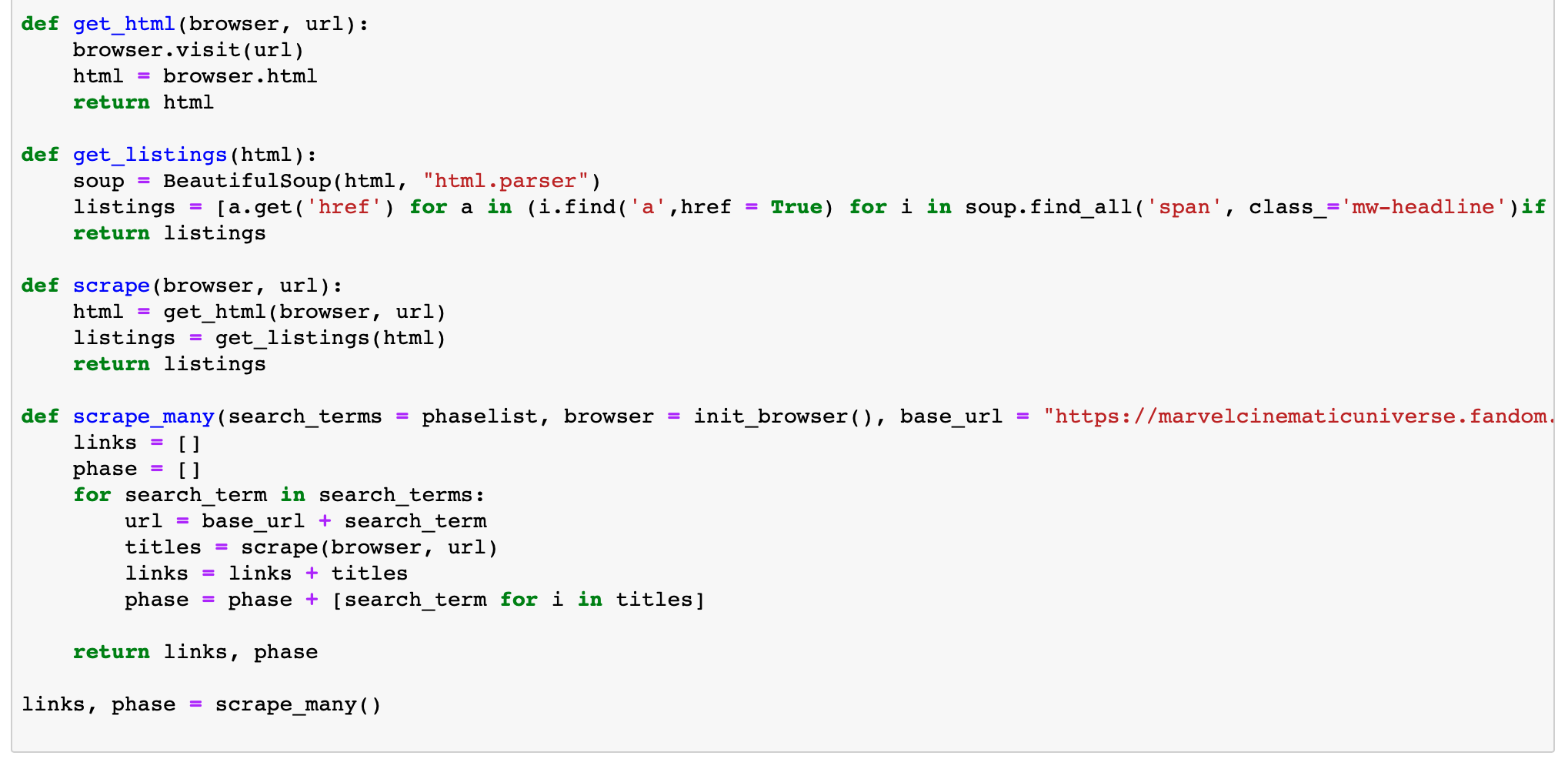


Fig. 4: Use ‘phaselist’ as search term, then scrape for the links to movie titles included in that phase

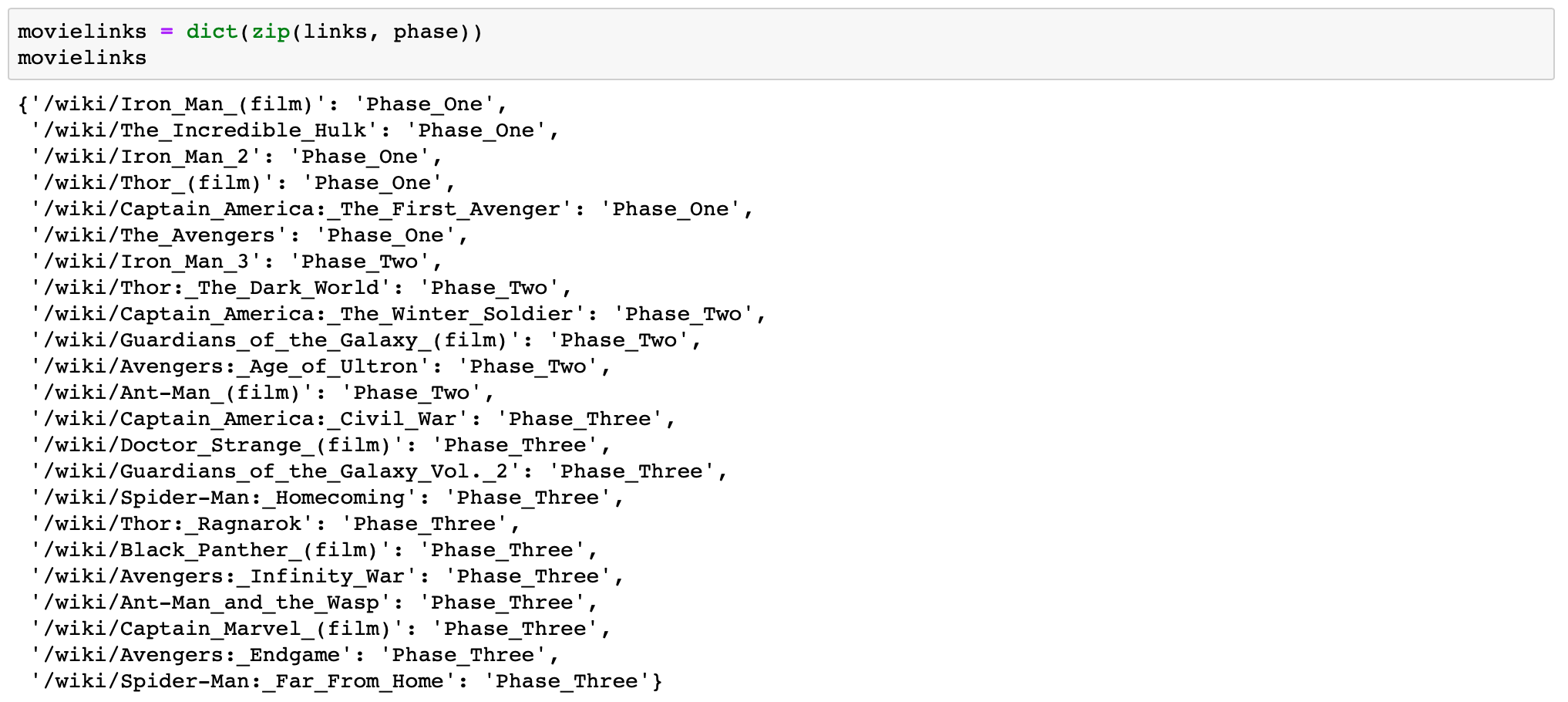
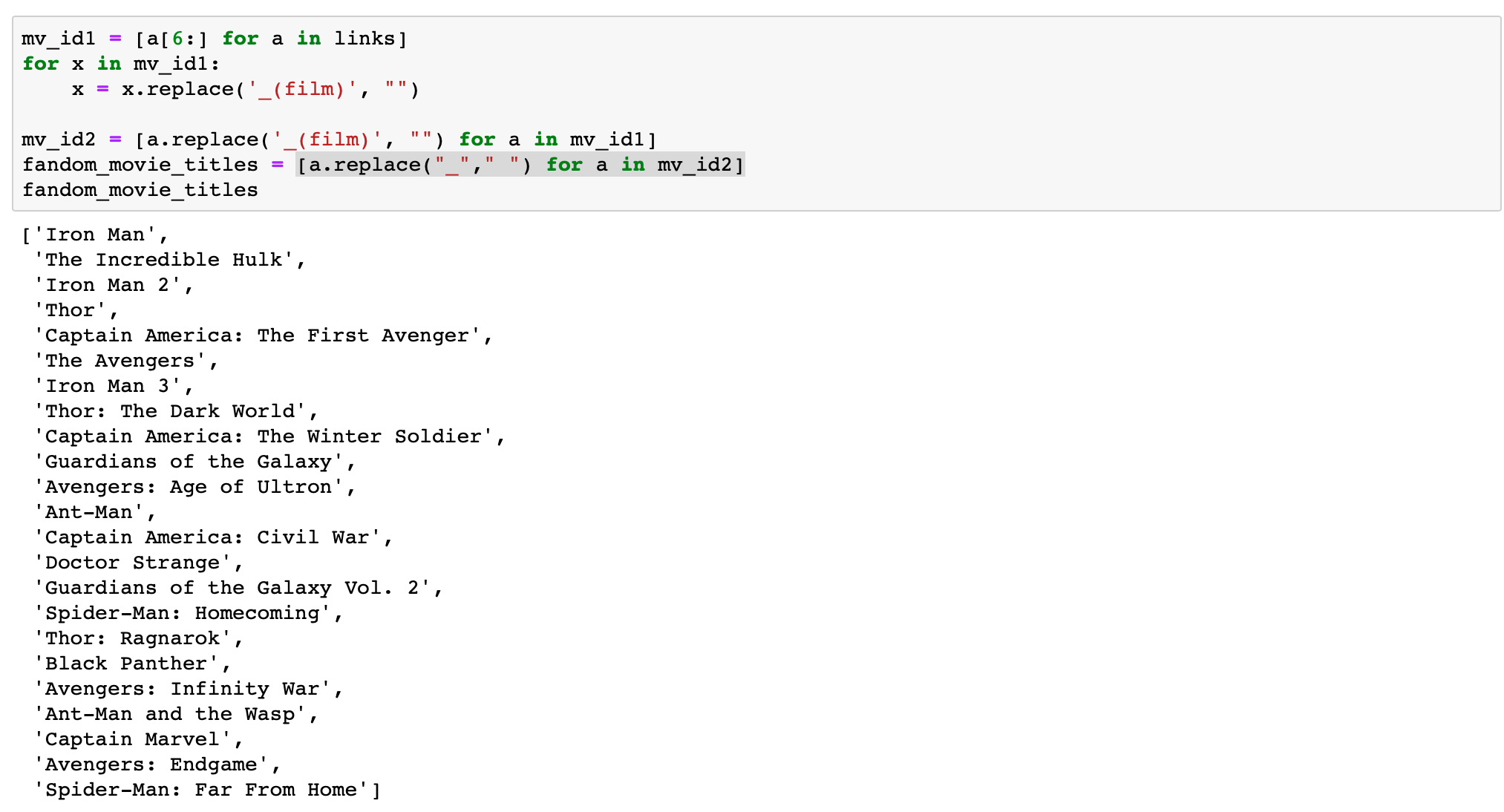
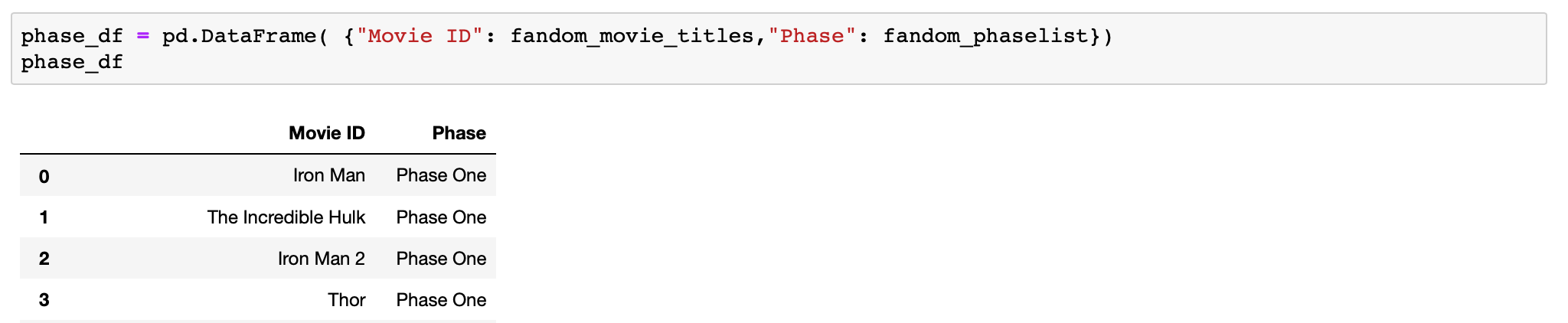


Fig. 5: Output of code in Fig. 4. Links can now be used as search terms added to a base url, along with its corresponding phase.

1. Prepare a data frame that can be merged with the Kaggle data frame







III. Merge Prepared Data Frames



Result of merging is a data frame of the Kaggle IMDb csv with additional information from the fandom for categorization and grouping for analyses.