Semester	T.E. Semester VI – Computer Engineering
Subject	Data Warehousing and Mining
Subject Professor In-charge	Prof. Kavita Shirsat
Assisting Teachers	Prof. Kavita Shirsat
Laboratory	Lab 312 A

Student Name	Deep Salunkhe	
Roll Number	21102A0014	
Grade and Subject Teacher's Signature		

Experiment Number	01	
Experiment Title	To find Mean, Median, Mode and Range of a particular numeric column on a live Dataset.	
Resources / Apparatus Required	Hardware: Computer system	Software: Python
Description	• The terms mean, median and mode are used to describe the central tendency of a large data set. Range provides context for the mean, median and mode. • When working with a large data set, it can be useful to represent the entire data set with a single value that describes the "middle" or "average" value of the entire set. In statistics, that single value is called the central tendency and mean, median and mode are all ways to describe it. • To find the mean, add up the values in the data set and then divide by the number of values that you added. • To find the median, list the values of the data set in numerical order and identify which value appears in the middle of the list. • To find the mode, identify which value in the data set occurs most often. • Range, which is the difference between the largest and smallest value in the data set, describes how well the central tendency represents the data. If the range is large, the central tendency is not as representative of the data as it would be if the range was small.	
Implementation	: `import pandas as pd` : `data=pd.read_csv('netflix1.csv')`	

```
data['date_added'] = pd.to_datetime(data['date_added'], errors='coerce')
```

```
mean_date = data['date_added'].mean()
```

```
mean_date
```

```
Timestamp('2019-05-17 21:44:01.638225408')
```

```
for idx, x in enumerate(data['date_added']):
    if x=='########':  # Check for NaT values
        data.at[idx, 'date_added'] = mean_date
        print(idx)
```

```
data.to_csv('final.csv')
```

```
data['date_added']
```

Output

```
Timestamp('2019-05-17 21:44:01.638225408')
```

Conclusion:

IT professionals need to understand the definition of mean, median, mode and range to plan capacity and balance load, manage systems, perform maintenance and troubleshoot issues. These various tasks dictate that the administrator calculate mean, median, mode or range, or often some combination, to show a statistically significant quantity, trend or deviation from the norm. Finding the mean, median, mode and range is only the start. The administrator then needs to apply this information to investigate root causes of a problem, accurately forecast future needs or set acceptable working parameters for IT systems.

Thus, with the help of this experiment, an IT professional can apply this information for analyzing the data.