

SRM VALLIAMMAI ENGINEERING COLLEGE
(An Autonomous Institution)

SRM Nagar, Kattankulathur-603203

DEPARTMENT OF INFORMATION TECHNOLOGY

QUESTION BANK



I SEMESTER M.Tech., Data Science

DS3161-PYTHON FOR DATASCIENCE

Regulation- 2023

Academic Year 2026 – 2027 (Odd Semester)

Prepared by

Ms.I.Kavitha, Assistant Professor-O.G/IT



SRM VALLIAMMAI ENGINEERING COLLEGE
(An Autonomous Institution)
SRM Nagar, Kattankulathur-603203.



DEPARTMENT OF INFORMATION TECHNOLOGY
QUESTION BANK

SUBJECT : DS3161 – PYTHON FOR DATA SCIENCE

SEM/YEAR : I Semester / I Year

UNIT I		INTRODUCTION TO DATA SCIENCE	
Data Science: Benefits and uses – facets of data - Data Science Process: Overview – Defining research goals – Retrieving data – Data preparation - Exploratory Data analysis – build the model– presenting findings and building applications - Data Mining - Data Warehousing – Basic Statistical descriptions of Data			
PART– A			
Q.No.	Questions	BT Level	Competence
1.	Explain about roles and stages in data science project.	BTL1	Remembering
2.	Illustrate data cleaning process.	BTL1	Remembering
3.	Demonstrate data exploration process.	BTL1	Remembering
4.	Describe the facets of data.	BTL2	Remembering
5.	Describe data mining.	BTL1	Remembering
6.	Outline the data warehouse.	BTL1	Remembering
7.	Define data lake.	BTL2	Understanding
8.	What is machine - generated data ?	BTL2	Understanding
9.	Define streaming data.	BTL1	Remembering
10.	What is outlier detection?	BTL2	Understanding
11.	Explain the benefits of data science.	BTL3	Applying
12.	List out data science process steps.	BTL3	Applying
13.	Define data preparation.	BTL3	Applying
14.	How to retrieving the data from the data store.	BTL4	Analyzing
15.	Defining research goals of the project.	BTL4	Analyzing
16.	Explain statistical descriptions of data.	BTL4	Analyzing
17.	What are the tools are used in data science.	BTL5	Evaluating
18.	Illustrate data modeling process.	BTL5	Evaluating
19.	Explain the uses of data science.	BTL2	Understanding
20.	Explain in detail about structured and unstructured data.	BTL2	Understanding
21.	Define data cleansing.	BTL1	Remembering
22.	Explain basic statistical descriptions of data.	BTL2	Understanding
23.	Explain how to retrieving data.	BTL2	Understanding

24.	List out the research goals in data science.	BTL4	Analyzing
PART– B			
1.	Examine the different types of data with the challenges in their processing. (16)	BTL1	Remembering
2.	Explain in detail about Basic Statistics Concepts for Data Science. (16)	BTL1	Remembering
3.	Explain the various steps associated with data science process and explain any three steps of it with suitable diagrams and example. (16)	BTL2	Understanding
4.	How do work together in data warehouses, database, and data lakes. (16)	BTL4	Analyzing
5.	What is data mining? How its work benefits and technique? (16)	BTL4	Analyzing
6.	Demonstrate Retrieving data and Data preparation. (16)	BTL2	Understanding
7.	Explain about Exploratory Data analysis and methods to build the model? (16)	BTL2	Understanding
8.	Define Data warehouse architecture and its concepts. (16)	BTL1	Remembering
9.	Examine the detail in data warehousing. (16)	BTL3	Applying
10.	How to build the Data Science Model? (16)	BTL6	Creating
11.	What are the data processing methods in data science? (16)	BTL4	Analyzing
12.	Explain in detail about i) Data Lake (4) ii) Data Mart (6) iii) Data Base (6)	BTL2	Understanding
13.	What is data warehousing? How its work benefits and technique? (16)	BTL2	Understanding
14.	Explain in detail about Exploratory Data analysis. (16)	BTL1	Remembering
15.	Explain detail about Data Mining. (16)	BTL3	Applying
16.	Explain about Data Science and its benefits. (16)	BTL1	Remembering
17.	Define research goals in Data Science. (16)	BTL4	Analyzing
UNIT II MATHEMATICAL FOUNDATIONS			
Linear Algebra: Vectors, Matrices- Statistics: Describing a Single Set of Data, Correlation, Simpson's Paradox-Correlation and Causation- Probability: Dependence and Independence, Conditional Probability, Bayes's-Theorem, Random Variables-Continuous Distributions- The Normal Distribution-The Central Limit Theorem.			
PART–A			
1.	Explain Simple linear correlations.	BTL2	Understanding
2.	Define Yule – Simpson method.	BTL5	Evaluating
3.	Define Bivariate normal distribution.	BTL1	Remembering
4.	List out Central Limit Theorem formula's.	BTL1	Remembering
5.	How to apply Bayes's Theorem or Bayes rule in Machine Learning.	BTL1	Remembering
6.	Explain Perfect Negative Linear Relationship.	BTL1	Remembering
7.	Difference between Vector and Matrices.	BTL2	Understanding
8.	Describe in detail about Sampling.	BTL2	Understanding
9.	Describing a Single Set of Data.	BTL2	Understanding
10.	Explain Dependence and Independence.	BTL2	Understanding
11.	Define Simpson's Paradox.	BTL3	Applying
12.	Explain Correlation and its types.	BTL3	Applying

13.	Illustrate Continuous Distributions.	BTL3	Applying
14.	Explain Central limit theorem.	BTL1	Remembering
15.	Define Bayes's theorem.	BTL4	Analyzing
16.	Define Normal Distribution.	BTL4	Analyzing
17.	Define Conditional Probability.	BTL5	Evaluating
18.	Difference between Causation and Correlation.	BTL3	Applying
19.	Define Linear Algebra.	BTL1	Remembering
20.	Demonstrate in detail Statistics.	BTL4	Analyzing
21.	Explain Probability and its types.	BTL4	Analyzing
22.	Explain in detail about Vectors.	BTL5	Evaluating
23.	Explain in detail Causation.	BTL2	Understanding
24.	Brief explain about Statistics and its types.	BTL2	Understanding
PART-B			
1.	Brief explain about Linear Algebra. (16)	BTL3	Applying
2.	Brief explain about Probability. (16)	BTL2	Understanding
3.	Explain the brief about Central Limit Theorem. (16)	BTL2	Understanding
4.	Explain in detail about Distribution concepts. (16)	BTL4	Analyzing
5.	Demonstrate the Normal Distribution. (16)	BTL2	Understanding
6.	Define Simpson's Paradox. (16)	BTL2	Understanding
7.	Difference between Correlation and Causation. (16)	BTL2	Understanding
8.	Briefly explain in detail about the Statistics. (16)	BTL3	Applying
9.	Explain Bayes's –Theorem. (16)	BTL3	Applying
10.	Explain Random Variables definition types and examples & formula. (16)	BTL3	Applying
11.	Explain a Single Set of Data and Correlation. (16)	BTL4	Analyzing
12.	Explain Distribution Concepts. (16)	BTL4	Analyzing
13.	Brief explain about Random variables. (16)	BTL5	Evaluating
14.	Explain the brief about (i) Correlation (8) (ii) Causation (8)	BTL6	Creating
15.	Difference between Continuous Distributions and Normal distribution. (16)	BTL6	Creating
16.	Define Vector and Matrices. (16)	BTL6	Creating
17.	Explain Dependence and Independence Probability. (16)	BTL5	Evaluating
UNIT III MACHINE LEARNING			
Overview of Machine learning concepts –Types of Machine learning - Linear Regression- model assumptions-Classification and Regression algorithms- Naïve Bayes, K-Nearest Neighbors, logistic regression- Support Vector Machines (SVM), Decision Trees, And Random Forest.			
PART-A			
1.	What is Machine Learning?	BTL1	Remembering
2.	Infer the term KNN.	BTL2	Understanding
3.	List the types of ML Classification Algorithms.	BTL1	Remembering
4.	Show the types of Machine Learning.	BTL2	Understanding
5.	What is Logistic Regression?	BTL1	Remembering

6.	Illustrate the difference between Regression vs Classification.	BTL2	Understanding
7.	Define Supervised machine Learning algorithm.	BTL2	Understanding
8.	Define Reinforcement Learning.	BTL1	Remembering
9.	What is hypothesis function for Linear Regression?	BTL2	Understanding
10.	Infer the Examples of Machine Learning Classification in Real Life.	BTL2	Understanding
11.	Recall the types of Classification tasks in Machine Learning.	BTL1	Remembering
12.	What is the difference between a feature and a label?	BTL1	Remembering
13.	What is the purpose of a model in machine learning?	BTL2	Understanding
14.	How does a machine learning model learn from data?	BTL1	Remembering
15.	How does K-Nearest Neighbors algorithm work?	BTL2	Understanding
16.	Name two common evaluation metrics for classification problems.	BTL1	Remembering
17.	What is the role of a kernel function in SVM?	BTL2	Understanding
18.	What is the ROC curve?	BTL1	Remembering
19.	What is the difference between precision and recall?	BTL2	Understanding
20.	What is the purpose of a confusion matrix?	BTL1	Remembering
21.	What are the assumptions of linear regression?	BTL2	Understanding
22.	How does Naive Bayes algorithm work?	BTL1	Remembering
23.	What is the difference between a decision tree and a random forest?	BTL2	Understanding
24.	How does a decision tree make predictions?	BTL1	Remembering

PART-B

1.	Describe the overview of Machine learning concepts. (16)	BTL5	Evaluating
2.	Analyze the concepts of Linear Regression and its types. (16)	BTL3	Applying
3.	Examine in detail about Random Forest and how can you use random forests to identify important features in a dataset? (16)	BTL4	Analyzing
4.	Compare Classification and Regression algorithms. (16)	BTL5	Evaluating
5.	Explain in detail about different types of Classification tasks in Machine Learning. (16)	BTL4	Analyzing
6.	Demonstrate in detail about model assumptions in machine learning. (16)	BTL3	Applying
7.	Analyze the advantages of SVM and how optimal Hyperplane differs from Hyperplane. (16)	BTL5	Evaluating
8.	Explain Commonly used Machine Learning algorithm. (16)	BTL4	Analyzing
9.	Explain (i) Regression and Correlation (8) (ii) Limitation of Regression Model. (8)	BTL3	Applying
10.	Explain the supervised and unsupervised learning with example. (16)	BTL3	Applying
11.	Explain KNN and its importance with example. (16)	BTL4	Analyzing
12.	Explain hyper-parameter tuning with example. (16)	BTL4	Analyzing

13.	Explain the various methods of measuring classifier performance with suitable examples. (16)	BTL5	Evaluating
14.	Explain the process of constructing CART (Classification and Regression Tree) with a suitable example. (16)	BTL4	Analyzing
15.	Define clustering. Explain the different types of clustering algorithms in detail. (16)	BTL5	Evaluating
16.	Explain the evaluation technology of machine learning algorithm. (16)	BTL5	Evaluating
17.	Examine in detail about Decision Tree. (16)	BTL5	Evaluating

UNIT IV PYTHON LIBRARIES FOR DATA WRANGLING

Basics of Numpy arrays –aggregations –computations on arrays –comparisons, masks, boolean logic – fancy indexing – structured arrays – Data manipulation with Pandas – data indexing and selection – operating on data – missing data – Hierarchical indexing – combining datasets – aggregation and grouping – pivot tables

1.	What is NumPy?	BTL1	Remembering
2.	How do you create a NumPy array?	BTL2	Understanding
3.	What are the basic data types in NumPy?	BTL1	Remembering
4.	How do you access elements in a NumPy array?	BTL2	Understanding
5.	What is the shape and size of a NumPy array?	BTL1	Remembering
6.	How do you calculate the sum, mean, and standard deviation of a NumPy array?	BTL2	Understanding
7.	What is the difference between axis=0 and axis=1 in NumPy aggregations?	BTL2	Understanding
8.	How do you perform element-wise operations on NumPy arrays?	BTL1	Remembering
9.	How do you perform matrix multiplication in NumPy?	BTL2	Understanding
10.	What are universal functions in NumPy?	BTL2	Understanding
11.	What is a structured array in NumPy?	BTL1	Remembering
12.	How do you create a structured array?	BTL1	Remembering
13.	How do you access fields in a structured array?	BTL2	Understanding
14.	How do you sort a structured array by a specific field?	BTL1	Remembering
15.	What is fancy indexing in NumPy?	BTL2	Understanding
16.	How do you select specific elements using fancy indexing?	BTL1	Remembering
17.	What is Pandas?	BTL2	Understanding
18.	How do you create a Pandas DataFrame?	BTL1	Remembering
19.	How do you access rows and columns in a Pandas DataFrame?	BTL2	Understanding
20.	What is the difference between loc and iloc indexing?	BTL1	Remembering
21.	How do you filter rows and columns based on conditions?	BTL2	Understanding
22.	What is Pandas?	BTL1	Remembering
23.	How do you add, subtract, multiply, and divide columns in a Pandas DataFrame?	BTL2	Understanding

24.	How do you apply functions to columns in a Pandas DataFrame?	BTL1	Remembering
PART-B			
1.	What is an aggregate function? Evaluate about the aggregate functions in Numpy. (16)	BTL5	Evaluating
2.	(i) What is broadcasting? Identify the rules of broadcasting with an example. (8) (ii) Organize the concepts of mapping between Python operators and Pandas methods. (8)	BTL3	Applying
3.	Analyze grouping in python with example. (16)	BTL4	Analyzing
4.	Explain the following in python (i) Data Indexing. (8) (ii) Operation on missing data. (8)	BTL5	Evaluating
5.	Examine the following (i) Fancy Indexing with Example. (8) (ii) Explain about Comparison, Masks and Boolean Logic. (8)	BTL4	Analyzing
6.	Identify the various methods of handling the missing data in Pandas. (16)	BTL3	Applying
7.	(i) Briefly explain about Hierarchical Indexing. (8) (ii) Determine different ways of creating Pandas data frame. (8)	BTL5	Evaluating
8.	Imagine you have a series of data that represents the amount of precipitation each day for a year in a given city. Load the daily rainfall statistics for the city of Chennai in 2021. Which is given in a csv file Chennai rainfall 2021.csv using Pandas generate a histogram for rainy days and find out the days that have high rainfall. (16)	BTL4	Analyzing
9.	Consider that an E commerce organization like Amazon have different region sales as Northsales, Southsales, Westsales.csv files. They want combine North and west region ales and south and east sales to find the aggregate sales of this collaborating region help them to do so using python code. (16)	BTL3	Applying
10.	Experiment with the basics of Numpy arrays with example. (16)	BTL3	Applying
11.	What is universal function? Analyze each type of function with illustration. (16)	BTL4	Analyzing
12.	Examine the data objects in pandas. (16)	BTL4	Analyzing
13.	Evaluate the concepts of pivot table with example. (16)	BTL5	Evaluating
14.	Consider the following NumPy array: <pre>import numpy as np arr = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])</pre> (i) Reshape the array to a 1D array. (4) (ii) Calculate the mean of each row. (4) (iii) Find the maximum value in the entire array and create a new array by adding 2 to each element of the original array. (8)	BTL4	Analyzing
15.	(i) Create two NumPy arrays of different shapes that can be broadcasted together. (8) (ii) Discuss the advantages and disadvantages of using NumPy arrays over Python lists. (8)	BTL5	Evaluating
16.	Given a CSV file containing information about students (name, age, grade), create a Pandas DataFrame and perform the following tasks: (i) Display the first 5 rows of the DataFrame. (4 marks) (ii) Sort the DataFrame by age in descending order. (4 marks)	BTL5	Evaluating

	(iii) Calculate the average age of students in each grade. (4 marks) (iv) Filter the DataFrame to show only students with a grade higher than 85. (4 marks)																						
17.	Assess the following, (i) Create a sample DataFrame with a hierarchical index representing different products sold in various stores. (8) (ii) Discuss the advantages of using hierarchical indexing in data analysis. (8)	BTL5	Evaluating																				
UNIT V PYTHON FOR DATA VISUALIZATION																							
Visualization with matplotlib lib – line plots – scatter plots – visualizing errors – density and contour plots – histograms, binnings, and density – three dimensional plotting – geographic data – data analysis using stat models and seaborn – graph plotting using Plotly – interactive data visualization using Bokeh																							
PART-A																							
1.	What is a line plot used for?	BTL1	Remembering																				
2.	How do you customize the line style and color in a Matplotlib line plot?	BTL2	Understanding																				
3.	What is a scatter plot used for?	BTL1	Remembering																				
4.	How do you add markers and labels to a scatter plot?	BTL2	Understanding																				
5.	How do you add error bars to a line plot?	BTL1	Remembering																				
6.	What are the different types of error bars?	BTL2	Understanding																				
7.	What is a density plot?	BTL2	Understanding																				
8.	What is a contour plot?	BTL1	Remembering																				
9.	How do you create a histogram in Matplotlib?	BTL2	Understanding																				
10.	What is binning in the context of histograms?	BTL2	Understanding																				
11.	How do you plot a probability density function (PDF) over a histogram?	BTL1	Remembering																				
12.	What are the basic 3D plot types in Matplotlib?	BTL1	Remembering																				
13.	How do you add labels and a colorbar to a 3D plot?	BTL2	Understanding																				
14.	How do you plot geographical data on a map using Matplotlib?	BTL1	Remembering																				
15.	What is Basemap toolkit, and how is it used for geographic plotting?	BTL2	Understanding																				
16.	What is the role of Statsmodels in data analysis?	BTL1	Remembering																				
17.	How does Seaborn simplify data visualization compared to Matplotlib?	BTL2	Understanding																				
18.	What is Plotly?	BTL1	Remembering																				
19.	How do you create interactive plots using Plotly?	BTL2	Understanding																				
20.	What is Bokeh?	BTL1	Remembering																				
21.	How does Bokeh differ from Matplotlib?	BTL2	Understanding																				
22.	How do you create interactive dashboards using Bokeh?	BTL1	Remembering																				
23.	What is the purpose of errorbar function in Matplotlib? Give an example.	BTL2	Understanding																				
24.	Write a python code snippet that draws a histogram for the following list of positive numbers. <table><tr><td>7</td><td>18</td><td>9</td><td>44</td><td>2</td><td>5</td><td>89</td><td>91</td><td>11</td><td>6</td></tr><tr><td>77</td><td>85</td><td>91</td><td>6</td><td>55</td><td></td><td></td><td></td><td></td><td></td></tr></table>	7	18	9	44	2	5	89	91	11	6	77	85	91	6	55						BTL1	Remembering
7	18	9	44	2	5	89	91	11	6														
77	85	91	6	55																			

PART-B

1.	Appraise the following with appropriate Python code: (i) Histograms (6 marks) (ii) Binnings (6 marks) (iii) Density (4 marks)	BTL5	Evaluating
2.	How text and image annotations are done using Python? Give an example on your own with appropriate Python code. (16)	BTL3	Applying
3.	Explain about various visualization charts like line plots, scatter plots and histograms using Matplotlib with an example. (16)	BTL4	Analyzing
4.	Evaluate any two three-dimensional plotting in Matplotlib with an example. (16)	BTL5	Evaluating
5.	Analyze the different types of joins in python. (16)	BTL4	Analyzing
6.	Explain various features of Matplotlib platform used for data visualization and illustrate its challenges. (16)	BTL3	Applying
7.	Write a code snippet that projects our globe as a 2-D flat surface (using cylindrical project) and convey information about the location of any three major Indian cities in the map (using scatter plot) (16)	BTL5	Evaluating
8.	(i) Write a working code that performs a simple Gaussian process regression (GPR) using the Scikit-Learn API. (8) (ii) Briefly explain about visualization with Seaborn. Give an example working code segment that represents a 2D kernel density plot for any data. (8)	BTL4	Analyzing
9.	Perform an exploratory data analysis for the following data with different types of plots: The dataset contains cases from a study that was conducted between 1958 and 1970 at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer. (16) Data attributes: Age of patient at the time of operation (numerical) Patient's year of operation (year-1990, numerical) Number of positive axillary nodes detected (numerical) Survival status (class attribute) 1 = the patient survived 5 years or longer, 2 = the patient died within 5 year.	BTL3	Applying
10.	(i) How can you overlay a probability density function (PDF) on a histogram? (8) (ii) Create a histogram and a kernel density estimate (KDE) plot for a given dataset. (8)	BTL3	Applying
11.	(i) Discuss the use of subplots to create multiple plots in a single figure. (8) (ii) How can you customize the layout and spacing of subplots? (8)	BTL4	Analyzing
12.	(i) Explain the role of Statsmodels in statistical modeling and data analysis. (8) (ii) Analyze how can you use Statsmodels to perform linear regression and hypothesis testing? (8)	BTL4	Analyzing
13.	(i) Evaluate how does Seaborn simplify data visualization compared to Matplotlib. (8) (ii) Create a pair plot and a heatmap using Seaborn to visualize relationships between variables in a dataset. (8)	BTL5	Evaluating
14.	(i) Inspect the advantages of using interactive visualization tools like Plotly and Bokeh. (8) (ii) Analyze how can you create interactive line plots and scatter plots using Plotly. (8)	BTL4	Analyzing

15.	(i) (ii)	Explain the concept of interactive dashboards. (8) Evaluate how can you build interactive dashboards using Bokeh to visualize complex datasets and allow users to explore data dynamically? (8)	BTL5	Evaluating
16.	(i) (ii)	How can you visualize time series data using Matplotlib? (8) Explain the importance of time series analysis and how Matplotlib can be used to identify trends, seasonality, and cyclic patterns. (8)	BTL5	Evaluating
17.	(i) (ii)	Discuss the role of Basemap toolkit in visualizing geographic data with Matplotlib. (8) How can you create maps with markers, lines, and polygons to represent geographic features and data? (8)	BTL5	Evaluating