QUESTION BANK

Subject Code &Name : CS3352 FOUNDATIONS OF DATA SCIENCE

Year / Sem : II / III

UNIT I–INTRODUCTION			
Q.No	**PART – A**	**BT Level**	**Competence**
1.	What is data science ?	BTL1	Remembering
2.	What is the role of data science in business, medical research, healthcare, education, social media, technology and financial institutions?	BTL1	Remembering
3.	Write the main types/categories of data?	BTL1	Remembering
4.	What isNLP ? Is natural language structured data?	BTL1	Remembering
5.	What is machine generated data with an example?	BTL1	Remembering
6.	What is graph-based or network data?	BTL2	Understanding
7.	List the steps involved in data science processing?	BTL1	Remembering
8.	What are outliers?	BTL2	Understanding
9.	What are the different ways of combining data?	BTL2	Understanding
10.	What is big data?	BTL2	Understanding
11.	What is data Cleansing?	BTL1	Remembering
12.	Define the term retrieving data.	BTL2	Understanding
13.	Define data exploration process.	BTL2	Understanding
14.	What is data preparation and process?	BTL2	Understanding
15	What is data modeling?	BTL2	Understanding

16	Write about presentation and automation process.	BTL2	Understanding
17	List the V's of big data	BTL1	Remembering
18	Identify the components of data science	BTL1	Remembering
19	List few applications of data science	BTL1	Remembering
20	Enumerate the categories of data used in data science	BTL1	Remembering
21	Give example of unstructured data science	BTL1	Remembering
22	Mention the significance of setting goals in data science project	BTL2	Understanding
23	What is project charter?	BTL2	Understanding
24	Identify the importance of project charter	BTL2	Understanding
25	Define data warehousing, data mart and data lake	BTL1	Remembering
26	Mention any 4 common errors that occur in data	BTL3	Understanding
27	How can you handle missing values in dataset?	BTL1	Remembering
28	What are the implications of erroneous data for analysis?	BTL2	Understanding
29	Give example for any 3 type of virtualization methods used for data exploration.	BTL2	Understanding
30	Mention any 3 types of data science models	BTL2	Understanding
31	What is confusion matrix?	BTL2	Understanding
32	List the common evaluation metrics used to measure the performance of models.	BTL2	Understanding
33	Differentiate between data science and data mining.	BTL2	Understanding

PART– B			
1.	Explain the benefits of data science.	BTL1	Remembering
2.	List the facets of data with example	BTL1	Remembering
3.	Briefly explain the steps in data science process with diagram	BTL1	Remembering
4.	Briefly explain the architecture of data mining	BTL1	Remembering
5.	Briefly explain the architecture of data warehousing	BTL2	Understanding
6.	How do you set the research goal, retrieving data and data preparation process in data science process?	BTL2	Understanding

UNIT II-DESCRIBING DATA			
Q.No	**Question**	**BT Level**	**Competence**
PART – A			
1.	What is frequency distribution?	BTL1	Remembering
2.	what are the types and uses of frequency distribution?	BTL1	Remembering
3.	what is grouped frequency distribution?	BTL1	Remembering
4.	What is the ungrouped frequency distribution?	BTL3	Applying
5.	What is cumulative frequency distribution?	BTL1	Remembering
6.	What is relative frequency distribution?	BTL1	Remembering
7.	Define percentile ranks?	BTL2	Understanding
8.	What is a histogram?	BTL2	Understanding
9.	Explain any three features of histogram?	BTL2	Understanding

10.	What is frequency polygon?	BTL2	Understanding
11.	What if distribution have more than one mode or no mode at all?	BTL2	Understanding
12.	Explain range, variance and SD ?	BTL1	Remembering
13.	What is degree of freedom?	BTL2	Understanding
14.	What is interquartile range (IQR)?	BTL3	Applying
15.	Define normal curve and its property?	BTL2	Understanding
16.	What is z-score?	BTL2	Understanding
17.	Define data. What are the types of data	BTL1	Remembering
18.	What is qualitative data. Give an example	BTL1	Remembering
19.	What is quantitative data. Give an example	BTL1	Remembering
20.	Compare discrete and continuous variables	BTL3	Applying
21.	Differentiate between bar graph and a histogram	BTL1	Remembering
22.	What are the measures of central tendency?	BTL1	Remembering
23.	What is positively skewed distribution?	BTL2	Understanding
24.	What is negatively skewed distribution?	BTL2	Understanding
25.	What is a normal curve?	BTL2	Understanding
25.	How will you convert a z-score to original score?	BTL2	Understanding

	PART – B		
1.	Explain the types of frequency distribution with examples.	BTL3	Apply
2.	Describe mean median mode and averages with example	BTL3	Apply
3.	Specify the real limits for the lowest class interval in this frequency distribution for the given dataset 91 85 84 79 80 87 96 75 86 104 95 71 105 90 77 123 80 100 93 108 98 69 99 95 90 110 109 94 100 103 112 90 90 98 89	BTL3	Apply
4.	Analyze how graphs are used to represent qualitative and quantitative data?	BTL3	Apply
5.	Generate the grouped and ungrouped frequency table for the following data 90,92,87,88,87,92,98,90,90,87,87,88,88,89,90,87,89,92,92,92,98,90,9 5,87,87 i)How many people scored 98? ii)How many people scored 90 or less? iii)What proportion scored 87?	BTL3	Apply
6.	(i) Calculate the sum of square population standard deviation of the given x data value 13,10,11,7,9,11,9 (ii) Calculate the sample standard deviation for the given data 7,3,1,0,4	BTL3	Apply

7.	Suppose the IQ score have a bell shaped distribution with a mean of 100 and standard deviation of 15 then calculate the following : (i) what percentage of people should have an IQ score between 85 and 115? (ii) what percentage of people should have an IQ score between 70 and 130?	BTL3	Apply
	(iii) what percentage of people should have an IQ score more than 130? (iv) A person with an IQ score greater than 145 is considered a genius. Does the empirical rule support this statement		

UNIT III-DESCRIBING RELATIONSHIPS

Q.No	Question	BT Level	Competence
	PART - A		
1.	What is correlation and its types?	BTL1	Remembering
2.	Define Scatterplots?	BTL4	Understanding
3.	What is a correlation coefficient?	BTL1	Remembering
4.	Define Regression.	BTL1	Remembering
5.	Write the types of regression analysis.	BTL1	Remembering
6.	Define single and multiple linear regression.	BTL1	Remembering
7.	What is ridge regression?	BTL2	Understanding
8.	What is decision tree?	BTL3	Understanding
9.	What is the need for correlation?	BTL1	Remembering
10.	What is causation?	BTL4	Understanding
11.	What is linear relationship and non-linear relationship?	BTL1	Remembering
12.	List the types of nonlinear relationship	BTL1	Remembering

13.	What is curvilinear relationship	BTL1	Remembering
14.	What are the key properties of correlation coefficient r ?	BTL1	Remembering
15.	Compare correlation and regression	BTL2	Understanding
16.	What is restricted range?	BTL3	Understanding
17.	What is interpretation of r2 ?	BTL1	Remembering
18.	What is regression towards mean?	BTL4	Understanding
19.	When does regression fallacy occur?	BTL1	Remembering
	PART – B		
1.	How correlation coefficient can be calculated for the quantitative data?	BTL2	Understanding
2.	Explain the different types of regression analysis in detail.	BTL2	Understanding
3.	Briefly explain in detail about the standard error of estimation	BTL2	Understanding
4.	Calculate the value of r using computation formula for the following data	BTL 3	Apply

FRIENDS	SENT	RECEIVED
Dories	13	14
Steve	9	18
Mike	7	12
Andrea	5	10
John	1	6

5.	How correlation coefficient can be calculated for the quantitative data?	BTL 3	Apply

UNIT IV – PYTHON LIBRARIES FOR DATA WRANGLING

Q.No	Question	BT Level	Competence
	PART - A		
1.	What is NumPy in python used for?	BTL2	Understanding
2.	Write a python library create an array?	BTL2	Understanding
3.	Write the output for the following numpy code? (i) np.array([3,14,4,2,3]) (ii) np.array([1,2,3,4],dtype='float32') (iii) np.array([range(i,i+3) for i in [2,4,6]]) (iv) np.zeros(10,dtype=int) (v) np.ones((3,5), dtype=float) (vi) np.full((3,5),3.14) (vii) np.arrange(0,20,20) (viii) np.linespace(0,1,50 (ix) np.random.random((3,3)) (x) np.random.normal(0,1,(3,3))	BTL2	Understanding
4.	Define series object.	BTL2	Understanding
5.	What is Data frame?	BTL2	Understanding
6.	How a pandas data frame can be constructed?	BTL2	Understanding
7.	What are indexers?	BTL2	Understanding
8.	How missing data can be handled in python?	BTL2	Understanding
9.	How the operations can be performed on null values in pandas data science?	BTL2	Understanding
10.	Define Hierarchical indexing.	BTL2	Understanding
11.	What is pivot table?	BTL2	Understanding
12.	Identify the details maintained by python to store an integer	BTL2	Understanding
13.	Write python code to create 1D,2D and 3D numpy arrays.	BTL2	Understanding
14.	How do you verify the shape of 1D, 2D and 3D/ND array respectively?	BTL2	Understanding

15.	Compare python list with arrays	BTL2	Understanding
16.	Write short note on python array object	BTL2	Understanding
17.	How to perform slicing to access the elements of numpy arrays	BTL2	Understanding
18.	List is mutable. Justify this with suitable example.	BTL2	Understanding
19.	What is indexing and negative indexing in tuple.	BTL2	Understanding
20.	Write the list of aggregate functions of numpy	BTL2	Understanding
21.	What is fancy indexing?	BTL2	Understanding
22.	Write short note on pandas.	BTL2	Understanding
23.	Explain reindexing in pandas.	BTL2	Understanding
24.	What is universal function?	BTL2	Understanding
	PART – B		
1.	Briefly explain the basics of numpy arrays with example	BTL 3	Apply
2.	Describe about fancy indexing with an example.	BTL 3	Apply
3.	Explain structured data in a numpy array.	BTL2	Understanding
4.	What is a universal function? Explain clearly each function with examples.	BTL 3	Apply
5.	Explain aggregate function with example	BTL 3	Apply
6.	What is broadcasting and explain rules with examples	BTL 3	Apply
7.	Explain data objects in pandas.	BTL1	Remembering
8.	Briefly explain the hierarchical indexing with examples	BTL 3	Apply
9.	What is a pivot table? Explain in detail.	BTL1	Remembering

UNIT V – DATA VISUALIZATION

Q.No	Question	BT Level	Competence
	PART - A		
1.	What is purpose of matplotlib?	BTL2	Understanding
2.	Write the dual interface of matplotlib?	BTL2	Understanding
3.	How to draw a simple line plot using matplotlib?	BTL2	Understanding
4.	What functions can be used to draw scatter plots?	BTL2	Understanding
5.	Write the difference between plot and scatter functions?	BTL2	Understanding
6.	Define contour plots?	BTL2	Understanding
7.	What functions can be used to draw contour plots?	BTL2	Understanding
8.	What is the purpose of histogram?	BTL2	Understanding
9.	Write a source code to draw a simple histogram	BTL2	Understanding
10.	How to create a 3-D wireframe plot?	BTL2	Understanding
11.	Define surface plot?	BTL2	Understanding
12.	What is the use of seaborn?	BTL2	Understanding
13.	What is pair plots?	BTL2	Understanding
14.	What is density plot?	BTL2	Understanding
15.	Mention the significance of subplots?	BTL2	Understanding
16.	Brief on basemap tool kit.	BTL2	Understanding
17.	Write python code to plot sine and cos wave.	BTL2	Understanding
18.	How can you set different colors for line plot.	BTL2	Understanding
19.	List the applications of lineplot.	BTL2	Understanding
20.	Write the syntax of scatter() method.	BTL2	Understanding

PART – B			
1.	What is matplotlib? Specify the two interfaces used by it.	BTL1	Remembering
2.	Briefly explain about the line plot and scatter plot.	BTL1	Remembering
3.	Explain contour plot and histogram.	BTL2	Understanding
4.	What is 3D plotting? Explain it with examples.	BTL 3	Apply
5.	How graphical data can be projected using matplotlib? Explain with examples.	BTL 3	Apply