

GUJARAT TECHNOLOGICAL UNIVERSITY
DIPLOMA IN ENGINEERING – SEMESTER 4 – EXAMINATION – SUMMER 2026

Subject Code: DI04000071

Date: 21/05/2026

Subject Name: Introduction to Data Analysis

Time: 10:30 AM TO 01:00 PM

Total Marks: 70

Instructions :

1. Attempt all questions.
2. Make Suitable assumptions wherever necessary
3. Figures to the right indicate full marks.
4. Use of simple calculators and non-programmable scientific calculators are permitted.
5. English version is authentic.

- Q.1** (a) Define Data Analysis. List any four applications of data analysis. **03**
પ્રશ્ન ૧ (અ) ડેટા વિશ્લેષણ વ્યાખ્યાયિત કરો. ડેટા વિશ્લેષણના કોઈ પણ ચાર ઉપયોગો લખો. **03**
- (b) Explain the difference between structured and unstructured data with examples. **04**
(બ) ઉદાહરણ સાથે structured અને unstructured ડેટા વચ્ચેનો તફાવત સમજાવો. **04**
- (c) Explain the various phases of the Data Analytics Lifecycle. **07**
(ક) ડેટા એનાલિટિક્સ લાઇફ સાઇકલના તબક્કા સમજાવો. **07**
- (ક) ડેટા એનાલિટિક્સ લાઇફ સાઇકલના તબક્કા સમજાવો. **09**

અથવા

OR

- (c) Explain the concept of Big Data. What are the characteristics of Big Data (5 V's)? **07**
(ક) Big Data ને સમજાવો. Big Data ની વિશેષતાઓ (5 V's) શું છે? **09**

- Q.2** (a) What is NumPy? State the difference between a NumPy array and a Python list. **03**
પ્રશ્ન ૨ (અ) NumPy શું છે? NumPy array અને Python list વચ્ચેનો તફાવત જણાવો. **03**
- (b) Write a Python code to create a 2D NumPy array of shape (3,3) with random numbers. Also demonstrate vertical and horizontal stacking of two arrays. **04**
(બ) Random Numbers સાથે (3,3) આકાર નો 2D NumPy array બનાવવા માટે Python code લખો. બે arrays ની vertical અને horizontal stacking પણ દર્શાવો. **04**
- (c) Explain Pandas DataFrame with an example. Write Python code to: (i) Read a CSV file, (ii) Handle missing values, (iii) Find top 3 rows using nlargest(). **07**
(ક) ઉદાહરણ સાથે Pandas DataFrame સમજાવો. Python code લખો: (i) CSV ફાઈલ વાંચો, (ii) missing values handle કરો, (iii) nlargest() વડે top 3 rows મેળવો. **09**

અથવા

OR

- Q.2** (a) What is Web Scraping? Mention its purpose. **03**
પ્રશ્ન ૨ (અ) Web Scraping શું છે ? તેનો હેતુ જણાવો. **03**
- (b) Explain Data Normalization and Data Transformation with suitable examples. **04**

- (બ) ઉચિત ઉદાહરણ સાથે Data Normalization અને Data Transformation સમજાવો. 08
- (c) Explain BeautifulSoup library for web scraping. Write a Python program that sends a GET request to a website and extracts all paragraph tags using BeautifulSoup. 07
- (ફ) Web scraping માટે BeautifulSoup library સમજાવો. Python program લખો જે website પર GET request મોકલે અને BeautifulSoup વડે paragraph tags extract કરે. 09

Q.3
પ્રશ્ન 3

- (a) Define Mean, Median and Mode with examples. 03
- (અ) ઉદાહરણ સાથે Mean, Median અને Mode વ્યાખ્યાયિત કરો. 03
- (b) Explain Pearson's Correlation Coefficient. What does a value of +1, 0, and -1 indicate? 04
- (બ) Pearson's Correlation Coefficient સમજાવો. +1, 0 અને -1 ની value શું સૂચવે છે? 08
- (c) Describe Normal, Poisson, Exponential, and Bernoulli probability distributions with their definitions and importance. 07
- (ફ) Normal, Poisson, Exponential અને Bernoulli probability distributions ની વ્યાખ્યા અને મહત્વ વર્ણવો. 09

અથવા
OR

Q.3
પ્રશ્ન 3

- (a) What is the Method of Least Squares? State its use. 03
- (અ) Least Squares Method શું છે? તેનો ઉપયોગ જણાવો. 03
- (b) Write a short note on scatterplots and state its use in identifying correlation between variables. 04
- (બ) સ્કેટરપ્લોટ્સ પર ટૂંકી નોંધ લખો અને variables વચ્ચેના correlation ને ઓળખવામાં તેનો ઉપયોગ જણાવો. 08
- (c) Explain Descriptive Statistics in detail. Write a Python program to calculate Mean, Median, Mode, and Pearson's Correlation Coefficient for a given dataset. 07
- (ફ) Descriptive Statistics વિગતવાર સમજાવો. Python program લખો જે dataset માટે Mean, Median, Mode અને Pearson's Correlation Coefficient ગણે. 09

Q.4
પ્રશ્ન 4

- (a) What is Data Visualization? State its importance. 03
- (અ) Data Visualization શું છે? તેનું મહત્વ જણાવો. 03
- (b) Differentiate between Matplotlib and Seaborn libraries for data visualization. 04
- (બ) Data visualization માટે Matplotlib અને Seaborn libraries વચ્ચે તફાવત કરો. 08
- (c) Write a Python program using Matplotlib to plot a line chart, bar chart, and scatter plot for sample data. 07
- (ફ) Matplotlib વાપરીને sample data માટે line chart, bar chart અને scatter plot દોરવા માટે Python program લખો. 09

અથવા
OR

Q.4 પ્રશ્ન ૪	(a) What is Plotly? State any two advantages of using Plotly for visualization.	03
	(અ) Plotly શું છે? Visualization માટે Plotly ઉપયોગ કરવાના કોઈ પણ બે ફાયદા જણાવો.	03
	(b) Explain Time Series Data Visualization with a suitable example.	04
	(બ) ઉચિત ઉદાહરણ સાથે Time Series Data Visualization સમજાવો.	0૪
	(c) Write a Python program using Seaborn to:	07
	1. Create a bar plot showing the average income across regions using the 'tips' dataset.	
	2. Create a box plot to visualize the distribution of income across regions.	
	(ક) Seaborn નો ઉપયોગ કરીને Python પ્રોગ્રામ લખો:	09
	૧. 'tips' dataset નો ઉપયોગ કરીને વિવિધ પ્રદેશોમાં સરેરાશ આવક (income) દર્શાવતો bar plot બનાવો.	
	૨. વિવિધ પ્રદેશોમાં આવકનું વિતરણ (distribution) દર્શાવવા માટે box plot બનાવો.	
Q.5 પ્રશ્ન ૫	(a) What is Generative AI? How is it different from Predictive AI?	03
	(અ) Generative AI શું છે? તે Predictive AI થી કેવી રીતે અલગ છે?	03
	(b) Explain any four challenges of Big Data Visualization.	04
	(બ) Big Data Visualization ના કોઈ પણ ચાર પડકારો સમજાવો.	0૪
	(c) Explain Power BI and its key features.	07
	(ક) Power BI અને તેના મુખ્ય features સમજાવો.	09
	અથવા OR	
Q.5 પ્રશ્ન ૫	(a) List any three uses of Generative AI in Analytics.	03
	(અ) એનલિટિક્સ માં Generative AI ના કોઈપણ ત્રણ ઉપયોગોની યાદી આપો.	03
	(b) Explain bias, fairness, and privacy challenges of Generative AI.	04
	(બ) Generative AI ના bias, fairness અને privacy challenges સમજાવો.	0૪
	(c) Explain how to visualize large datasets using Power BI. Describe the role of AI copilots in Business Intelligence.	07
	(ક) Power BI વાપરીને large datasets visualize કેવી રીતે કરવા તે સમજાવો. Business Intelligence માં AI copilots ની ભૂમિકા વર્ણવો.	09
