CS432/532: Final Project Report

Project Title: Exploration of Disney Plus Movies and TV Shows

Team Member(s): Sumeet Patle

I. Problem

In this project exploration and analysis is done on Dataset of Disney plus. Disney plus dataset contain all data of different countries. Once Process all the data then able to view Relation between Type (movie and shows) and Rating. Also it difficult for marvel fan to get all movies according to marvel universe so here marvel movie according to their time also available.

II. Software Design and Implementation

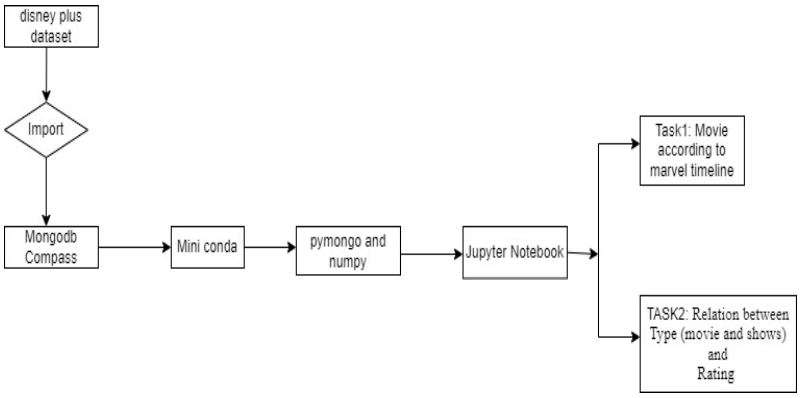
A. Software Design and NoSQL-Databse and Tools Used

1. Pick the most updated dataset from public dataset(kaggle).
2. Based on requirement use Mongodb for nosql dataset.
3. Installed python by mini-conda so to install library like pandas and pymongo (for data processing)
4. For data visualization use matplotlib and seaborn library.
5. For implementation of code use jupter notbook.

B. Parts that you have implemented

1. First start mongodb server on local host.
2. With the help Mongodb compass, Import
   dataset under 'project3' database and collection 'disney_plus'
3. Then perform task on python

III. Project Outcome



References

[1] Disney_plus_tv_shows_and_Movies, https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows

[2] For data processing and data visualization command Tutorial — PyMongo 4.1.1 documentation
https://www.mongodb.com/languages/python
https://matplotlib.org/2.0.2/users/pyplot_tutorial.html

Exploration of DISNEY PLUS MOVIES AND TV SHOWS

Disney Plus dataset consists of TV shows and movies data. The dataset is taken from Kaggle.
For data processing I used Pandas and Numpy library.
Also for visualize I used matplotlib and seaborn library.
For nosql dataset use mongodb

In [1]:
```
import sys
from pymongo import MongoClient
client = MongoClient('mongodb://localhost:27017/')
db = client.Project3
collection = db.disney_plus
print("Connected to MongoDB Server")
```

```
Connected to MongoDB Server
```

In [2]:
```
import numpy
import pandas
import matplotlib.pyplot
import seaborn
import pprint
import jovian
```

TASK
1. Display all movies according to marvel universe timeline irrespective of year they released

In [3]:
```
cursor_4 = collection.find({"rating": {"$in" : ["PG-13", "PG"]},"country": "India
df = pandas.DataFrame(list(cursor_4))
df.head(17)
```

Out[3]:

	_id	show_id	type	title	director	cast	country	date_
0	627443e6fd923b2544adef08	s1042	Movie	Marvel Studios' Captain America: The First Ave...	Joe Johnston	Chris Evans, Tommy Jones, Hugo Weaving, Hayley...	India	12-I
1	627443e6fd923b2544adef0a	s1044	Movie	Marvel Studios' Captain Marvel	Anna Boden, Ryan Fleck	Brie Larson, Samuel Jackson, Ben Mendelsohn, D...	India	12-I
2	627443e6fd923b2544adef0e	s1048	Movie	Marvel Studios' Iron Man	Jon Favreau	Robert Downey Jr., Terrence Howard, Jeff Bridg...	India	12-I
3	627443e6fd923b2544adef0f	s1049	Movie	Marvel Studios' Iron Man 2	Jon Favreau	Robert Downey Jr., Gwyneth Paltrow, Don Cheadl...	India	12-I
4	627443e6fd923b2544adef02	s1036	Movie	Marvel Studios: The Incredible Hulk	Louis Leterrier	Edward Norton, Liv Tyler	India	12-I
5	627443e6fd923b2544adef12	s1052	Movie	Marvel Studios' Thor	Kenneth Branagh	Chris Hemsworth, Natalie Portman, Tom Hiddlest...	India	12-I
6	627443e6fd923b2544adef11	s1051	Movie	Marvel Studios' The Avengers	Joss Whedon	Robert Downey Jr., Chris Evans, Mark Ruffalo, ...	India	12-I
7	627443e6fd923b2544adef13	s1053	Movie	Marvel Studios' Thor: The Dark World	Alan Taylor	Chris Hemsworth, Natalie Portman, Tom Hiddlest...	India	12-I
8	627443e6fd923b2544adef10	s1050	Movie	Marvel Studios' Iron Man 3	Shane Black	Robert Downey Jr., Gwyneth Paltrow, Don Cheadl...	India	12-I

	_id	show_id	type	title	director	cast	country	date_
9	627443e6fd923b2544adef09	s1043	Movie	Marvel Studios' Captain America: The Winter So...	Anthony Russo, Joe Russo	Chris Evans, Scarlett Johansson, Sebastian Sta...	India	12-I
10	627443e6fd923b2544adef0c	s1046	Movie	Marvel Studios' Guardians of the Galaxy	James Gunn	Chris Pratt, Zoë Saldana, Dave Bautista, Vin D...	India	12-I
11	627443e6fd923b2544adef0d	s1047	Movie	Marvel Studios' Guardians of the Galaxy Vol. 2	James Gunn	Chris Pratt, Bradley Cooper, Vin Diesel, Pom K...	India	12-I
12	627443e6fd923b2544adef05	s1039	Movie	Marvel Studios' Avengers: Age of Ultron	Joss Whedon	Robert Downey Jr., Chris Hemsworth, Mark Ruffa...	India	12-I
13	627443e6fd923b2544adef04	s1038	Movie	Marvel Studios' Ant-Man	Peyton Reed	Paul Rudd, Evangeline Lilly, Corey Stoll, Bobb...	India	12-I
14	627443e6fd923b2544adef07	s1041	Movie	Marvel Studios' Captain America: Civil War	Anthony Russo, Joe Russo	Chris Evans, Robert Downey Jr., Scarlett Johan...	India	12-I
15	627443e6fd923b2544adef0b	s1045	Movie	Marvel Studios' Doctor Strange	Scott Derrickson	Benedict Cumberbatch, Chiwetel Ejiofor, Rachel...	India	12-I
16	627443e6fd923b2544adeda4	s686	Movie	Marvel Studios' Thor: Ragnarok	Taika Waititi	Chris Hemsworth, Tom Hiddleston, Cate Blanchet...	India	5-I

For futher task making copy of data first place curser to make copy of orginal data

In [4]:
```
cursor = collection.find()
df = pandas.DataFrame(list(cursor))
df.shape
```

Out[4]:
```
(1449, 14)
```

In [5]:
```
disney_plus_df = df.copy()
```

```
removing unwanted column which is not releated to this task
```

In [6]:
```
disney_plus_df.drop(['director','cast','date_added','country','marvel_t'],axis=1,
```

In [7]:
```
disney_plus_df['rating']
```

Out[7]:
```
0        TV-G
1          PG
2        TV-G
3       TV-PG
4         NaN
        ...
1444       PG
1445    PG-13
1446       PG
1447    PG-13
1448    PG-13
Name: rating, Length: 1449, dtype: object
```

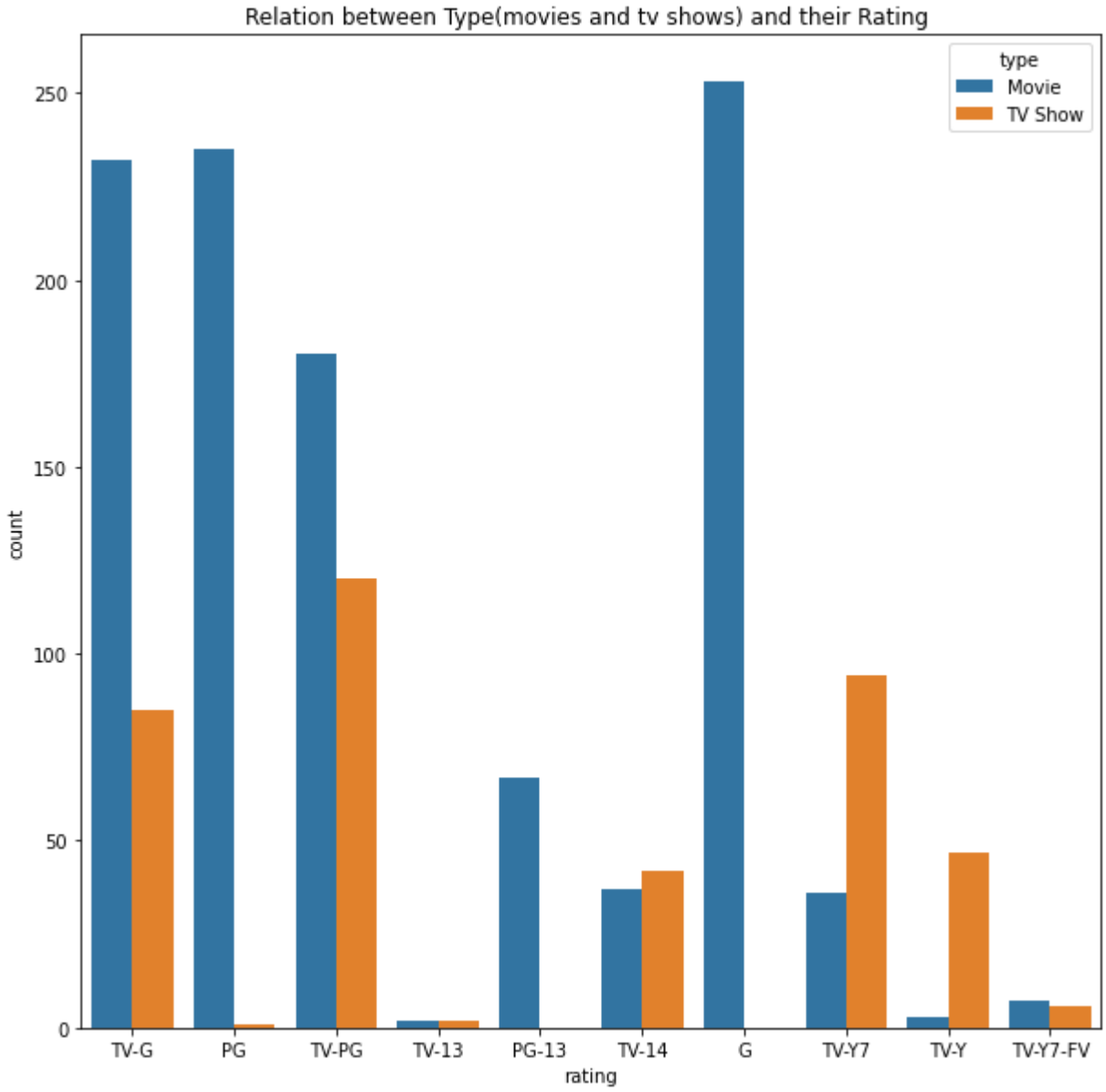
In [8]:
```
disney_plus_df['rating'].replace(numpy.NaN,'TV-13',inplace=True)
```

In [9]:
```
disney_plus_df.isnull().sum()
```

Out[9]:
```
_id             0
show_id         0
type            0
title           0
release_year    0
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

```
Task 2: Releation between Type(movies or TV shows) and rating
```

In [10]:
```
matplotlib.pyplot.figure(figsize=(10,10))
seaborn.countplot(x='rating',hue='type', data=disney_plus_df)
matplotlib.pyplot.title('Relation between Type(movies and tv shows) and their Rat
matplotlib.pyplot.show();
```



In [ ]: