

GUJARAT TECHNOLOGICAL UNIVERSITY**BE (Minor) - SEMESTER– V EXAMINATION – WINTER 2023****Subject Code: 115AG01****Date: 07-02-2024****Subject Name: Introduction to Data Analytics****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. Simple and non-programmable scientific calculators are allowed.
5. There is no need of graph paper, one can draw rough sketch whenever needed in answer sheet only.

		MARKS
Q.1	(a) Usefulness of data science is everywhere. Justify it with listing sectors in which data science is widely used, also show them in ascending order, heavy usage listed first.	03
	(b) List specific Components of Python used in data science. Explain any one of them briefly.	04
	(c) Python is used in Data Science, give reasons and justify your answer.	07
Q.2	(a) Define and explain the following in the context of the Data Analytics Process. i. Data Collections ii. Data Wrangling	03
	(b) Domain Knowledge in Data Science is very important for any data analytics, justify it using one, to the point example.	04
	(c) What do you mean by Exploratory Data Analysis (EDA), why it is important in data analytics? What are the goals of EDA?	07
	OR	
	(c) What are Univariate and Multivariate EDA? Explain it with an example.	07
Q.3	(a) Where and why Pearson correlation coefficient is being used?	03
	(b) Taking one relevant example, explain the process of feature selection.	04
	(c) What is Feature selection and its significance in data analytics? Explain feature selection techniques applicable for univariate and multivariate cases.	07
	OR	
Q.3	(a) What is the significance of Correlation and covariance in the feature selection process?	03
	(b) Describe importance of brainstorming and role of domain expertise, in context of Feature Generation, in brief.	04
	(c) Taking example of Customer Retention Analytics, describe types of analytics and strategies to derive such analytics.	07
Q.4	(a) Following text shows code to visualize data. Draw rough sketch of the resulting graph, along with the resulting nomenclature.	03
	<pre> from matplotlib import pyplot as plt years = [1950, 1960, 1970, 1980, 1990, 2000, 2010] gdp = [300.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 10000.3] plt.plot(years, gdp, color='green', marker='o', linestyle='solid') plt.title("Nominal GDP") plt.ylabel("Billions of \$") plt.show()</pre>	

- (b) What is Scatterplots? In which types of application of visualization it is useful? **04**
- (c) Describe relationship between Seaborn and matplotlib. What are features of seaborn for visualization of data. **07**
- OR**
- Q.4** (a) Following text shows code to visualize data. Draw rough sketch of the resulting graph, along with the resulting nomenclature. **03**
- ```

movies = ["Annie Hall", "Ben-Hur", "Casablanca", "Gandhi", "West Side
Story"]
num_oscars = [5, 11, 3, 8, 3]
plt.bar(range(len(movies)), num_oscars)
plt.title("My Favorite Movies")
plt.ylabel("# of Academy Awards")
plt.xticks(range(len(movies)), movies)
plt.show()

```
- (b) What are Density and Contour Plots? In which types of application of visualization it is useful? **04**
- (c) Describe how Three-Dimensional Plotting is achieved. Describe it specifically how one can do it in Matplotlib **07**
- Q.5** (a) How will next generation data scientists be different compared to now? **03**
- (b) Clearly distinguish between privacy and security in the context of data science. **04**
- (c) Taking specific applications of face detection, explain the entire data analytics process. **07**
- OR**
- Q.5** (a) What are the important characteristics of “next-generation data scientists”? **03**
- (b) What are Ethical Issues related to Data Science? What changes are expected due to this? **04**
- (c) Taking specific applications of stock market prices, explain the entire data analytics process. **07**
- .

\*\*\*\*\*