

GUJARAT TECHNOLOGICAL UNIVERSITY

Diploma Engineering – SEMESTER – 6 (NEW) – EXAMINATION – Summer-2024

Subject Code: 4360707**Date: 21-05-2024****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 programmable & Communication aids are strictly prohibited.
5. Use of non-programmable scientific calculator is permitted.
6. English version is authentic.

		Marks
Q.1	(a) List the difference between structured, semi-structured, and unstructured data.	03
પ્રશ્ન.1	(અ) સ્ટ્રક્ચર્ડ, સેમીસ્ટ્રક્ચર્ડ અને અનસ્ટ્રક્ચર્ડ ડેટા વચ્ચેના તફાવતની સૂચિ બનાવો.	૦૩
	(b) What is the purpose of web scraping? Discuss the legality and ethical considerations associated with it.	04
	(બ) વેબ સ્ક્રેપિંગનો હેતુ શું છે? તેની સાથે સંકળાયેલ કાયદેસરતા અને નૈતિક બાબતોની ચર્ચા કરો.	૦૪
	(c) How does a Big Data platform differ from traditional data processing systems?	07
	(ક) બિગ ડેટા પ્લેટફોર્મ પરંપરાગત ડેટા પ્રોસેસિંગ સિસ્ટમથી કેવી રીતે અલગ છે?	૦૭
OR		
	(c) Describe briefly all phases of the data analytics lifecycle.	07
	(ક) ડેટા એનાલિટિક્સ જીવનચક્રના તમામ તબક્કાઓનું ટૂંકમાં વર્ણન કરો.	૦૭
Q.2	(a) List the difference between a NumPy array and a Python list.	03
પ્રશ્ન.2	(અ) NumPy array અને Python list વચ્ચેના તફાવતની સૂચિ બનાવો.	૦૩
	(b) What is data normalization, and how does it differ from rescaling?	04
	(બ) ડેટા normalization શું છે અને તે rescaling થી કેવી રીતે અલગ પડે છે?	૦૪
	(c) Find the top 5 and bottom 5 records of a DataFrame based on a specific column by using Pandas' nlargest() and nsmallest() functions	07
	(ક) Pandas ની nlargest() અને nsmallest() ફંક્શનનો ઉપયોગ કરીને ચોક્કસ કોલમ પર આધારિત ડેટાફ્રેમના ટોચના 5 અને નીચેના 5 રેકૉર્ડ્સ શોધો.	૦૭
OR		
Q.2	(a) Name three common file formats that can be read using Python's file reading methods.	03
પ્રશ્ન.2	(અ) Python's file વાંચન પદ્ધતિઓનો ઉપયોગ કરીને વાંચી શકાય તેવા ત્રણ સામાન્ય ફાઇલ ફોર્મેટના નામ આપો.	૦૩
	(b) Describe any four functionalities provided by BeautifulSoup for navigating and manipulating HTML elements.	04
	(બ) HTML તત્વો નેવિગેટ કરવા અને હેરફેર કરવા માટે BeautifulSoup દ્વારા પૂરી પાડવામાં આવેલ કોઈપણ ચાર કાર્યોનું વર્ણન કરો.	૦૪

	(c)	Write a Python script using Selenium to automate the process of logging into a website and extracting specific data from a dynamically loaded page.	07
	(ક)	વેબસાઇટમાં લોગ ઇન કરવાની પ્રક્રિયાને સ્વયંચાલિત કરવા માટે Selenium નો ઉપયોગ કરીને Python script લખો અને ગતિશીલ રીતે લોડ થયેલા પૃષ્ઠમાંથી ચોક્કસ ડેટા કાઢવા.	૦૭
Q. 3	(a)	List differences between Simple Linear Regression and Multiple Linear Regression.	03
પ્રશ્ન.3	(અ)	Simple Linear Regression અને Multiple Linear Regression વચ્ચેના તફાવતોની યાદી બનાવો.	૦૩
	(b)	Describe briefly any two probability distributions.	04
	(બ)	કોઈપણ બે probability distributions નું ટૂંકમાં વર્ણન કરો.	૦૪
	(c)	The manufacturing plant produces electronic components, and you are tasked with improving the production process to reduce defects. Describe how you would use the method of least squares to analyze the relationship between process parameters (such as temperature, pressure, and cycle time) and the occurrence of defects in the manufactured components.	07
	(ક)	મેન્યુફેક્ચરિંગ પ્લાન્ટ ઇલેક્ટ્રોનિક ઘટકોનું ઉત્પાદન કરે છે, અને તમને ખામી ઘટાડવા માટે ઉત્પાદન પ્રક્રિયામાં સુધારો કરવાનું કામ સોંપવામાં આવે છે. પ્રક્રિયાના પરિમાણો (જેમ કે તાપમાન, દબાણ અને ચક્ર સમય) અને ઉત્પાદિત ઘટકોમાં ખામીઓની ઘટના વચ્ચેના સંબંધનું વિશ્લેષણ કરવા માટે તમે least squares ની પદ્ધતિનો ઉપયોગ કેવી રીતે કરશો તેનું વર્ણન કરો.	૦૭
OR			
Q. 3	(a)	List the differences between mean, median, and mode as measures of central tendency.	03
પ્રશ્ન.3	(અ)	Central tendency ના માપદંડ તરીકે mean, median અને mode વચ્ચેના તફાવતોની સૂચિ બનાવો.	૦૩
	(b)	How does multivariate analysis differ from univariate analysis in terms of scope and methodology?	04
	(બ)	અવકાશ અને કાર્યપદ્ધતિની દ્રષ્ટિએ multivariate analysis થી univariate analysis કેવી રીતે અલગ પડે છે?	૦૪
	(c)	As a data analyst for a retail company, you have been tasked with analyzing the relationship between store foot traffic and daily sales revenue to optimize staffing levels and improve customer service. Using scatterplots and other graphical techniques, develop a comprehensive analysis plan to identify the correlation between store foot traffic (independent variable) and daily sales revenue (dependent variable).	07
	(ક)	રિટેલ કંપની માટે ડેટા વિશ્લેષક તરીકે, તમને સ્ટાફિંગ સ્તરને ઓપ્ટિમાઇઝ કરવા અને ગ્રાહક સેવામાં સુધારો કરવા માટે સ્ટોર ફૂટ ટ્રાફિક અને દૈનિક વેચાણની આવક વચ્ચેના સંબંધનું વિશ્લેષણ કરવાનું કામ સોંપવામાં આવ્યું છે. Scatterplots અને અન્ય ગ્રાફિકલ તકનીકોનો ઉપયોગ કરીને, સ્ટોર ફૂટ ટ્રાફિક (સ્વતંત્ર ચલ) અને દૈનિક વેચાણ આવક (આશ્રિત ચલ) વચ્ચેના સંબંધને ઓળખવા માટે એક વ્યાપક વિશ્લેષણ યોજના વિકસાવો.	૦૭
Q. 4	(a)	Define data visualization. List down importance of it.	03
પ્રશ્ન.4	(અ)	ડેટા વિઝ્યુલાઇઝેશન વ્યાખ્યાયિત કરો. તેનું મહત્વ લખો.	૦૩
	(b)	Describe Pearson's correlation coefficient and its significance in statistical analysis.	04
	(બ)	Pearson's correlation coefficient અને આંકડાકીય વિશ્લેષણમાં તેનું મહત્વ વર્ણવો.	૦૪
	(c)	For a retail company, as a data analyst analyzing sales performance across different product categories over the past year is tasked to you. Using	07

Seaborn, develop an interactive data visualization dashboard that provides insights into sales trends, seasonality, and product performance.

- (ક) રિટેલ કંપની માટે, ડેટા વિશ્લેષક તરીકે પાછલા વર્ષમાં વિવિધ પ્રોડક્ટ કેટેગરીમાં વેચાણ પ્રદર્શનનું વિશ્લેષણ કરવાનું કામ તમને સોંપવામાં આવ્યું છે. Seaborn નો ઉપયોગ કરીને, એક interactive ડેટા વિઝ્યુલાઇઝેશન ડેશબોર્ડ વિકસાવો જે વેચાણના વલણો, મોસમ અને ઉત્પાદન પ્રદર્શનમાં આંતરદૃષ્ટિ પ્રદાન કરે છે. ૦૭

OR

- Q. 4** (a) List down any six common types of data visualizations used in practice. 03
પ્રશ્ન.4 (અ) વ્યવહારમાં ઉપયોગમાં લેવાતા કોઈપણ છ સામાન્ય પ્રકારના ડેટા ૦૩

વિઝ્યુલાઇઝેશનની યાદી બનાવો.

- (b) Describe the role of scatterplots and other graphical techniques in identifying the correlation between variables in a dataset. 04
(બ) ડેટાસેટમાં variable વચ્ચેના સહસંબંધને ઓળખવામાં scatterplots અને અન્ય ગ્રાફિકલ તકનીકોની ભૂમિકાનું વર્ણન કરો. ૦૪
(c) For a marketing firm, as a data analyst you have to analyze the effectiveness of different advertising campaigns across various channels. Using Matplotlib, develop an interactive data visualization dashboard that provides insights into advertising performance, audience engagement, and return on investment (ROI). 07
(ક) માર્કેટિંગ ફર્મ માટે, ડેટા વિશ્લેષક તરીકે તમારે વિવિધ ચેનલોમાં વિવિધ જાહેરાત ઝુંબેશની અસરકારકતાનું વિશ્લેષણ કરવું પડશે. Matplotlib નો ઉપયોગ કરીને, એક ઇન્ટરેક્ટિવ ડેટા વિઝ્યુલાઇઝેશન ડેશબોર્ડ વિકસાવો જે જાહેરાત પ્રદર્શન, પ્રેક્ષકોની engagement અને રોકાણ પર વળતર (ROI) માં આંતરદૃષ્ટિ પ્રદાન કરે છે. ૦૭

- Q.5** (a) List down two common types of plots used for visualizing time series data. Describe briefly each plot type. 03
પ્રશ્ન.5 (અ) Time series ના ડેટાને વિઝ્યુઅલાઇઝ કરવા માટે ઉપયોગમાં લેવાતા બે સામાન્ય પ્રકારના પ્લોટની યાદી બનાવો. દરેક પ્લોટ પ્રકારનું ટૂંકમાં વર્ણન કરો. ૦૩

- (b) Describe the process of creating an interactive data visualization using Plotly in Python. 04
(બ) Python માં Plotly નો ઉપયોગ કરીને ઇન્ટરેક્ટિવ ડેટા વિઝ્યુલાઇઝેશન બનાવવાની પ્રક્રિયાનું વર્ણન કરો. ૦૪
(c) List recent trends in data collection and analysis techniques and give brief description of each technique. 07
(ક) Data collection અને analysis તકનીકોમાં તાજેતરના વલણોની સૂચિ બનાવો અને દરેક તકનીકનું સંક્ષિપ્ત વર્ણન આપો. ૦૭

OR

- Q.5** (a) Name the Python library which is commonly used for data visualization, particularly for creating statistical graphics and exploring datasets. Describe it briefly. 03
પ્રશ્ન.5 (અ) Python લાઇબ્રેરીને નામ આપો જેનો ઉપયોગ સામાન્ય રીતે ડેટા વિઝ્યુલાઇઝેશન માટે લેવાય છે, ખાસ કરીને આંકડાકીય ગ્રાફિક્સ બનાવવા અને ડેટાસેટ્સનું શોધખોળ કરવા માટે. તેનું ટૂંકમાં વર્ણન કરો. ૦૩

- (b) How Box plot can be used for visualizing data distributions and identifying outliers? 04
(બ) ડેટા ડિસ્ટ્રિબ્યુશનને વિઝ્યુઅલાઇઝ કરવા અને આઉટલાયર્સને ઓળખવા માટે બોક્સ પ્લોટનો ઉપયોગ કેવી રીતે કરી શકાય? ૦૪

- (c) List challenges associated with big data visualization. Briefly explain any three challenges and provide a potential solution for these three challenges. **07**
- (ક) Big data વિઝ્યુલાઇઝેશન સાથે સંકળાયેલ પડકારોની સૂચિ બનાવો. કોઈપણ ત્રણ પડકારોને સંક્ષિપ્તમાં સમજાવો અને આ ત્રણ પડકારો માટે સંભવિત ઉકેલ પ્રદાન કરો. **૦૭**