NARAYANA BUSINESS SCHOOL
AICTE APPROVED

An Assignment On

"Data Visualization using Python"

Submitted in partial fulfilment of the requirement for the award of

POST GRADUATE DIPLOMA IN MANAGEMENT

From

NARAYANA BUSINESS SCHOOL, AHMEDABAD

Subject: **PGDM DATA SCIENCE AND ANALYTICS**

DSA2023 - Data Visualization Techniques using Python

Component : CEC

Submitted By

NAME : DEEP PRATEEK
BATCH : PGDMDSA 2023-25

ROLL NO : 004
SECTION : Data Science

DATE OF SUBMISSION: 24-1-2024

Under The Guidance Of

NAME : **Miss. Archana Singh and Miss. Sejal Bhavsar**
DEPARTMENT : **Data Visualization Techniques using Python**

First Import pandas: -

Input: -

```
import pandas as pd

Best = r"D:\pandas\bestsellers.csv"

A = pd. read_csv(Best)

Print(A)
```

Output: -

```
                                                  Name  \
0                         10-Day Green Smoothie Cleanse
1                                      11/22/63: A Novel
2                  12 Rules for Life: An Antidote to Chaos
3                                  1984 (Signet Classics)
4      5,000 Awesome Facts (About Everything!) (Natio...
..                                                   ...
545          Wrecking Ball (Diary of a Wimpy Kid Book 14)
546  You Are a Badass: How to Stop Doubting Your Gr...
547  You Are a Badass: How to Stop Doubting Your Gr...
548  You Are a Badass: How to Stop Doubting Your Gr...
549  You Are a Badass: How to Stop Doubting Your Gr...

                       Author  User Rating  Reviews  Price  Year         Genre
0                   JJ Smith          4.7    17350      8  2016  Non Fiction
1                Stephen King          4.6     2052     22  2011      Fiction
2           Jordan B. Peterson          4.7    18979     15  2018  Non Fiction
3                George Orwell          4.7    21424      6  2017      Fiction
4      National Geographic Kids          4.8     7665     12  2019  Non Fiction
..                          ...          ...      ...    ...   ...          ...
545                Jeff Kinney          4.9     9413      8  2019      Fiction
546                Jen Sincero          4.7    14331      8  2016  Non Fiction
547                Jen Sincero          4.7    14331      8  2017  Non Fiction
548                Jen Sincero          4.7    14331      8  2018  Non Fiction
549                Jen Sincero          4.7    14331      8  2019  Non Fiction

[550 rows x 7 columns]
```

Assignment -2

Q-1: Import the bestsellers.csv dataset and use it to answer the following questions:

Q (i)- Find the lowest User Rating in the DF

Input:-

```
Rating=A['User Rating'].min()
print('Lowest User Rating -',Rating)
```

Output:-

```
Lowest User Rating - 3.3
```

Q (ii)- Find the highest Price in the DF

Input:-

```
Highest_price = A['Price'].max()
print('Highest_price -',Highest_price)
```

Output:-

```
Highest_price - 105
```

Q(iii)- What is the average User Rating?

Input:-

```
Average = A['User Rating'].mean()
print('The average of user rating is :',Average)
```

Output:-

```
The average of user rating is : 4.618363636363637
```

Q-(iv)- What the average User Rating of the first 5 rows?

Input:-

```
First_5 = A.head(5)
da = A['User Rating'].mean()
print('The first 5 average of user rating is :',da)
```

Output:-

```
The first 5 average of user rating is : 4.618363636363637
```

Q-(v)- What User Rating score appeared the most?

Input:-

```
Most = A['User Rating'].mode()

print('The most appeared user rating is :',Most)
```

Output:-

```
The most appeared user rating is : 0    4.8
Name: User Rating, dtype: float64
```

Q-(vi)- What is the total (sum) of all the values in the Reviews column?

Input: -

```
Total= A['Reviews'].sum()

print('The sum of all the values in Reviews column is :',Total)
```

Output: -

```
The sum of all the values in Reviews column is : 6574305
```

Q- (vii)- How many different authors are featured in the dataset?

Input:-

```
Authors= A['Author'].nunique()

print('There are',Authors,'Unique Autors in the Data set.')
```

Output:-

```
There are 248 Unique Autors in the Data set.
```

Q- (viii)- Which author wrote the most number of books on the list? How many did they write?

Input:-

```
Book = A['Author'].value_counts().head(1)

print(Book,'Wrote the max no of books')
```

Output:-

```
Jeff Kinney    12
Name: Author, dtype: int64 Wrote the max no of books
```

Q-2: Work with the bestsellers.csv dataset to answer the following questions

Q- (i)- Retrieve a series that contains the book Names

Input:-

Book_name= A['Name']

print(Book_name)

Output:-

```
0                          10-Day Green Smoothie Cleanse
1                                      11/22/63: A Novel
2                 12 Rules for Life: An Antidote to Chaos
3                                   1984 (Signet Classics)
4      5,000 Awesome Facts (About Everything!) (Natio...
                              ...
545          Wrecking Ball (Diary of a Wimpy Kid Book 14)
546    You Are a Badass: How to Stop Doubting Your Gr...
547    You Are a Badass: How to Stop Doubting Your Gr...
548    You Are a Badass: How to Stop Doubting Your Gr...
549    You Are a Badass: How to Stop Doubting Your Gr...
Name: Name, Length: 550, dtype: object
```

Q-(ii)- Retrieve a series that contains the User Ratings

Input:-

User = A['User Rating']

print(User)

Output:-

```
0      4.7
1      4.6
2      4.7
3      4.7
4      4.8
      ...
545    4.9
546    4.7
547    4.7
548    4.7
549    4.7
Name: User Rating, Length: 550, dtype: float64
```

Q- (iii)- Retrieve the first 8 Authors

Input:-

First = A['Author'].head(8)

print(First)

Output:-

```
0                    JJ Smith
1                Stephen King
2          Jordan B. Peterson
3                George Orwell
4    National Geographic Kids
5          George R. R. Martin
6          George R. R. Martin
7                  Amor Towles
Name: Author, dtype: object
```

Q-(iv)- Find the unique Genres

Input:-

```
Unique = A['Genre'].unique()
print(Unique,': Are the unique Genre in the data set')
```

Output:-

```
['Non Fiction' 'Fiction'] : Are the unique Genre in the data set
```

Q- (v)- Find the number of unique Authors

Input:-

```
Authors = A['Author'].unique()
print(Authors,': Are the unique Authors in the Data set')
```

Output:-

```
['JJ Smith' 'Stephen King' 'Jordan B. Peterson' 'George Orwell'
 'National Geographic Kids' 'George R. R. Martin' 'Amor Towles'
 'James Comey' 'Fredrik Backman' 'Larry Schweikart' 'Jaycee Dugard'
 "Madeleine L'Engle" 'Steve Harvey' 'Adult Coloring Book Designs'
 'Blue Star Coloring' 'Coloring Books for Adults' 'Ron Chernow'
 'Anthony Doerr' 'Veronica Roth' 'Chris Kyle' 'Khaled Hosseini'
 'Glenn Beck' 'Neil deGrasse Tyson' 'Mark Twain' 'DK' 'Angie Grace'
 'Ina Garten' 'Michelle Obama' 'Atul Gawande' 'Ta-Nehisi Coates'
 'Bruce Springsteen' 'Stephenie Meyer' 'Bill Martin Jr.' 'Jeff Kinney'
 "Sasha O'Hara" 'David Goggins' 'Thomas Piketty' 'Suzanne Collins'
 'Chrissy Teigen' 'Francis Chan' 'Kevin Kwan' 'Marjorie Sarnat'
 'Abraham Verghese' 'Brené Brown' 'Malcolm Gladwell' 'Charlaine Harris'
 'Rod Campbell' 'George W. Bush' 'Tony Hsieh'
 'American Psychiatric Association' 'M Prefontaine' 'Zhi Gang Sha'
 'Dav Pilkey' 'Marty Noble' 'Daniel H. Pink' 'David Zinczenko'
 'Joel Fuhrman MD' 'Tara Westover' 'Johanna Basford' 'Ray Bradbury'
 'J.K. Rowling' 'Bob Woodward' 'E L James' 'Michael Wolff' 'Roger Priddy'
 'Michael Pollan' 'RH Disney' 'John Heilemann' 'George R.R. Martin'
 'Peter A. Lillback' 'Brian Kilmeade' 'Giles Andreae' 'Rachel Hollis'
 'Harper Lee' 'Adam Mansbach' 'Sarah Palin' 'Gillian Flynn'
 'Pretty Simple Press' 'Jim Collins' 'Margaret Wise Brown'
 'Sherri Duskey Rinker' 'David Perlmutter MD' 'Raina Telgemeier'
 'Lin-Manuel Miranda' 'Phil Robertson' 'J. K. Rowling' 'Scholastic'
 'Mitch Albom' 'Todd Burpo' 'J. D. Vance' 'Joanna Gaines' 'Dale Carnegie'
 'Howard Stern' 'Brandon Stanton' 'Allie Brosh' 'Hopscotch Girls'
 'James Patterson' 'Ann Whitford Paul' 'Gayle Forman' 'Eric Larson'
 'Dan Brown' 'Christopher Paolini' 'Jennifer Smith' 'Lysa TerKeurst'
 'Sarah Young' 'David Grann' "Bill O'Reilly" 'Anthony Bourdain'
 'Rob Elliott' 'Jill Twiss' 'Sheryl Sandberg' 'Walter Isaacson'
 'Paper Peony Press' 'Mark R. Levin' 'Keith Richards' 'Chris Cleave'
 'Alice Schertle' 'Celeste Ng' 'John Green' 'Rob Bell' 'Robert Munsch'
 'Admiral William H. McRaven' 'Julia Child' 'Rupi Kaur' 'Adam Gasiewski'
 'Carol S. Dweck' 'Crispin Boyer' 'Amy Shields' 'Elie Wiesel' 'Mark Owen'
 'Pete Souza' 'Dr. Seuss' 'Elizabeth Strout' 'Ann Voskamp'
```

Q- (vi)- Find the average Price

Input:-

```
Average = A['Price'].mean()
print(Average,'is a average price')
```

Output:-

```
13.1 is a average price
```

Q- (vii)- Find the 10 highest prices

Input:-

```
Highest = A.nlargest(10,'Price')['Price']
print('The top 10 Highest prices are :',Highest)
```

Output:-

```
The top 10 Highest prices are : 69     105
70     105
473     82
151     54
346     53
159     52
271     46
272     46
273     46
274     46
Name: Price, dtype: int64
```

Q-(viii)- Find the top 3 most common book titles in the dataset

Input:-

```
Top_3 = A['Name'].value_counts().head(3)
print(Top_3)
```

Output:-

```
Publication Manual of the American Psychological Association, 6th Edition    10
StrengthsFinder 2.0                                                           9
Oh, the Places You'll Go!                                                     8
Name: Name, dtype: int64
```

Q-(ix)- Create a new dataframe with only Author and User Rating.

Input:-

```
New = A[['Author','User Rating']]
df = pd.DataFrame(New)
print(df)
```

Output:-

	Author	User Rating
0	JJ Smith	4.7
1	Stephen King	4.6
2	Jordan B. Peterson	4.7
3	George Orwell	4.7
4	National Geographic Kids	4.8
..	...	...
545	Jeff Kinney	4.9
546	Jen Sincero	4.7
547	Jen Sincero	4.7
548	Jen Sincero	4.7
549	Jen Sincero	4.7

[550 rows x 2 columns]

Q- (x)- Using the new dataframe, find the most common combination of Author and User Review Score.

Input:-

```
Str = New.groupby(['Author', 'User Rating']).size().idxmax()
```

print(Str)

Output:-

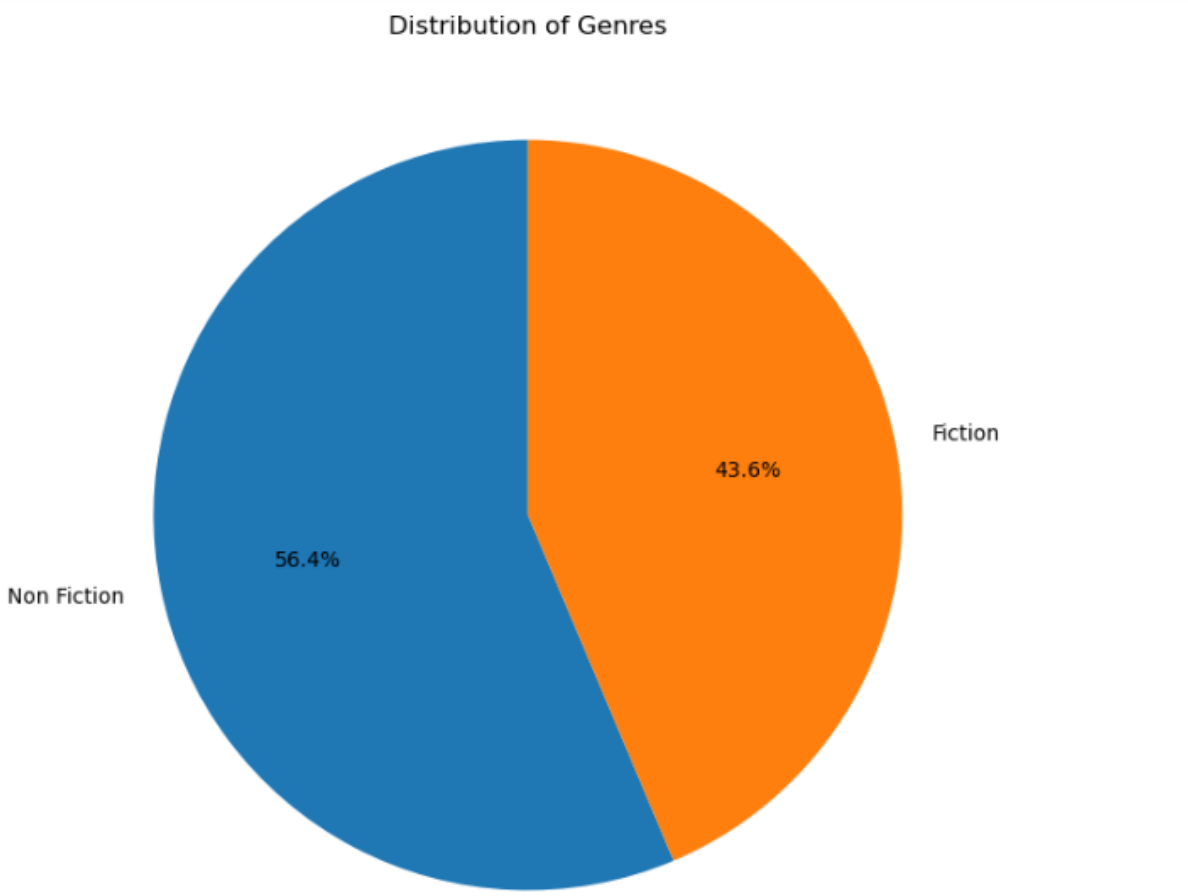
```
('American Psychological Association', 4.5)
```

Q-(xi)- Create a pie chart showing the total number of times each Genre appears in the dataset.

Input:-

```
import matplotlib.pyplot as plt
genre_counts = A['Genre'].value_counts()
plt.figure(figsize=(10, 8))
genre_counts.plot.pie(autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Genres')
plt.ylabel('')
plt.show()
```

Output:-

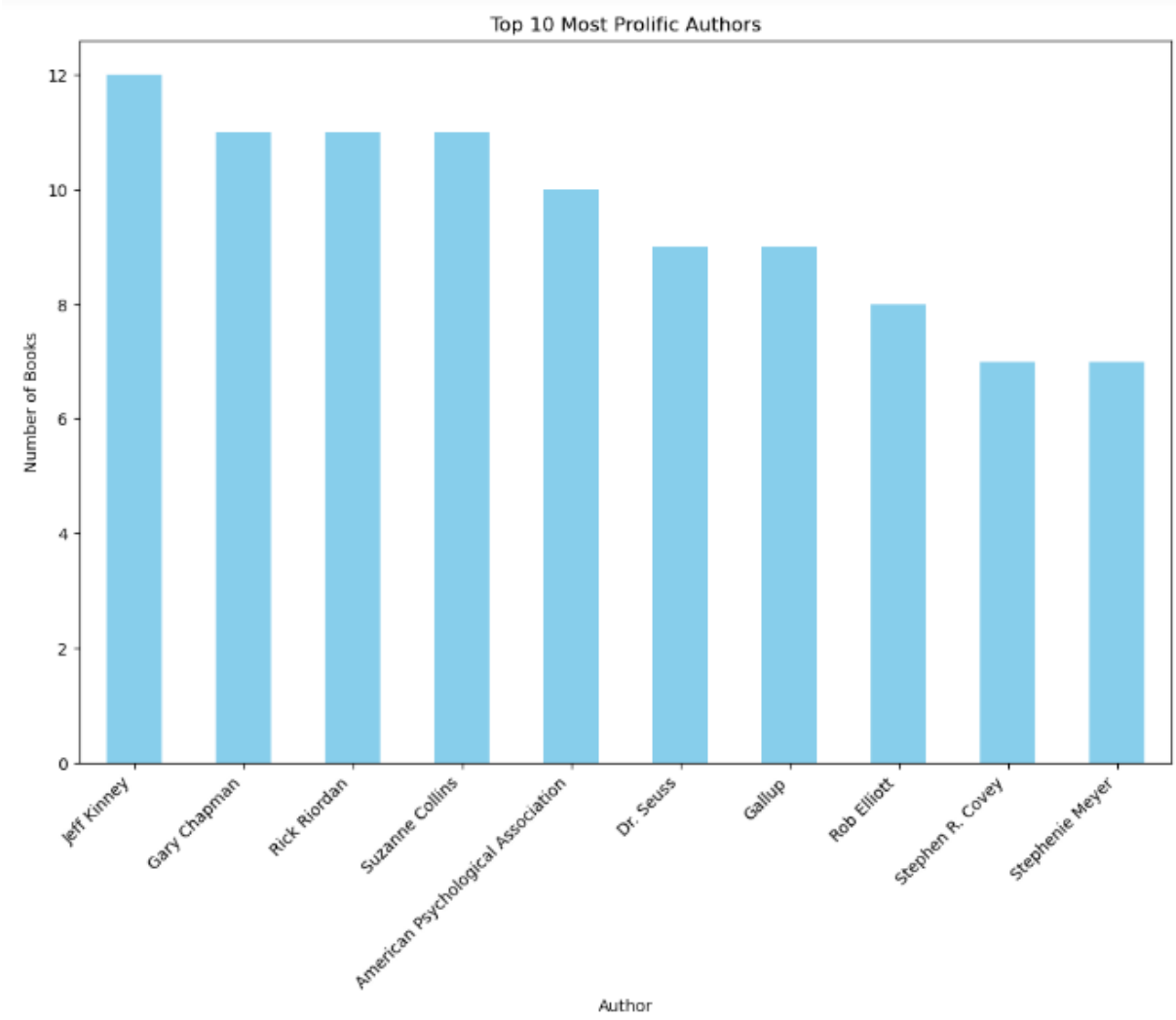


Q- (xii)- Find out the top 10 most prolific authors and plot their number of books as a bar plot

Input:-

```
author_counts = df['Author'].value_counts().head(10)
plt.figure(figsize=(12, 8))
author_counts.plot(kind='bar', color='skyblue')
plt.title('Top 10 Most Prolific Authors')
plt.xlabel('Author')
plt.ylabel('Number of Books')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Output:-



Q- (xiii)- BONUS: create a histogram showing the distribution of user review scores

Input:-

```
user_review_scores = A['Reviews']
plt.hist(user_review_scores, bins=(20), edgecolor='black',color='skyblue')
plt.xlabel('User Review Scores')
```

```
plt.ylabel('Frequency')
plt.title('Distribution of User Review Scores')
plt.show()
```

Output:-