DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Graphing using Matplot in Python

EXPERIENTIAL LEARNING PROJECT
REPORT

Submitted by,

Machani Bhanu Teja	RVCE23BAI119
Garv Agarwalla	RVCE23BAI054
Aditya Ranjan	RVCE23BAI05l

ACKNOWLEDGEMENT

We are indebted to our faculty, **Mr. Somesh Nandi,** Assistant Professor, Department of Artificial Intelligence and Machine Learning, RVCE for all his support, encouragement, and invaluable advice throughout our Experiential Learning Project.

We are sincerely grateful to our beloved Principal, **Dr. K. N. Subramanya** for his appreciation towards this Experiential Learning Project.

Lastly, we thank our family members and friends who supported us throughout the project work.

CERTIFICATE

This is to certify that the Experiential Learning Project titled 'GRAPHING USING MATPLOT IN PYTHON' carried out by Garv Agarwalla , Machani Bhanu Teja , Aditya Ranjan who are bonafide students of RV College of Engineering, Bengaluru , during the first semester, in the year 2023-2024. It is also certified that all corrections/suggestions indicated for the Internal Assessment have been incorporated in the report. The report has been approved as it satisfies the academic requirements in respect of Experiential Learning.

 Marks awarded:

                                                         Signature of Staff In-charge:

ABSTRACT

This study explores a diverse Netflix rating dataset, investigating the influence of genre, release year, and content type on user preferences. Utilizing advanced statistical methods, sentiment analysis, and SQL for enhanced data exploration, the research reveals nuanced insights for optimizing user experiences. The project, featuring a Python-based solution integrated with SQL capabilities, contributes valuable findings to the digital entertainment landscape, offering content creators and platforms actionable strategies to align with audience expectations and refine content recommendations.

TABLE OF CONTENTS

INTRODUCTION

❑ Netflix's dataset captures global user preferences across a wide array of content, providing a rich source for understanding viewer choices.

❑ With a vast international subscriber base, the dataset reflects diverse audience tastes, enabling a comprehensive analysis of cross-cultural preferences.

❑ By examining user ratings, we can identify popular content and uncover factors contributing to the success of specific titles.

❑ The dataset spans different release years, offering insights into the evolution of user preferences over time.

❑ Going beyond numerical ratings, sentiment analysis of user reviews adds depth, allowing a nuanced exploration of viewer satisfaction within the Netflix ecosystem.

DATASET CHOSEN AND ITS RELEVANCE

❑ Netflix uses recommendation algorithms to suggest movies and TV shows to its users. dataset containing user ratings, viewing history, and other relevant information can be crucial for improving and developing more effective recommendation systems.

❑ Understanding user preferences through ratings can help Netflix in making informed decisions.

❑ Analyzing ratings can provide insights into user satisfaction and dissatisfaction. Netflix can see this information to enhance the user experience, identify areas of improvement, and address any issues with their platform.

❑ Researchers and data scientists can use Netflix rating datasets for academic and analytical purposes. The dataset can be employed to study user behavior, trends in content consumption, and the impact of various factors on user engagement.

❑ Netflix may use rating datasets to benchmark its content against competitors, evaluate the success of different genres, and assess how well their original productions are received compared to licensed content.

❑ The data can be leveraged to create more personalized user experiences. By understanding individual preferences, Netflix can provide users with tailored content recommendations and improve overall satisfaction.

NumPy and Pandas Libraries

NumPy Library :

❑ NumPy provides array creation, indexing, element-wise operations, and linear algebra functions, making it fundamental for numerical computations.

❑ Usage in Code: In the code, NumPy is imported as 'np' and used implicitly for numerical operations, though not explicitly shown. It's valuable for calculations involving data.

Pandas Library :

❑ Pandas excels at data exploration, offering functions like read_csv, head(), describe() and dtype() for effective data loading, cleaning, and analysis .

❑ Usage in Code: pandas is used to read CSV data into a DataFrame . DataFrames are central for data exploration, and functions like describe(), head() and dtypes() provide insights into the dataset.

Matplot and Seaborn Libraries

Matplot Library :

❑Matplotlib is a versatile Python library for creating various types of static and interactive data visualizations.

❑It allows fine-grained control over plot elements, enabling customization of colors, markers, labels, and titles.

❑Usage in Code: Matplotlib is imported as 'plt' in the code. It's used to create and customize plots like boxplots and barplots to effectively visualize data.

Seaborn Library :

❑Seaborn is a Python library based on Matplotlib, specializing in creating aesthetically pleasing and informative statistical graphics.

❑It offers built-in plots like box plots, violin plots, pair plots, and heatmaps optimized for visualizing statistical relationships and distributions.

❑Usage in Code: Seaborn is imported as 'sns' in the code. It's used to create specific types of plots, such as boxplots and barplots, to visualize data relationships and distributions within the dataset.

SQL Libraries

SQL Library :

❏ SQL is a standardized language for managing relational databases, supporting operations like table creation, data insertion, and information retrieval.

❏ Widely used across various database management systems, SQL's declarative nature makes it efficient for handling large datasets, essential for developers, data analysts, and administrators in critical business applications.

❏ Usage in Python Code: In python Jupyter notebook, 'import sqlite3' imports the SQLite module, enabling interaction with SQLite databases. 'conn' represents a connection to a database, created using sqlite3.connect(). It establishes a connection to a database file ('Netflix.db' in this case). The '%%sql' is a magic command used in Jupyter notebooks to execute SQL queries directly within cells.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sqlite3
import seaborn as sns
```

```python
data= pd.read_csv('Netflix.csv')
arr = np.array(data)
```

NUMPY

```python
print(np.mean(data.imdb_votes, axis=0))
print()
print(np.mean(data.imdb_score, axis=0))
print()
print(np.max(data.imdb_score, axis=0))
print()

sorted_ind = np.argsort(arr[:,9])
sorted_data = arr[sorted_ind]
print(sorted_data[:10])
print()
print(sorted_data[-10:])
```

23407.194987659008

6.5334469051675175

9.6

[[1664 'tm549892' 'Aerials' 'MOVIE'
  'Earth is invaded by Aliens from outer space. An intermarried couple living in the city of Dubai are confined
to their home due to the uncertainty of the situation. Disconnected from the world outside due to the loss of c
ommunication, they explore around their cultural differences in Science in order to understand the reason behin
d Aliens coming to our planet; only to find themselves caught between a series of extraterrestrial encounters a
t their very home.'
  2016 nan 90 'tt5314138' 1.5 382.0]
 [4648 'ts319303' "He's Expecting" 'SHOW'
  "When a successful ad executive who's got it all figured out becomes pregnant, he's forced to confront social
inequities he'd never considered before."
  2022 'TV-PG' 26 'tt14719276' 1.6 2735.0]
 [1101 'tm171891' 'Himmatwala' 'MOVIE'
  'When a temple priest commits suicide after being dishonored by an evil landlord, his son returns to his nati
ve village on a mission of vengeance.'
  2013 nan 150 'tt2802286' 1.7 8550.0]
 [3355 'tm839411' "Me Against You: Mr. S's Vendetta" 'MOVIE'
  'A young couple who makes popular YouTube videos for children sets out to win an award, but an evil mastermin
d stands in the way of their success.'
  2020 nan 64 'tt11559652' 1.7 362.0]
 [4782 'ts280693' 'Thomas & Friends: All Engines Go!' 'SHOW'
  'In a reboot of the classic TV series, a younger Thomas The Tank Engine goes on adventures with all of his fr
iends, as they work on The Island Of Sodor.'
  2021 'TV-Y' 11 'tt14556544' 1.7 509.0]
 [5035 'ts287941' 'Racket Boys' 'SHOW'
  "A story of a boys' badminton team at a middle school in Haenam as they compete in a junior athletic competit
ion. The 16 year old boys and girls in Haenam grow as people during this time."
  2021 'TV-14' 70 'tt14482414' 1.8 27484.0]
 [1556 'tm232930' 'Kyaa Kool Hain Hum 3' 'MOVIE'
  "In order to impress his girlfriend's conservative father, a porn actor uses his fellow adult movie stars in
creating a fake traditional family."
  2016 nan 124 'tt5290620' 1.9 4341.0]
 [801 'tm53395' 'FRED 3: Camp Fred' 'MOVIE'
  "Schools out, and Fred Figglehorn's dream of water slides, horseback riding and monkey butlers during the sum
mer turns into a nightmare of gruel and poisonous berries when his mom signs him up to an unsanitary camp."
  2012 'PG' 81 'tt2395219' 2.0 2625.0]
 [26 'tm94651' 'Dostana' 'MOVIE'
  "Vijay and Ravi are best friends (hence the name 'Dostana') who do not even question each other about their c
areers. Vijay, a police officer and Ravi, a lawyer, while the former catches the criminals who smuggle, Ravi on
the other hand bail those criminals, employed by Daaga. One day, both Vijay and Ravi meet Sheetal at different
places and times, and both fancy the same girl, Sheetal, but Vijay's love was more serious. Ravi admits his lov
e for Sheetal to Vijay and is devastated but Vijay decides to sacrifice his love and give Sheetal's hand. Howev
er, there was a misunderstanding when Daaga decided to stir up Vijay and Ravi's friendship by showing a photo o
f Vijay and Sheetal to Ravi. They became rivals for the first time and Ravi started to turn to be an enemy of V
ijay. Do their friendship recover? Does Ravi get to know the truth? Do Sheetal get back with Vijay?"
  1980 nan 161 'tt0080653' 2.1 25.0]
 [1710 'tm350727' 'Hajwala: The Missing Engine' 'MOVIE'
  'Khalid and Kehailan are two team leaders who are passionate about car racing and challenges. They enter into
a crazy competition to win the challenge involving a lot of surprises, which unexpectedly change both their liv
es.'
  2016 'R' 96 'tt6200274' 2.1 250.0]]

[[3026 'ts81120' 'The Last Dance' 'SHOW'
  'A 10-part documentary chronicling the untold story of Michael Jordan and the Chicago Bulls dynasty with rare
, never-before-seen footage and sound from the 1997-98 championship season â€" plus over 100 interviews with fa
mous figures and basketballâ€™s biggest names.'
  2020 'TV-MA' 50 'tt8420184' 9.1 108321.0]
```

```
[1953 'ts78298' 'My Mister' 'SHOW'
 'In a world that is less than kind, a young woman and a middle-aged man develop a sense of kinship as they fi
nd warmth and comfort in one another.'
 2018 'TV-14' 79 'tt7923710' 9.2 5481.0]
[4628 'ts296563' 'Who Rules The World' 'SHOW'
 'A legendary wuxia romance that focuses on the decade-long romance between Hei Fengxi and Bai Fengxi, and the
adventures they embark on together.'
 2022 'TV-PG' 44 'tt14950036' 9.2 81.0]
[1346 'ts37660' 'Reply 1988' 'SHOW'
 'A nostalgic trip back to the late 1980s through the lives of five families and their five teenage kids livin
g in a small neighborhood in Seoul.'
 2015 'TV-14' 95 'tt5182866' 9.2 6557.0]
[245 'ts3371' 'Avatar: The Last Airbender' 'SHOW'
 'In a war-torn world of elemental magic, a young boy reawakens to undertake a dangerous mystic quest to fulfi
ll his destiny as the Avatar, and bring peace to the world.'
 2005 'TV-Y7' 24 'tt0417299' 9.3 297336.0]
[3597 'ts90621' 'Kota Factory' 'SHOW'
 "Dedicated to Shrimati SL Loney ji, Shri Irodov ji and Maanniya HC Verma ji, 'Kota Factory' is TVF's latest o
riginal. India's first 'Black and White' show highlights the problems present day IIT-JEE aspirants face in the
ir day-to-day lives."
 2019 'TV-MA' 42 'tt9432978' 9.3 66985.0]
[3144 'ts85398' 'Our Planet' 'SHOW'
 "Experience our planet's natural beauty and examine how climate change impacts all living creatures in this a
mbitious documentary of spectacular scope."
 2019 'TV-G' 50 'tt9253866' 9.3 41386.0]
[229 'ts4' 'Breaking Bad' 'SHOW'
 "When Walter White, a New Mexico chemistry teacher, is diagnosed with Stage III cancer and given a prognosis
of only two years left to live. He becomes filled with a sense of fearlessness and an unrelenting desire to sec
ure his family's financial future at any cost as he enters the dangerous world of drugs and crime."
 2008 'TV-MA' 48 'tt0903747' 9.5 1727694.0]
[622 'ts160526' 'Khawatir' 'SHOW'
 'A TV show devoted to help young people to be mature in their faith, work & knowledge'
 2005 'TV-14' 20 'tt3062514' 9.6 3046.0]
[2787 'ts265844' '#ABtalks' 'SHOW'
 "#ABtalks is a YouTube interview show hosted by Anas Bukhash, showcasing the 'raw' side of celebrities, athle
tes, entrepreneurs, and influencers as human beings."
 2018 'TV-PG' 68 'tt12635254' 9.6 7.0]]
```

PANDAS

In [4]: `data.head(10)`

Out[4]:

	index	id	title	type	description	release_year	age_certification	runtime	imdb_id	imdb_score	imdb_votes
0	0	tm84618	Taxi Driver	MOVIE	A mentally unstable Vietnam War veteran works ...	1976	R	113	tt0075314	8.3	795222.0
1	1	tm127384	Monty Python and the Holy Grail	MOVIE	King Arthur, accompanied by his squire, recrui...	1975	PG	91	tt0071853	8.2	530877.0
2	2	tm70993	Life of Brian	MOVIE	Brian Cohen is an average young Jewish man, bu...	1979	R	94	tt0079470	8.0	392419.0
3	3	tm190788	The Exorcist	MOVIE	12-year-old Regan MacNeil begins to adapt an e...	1973	R	133	tt0070047	8.1	391942.0
4	4	ts22164	Monty Python's Flying Circus	SHOW	A British sketch comedy series with the shows ...	1969	TV-14	30	tt0063929	8.8	72895.0
5	5	tm14873	Dirty Harry	MOVIE	When a madman dubbed 'Scorpio' terrorizes San ...	1971	R	102	tt0066999	7.7	153463.0
6	6	tm185072	My Fair Lady	MOVIE	A snobbish phonetics professor agrees to a wag...	1964	G	170	tt0058385	7.8	94121.0
7	7	tm98978	The Blue Lagoon	MOVIE	Two small children and a ship's cook survive a...	1980	R	104	tt0080453	5.8	69053.0
8	8	tm119281	Bonnie and Clyde	MOVIE	In the 1930s, bored waitress Bonnie Parker fal...	1967	R	110	tt0061418	7.7	111189.0
9	9	tm67378	The Professionals	MOVIE	An arrogant Texas millionaire hires four adven...	1966	PG-13	117	tt0060862	7.3	16168.0

In [5]: `data.tail(10)`

	index	id	title	type	description	release_year	age_certification	runtime	imdb_id	imdb_score	imdb_votes
5273	5273	tm1097142	My Bride	MOVIE	The story follows a young man and woman who go...	2021	NaN	93	tt14216488	4.9	281.0
5274	5274	tm878575	The Heartbreak Club	MOVIE	Coping with heartbreak, the shy owner of floun...	2021	NaN	101	tt11841144	6.2	188.0
5275	5275	ts286386	The Big Day	SHOW	For six engaged couples, happily ever after be...	2021	TV-MA	45	tt13887518	4.6	327.0
5276	5276	tm985215	Princess 'Daya'Reese	MOVIE	Reese is a con artist from Manila who dreams o...	2021	NaN	115	tt13399802	7.2	45.0
5277	5277	tm1004011	Time to Dance	MOVIE	When a ballroom dancer's shot at a crucial t...	2021	NaN	107	tt8622232	2.2	950.0
5278	5278	tm1040816	Momshies! Your Soul is Mine	MOVIE	Three women with totally different lives accid...	2021	NaN	108	tt14412240	5.8	26.0
5279	5279	tm1014599	Fine Wine	MOVIE	A beautiful love story that can happen between...	2021	NaN	100	tt13857480	6.9	39.0
5280	5280	tm1045018	Clash	MOVIE	A man from Nigeria returns to his family in Ca...	2021	NaN	88	tt14620732	6.5	32.0
5281	5281	tm1098060	Shadow Parties	MOVIE	A family faces destruction in a long-running c...	2021	NaN	116	tt10168094	6.2	9.0
5282	5282	ts271048	Mighty Little Bheem: Kite Festival	SHOW	With winter behind them, Bheem and his townspe...	2021	NaN	0	tt13711094	8.8	16.0

```
In [6]: print(data.isnull().sum())
        print()
        print(data.info())
```

```
index                0
id                   0
title                0
type                 0
description          5
release_year         0
age_certification 2285
runtime              0
imdb_id              0
imdb_score           0
imdb_votes          16
dtype: int64

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5283 entries, 0 to 5282
Data columns (total 11 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   index              5283 non-null   int64
 1   id                 5283 non-null   object
 2   title              5283 non-null   object
 3   type               5283 non-null   object
 4   description        5278 non-null   object
 5   release_year       5283 non-null   int64
 6   age_certification  2998 non-null   object
 7   runtime            5283 non-null   int64
 8   imdb_id            5283 non-null   object
 9   imdb_score         5283 non-null   float64
 10  imdb_votes         5267 non-null   float64
dtypes: float64(2), int64(3), object(6)
memory usage: 454.1+ KB
None
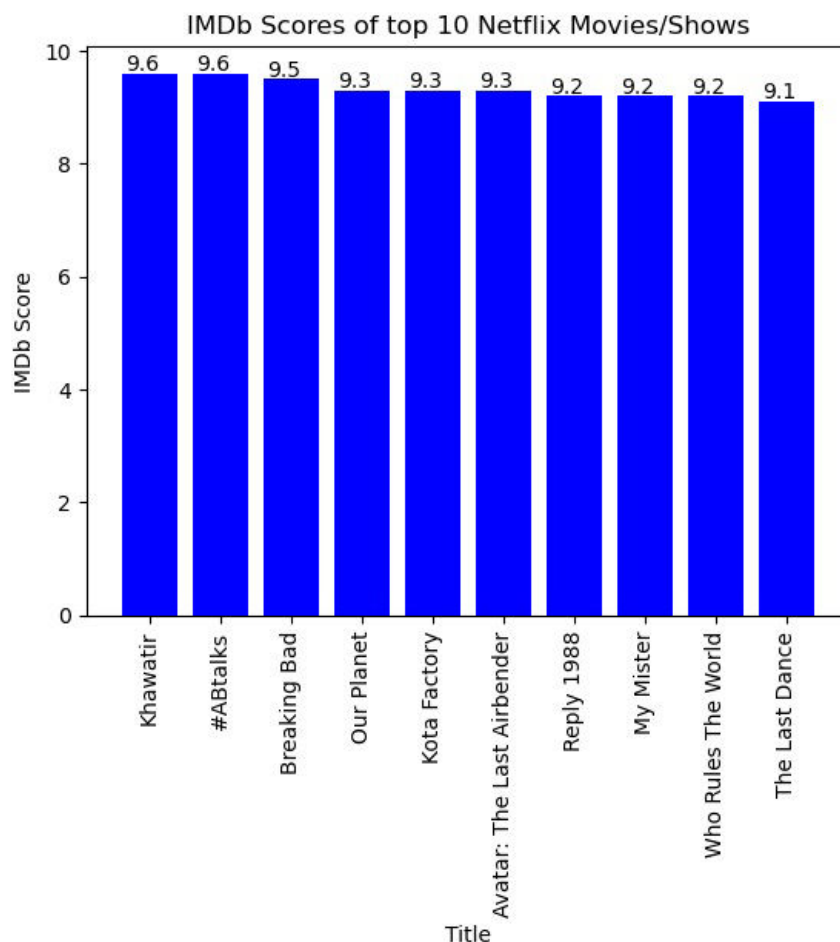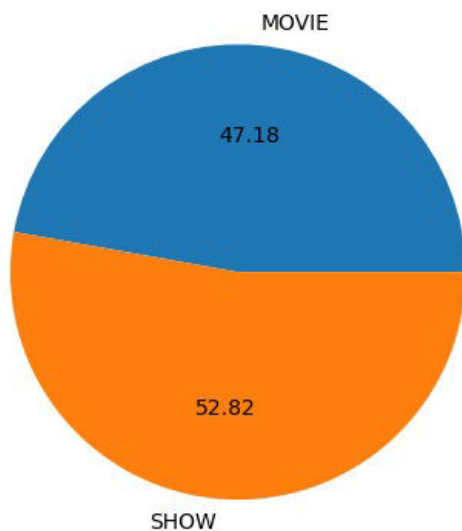
MATPLOTLIB
```

```
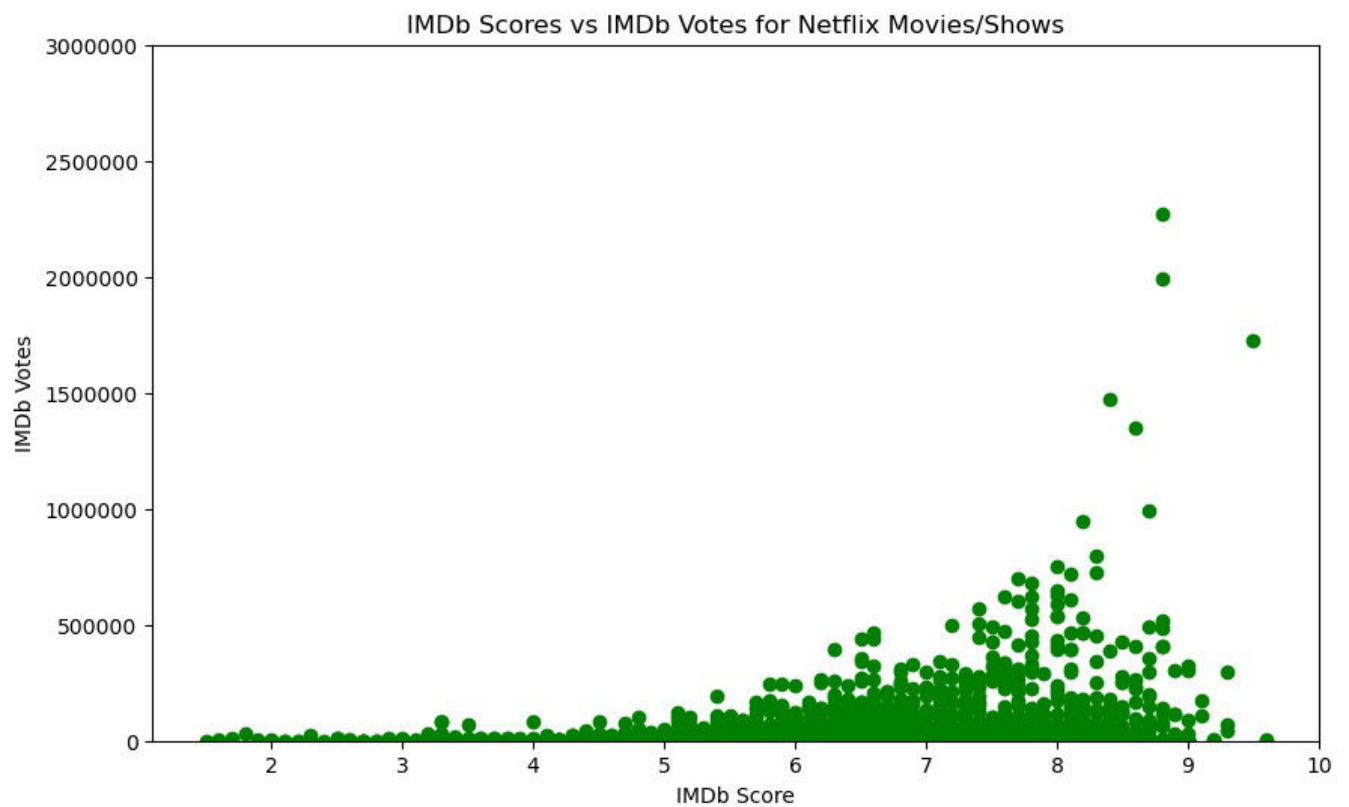In [7]: df = pd.DataFrame(data)
```

```
grouped = df.groupby('type')['imdb_score'].mean()
plt.pie(grouped.values, labels = grouped.index, autopct='%.2f')
plt.show()

top10_imdb_score = df.sort_values(by = 'imdb_score',ascending= False).head(10)
bars = plt.bar(top10_imdb_score['title'], top10_imdb_score['imdb_score'], color='blue')
plt.xlabel('Title')
plt.ylabel('IMDb Score')
plt.title('IMDb Scores of top 10 Netflix Movies/Shows')
for bar, rating in zip(bars, top10_imdb_score['imdb_score']):
    plt.text(bar.get_x() + bar.get_width() / 2 - 0.1, bar.get_height() + 0.05, f'{rating:.1f}', ha='center', co
plt.xticks(rotation=90)
plt.show()

plt.figure(figsize=(10, 6))
plt.scatter(df['imdb_score'], df['imdb_votes'], color='green')
plt.xlabel('IMDb Score')
plt.ylabel('IMDb Votes')
plt.title('IMDb Scores vs IMDb Votes for Netflix Movies/Shows')
plt.ylim(0,3000000)
plt.ticklabel_format(style='plain',axis='y')
plt.show()
```

IMDb Scores vs IMDb Votes for Netflix Movies/Shows

SQL

```
In [8]:  cnn = sqlite3.connect('Netflix.db')
         %load_ext sql
         %sql sqlite:///Netflix.db
```

```
In [9]:  %%sql
         select title, type, release_year,runtime, imdb_score, imdb_votes
         from Netflix where imdb_score>8.5 order by imdb_votes desc limit 10
```

```
 * sqlite:///Netflix.db
Done.
```

Out[9]:

title	type	release_year	runtime	imdb_score	imdb_votes
Inception	MOVIE	2010	148	8.8	2268288.0
Forrest Gump	MOVIE	1994	142	8.8	1994599.0
Breaking Bad	SHOW	2008	48	9.5	1727694.0
Saving Private Ryan	MOVIE	1998	169	8.6	1346020.0
Stranger Things	SHOW	2016	52	8.7	989090.0
Black Mirror	SHOW	2011	59	8.8	515577.0
House of Cards	SHOW	2013	52	8.7	494092.0
Peaky Blinders	SHOW	2013	58	8.8	485506.0
The Queen's Gambit	SHOW	2020	56	8.6	406350.0
Better Call Saul	SHOW	2015	49	8.8	404920.0

```
In [10]: %%sql
         select * from Netflix group by type having max(imdb_score)
```

```
 * sqlite:///Netflix.db
Done.
```

level_0	index	id	title	type	description	release_year	age_certification	runtime	imdb_id	imdb_score	imdb_votes
24	24	tm76557	No Longer Kids	MOVIE	By coincidence, Ahmad discovers that his father, Ramada Al Sokkary has a love letter in his suit case, which reveals his love to a woman other than his mother. Ahmad tries with his brothers and sister to find out the reasons that made their father look for another woman. The play shows the amusing acts the children undertake to save their family.	1979	None	235	tt8312792	9.0	943.0
622	622	ts160526	Khawatir	SHOW	A TV show devoted to help young people to be mature in their faith, work & knowledge	2005	TV-14	20	tt3062514	9.6	3046.0

In [11]:
```sql
%%sql
select title, type, release_year,runtime, imdb_score, imdb_votes
from Netflix where release_year>2000 and type ="MOVIE" order by imdb_score desc limit(20)
```

 * sqlite:///Netflix.db
Done.

title	type	release_year	runtime	imdb_score	imdb_votes
Chhota Bheem & Krishna in Mayanagari	MOVIE	2011	66	9.0	5.0
C/o Kancharapalem	MOVIE	2018	152	9.0	6562.0
David Attenborough: A Life on Our Planet	MOVIE	2020	83	9.0	31180.0
Inception	MOVIE	2010	148	8.8	2268288.0
Sky Tour: The Movie	MOVIE	2020	94	8.8	1036.0
A Lion in the House	MOVIE	2006	225	8.7	312.0
Anbe Sivam	MOVIE	2003	160	8.7	20595.0
Rubaru Roshni	MOVIE	2019	110	8.7	584.0
Bo Burnham: Inside	MOVIE	2021	87	8.7	44074.0
Chhota Bheem & Krishna vs Zimbara	MOVIE	2013	64	8.6	13.0
Chhota Bheem and the ShiNobi Secret	MOVIE	2013	63	8.6	6.0
A Second Chance	MOVIE	2015	130	8.6	47.0
Merku Thodarchi Malai	MOVIE	2018	122	8.6	1810.0
Best Wishes, Warmest Regards: A Schitt's Creek Farewell	MOVIE	2020	44	8.6	1357.0
In Our Mothers' Gardens	MOVIE	2021	84	8.6	124.0
Chhota Bheem Neeli Pahaadi	MOVIE	2013	63	8.5	9.0
Springsteen On Broadway	MOVIE	2018	153	8.5	2178.0
Ani... Dr. Kashinath Ghanekar	MOVIE	2018	159	8.5	1071.0
Emicida: AmarElo - It's All for Yesterday	MOVIE	2020	89	8.5	1472.0
K. D.	MOVIE	2019	119	8.5	1750.0

SEABORN

In [12]:
```python
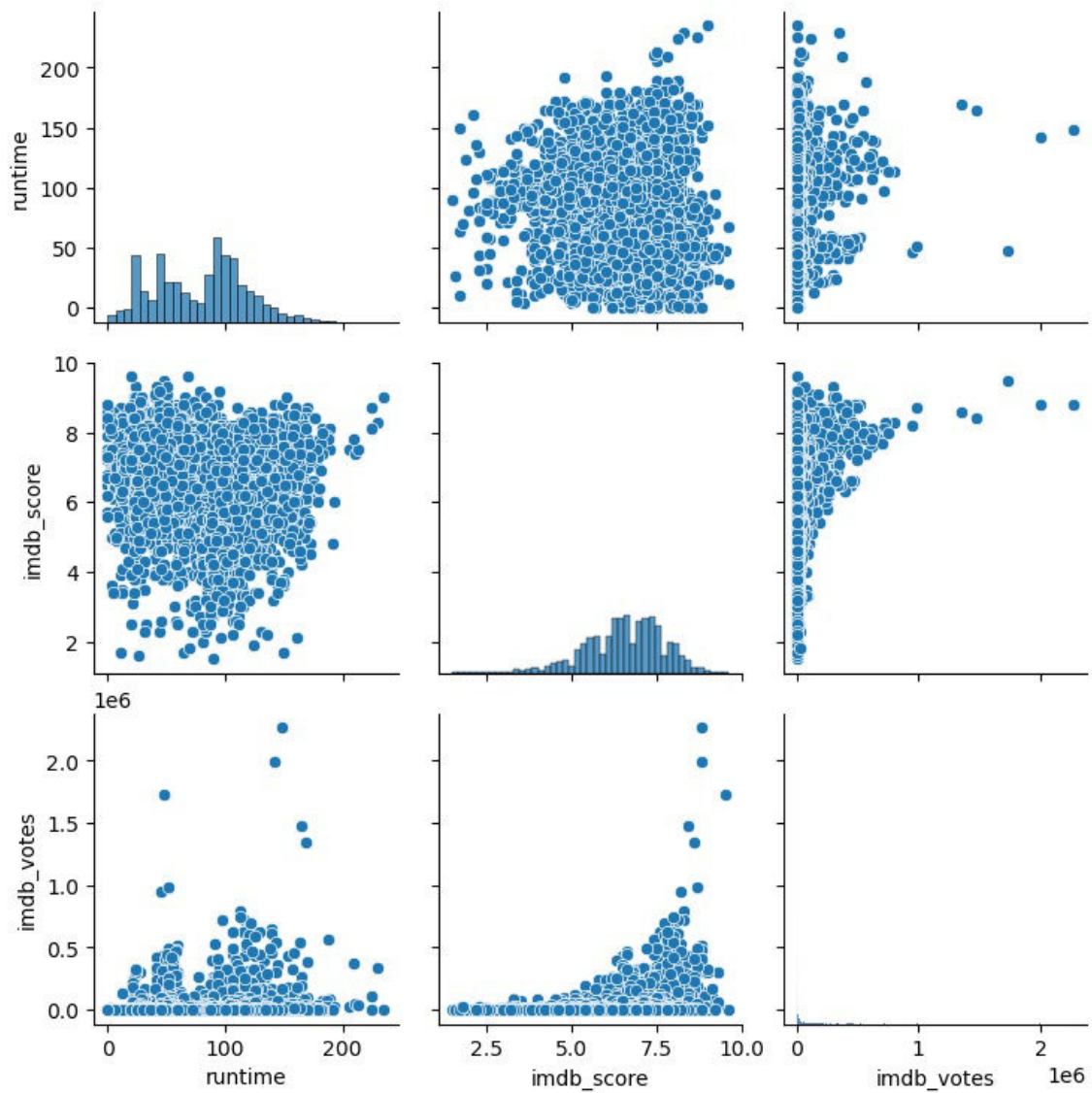numeric_columns= ['runtime','imdb_score','imdb_votes']
sns.pairplot(data[numeric_columns])
plt.suptitle('Pair Plot for Netflix Movie/Show Data',y=1.02)
plt.show()
```

C:\Users\garva\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)

Pair Plot for Netflix Movie/Show Data

CONCLUSION

We would like to thank all the professors and faculty for giving us this opportunity to work on the data relevant to our branch and also enhance our knowledge in python.

REFERENCES

[1] Kaggle
  (https ://www.kaggle.com/datasets/thedevastator/netflix-imdb-scores)

[2] NumPy Documentation
  (https://numpy.org/doc/stable/)

[3] Matplotlib Documentation
  (https://matplotlib.org/stable/contents.html)