Day 14

DIY

Q1. Problem Statement: Measures of Central Tendency- I

Write a Python program to declare a 1-D array. With the help of the statistics library, calculate the Mean, Median, and Mode.

Note: You can write any data of your choice inside the array

Input Format:

```
The Array is:
[1 2 3 4 5 6 2 3 4 5 5 5 5]
```

Sample Output:

```
Mean of the Array is:
3
Median of the Array is:
4
Mode of the Array is:
5
```

Q2. Problem Statement: Measures of Central Tendency- II

Write a Python program to declare a DataFrame as shown in the dataset section below and calculate the Mean, Median, and Mode of the entire DataFrame.

Note: You can write any data of your choice inside the DataFrame

Dataset:

	Apple	Orange	Banana	Pear
Bag1	10	20	30	40
Bag2	7	14	21	28
Bag3	55	15	8	12
Bag4	15	14	1	8
Bag5	7	1	1	8
Bag6	5	4	9	2

Sample Output:

```
Calculating the Mean
Apple      16.500000
Orange     11.333333
Banana     11.666667
Pear       16.333333
dtype: float64
```

```
Calculating the Median
Apple       8.5
Orange     14.0
Banana      8.5
Pear       10.0
dtype: float64
```

```
Calculating the Mode
   Apple  Orange  Banana  Pear
0      7      14       1     8
```

Q3. Problem Statement: Measures of Central Tendency- III

Load the *"Books.csv"* dataset into a DataFrame and perform the following tasks:

1. Perform initial analysis on the dataset using the `info()` and the `describe()` functions
2. Group the data based on the *"Author"* column and calculate the average user ratings and average price, and write them in new columns
3. Find out which of the authors have written the maximum number of books

Dataset:

	Name	Author	User Rating	Reviews	Price	Year	Genre
0	Act Like a Lady, Think Like a Man: What Men Re...	Steve Harvey	4.6	5013	17	2009	Non Fiction
1	Arguing with Idiots: How to Stop Small Minds a...	Glenn Beck	4.6	798	5	2009	Non Fiction
2	Breaking Dawn (The Twilight Saga, Book 4)	Stephenie Meyer	4.6	9769	13	2009	Fiction
3	Crazy Love: Overwhelmed by a Relentless God	Francis Chan	4.7	1542	14	2009	Non Fiction
4	Dead And Gone: A Sookie Stackhouse Novel (Sook...	Charlaine Harris	4.6	1541	4	2009	Fiction

Sample Output:

1. Perform initial analysis of the dataset using the info() and the describe() functions

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Name         700 non-null    object 
 1   Author       700 non-null    object 
 2   User_Rating  700 non-null    float64
 3   Reviews      700 non-null    int64  
 4   Price        700 non-null    int64  
 5   Year         700 non-null    int64  
 6   Genre        700 non-null    object 
dtypes: float64(1), int64(3), object(3)
memory usage: 38.4+ KB
```

2. Group the data based on the "Author" column and calculate the average user ratings and average price, and write them in new columns

	Author	Average_User_Rating
0	Abraham Verghese	4.6
1	Adam Gasiewski	4.4
2	Adam Mansbach	4.8
3	Adam Silvera	4.7
4	Adam Wallace	4.8
...	...	...
300	William P. Young	4.6
301	Wizards RPG Team	4.8
302	Wonder House Books	4.6
303	Workman Publishing	4.8
304	Zhi Gang Sha	4.6

305 rows × 2 columns

	Author	Average_Price
0	Abraham Verghese	11.000000
1	Adam Gasiewski	6.000000
2	Adam Mansbach	9.000000
3	Adam Silvera	8.000000
4	Adam Wallace	6.250000
...	...	...
300	William P. Young	8.000000
301	Wizards RPG Team	27.000000
302	Wonder House Books	16.000000
303	Workman Publishing	7.666667
304	Zhi Gang Sha	11.500000

305 rows × 2 columns

3. Find out which of the authors have written the maximum number of books

```
Maximum number of books are written by:
0    Gary Chapman
1     Jeff Kinney
dtype: object
```