Web Scraping: Top Rated Indian Movies

Introduction

This is a beginner project where I scraped data on top rated Indian movies from the IMDB movie website. http://www.imdb.com/ (http://www.imdb.com/) and put it in a pandas dataframe for analysis. Simple analysis is done to practice and revise simple concepts.

Method

I used Python 3.6.3 and jupyter notebook for the project. Beautiful Soup and requests libraries were used to scrape the data from IMDB movie website.Request library was used to request the web page and Beautiful soup to go through the html code and pull out the attributes. I stored the data in Python lists and merged the list togather to make a pandas dataframe.Next step I checked for any discrepencies. Year attribute was within parentheses which was removed. Then I did a simple analysis using libraries pandas, numpy and seaborn.

Web Scraping

Top Rated Movies

Movie	Rating	Year
Kantara	8.5	2022
Ramayana: The Legend of Prince Rama	8.5	1993
Rocketry: The Nambi Effect	8.4	2022
Anbe Sivam	8.4	2003
Nayakan	8.4	1987
Hanky Panky	8.4	1979
Jai Bhim	8.4	2021
777 Charlie	8.4	2022
Pariyerum Perumal	8.4	2018
Manichithrathazhu	8.4	1993

Analysis

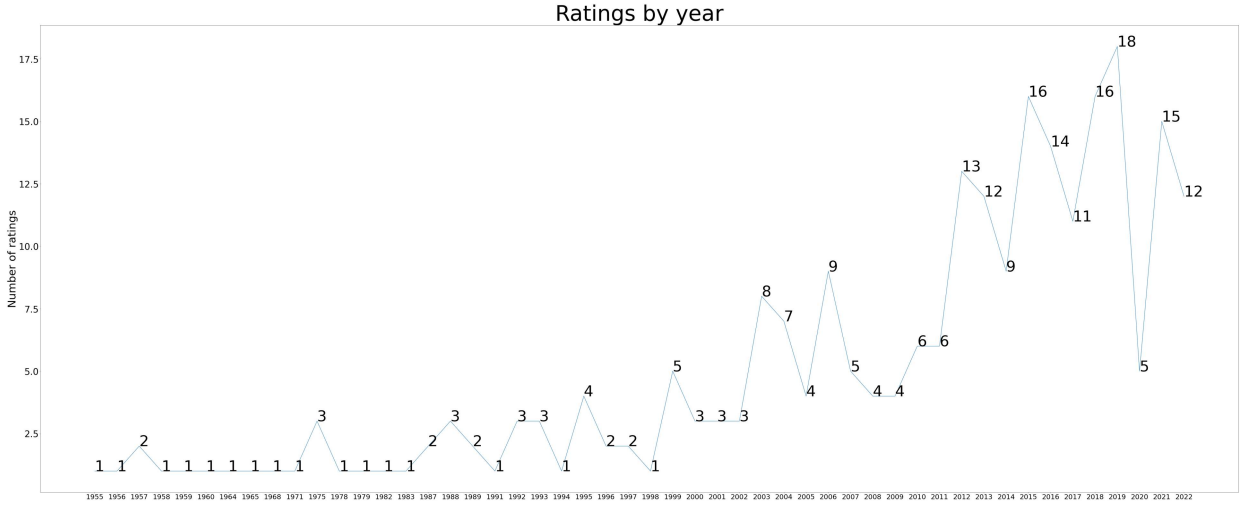
Years with highest rating

Year	Number_of_ratings
2019	18
2018	16
2015	16
2021	15
2016	14
2012	13
2022	12
2013	12
2017	11
2006	9
2014	9
2003	8

Years with least ratings

Year	Number_of_ratings
1991	1
1965	1
1956	1
1998	1
1958	1
1959	1
1960	1
1964	1
1968	1
1994	1
1971	1
1978	1
1979	1
1982	1
1983	1
1955	1

Ratings by year



2019 had the greatest number of movie rating, 54. 16 years between 1955 and 1998 had the least number of ratings of 3 with some increase in between. Number of ratings by year plot has almost a zig-zag pattern.Ratings increased a lot after 2002, you could say rapid increase in technology could be responsible for this.

Minimum rating given is 7.6
Maximum rating given is 8.5
Average rating given is 7.96

When and which - Minimum rating

Movie	Rating	Year
Ennu Ninte Moideen	7.6	2015
Velaiyilla Pattathari	7.6	2014

Among the top rated 250 movies two movies Ennu Ninte Moideen and Velaiyilla Pattathari had the minimum rating of 7.6. Velaiyilla Pattathari was rated three times in 2014 with rating 7.6 and Ennu Ninte Moideen was rated three times in 2015 with rating 7.6.

When and which - Maximum rating

Movie	Rating	Year
Kantara	8.5	2022
Ramayana: The Legend of Prince Rama	8.5	1993

Among the top 250 movies that were rated from 1955 to 2022 Kantara and Ramayana:The Legend of Prince Rama got the maximum rating of 8.5 in 2022 and 1993 respectively

Most Rated Movies

Movies	Number of Ratings
Drishyam	2
Pelli Choopulu	1
Hey Ram	1
Kahaani	1
Shershaah	1
Paan Singh Tomar	1
Mother India	1
Oru CBI Diary Kurippu	1
Ayirathil Oruvan	1
Hera Pheri	1
A Northern Story of Valor	1
Jigarthanda	1

Movie	Rating	Year
Drishyam	8.2	2013
Drishyam	8.1	2015

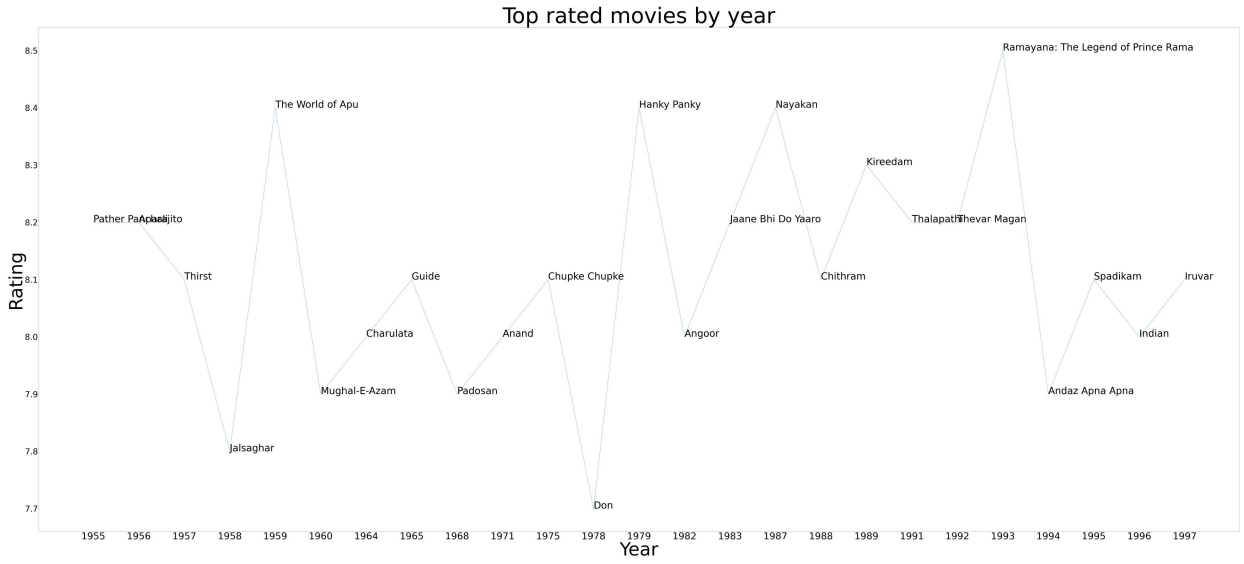
Among the 250 movies only one movie 'Drishyam' got the rating greatest number of times, 2. All other top rated movies rated from 1955 t0 2022 were rated only one time. 'Drishyam' was rated two times in 2013 and 2015.

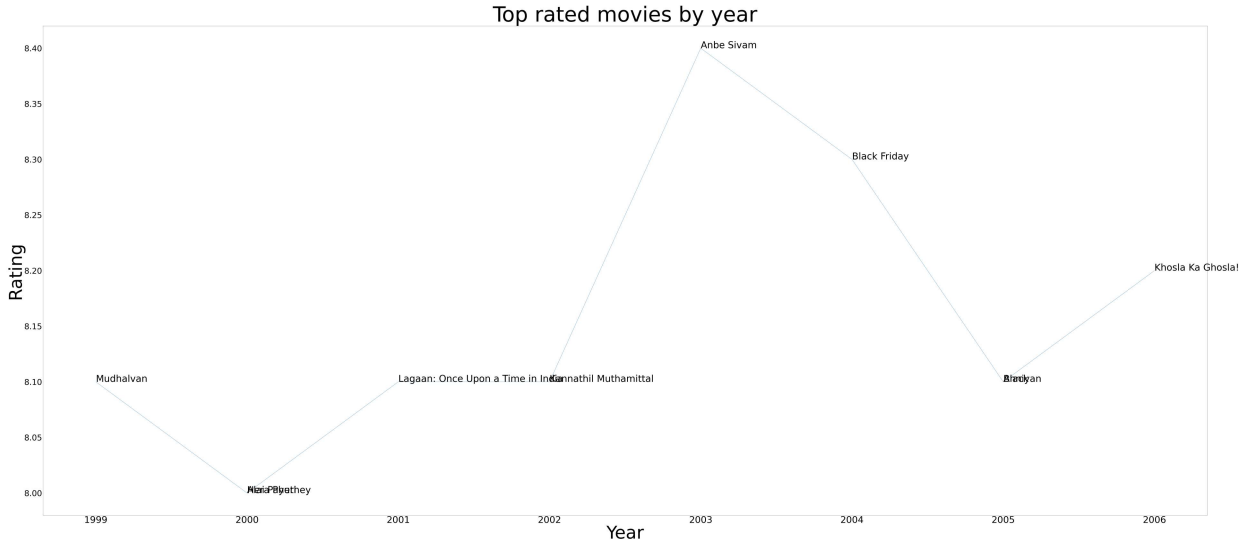
Most rated by year

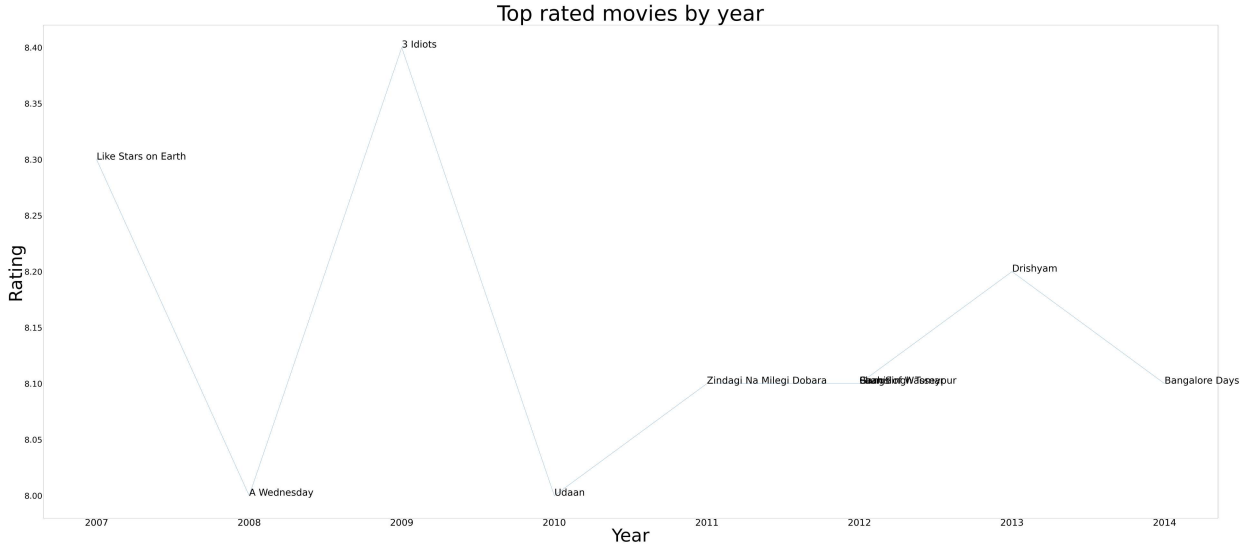
Year	Movie	Number of rating
1955	Pather Panchali	1
1956	Aparajito	1
1957	Mother India	1
1957	Thirst	1
1958	Jalsaghar	1
1959	The World of Apu	1
1960	Mughal-E-Azam	1
1964	Charulata	1
1965	Guide	1
1968	Padosan	1

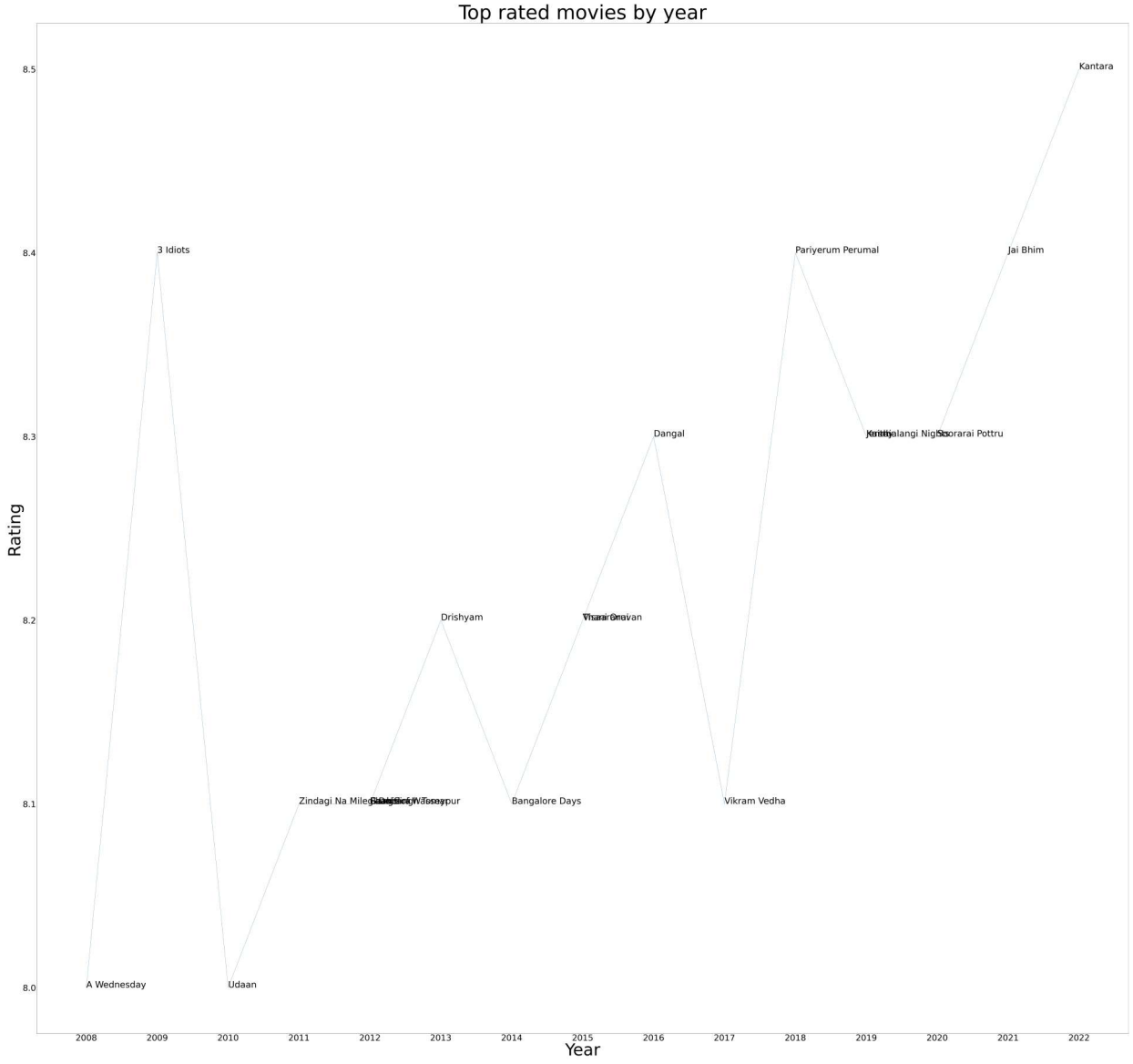
Year	Movie	Number of rating
2022	Hridayam	1
2022	Jana Gana Mana	1
2022	K.G.F: Chapter 2	1
2022	Kantara	1
2022	Kaun Pravin Tambe?	1
2022	Major	1
2022	Ponniyin Selvan: Part I	1
2022	RRR	1
2022	Rocketry: The Nambi Effect	1
2022	Sita Ramam	1

All top rated movies that are rated are rated once except one, Drishyam that was rate twice in a year






Top rated movies by year
Rating
Year
8.40
8.35
8.30
8.25
8.20
8.15
8.10
8.05
8.00
2007
2008
2009
2010
2011
2012
2013
2014
Like Stars on Earth
A Wednesday
3 Idiots
Udaan
Zindagi Na Milegi Dobara
Drishyam
Bangalore Days


Top rated movies by year
Rating
Year
8.5
8.4
8.3
8.2
8.1
8.0
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Kantara
3 Idiots
Pariyerum Perumal
Jai Bhim
Dangal
Drishyam
Bangalore Days
Vikram Vedha
Zindagi Na Milegi Dobara
A Wednesday
Udaan

Summary:

- All top rated Indian movies in this data that were rated above 7.6
- All except one top rated Indian movie is rated only once
- Maximum times a top rated Indian movie is rated is 2. It was for only one movie Drishyam
- 2019 had the greatest number of movie rating, 54. Ratings increased a lot after 2002.
- Maximum rating given to an Indian movie in this top rated movie dataset is 8.5.Kantara and Ramayana:The Legend of Prince Rama got the maximum rating of 8.5 in 2022 and 1993 respectively.