

VIT Vidyalankar Institute of Technology Workstedulin Department of Computer Engineering Exp. No.1

Semester	T.E. Semester V – Computer Engineering
Subject	Data Warehousing and Mining
Subject Professor In-charge	Prof. Kavita Shirsat
Assisting Teachers	Prof. Kavita Shirsat
Laboratory	M-313A

Student Name	Vibodh Bhosure	
Roll Number	20102A0032	
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	 central tendency of a large d for the mean, median and m When working with a large d represent the entire data set describes the "middle" or "av In statistics, that single value and mean, median and mode To find the mean, add up the divide by the number of value To find the median, list the v numerical order and identify middle of the list. To find the mode, identify whost often. Range, which is the difference smallest value in the data set tendency represents the data central tendency is not as repwould be if the range was smallest was smalle	lata set, it can be useful to t with a single value that verage" value of the entire set. e is called the central tendency e are all ways to describe it. e values in the data set and then les that you added. values of the data set in which value appears in the hich value in the data set occurs ce between the largest and t, describes how well the central a. If the range is large, the presentative of the data as it	
Program	<pre># -*- coding: utf-8 -*- """Expt - 1.ipynb Automatically generated by</pre>	y Colaboratory.	

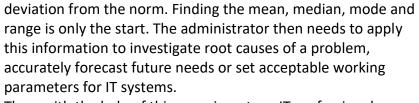


Department of Computer Engineering Exp. No.1

```
Original file is located at
                         https://colab.research.google.com/drive/1DVrFC96o0fx
                         JubIQ6PzllCe0QPdP0g7J
                         from google.colab import files
                         upload_data = files.upload()
                         import pandas as pd
                         df = pd.read_csv('TSLA (1).csv')
                          """Attribute - Cost"""
                         from statistics import mode
                         if df['Cost'].mean() > df['Cost'].median():
                           print("Dataset is positively skewed for Cost")
                         elif df['Cost'].mean() < df['Cost'].median():</pre>
                           print("Dataset is negatively skewed for Cost")
                           print("Dataset is normally distributed for Cost")
                         print("Range: ",max(df['Cost']) - min(df['Cost']))
                         print("Mid-Range: ",(max(df['Cost']) +
                         min(df['Cost']))/2)
                         print("Mean: ",df['Cost'].mean())
                         print("Median: ",df['Cost'].median())
                         print("Mode: ",mode(df['Cost']))
                         import matplotlib.pyplot as plt
                         import numpy as np
                         df.plot(x="Date",y="Cost")
                         """Different Attribute"""
                         pd.isnull(df['Cost'])
Output
                          Dataset is negatively skewed for Cost
                          Range: 1188.45003492
                          Mid-Range: 640.18501654
                          Mean: 531.0040880730204
                          Median: 603.8800049
                          Mode: 855.0
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
```



Department of Computer Engineering Exp. No.1



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