Name- Arya Nair
Batch- A2
Roll Number- 16010421063

```
import pandas as pd

df= pd.read_csv("data.csv")
follower_list=df["followers"].tolist()

df.head()
```

[43] ✓ 0.1s Python

	Country	Channel Name	Category	Main Video Category	username	followers	Main topic	More topics	Likes	Boost Index	...	Views	Views Avg.	Av
0	IN	T-Series	Gaming & Apps	Music	T-Series	220000000	Music of Asia	Entertainment,Music of Asia,Music,Movies	1.602680e+09	83	...	195660744416	2.095329e+06	1.522
1	US	ABCkidTV - Nursery Rhymes	Gaming & Apps	Education	ABCkidTV - Nursery Rhymes	138000000	Movies	Entertainment,Music,Movies	2.209901e+08	63	...	133025325473	7.027126e+07	1.837
2	IN	SET India	Gaming & Apps	Shows	SET India	137000000	Movies	Entertainment,TV shows,Music,Movies	1.748752e+08	79	...	121741739317	1.095729e+05	
3	US	PewDiePie	Gaming & Apps	Gaming	PewDiePie	111000000	Lifestyle	Gaming,Action game,Lifestyle,Action-adventure ...	2.191406e+09	88	...	28424113942	7.718345e+06	
4	US	MrBeast	Gaming & Apps	Entertainment	MrBeast	98100000	Lifestyle	Entertainment,Lifestyle,Technology	1.731833e+09	60	...	16242634269	9.876250e+07	

5 rows × 22 columns

```
sum=0
for i in follower_list:
    sum=sum+i
mean=sum/len(follower_list)
print(f"Mean- {mean}")
```

[44] ✓ 0.6s Python

```
Mean- 49529754.95915986
```

```
follower_list.sort()
if(len(follower_list)%2==0):
    median=(follower_list[len(follower_list)//2]+follower_list[len(follower_list)//2-1])/2
else:
    median=follower_list[len(follower_list)//2]
print(f"Median- {median}")
```

[45] ✓ 0.4s Python

```
Median- 41300000
```

```
frequency={}
for i in follower_list:
    if i in frequency:
        frequency[i]=frequency[i]+1
    else:
        frequency[i]=1

m=0
for i in frequency:
    if(frequency[i]>m):
        m=frequency[i]
        element=i

print(f"Mode- {element}")
```

[46] ✓ 0.6s Python

```
Mode- 37400000
```

```
# find variance of follower_list
variance=0
for i in follower_list:
    variance=variance+(i-mean)**2
variance=variance/len(follower_list)
print(f"Variance- {variance}")
```

[47] ✓ 0.4s Python

```
Variance- 818092591888613.8
```

```
# find standard deviation of follower_list
standard_deviation=variance**0.5
print(f"Standard Deviation- {standard_deviation}")
```

[48] ✓ 0.4s Python

```
Standard Deviation- 28602317.946079366
```

```python
# find median deviation of follower_list
median_deviation=0
for i in follower_list:
    median_deviation=median_deviation+abs(i-median)
median_deviation=median_deviation/len(follower_list)
print(f"Median Deviation- {median_deviation}")
```

[49] ✓ 0.4s Python

```
Median Deviation- 16742707.117852975
```

```python
standard_error=standard_deviation/(len(follower_list)**0.5)
print(f"Standard Error- {standard_error}")
```

0] ✓ 0.4s Python

```
Standard Error- 977036.5690595189
```

Postlab Question-

1. **What are the various applications of central tendency and variability of data?**
   Ans- Application of variability of data: -
   - It helps to compare different group
   - Helps to as-certain the measures of deviation

   Applications of central tendency: -
   - In insurance companies
   - In the construction business

2. **What are the outlier's data? What are the different ways to find out it? Give suitable example with its effect on central tendency and variability of data?**
   Ans- Outliers are data points that are far from other data points. In other words, they're unusual values in a dataset. Outliers are problematic for many statistical analyses because they can cause tests to either miss significant findings or distort real results

   Some of the most popular methods for outlier detection are:
   - Z-Score or Extreme Value Analysis (parametric)
   - Probabilistic and Statistical Modeling (parametric)
   - Linear Regression Models (PCA, LMS)
   - Proximity-Based Models (non-parametric)
   - Information Theory Models.

Measures of central tendency are mean, median and mode. Outliers affect the mean value of the data but have little effect on the median or mode of a given set of data. Example: The data shows Sara's scores for the last 5 math tests: 88, 90, 55, 94, and 89. Outlier-55 The outlier decreased the mean by 7.05 (with outlier-90.25 and without outlier-83.2). The outlier decreased the median by 0.5.(89.5 to 89). The mode did not change/ There is no mode.