

Webscrapping movies from imdb.com :

Primer and Analysis

Agenda

- Brief Introduction about IMDB
- What am i looking to scrape and analyze
- Brief description of the code
- Analysis

IMDB

IMDb (Internet Movie Database) is an [online database](#) of information related to films, television programs, home videos and video games, and internet streams, including fan reviews and ratings. The database is owned and operated by IMDb.com, Inc., a subsidiary of [Amazon](#).

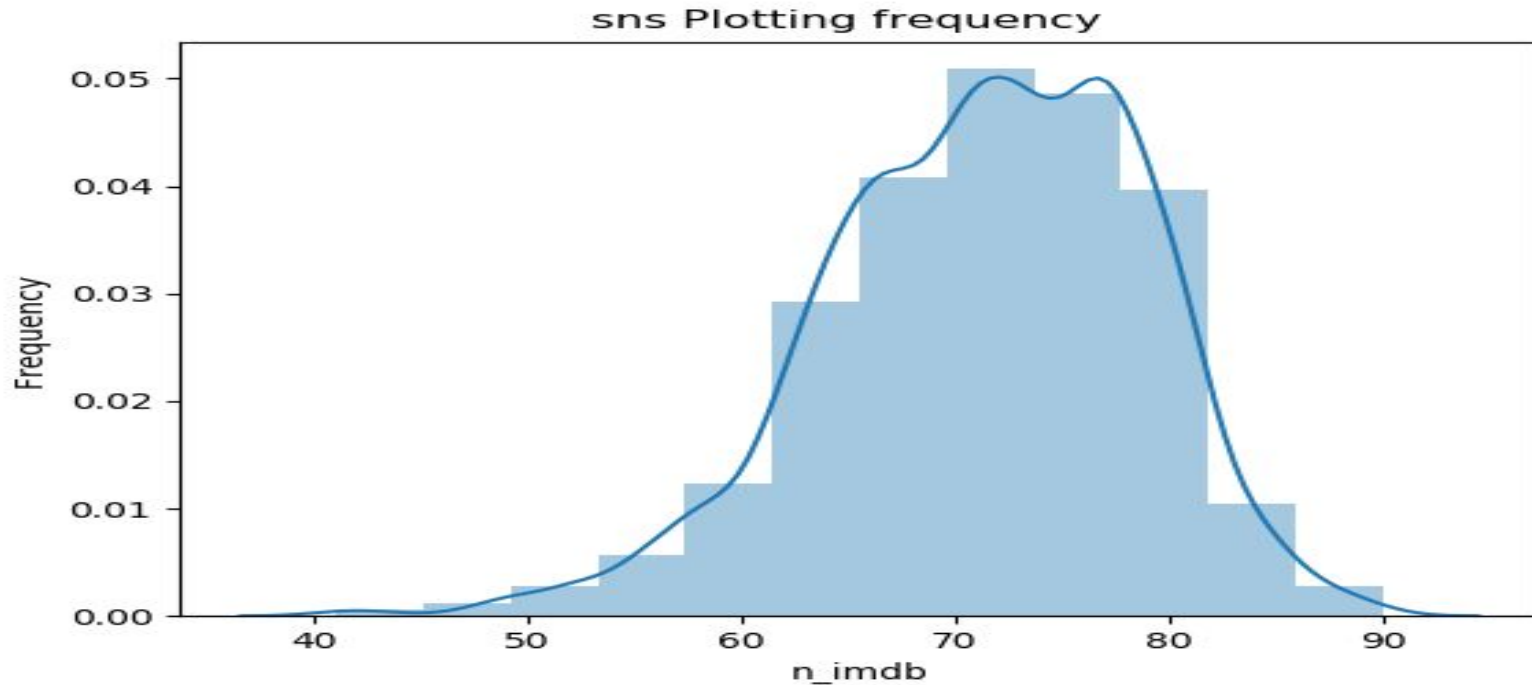
As of October 2018, IMDb has approximately 5.3 million titles (including episodes) and 9.3 million personalities in its database,^[2] as well as 83 million registered users.

This effort is to examine how the imdb movie rating compares with other ratings like metascore rating .

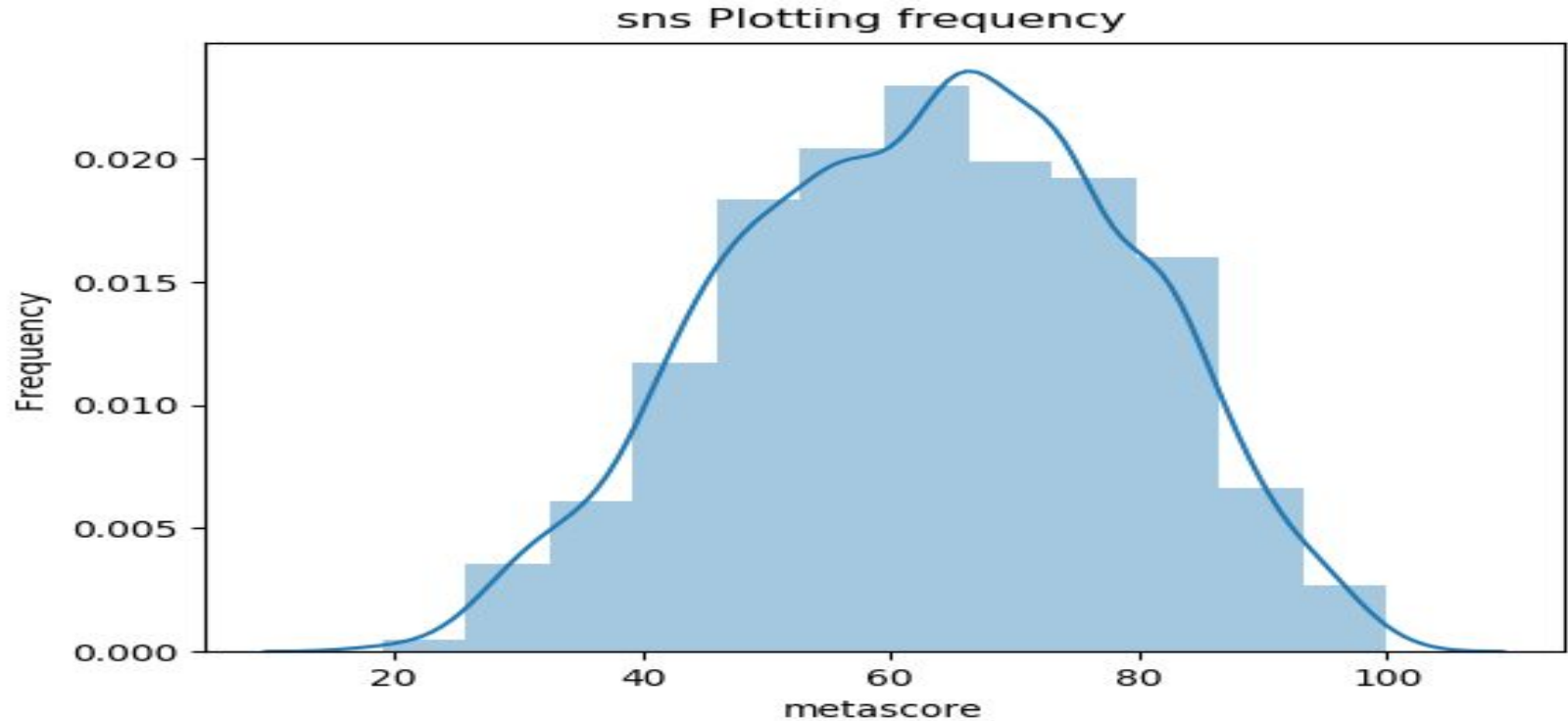
Goals and how was the code designed

- Scrape only what is needed . Identify the attributes needed to scrape
- Minimize the number of webpage requests
- Avoid burdening the website with high request rate .
- Establish a URL pattern which helps achieve the above goals
- Attributes selected are MovieName , Year , “Imdb score” , “Metascore” and Votes
- Chose years 1999-2018 to mine .
- For each of the years , decided to scrape 5 pages for a total of 95 pages
- Crawl rate was established to avoid a high request rate .

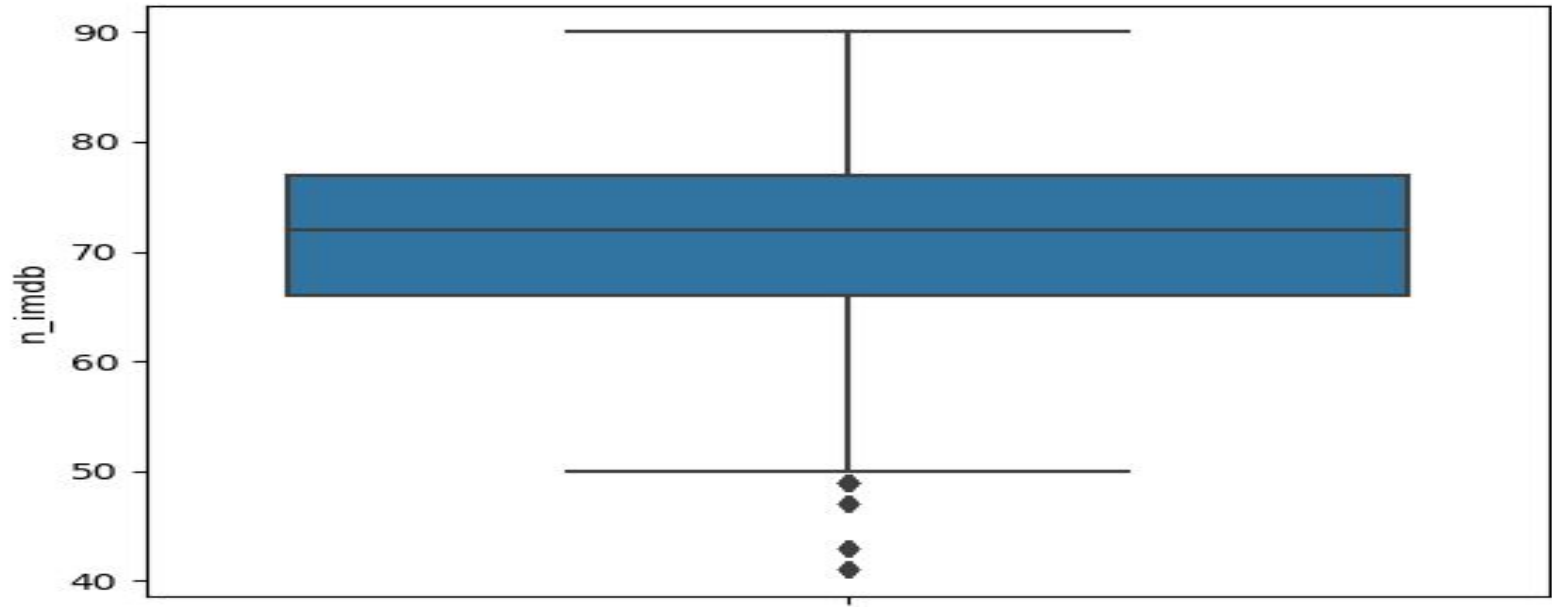
Seaborn distplot for “imdb score”



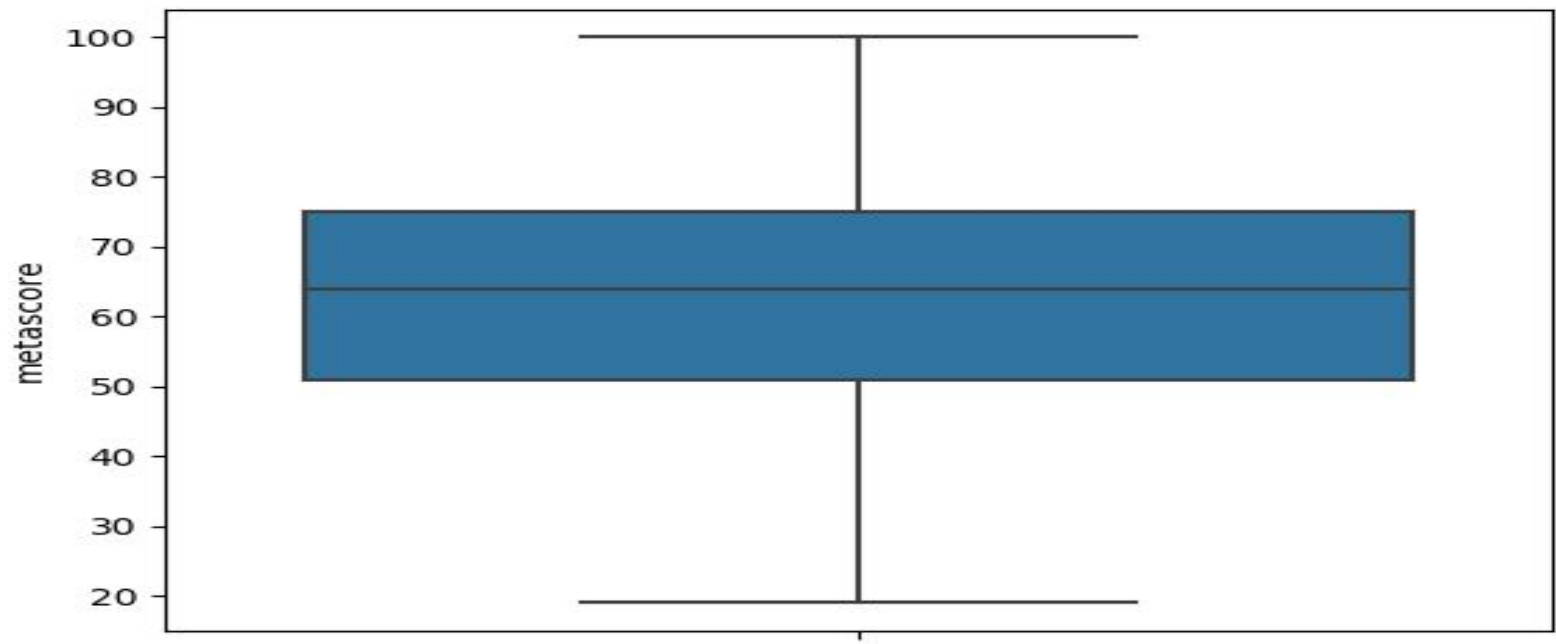
Seaborn distplot for metascore



Seaborn boxplot for imdb score



Seaborn boxplot for metascore



Seaborn boxplot for metascore

