

**Pak- Austria Fachhochschule Institute of Applied Sciences and Technology (PAF IAST)**

***Submitted By:***

*Name: Arshia Fajar*

*Department : IT & CS*

*Program: BS Data Science (Spring 23)*

*Registration Number: B23S0025DS001 (Reg No)*

***Submitted To:***

*Instructor Name: Sir Iqbal*

*Course Name: Intro To DSA*

*Lab No: 12*

*Topic: Project Report & Presentation*

*Date: 31, May ,2024*

# **PROJECT REPORT**

## **TITLE: NETFLIX MOVIE AND STOCK ANALYSIS**

### **Objective:**

To do the data analysis on a dataset and visualize it for better understanding of trends.

### **Environment Used:**

Jupyter Notebook

### **Performed Visualization:**

In this project , Exploratory Data Analysis has been done on netflix movies and tv shows data set. The project focuses on showing the trend among different variables of the dataset. It shows the total contribution by movies and tv shows into the dataset.

- Year in which highest content was released
- Heat Map showing null values
- Distribution of movies & Tv shows in whole data set
- Distribution of each country into content release. (g map)
- Ratings of movies & TV Shows
- Comparative graph of ratings for movies and tv shows in the US , India and France.
- Top directors of dataset
- Top countries producing content
- The highest duration for which movie or tv show has run
- Top Actor in the dataset
- Oldest movies in the dataset
- Top genres in movies and tv shows.
- Majority holders in stock
- Stock recommendation
- Line plot for high,open,close
- Volume of stock over years
- High ,open,close day,month and year wise
- Top highest stock values
- Top lowest stock values
- Visualization of high vs low overtime

### **Statistical Analysis :**

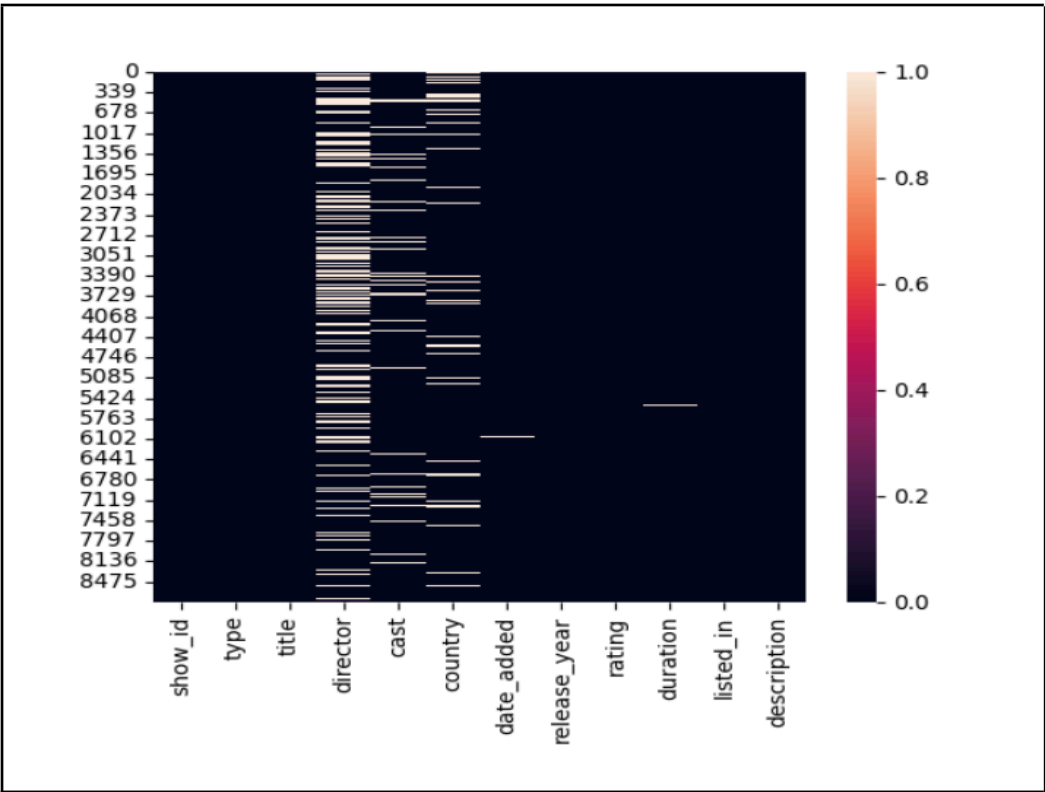
All the columns in data set are categorical or object type , except the release year. Therefore, the statistical analysis is given only for this column.

	release_year
count	8808.000000
mean	2014.180972
std	8.819111
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

Heat Map For Null Values:

Heat map is an easy way to visualize the null count because it gives the whole color a color and then shows small bars for the null values in the columns if it exists. So greater the number of bars greater the null count.

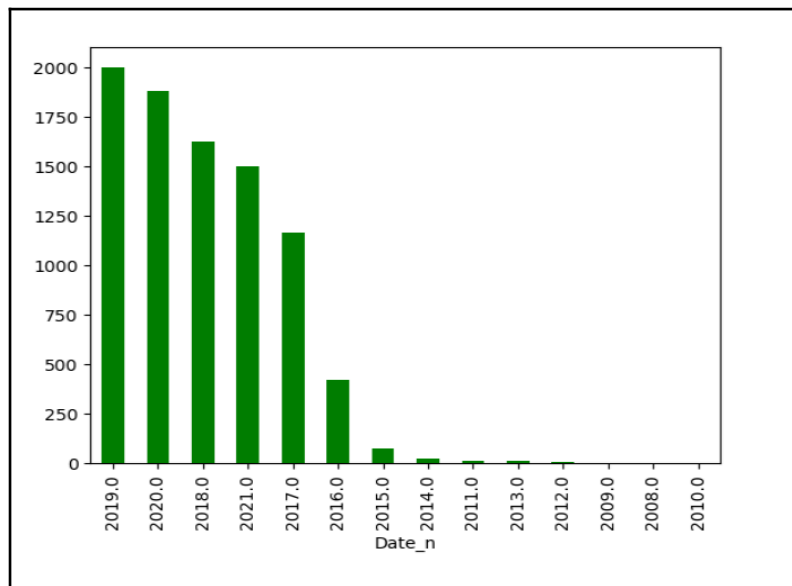
Such as in this dataset director column has the greatest null count



## Highest Content Released By Year:

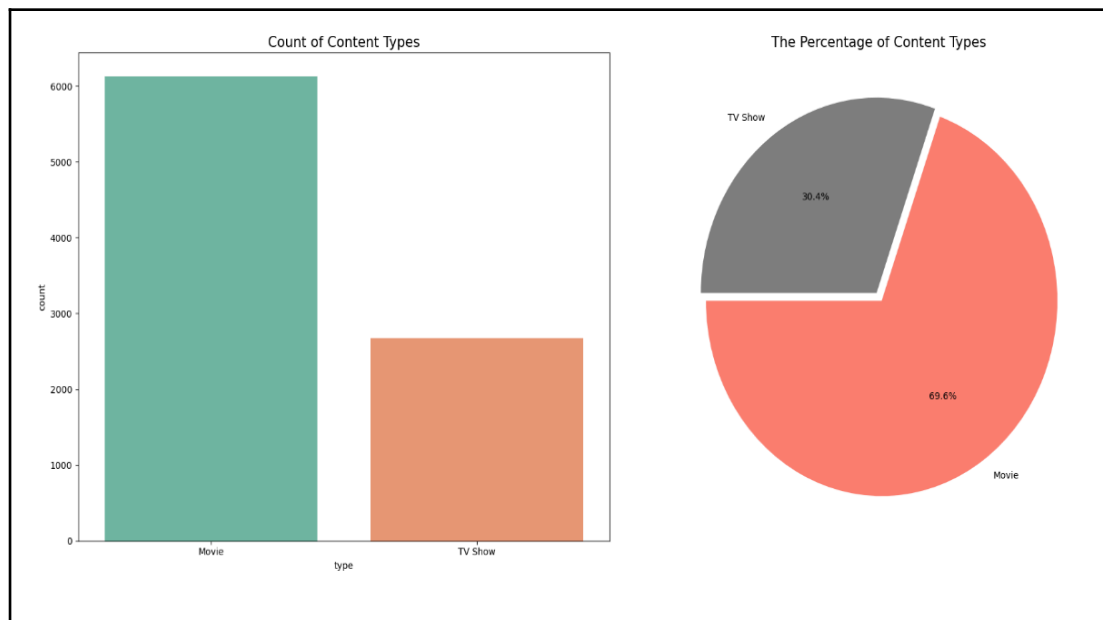
The greatest of numbers of content released (about 2000) is in the year 2019 then in 2020 and so on.

We counted the number of times a year has occurred in date column and counted its occurrence. That's how we got the year count.



## Distribution Of Movies & TV shows:

Through the percentage and bar graph it's obvious that the content is more in terms of movies as compared to TV shows.

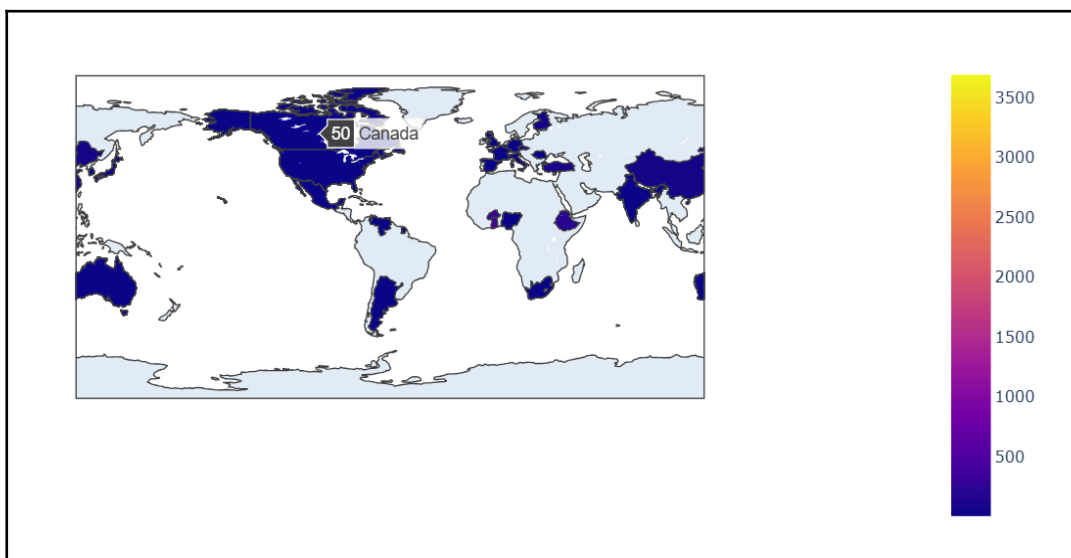


## Content Released By Country

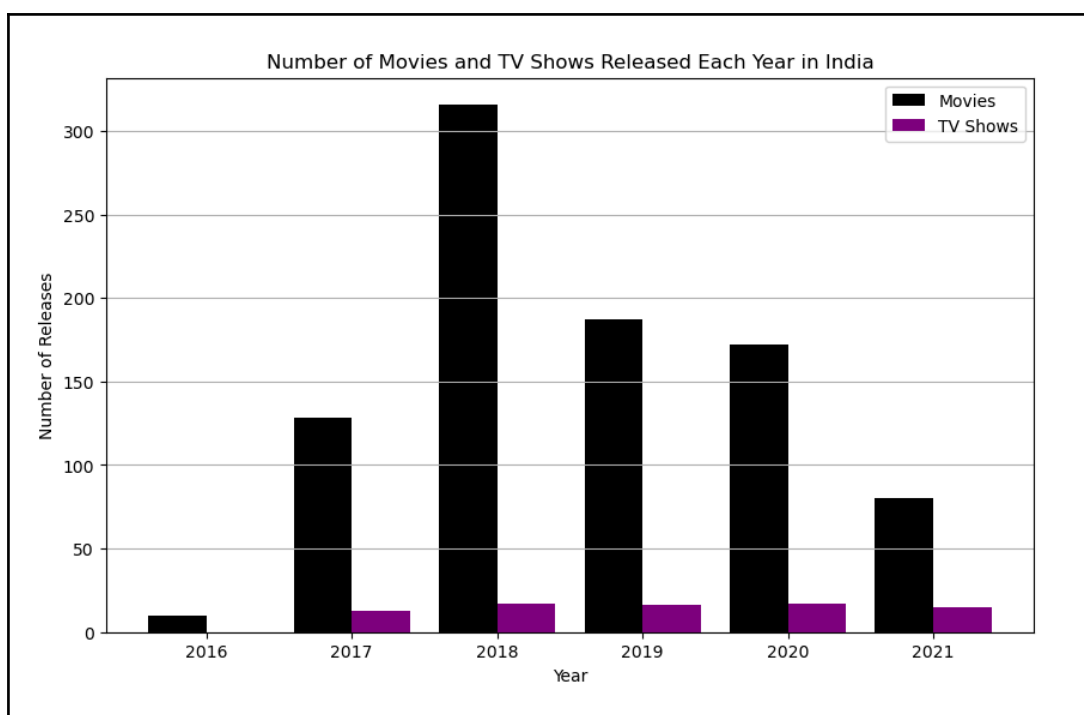
We made this graph using plotly because it allows us to zoom in , expand, scroll and see the count related to each value in the desired column we have made the graph for.

for example in our graph the region canada shows a count of 50 mean that total content released would be 50 either movies or TV shows

	country	count
0	United States	2818
1	India	972
2	United Kingdom	419
3	Japan	245
4	South Korea	199



## Movies & TV Shows Released By India Each Year

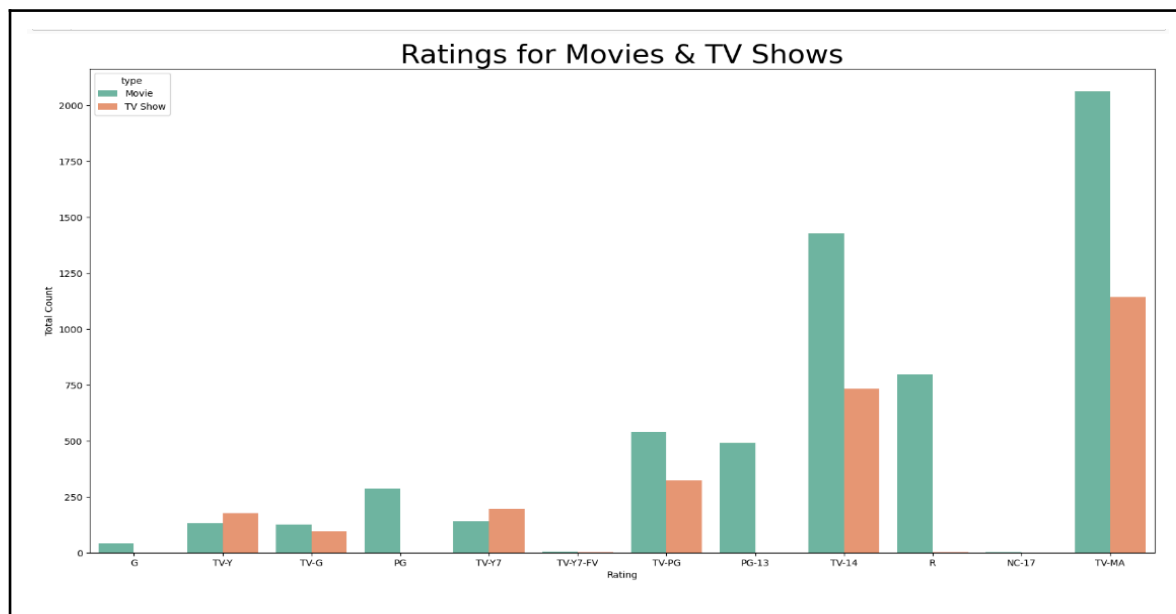


## Ratings For Movies & TV Shows

- G (General Audience)
- PG (Parental Guidance Suggested)
- PG-13 (Parents Strongly Cautioned)
- R (Restricted)
- TV-MA (Mature Audience)
- TV-Y (All Children)
- NC-17 content is restricted to viewers 17 years old and above.
- PG-13 content may be inappropriate for children under 13 years old.
- TV-14: Intended for viewers aged 14

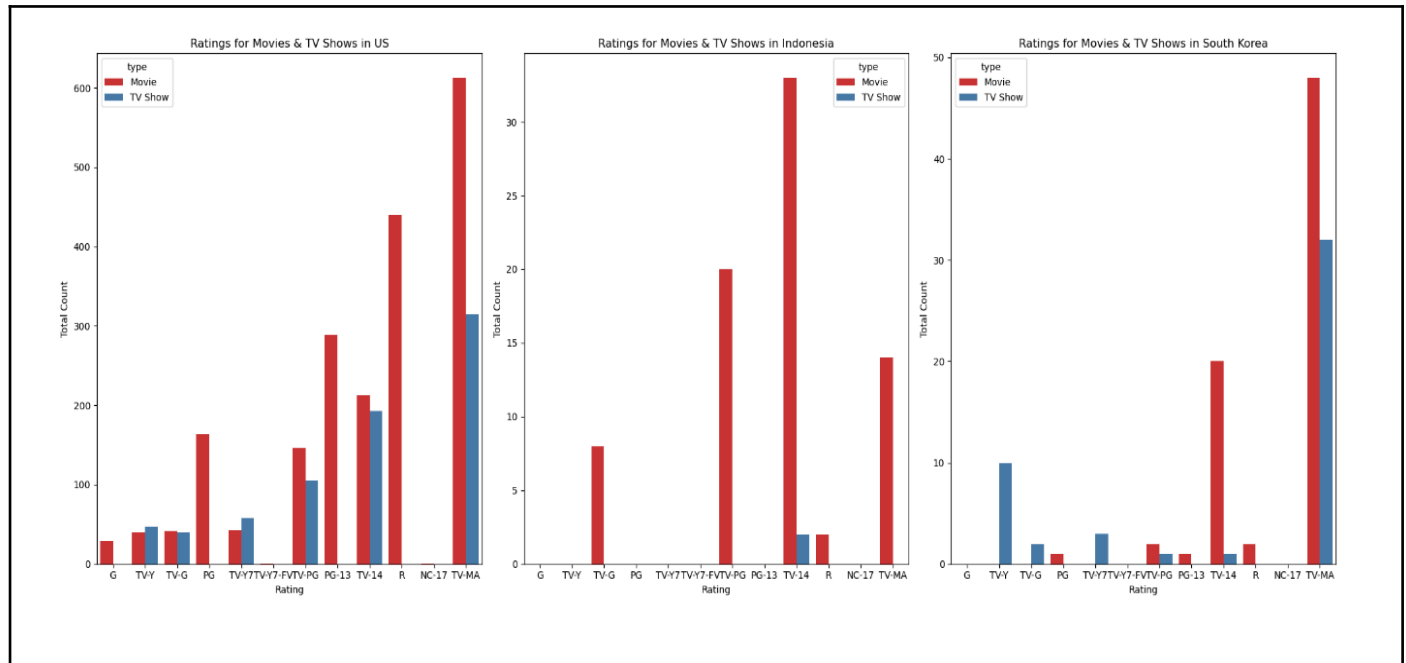
The count for movies and TV is the greatest for TV-MA ranking shows therefore it means that most of the movies and tv shows are for a mature audience.

The bar for General Audience is the smallest for movies which means that very less movies are for general audience while bar for tv in this shows does not exist meaning that no tv shows are for general audience

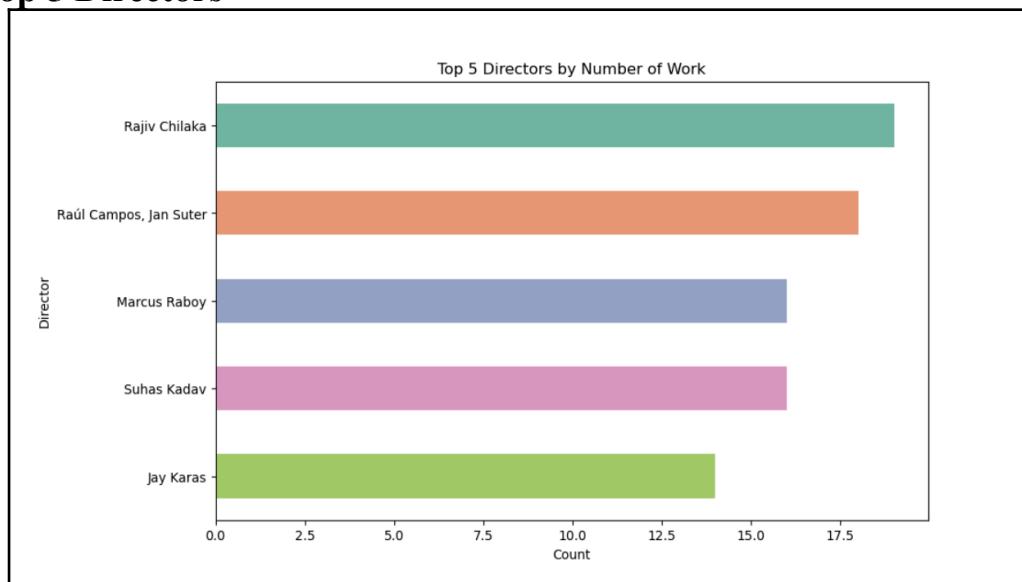


## Comparison Of Rating Between US, India and France

- The US has tv shows and movies for a more mature audience . ( previous graph)
- Indonesia has very less number of Tv shows and movies released while those which are released lies in the ranking TV-MA means mature audience
- France has the maximum number of movies and Tv shows that has TV-Ma ranking



## Top 5 Directors

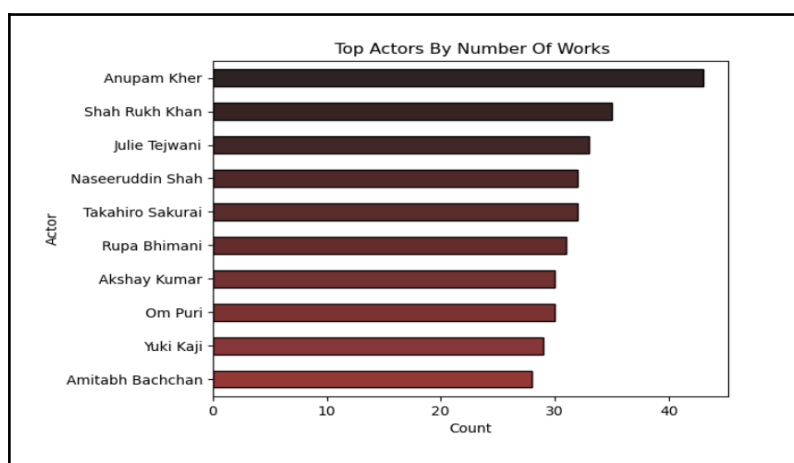


## Top 5 Content Producing Countries

	country	count
0	United States	2818
1	India	972
2	United Kingdom	419
3	Japan	245
4	South Korea	199

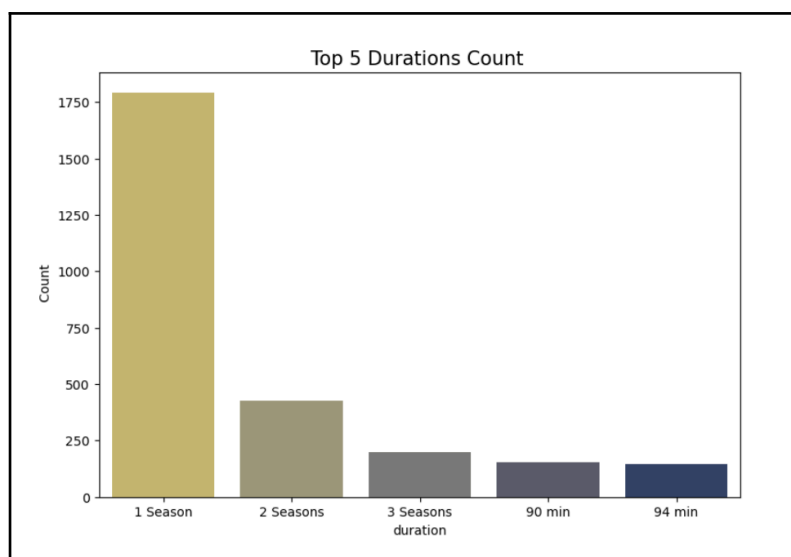
## Top Actors

The top 10 actors in the whole dataset are given .The count for the 1st top is 40 means that in the given dataset she has been casted in 40 movies.



## Top 5 Durations For Movies To Last

The longest a content has ever lasted is for 1 season, then a less number of content has 2 seasons, then 3 seasons and the shortest duration is 94 minutes. ( 94 minutes is 1hr and 30min the average time of a netflix movie.





## Oldest Movies

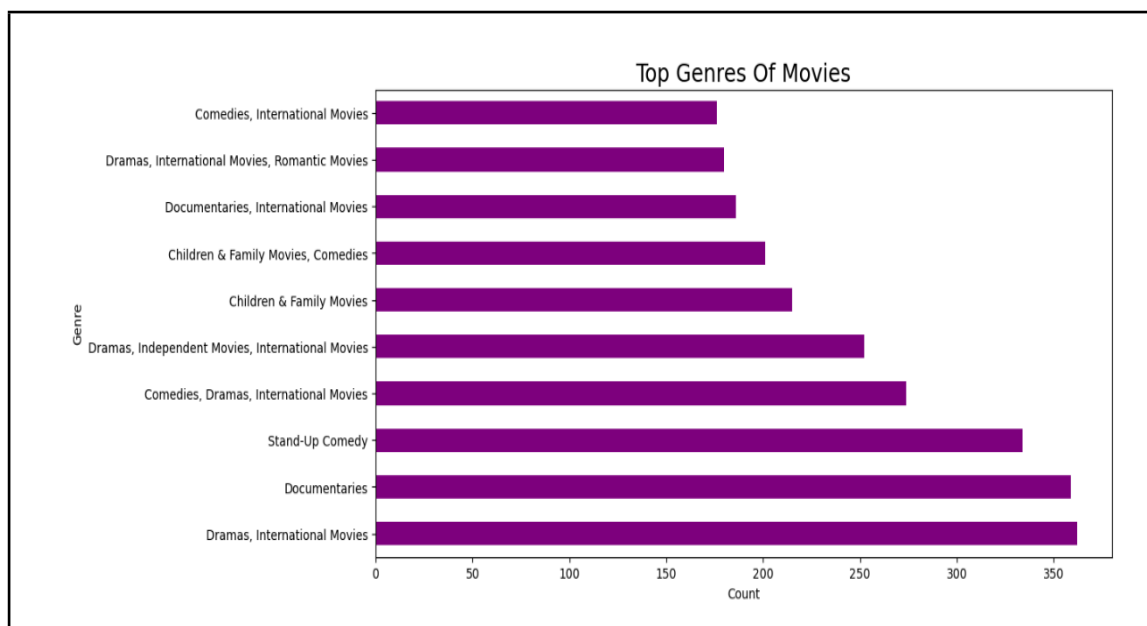
Through the released year column we have sorted in ascending order and automatically the oldest year would be printed first. The oldest movie was released in 1925.

	title	release_year
4250	Pioneers: First Women Filmmakers*	1925
7790	Prelude to War	1942
8205	The Battle of Midway	1942
8660	Undercover: How to Operate Behind Enemy Lines	1943
8739	Why We Fight: The Battle of Russia	1943
8763	WWII: Report from the Aleutians	1943
8640	Tunisian Victory	1944
8436	The Negro Soldier	1944
8419	The Memphis Belle: A Story of a Flying Fortress	1944
7930	San Pietro	1945
1331	Five Came Back: The Reference Films	1945
7219	Know Your Enemy - Japan	1945
7575	Nazi Concentration Camps	1945
7743	Pioneers of African-American Cinema	1946
7294	Let There Be Light	1946

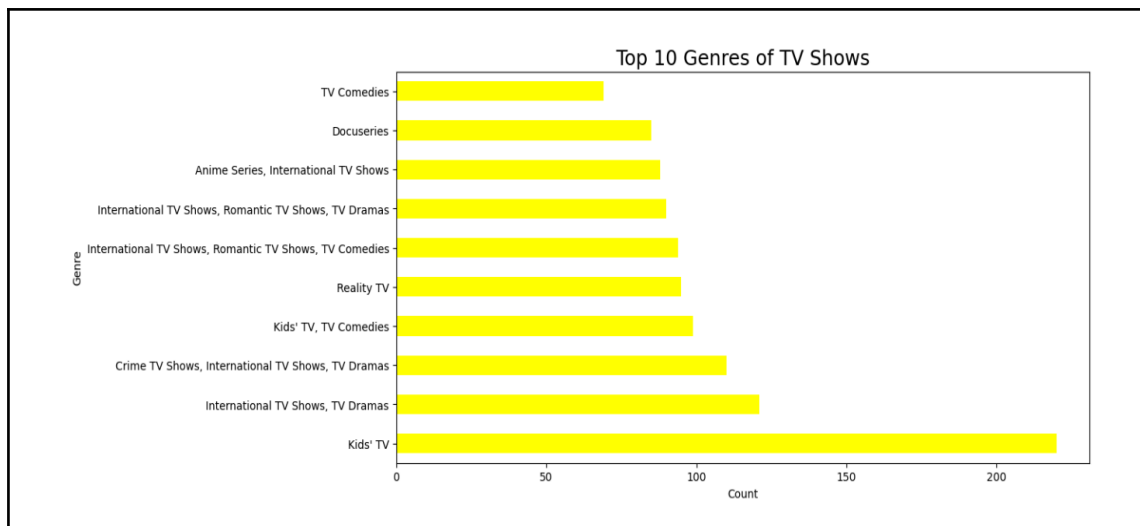
## Top Genres In Movies

Through the type column and listed\_in column we accessed all the genre of movies and tv shows. By the bar length:

- In movies the top genre is international movies and dramas with the count of almost 350+.
- In TV shows the top genre is Kids shows with the count of 250+



## Top Genre In Tv Shows



## Stock Analysis On Netflix DataSet

### Majority Holders

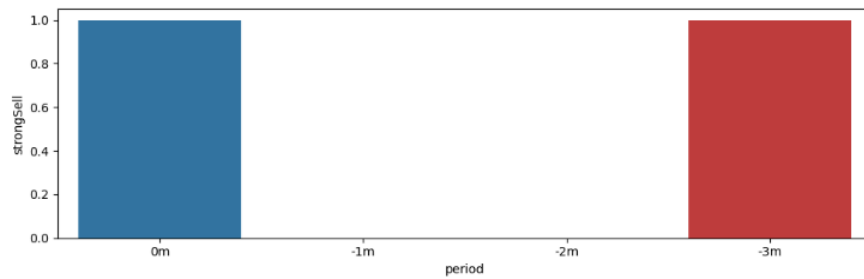
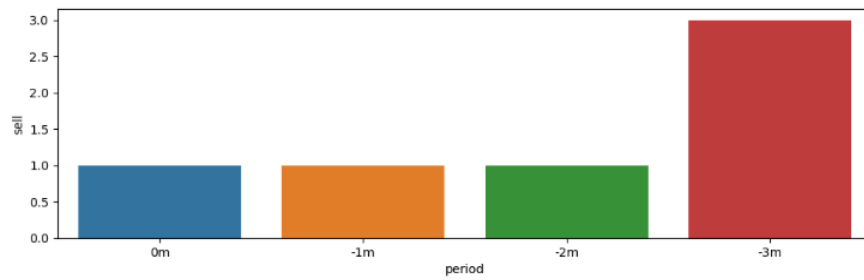
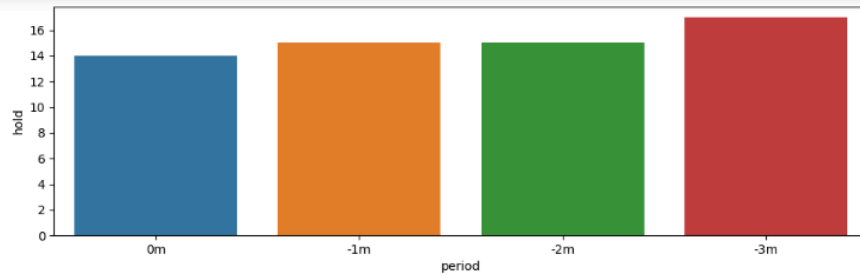
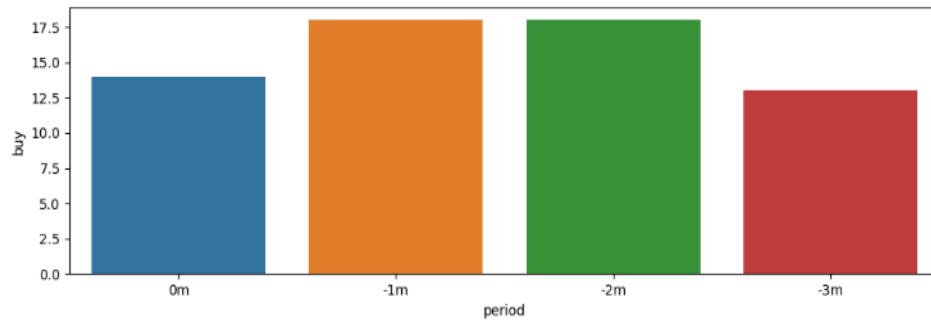
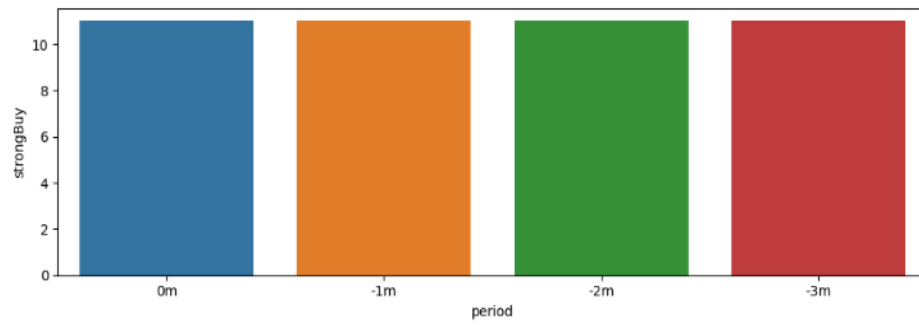
This represent the names of share holders ,their shares , and the current value of shares

	Date Reported	Holder	pctHeld	Shares	Value
0	2023-06-30	Vanguard Group Inc	0.0816	36179028	15936500043
1	2023-06-30	Blackrock Inc.	0.0675	29911686	13175798566
2	2023-06-30	FMR, LLC	0.0531	23551711	10374293178
3	2023-06-30	Capital Research Global Investors	0.0394	17479045	7699344532
4	2023-06-30	State Street Corporation	0.0377	16699906	7356141593
5	2023-06-30	Price (T.Rowe) Associates Inc	0.0279	12378010	5452389624
6	2023-06-30	Capital World Investors	0.0258	11431257	5035354395
7	2023-06-30	Capital International Investors	0.0252	11167223	4919050059
8	2023-06-30	JP Morgan Chase & Company	0.0216	9594129	4226117883
9	2023-06-30	Geode Capital Management, LLC	0.0189	8390820	3696072301

## Recommendations On Buying Stocks

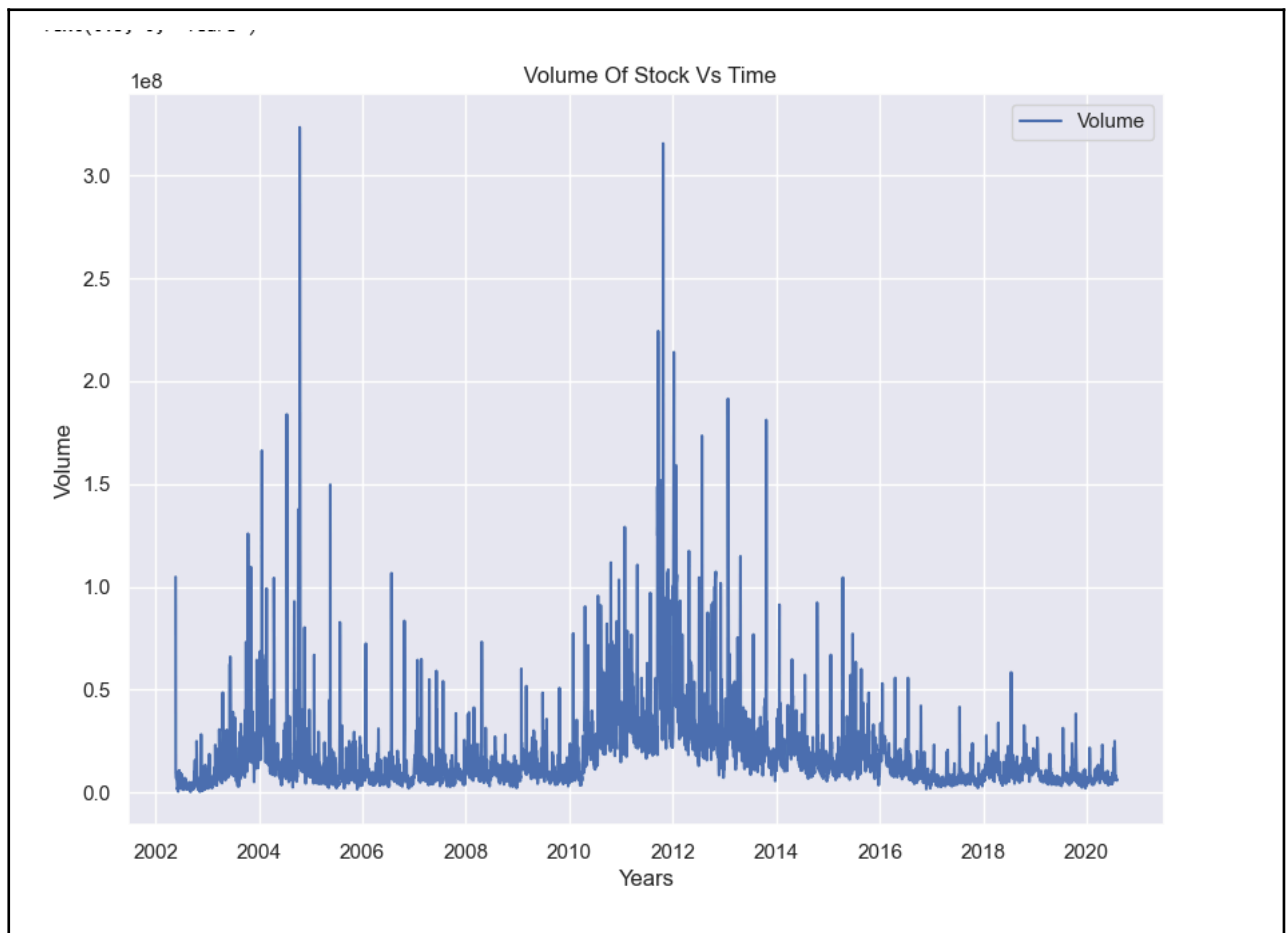
- The recommendations are actually determined by analyst and are based on months.
- 0 in period means current month, -1 means one month ago what were their analysis on selling or buying or holding....stocks
- strong sell means who strongly recommend selling the stock

## Netflix Recommendations On Stocks



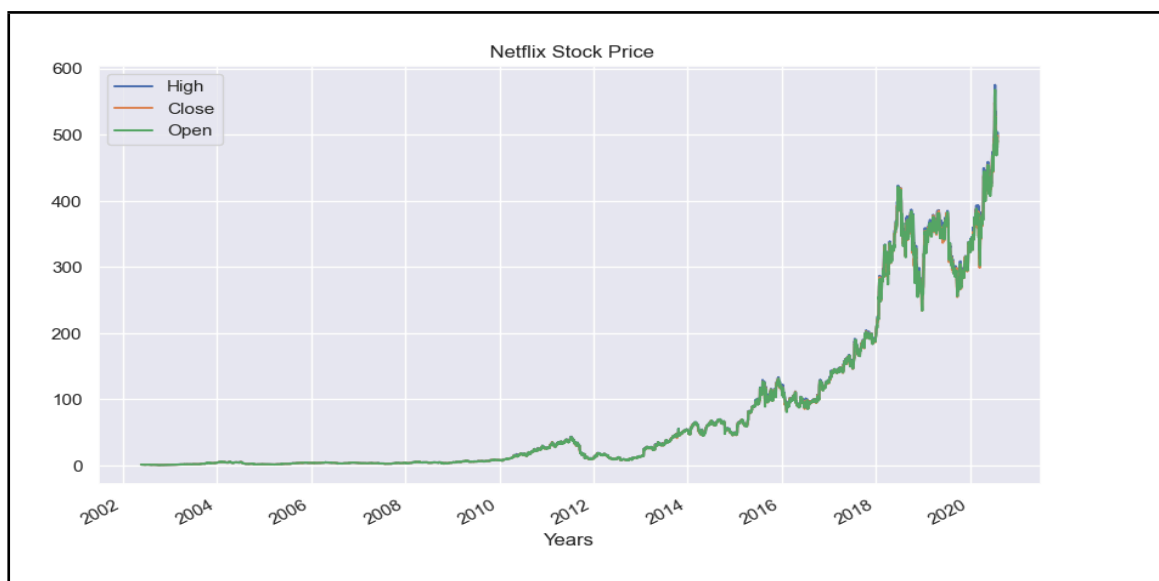
## Stock Vs Volume Over Years

It shows in which year stock had the maximum volume. In the year 2005 and 2017 the stock had the maximum volumes



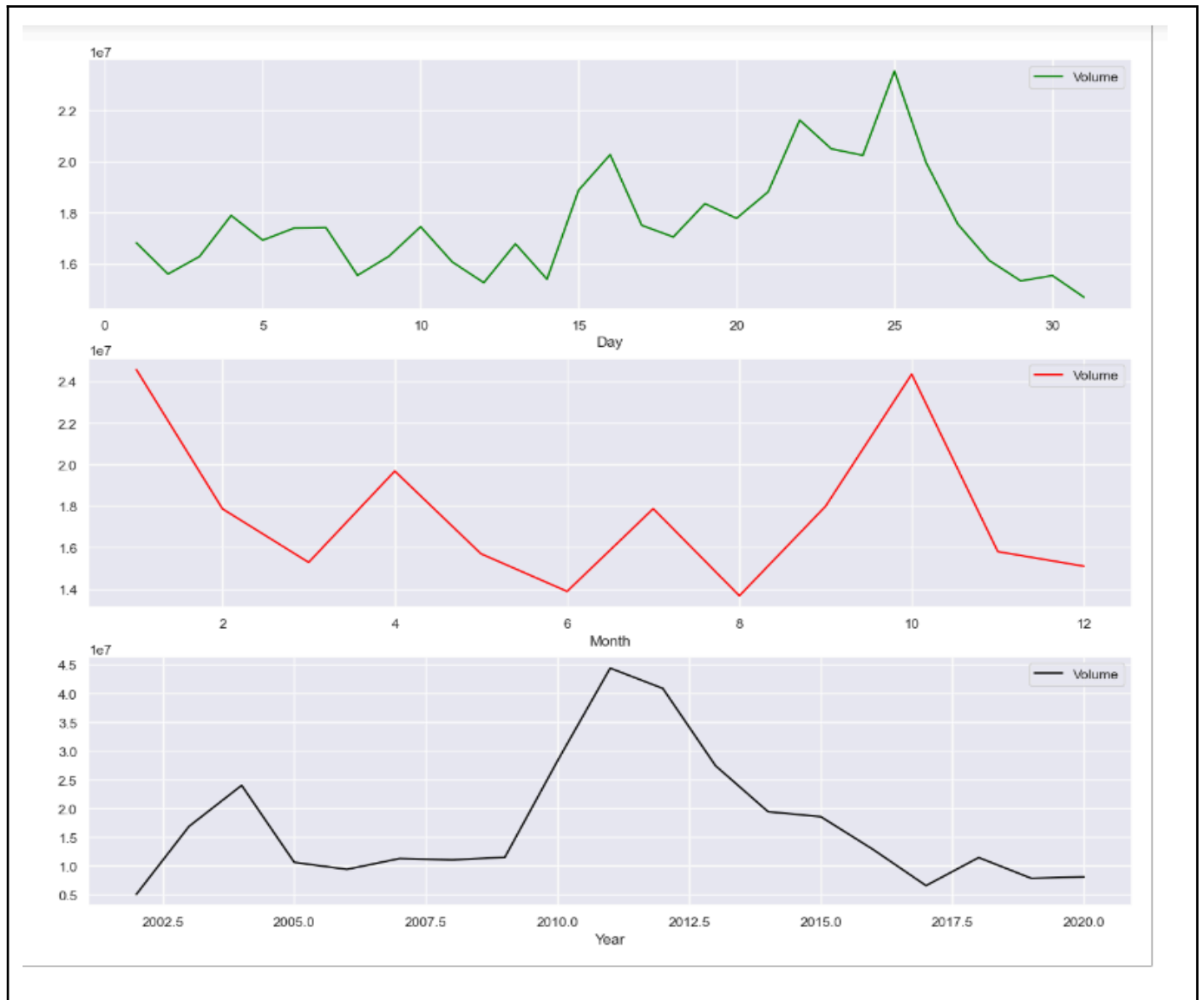
## Close Open & High Over The Years

This shows combined line plot for high open and close column by years. Because the data points for three are so close to each other they overlap and colored lines are not separated.



## Stock For day, month and year wise

By accessing the day, month and year column we have plotted the Volume .



## Top 5 Dates With Highest Stock Values

Extracting high column from data set, applying sort to it and setting ascending False will print the highest values first for high column

```
Date
2020-07-13    575.369995
2020-07-10    555.880005
2020-07-16    535.539978
2020-07-15    529.000000
2020-07-14    525.500000
Name: High, dtype: float64
```

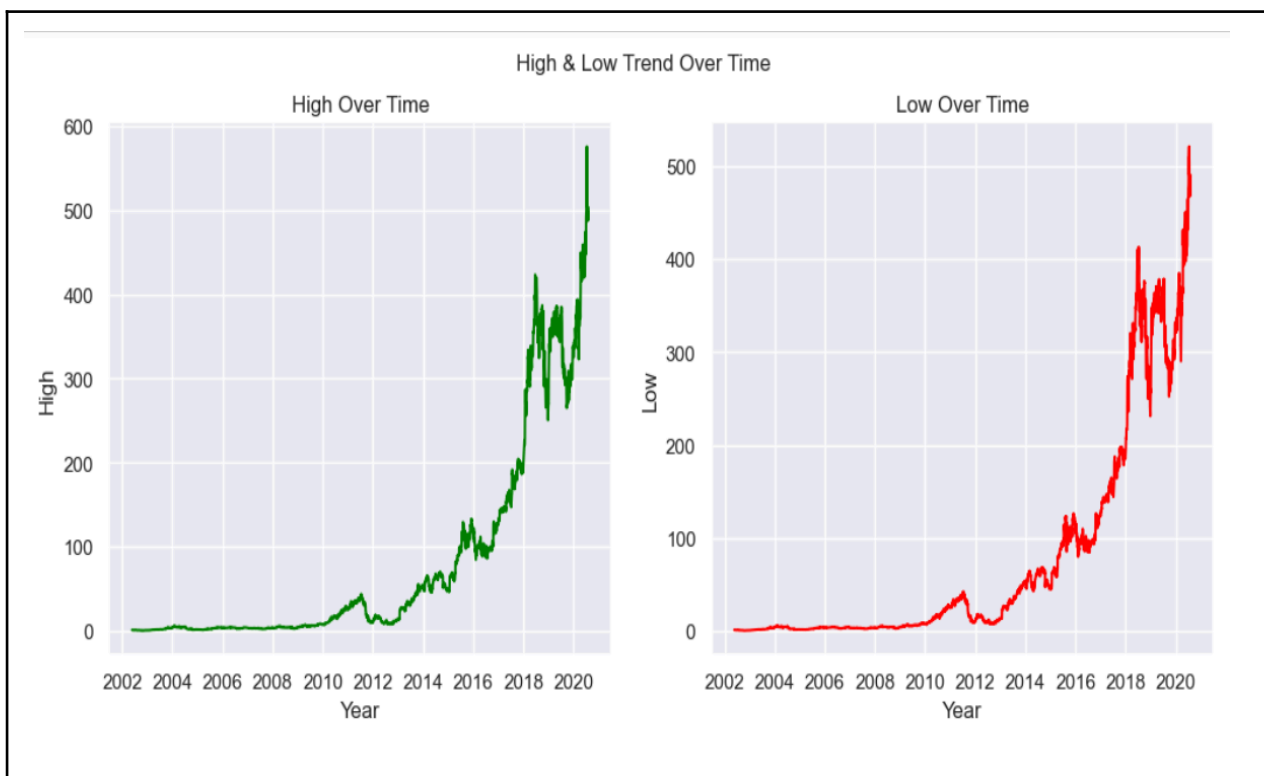
## Top 5 dates with Lowest Stock Values

Extracting low column from data set, applying sort to it and setting ascending True will print the values from lowest to highest manner for low column

```
Date
2002-10-10    0.346429
2002-10-09    0.347143
2002-10-07    0.382143
2002-10-08    0.390714
2002-10-16    0.442857
Name: Low, dtype: float64
```

## Graphs For High & Lower Over Years

This shows the plot for the actual high and low column ( not the top 5 for open and low ) in our data set by years. Since in our dataset the values for these column are so close to each other that's why both are showing the same trend



## Conclusion:

The task was to know the trend for for different columns of dataset which we have done perfectly. We have discovered the insights from both stock nad movies and tv shows datasets and did the visualization for each possible analysis.

=====