

IndianMovieCelebritiesPersonalityTraits

April 26, 2020

```
In [1]: # Assignment Requirements
'''
    Crawl popular websites & create a database of Indian movie celebrities with their images
'''

Out[1]: '\n Crawl popular websites & create a database of Indian movie celebrities with their images'

In [1]: %%time

    # Import packages
    from requests import get
    from bs4 import BeautifulSoup
    import pandas as pd
    import shutil
    import json
    import os

    # Declare global variables
    actors, images, perTraits = [], [], []

    # Top 100 celebrities list
    celebrities_url = "https://www.imdb.com/list/ls068010962/"
    imdb_url = "https://www.imdb.com/"
    personality_url = "http://nuclear.ucdavis.edu/~rpicha/personal/astrology/"

    # Soup objects
    actors_soup = BeautifulSoup(get(celebrities_url).text, 'lxml').findAll('div', class_='')
    personality_soup = BeautifulSoup(get(personality_url).text, 'lxml').findAll('td')

    # Create empty directory for downloading images
    try:
        os.mkdir('Images')
    except:
```

```

shutil.rmtree('Images')
os.mkdir('Images')

for item in actors_soup:
    temp_url = imdb_url + item.find('a')['href']
    temp_soup = BeautifulSoup(get(temp_url).text, 'lxml')

    # Actor name
    name = temp_soup.find('h1', class_="header").find('span').text.strip()
    actors.append(name if name is not None else None)

    # Image Urls
    try:
        img = temp_soup.findAll('img', id="name-poster")[0].get('src')
        images.append(img if img is not None else None)

        # Download images of celebrities to Images folder
        if img is not None:
            response = get(img, stream=True)
            with open('./Images/' + name + '.png', 'wb') as out_file:
                shutil.copyfileobj(response.raw, out_file)

            del response

    except:
        images.append(None)

    # Personality trait
    try:
        per_key = temp_soup.find('div', id="dyk-star-sign").find('a').text.strip()
        for obj in personality_soup:
            if obj.text.split()[0].strip() == per_key:
                traits = [ele for ele in obj.text.split('\n') if ele is not ''][2:]
                perTraits.append(traits)

    except:
        perTraits.append([None])

    print(f"Downloaded data of celebrity: {name}")

# Create dataframe with actors, imageUrls and Personality traits
assert len(actors) == len(images) == len(perTraits)
df = pd.DataFrame()
df['Actors'] = actors
df['ActorImageUrl'] = images

```

```
df['Personality-Traits'] = perTraits
df.to_csv("IndianMovieCelebritiesPersonalityTraits.csv")
```

```
Downloaded data of celebrity: Aamir Khan
Downloaded data of celebrity: Akshay Kumar
Downloaded data of celebrity: Ajay Devgn
Downloaded data of celebrity: Amjad Khan
Downloaded data of celebrity: Amitabh Bachchan
Downloaded data of celebrity: Amol Palekar
Downloaded data of celebrity: Abhay Deol
Downloaded data of celebrity: Akshaye Khanna
Downloaded data of celebrity: Arshad Warsi
Downloaded data of celebrity: Anushka Shetty
Downloaded data of celebrity: Atul Kulkarni
Downloaded data of celebrity: Asrani
Downloaded data of celebrity: Ayushmann Khurrana
Downloaded data of celebrity: Aruna Irani
Downloaded data of celebrity: Annu Kapoor
Downloaded data of celebrity: Ashish Vidyarthi
Downloaded data of celebrity: Adil Hussain
Downloaded data of celebrity: Anupam Kher
Downloaded data of celebrity: Anil Kapoor
Downloaded data of celebrity: Amrish Puri
Downloaded data of celebrity: Amole Gupte
Downloaded data of celebrity: Boman Irani
Downloaded data of celebrity: Balraj Sahni
Downloaded data of celebrity: Binnu Dhillon
Downloaded data of celebrity: Bharat Bhushan
Downloaded data of celebrity: Chiranjeevi
Downloaded data of celebrity: Chunky Pandey
Downloaded data of celebrity: Danny Denzongpa
Downloaded data of celebrity: Darsheel Safary
Downloaded data of celebrity: Deepak Tijori
Downloaded data of celebrity: Dharmendra
Downloaded data of celebrity: Deepika Padukone
Downloaded data of celebrity: Dimple Kapadia
Downloaded data of celebrity: Dev Anand
Downloaded data of celebrity: Deepti Naval
Downloaded data of celebrity: Dilip Kumar
Downloaded data of celebrity: Farooq Shaikh
Downloaded data of celebrity: Farhan Akhtar
Downloaded data of celebrity: Feroz Khan
Downloaded data of celebrity: Farida Jalal
Downloaded data of celebrity: Guru Dutt
Downloaded data of celebrity: Govinda
Downloaded data of celebrity: Gulshan Grover
Downloaded data of celebrity: Gurpreet Ghuggi
Downloaded data of celebrity: Girish Karnad
```

Downloaded data of celebrity: Hrithik Roshan
Downloaded data of celebrity: Huma Qureshi
Downloaded data of celebrity: Inaamulhaq
Downloaded data of celebrity: Irrfan Khan
Downloaded data of celebrity: Jackie Shroff
Downloaded data of celebrity: Jimmy Sheirgill
Downloaded data of celebrity: Jeetendra
Downloaded data of celebrity: Jaspal Bhatti
Downloaded data of celebrity: Johnny Lever
Downloaded data of celebrity: Jaswinder Bhalla
Downloaded data of celebrity: Kamal Haasan
Downloaded data of celebrity: Kangana Ranaut
Downloaded data of celebrity: Konkona Sen Sharma
Downloaded data of celebrity: Kader Khan
Downloaded data of celebrity: Kay Kay Menon
Downloaded data of celebrity: Kajol
Downloaded data of celebrity: Kulbhushan Kharbanda
Downloaded data of celebrity: Kalki Koechlin
Downloaded data of celebrity: Kapil Sharma
Downloaded data of celebrity: Krishna Abhishek
Downloaded data of celebrity: Karuna Bannerjee
Downloaded data of celebrity: Kanwaljit Singh
Downloaded data of celebrity: Lara Dutta
Downloaded data of celebrity: Manjot Singh
Downloaded data of celebrity: Manu Rishi Chadha
Downloaded data of celebrity: Mehmood
Downloaded data of celebrity: Mukesh Tiwari
Downloaded data of celebrity: Mammootty
Downloaded data of celebrity: Makrand Deshpande
Downloaded data of celebrity: Madhavan
Downloaded data of celebrity: Mohanlal
Downloaded data of celebrity: Mita Vasisht
Downloaded data of celebrity: Manoj Bajpayee
Downloaded data of celebrity: Manoj Joshi
Downloaded data of celebrity: Madhuri Dixit
Downloaded data of celebrity: Mohnish Bahl
Downloaded data of celebrity: Mithun Chakraborty
Downloaded data of celebrity: Mukesh Khanna
Downloaded data of celebrity: Mukul Dev
Downloaded data of celebrity: Madhubala
Downloaded data of celebrity: Meena Kumari
Downloaded data of celebrity: Mala Sinha
Downloaded data of celebrity: Manoj Pahwa
Downloaded data of celebrity: Manoj Kumar
Downloaded data of celebrity: Nassar
Downloaded data of celebrity: Naseeruddin Shah
Downloaded data of celebrity: Nandita Das
Downloaded data of celebrity: Nagarjuna Akkineni

```
Downloaded data of celebrity: Nana Patekar  
Downloaded data of celebrity: Nawazuddin Siddiqui  
Downloaded data of celebrity: Nargis  
Downloaded data of celebrity: Nirupa Roy  
Downloaded data of celebrity: Nutan  
Downloaded data of celebrity: Neeraj Kabi  
Downloaded data of celebrity: Navin Nischol  
CPU times: user 16.6 s, sys: 375 ms, total: 17 s  
Wall time: 3min 2s
```

```
In [ ]:
```