a1944839-question5-assignment1

September 7, 2024

1 Assignment Question-5

1.1 Name: Shubharthak Sangharasha

1.2 Student ID: a1944839

1.2.1 Question-5: You should complete this question using a Jupyter Notebook. All of the code you will need to complete this question can be taken directly or generalised from the week 1 practical, or will be given to you in the question.

1.2.2 Download the file series data.csv1 . This file contains data about television series listed on IMDB. Use Python to do the following:

1. Using pandas, read the data into a dataframe and print out its head().
2. Create a histogram of the average IMDB rating (IMDB Rating) for all TV series in the dataset. Hint: Remember to add axis labels. Note that we are asking for the series' rating, not how many votes it received.
3. Calculate the mean IMDB rating for all TV series. Print out a statement showing the mean rating, rounded to the nearest 3 decimal places. Hint: In the week 1 practical, we saw how to print out a line containing text and numbers. The round() function might also be useful here.
4. Among TV series with an IMDB rating above 8, find the series with the most votes. Hint: Try creating a new dataframe that only contains series with an IMDB rating of more than 8.
5. Now find the "Crime" series with the most votes among series with an IMDB rating above 8. Hint: Many series have multiple genres. We want to include all series with "Crime" in the genre, not just series where "Crime" is the only genre. The string method .str.contains() might be useful here.

1.2.3 1. Using pandas, read the data into a dataframe and print out its head()

[1]:
```
# Making sure to install pandas
!pip install pandas
```

```
Requirement already satisfied: pandas in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
```

```
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from pandas)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.5 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from python-
dateutil>=2.8.2->pandas) (1.16.0)
```

[2]:
```
#adding path of environment libraries
import sys
sys.path.append('/home/shubharthak/miniconda3/lib/python3.12/site-packages')
```

[4]:
```
#import libraries
import pandas as pd

df = pd.read_csv('series_data.csv')
df.head(10) #printing first 10 data of the series_dataset
```

[4]:

	Series_Title	Runtime_of_Series	Certificate	Runtime_of_Episodes	\
0	Game of Thrones	(2011-2019)	A	57 min	
1	Breaking Bad	(2008-2013)	18	49 min	
2	The Walking Dead	(2010-)	18+	44 min	
3	Friends	(1994-2004)	13+	22 min	
4	Stranger Things	(2016-)	15	51 min	
5	Sherlock	(2010-2017)	UA	88 min	
6	The Big Bang Theory	(2007-2019)	U	22 min	
7	Dexter	(2006-2021)	A	53 min	
8	How I Met Your Mother	(2005-2014)	15+	22 min	
9	True Detective	(2014-)	A	55 min	

	Genre	IMDB_Rating	\
0	Action, Adventure, Drama	9.3	
1	Crime, Drama, Thriller	9.5	
2	Drama, Horror, Thriller	8.2	
3	Comedy, Romance	8.9	
4	Drama, Fantasy, Horror	8.7	
5	Crime, Drama, Mystery	9.1	
6	Comedy, Romance	8.1	
7	Crime, Drama, Mystery	8.6	
8	Comedy, Romance	8.3	
9	Crime, Drama, Mystery	9.0	

	Overview	Star1	\
0	Nine noble families fight for control over the…	Emilia Clarke	
1	A high school chemistry teacher diagnosed with…	Bryan Cranston	

2	Sheriff Deputy Rick Grimes wakes up from a com…	Andrew Lincoln
3	Follows the personal and professional lives of…	Jennifer Aniston
4	When a young boy disappears, his mother, a pol…	Millie Bobby Brown
5	A modern update finds the famous sleuth and hi…	Benedict Cumberbatch
6	A woman who moves into an apartment across the…	Johnny Galecki
7	By day, mild-mannered Dexter is a blood-spatte…	Michael C. Hall
8	A father recounts to his children - through a …	Josh Radnor
9	Seasonal anthology series in which police inve…	Vince Vaughn

	Star2	Star3	Star4	No_of_Votes
0	Peter Dinklage	Kit Harington	Lena Headey	1773458
1	Aaron Paul	Anna Gunn	Betsy Brandt	1468887
2	Norman Reedus	Melissa McBride	Danai Gurira	854698
3	Courteney Cox	Lisa Kudrow	Matt LeBlanc	829816
4	Finn Wolfhard	Winona Ryder	David Harbour	824966
5	Martin Freeman	Una Stubbs	Rupert Graves	808717
6	Jim Parsons	Kaley Cuoco	Simon Helberg	724187
7	Jennifer Carpenter	David Zayas	James Remar	647136
8	Jason Segel	Cobie Smulders	Neil Patrick Harris	603824
9	Colin Farrell	Rachel McAdams	Taylor Kitsch	500194

1.2.4 2. Create a histogram of the average IMDB rating (IMDB Rating) for all TV series in the dataset. Hint: Remember to add axis labels. Note that we are asking for the series' rating, not how many votes it received.

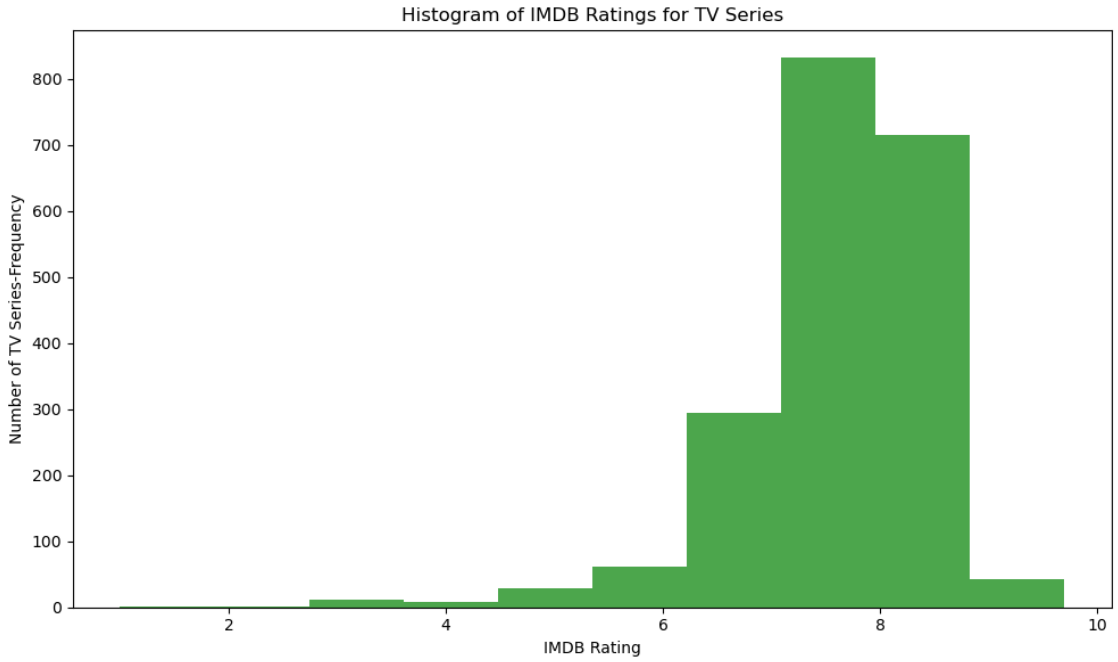
[5]:
```
#to create a histogram we will use matplotlib library
#first checking matplotlib is installed if not install it
!pip install matplotlib
```

```
Requirement already satisfied: matplotlib in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (3.9.1)
Requirement already satisfied: contourpy>=1.0.1 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(1.2.1)
Requirement already satisfied: cycler>=0.10 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(4.53.1)
Requirement already satisfied: kiwisolver>=1.3.1 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(1.4.5)
Requirement already satisfied: numpy>=1.23 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(1.26.4)
Requirement already satisfied: packaging>=20.0 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
```

```
(23.2)
Requirement already satisfied: pillow>=8 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from matplotlib)
(2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/home/shubharthak/miniconda3/lib/python3.12/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
```

[6]:
```
#import the library
import matplotlib.pyplot as plt
```

[7]:
```
plt.figure(figsize=(10, 6))
plt.hist(df['IMDB_Rating'], bins=10, color='green', alpha=0.7)
plt.xlabel('IMDB Rating')
plt.ylabel('Number of TV Series-Frequency')
plt.title('Histogram of IMDB Ratings for TV Series')
plt.tight_layout()
plt.show();
```



1.2.5 3. Calculate the Mean IMDB Rating for All TV Series

[16]:
```
imdb_mean_rating = df['IMDB_Rating'].mean() #calculating the mean using mean␣
↪function

print(f"The mean IMDB rating for all TV series is: {round(imdb_mean_rating,␣
↪3)}") #printing the mean of IMDB rating till 3 decimals-up
```

```
The mean IMDB rating for all TV series is: 7.591
```

1.2.6 4. Finding TV Series with most Votes (IMDB rating above 8)

[11]:
```
high_rating_series = df[df['IMDB_Rating'] > 8] #getting the tv series which␣
↪have imdb rating higher than 8
high_rating_series
```

[11]:

	Series_Title	Runtime_of_Series	Certificate	Runtime_of_Episodes	\
0	Game of Thrones	(2011-2019)	A	57 min	
1	Breaking Bad	(2008-2013)	18	49 min	
2	The Walking Dead	(2010-)	18+	44 min	
3	Friends	(1994-2004)	13+	22 min	
4	Stranger Things	(2016-)	15	51 min	
…	…	…	…	…	
1982	Velvet	(2013-2016)	NaN	43 min	
1985	Hajime no ippo	(2000-2002)	NaN	20 min	
1995	Shaman Kingu	(2001-2005)	NaN	23 min	
1996	Eerie, Indiana	(1991-1992)	NaN	30 min	
1998	The Cheat	(2017-)	NaN	20 min	

	Genre	IMDB_Rating	\
0	Action, Adventure, Drama	9.3	
1	Crime, Drama, Thriller	9.5	
2	Drama, Horror, Thriller	8.2	
3	Comedy, Romance	8.9	
4	Drama, Fantasy, Horror	8.7	
…	…	…	
1982	Adventure, Drama, Romance	8.1	
1985	Animation, Action, Comedy	8.8	
1995	Animation, Action, Adventure	8.1	
1996	Adventure, Comedy, Drama	8.2	
1998	Action, Drama, Sci-Fi	8.8	

	Overview	\
0	Nine noble families fight for control over the…	
1	A high school chemistry teacher diagnosed with…	
2	Sheriff Deputy Rick Grimes wakes up from a com…	
3	Follows the personal and professional lives of…	
4	When a young boy disappears, his mother, a pol…	

```
…                                                   …
1982  A feel-good, compelling Spanish story of a fas…
1985  Makunouchi Ippo is an ordinary high school stu…
1995  SHAMAN KING follows the adventures of a 13-yea…
1996  Teenage weirdness investigator Marshall Teller…
1998  Felix is a software developer who works for th…

                      Star1                Star2              Star3  \
0               Emilia Clarke        Peter Dinklage     Kit Harington
1               Bryan Cranston            Aaron Paul         Anna Gunn
2               Andrew Lincoln          Norman Reedus   Melissa McBride
3            Jennifer Aniston          Courteney Cox       Lisa Kudrow
4          Millie Bobby Brown          Finn Wolfhard      Winona Ryder
…                         …                    …                 …
1982  Miguel Ángel Silvestre   Aitana Sánchez-Gijón   Manuela Velasco
1985           Richard Epcar           D.C. Douglas   Richard Cansino
1995         Andrew Rannells  Michael Sinterniklaas  Megumi Hayashibara
1996        Justin Shenkarow    Mary-Margaret Humes      Francis Guinan
1998             Taner Sahin          Andac Ulukaya   Neslihan Ulusoy

               Star4  No_of_Votes
0       Lena Headey      1773458
1      Betsy Brandt      1468887
2      Danai Gurira       854698
3      Matt LeBlanc       829816
4     David Harbour       824966
…               …           …
1982            NaN         5159
1985            NaN         5155
1995            NaN         5131
1996            NaN         5128
1998            NaN         5111

[666 rows x 12 columns]
```

1.2.7 The above new high_rating_series dataframe have IMDB rating more than 8 Now, we will find which have highest votes among them using idxmax method.

[15]:
```
most_voted_series = high_rating_series.loc[high_rating_series['No_of_Votes'].
↪idxmax()]
most_voted_series
```

[15]:
```
Series_Title                                    Game of Thrones
Runtime_of_Series                                   (2011-2019)
Certificate                                                   A
Runtime_of_Episodes                                      57 min
Genre                                  Action, Adventure, Drama
```

```
IMDB_Rating                                                        9.3
Overview                Nine noble families fight for control over the…
Star1                                                     Emilia Clarke
Star2                                                    Peter Dinklage
Star3                                                     Kit Harington
Star4                                                       Lena Headey
No_of_Votes                                                     1773458
Name: 0, dtype: object
```

[12]:
```
# Find the series with the most votes
most_voted_series = high_rating_series.loc[high_rating_series['No_of_Votes'].
 ↪idxmax()] #idxmax (get us the location of maximum votes)

# Print the result
print(f"The series with the most votes among those with an IMDB rating above 8␣
 ↪is: {most_voted_series['Series_Title']}")
```

```
The series with the most votes among those with an IMDB rating above 8 is: Game
of Thrones
```

1.2.8 5. Now find the "Crime" series with the most votes among series with an IMDB rating above 8

[18]:
```
crime_series = df[(df['IMDB_Rating'] > 8) & (df['Genre'].str.contains('Crime',␣
 ↪case=False))] #getting the crime series which have IMDB rating more than 8
most_voted_crime_series = crime_series.loc[crime_series['No_of_Votes'].
 ↪idxmax()] #finding the series with most votes among crime series
print(f"The 'Crime' series with the most votes among those with an IMDB rating␣
 ↪above 8 is: {most_voted_crime_series['Series_Title']}") #printing the series
```

```
The 'Crime' series with the most votes among those with an IMDB rating above 8
is: Breaking Bad
```