Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 06/10/2026 **Time:** 09:30 AM - 11:00 AM

Course Code: MAT3051 **Course Name:** Data Analytics with Python

Semester: Seventh Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2	CO3
Marks	44	42	4

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Define mean and median with their mathematical expressions.	2	CO1	BT1
2	Explain the meaning of mode in a dataset.	2	CO1	BT2
3	Explain how correlation helps in interpreting the relationship between two variables	2	CO3	BT2
4	Define covariance and explain its significance in data analysis.	2	CO3	BT2
5	Explain how an outlier can affect the mean of a dataset	2	CO2	BT2

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	Apply the stages of the data analytics lifecycle to a real-world problem, from data collection through analysis to decision-making.	10	CO1	BT3
7	Apply vector operations and norms to analyze the following three student feature vectors, where the components represent Mathematics, Statistics, and Programming marks: A=(70,80,90),B=(60,85,88),C=(75,78,92) a) Calculate A+B, A−B, and 2A. b) Calculate the dot product of A and B. c) Calculate the L1 and L2 (Euclidean) norms of A. d) Calculate the Euclidean distance between A and B. e) Implement the above operations using NumPy and interpret the results in the context of student-performance analysis.	10	CO1	BT3

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	Apply appropriate data-cleaning techniques to a given dataset to handle missing values, duplicate records, and outliers, and prepare the dataset for further analysis.	10	CO1	BT3

9	Apply Pandas operations to the following student-performance data: **Student_ID Department Gender Marks Attendance** 101 CSE M 78 85 102 CSE F 85 90 103 ECE M NaN 82 104 ECE F 72 80 104 ECE F 72 80 105 CSE M 95 94 a) Create the dataset as a Pandas DataFrame. b) Identify and handle the missing Marks value using an appropriate method. c) Detect and remove the duplicate record. d) Use `groupby()` to calculate the average Marks for each Department. e) Construct a pivot table showing the average Marks for each Department and Gender. f) Interpret the results obtained from the `groupby()` and pivot table operations.	10	CO1	BT3

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
10	Apply the binomial probability distribution to the following situation: The number of telephone lines busy at any instant is a binomial random variable. The probability that a telephone line is busy is 0.10.1. If 10 lines are selected at random, calculate the probability that: (i) No line is busy (ii) All lines are busy (iii) At least one line is busy (iv) At most two lines are busy (v) Exactly two lines are busy Also, calculate the mean and variance of the distribution.	10	CO2	BT3
11	The marks of 1000 students in an examination follows a normal distribution with mean 70 and standard deviation 5. Find the number of students will be (i) less than 65 (ii) more than 75 (iii) between 65 and 75.	10	CO2	BT3

PART E: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
12	Fit a Poisson distribution for the data and calculate the theoretical frequencies x: 0 1 2 3 4 f: 122 60 15 2 1	10	CO2	BT3
13	Find the value of k such that the following distribution represents a finite probability distribution . Hence find it's mean and variance. Also find P(X<=1),P(X>1) and P(-1<X<=2) X: -3 -2 -1 0 1 2 3 p(x): k 2k 3k 4k 3k 2k k	10	CO2	BT3