

MSc IT (BI & A) Sem.-2 Examination

BIAMSC-12

Data Science Exploration

Time : 2-30 Hours]

May-2025

[Max. Marks : 70

Instructions:

- **Figures to the right indicate Full Marks.**
- **Do not write anything on the question paper.**
- **Write the answers according to the question marks**

Question 1: - Attempt any 10 questions.**[30]****1.**

- 1. What does the following line do?**
`df['Total_Female'].hist()`
 - a) History
 - b) Histogram
 - c) Both
 - d) None of the above
- 2. What does df.info() display?**
 - a) Correlation between columns
 - b) Summary of dataset including data types and non-null counts
 - c) Graphical data summary
 - d) Statistical outliers
- 3. What does sns.boxplot(x=df['Total_Male']) show?**
 - a) Trend of male population
 - b) Histogram of male population
 - c) Distribution and outliers in male population
 - d) Correlation between male and female
- 4. What is the purpose of this line?**
`df[['Total_Male', 'Total_Female', 'Railway_Station_Distance']]`
 - a) Filters categorical columns
 - b) Drops numeric columns
 - c) Selects specific numeric columns for analysis
 - d) Merges rows
- 5. What does df.isnull().sum().sum() return?**
 - a) Total number of missing values in the dataset
 - b) Only missing values in the first column
 - c) Maximum value of nulls in any row
 - d) Missing values in the 'sum' column
- 6. What does plt.title('Histogram of Total Male Population') do?**
 - a) Sets axis labels
 - b) Displays total values
 - c) Sets the title of the plot
 - d) Removes title from plot
- 7. What is sns.pairplot() used for?**
 - a) Pairing string columns
 - b) Plotting correlation heatmap

(P.T.O)

- c) Plotting scatter plots for pairwise combinations of columns
- d) Grouping columns by value
- 8. Which method gives the number of missing values in a column?
 - a) df.countna()
 - b) df.missing()
 - c) df.isnull().sum()
 - d) df.checknull()
- 9. Which line is used to rotate x-axis labels in a plot?
 - a) plt.xticks(45)
 - b) plt.xticks(rotation=45)
 - c) plt.xrotate(45)
 - d) plt.set_xticks(45)
- 10. What is the default number of bins in hist() if not specified?
 - a) 5
 - b) 10
 - c) 20
 - d) 100
- 11. What does inplace=True do in fillna() method?
 - a) Creates new copy of DataFrame
 - b) Ignores operation
 - c) Modifies DataFrame in place without returning a new object
 - d) Deletes the column
- 12. In the context of .groupby(), what does agg() do?
 - a) Adds new rows
 - b) Drops missing values
 - c) Aggregates statistics like sum, mean, etc.
 - d) Groups rows without calculating

Question 2: - Attempt this questions. [20]

- a) Could you give a few real-world examples of traditional data?
- b) What are the main stages in the Data Science Life Cycle?

OR

Question 2: - Attempt this questions. [20]

- a) What are some practical applications of Business Intelligence in the real world?
- b) What are the popular data structures commonly used in Data Science?

Question 3: - Attempt Any 2 questions. [20]

- a) Can you list out common data formatting techniques used in Data Science?
- b) What is the concept of Semi-Supervised Learning and Reinforcement Learning, along with their types, benefits, and drawbacks?
- c) How would you define Clustering and Association in the context of Data Science?
- d) What are the top use cases for Supervised Learning?

OR

Question 3: - Attempt Any 2 questions. [20]

- a) In what ways is Unsupervised Learning applied? Name the examples.
- b) What are the common applications of Reinforcement Learning in various industries?
- c) What is Logistic Regression, and how is it used in Data Science?
- d) Can you describe three popular feature selection techniques in Data Science?

*******ALL THE BEST*******