P P SAVANI UNIVERSITY

Seven Semester of B. Tech. Examination

December 2022

SEIT4013: Data Science

17.11.2022, Thursday **Time: 10:00 a.m. To 12:30 p.m.** **Maximum Marks: 60**

Instructions:

1. The question paper comprises of two sections.
2. Section I and II must be attempted in separate answer sheets.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculator is allowed.

SECTION – I

Q - 1	MCQ/Short Question/Fill in the Blanks (Any Five)	**[05]**	**CO**	**BTL**
(i)	Name two python packages used for handling big data.		**1**	**1**
(ii)	Name two tests that can be used in case of a small sample (n≤30).		**1**	**1**
(iii)	Write the python command for finding the descriptive analysis of the data set.		**1**	**1**
(iv)	What will be the outcome of following command executed on iris dataframe? iris[iris.species > 1]		**1**	**1**
(v)	What will be the outcome of following command executed on df dataframe? df[~df.name.str.contains('y')]		**1**	**1**
Q - 2 (a)	What are various data types used in data science and the operations that can be performed on them.	**[05]**	**2**	**2**
Q - 2 (b)	Explain the concept of 1-hot encoding and dummy encoding for nominal data type.	**[05]**	**3**	**2**
	OR			
Q - 2 (a)	Describe various non-graphical or statistical approaches used for exploratory data analysis.	**[05]**	**3**	**2**
Q - 2 (b)	Explain the Anova Hypothesis test with example. What are its merits?	**[05]**	**2**	**2**
Q - 3 (a)	How can outliers be treated. Explain the process based on Inter Quartile Range(IQR) value.	**[05]**	**2**	**2**
Q - 3 (b)	Perform the following methods on a dataframe named as df. What will be their effect? Assume whatever is necessary. Isnull(), replace(), dropna(), fillna(), interpolate()	**[05]**	**2**	**3**
	OR			
Q - 3 (a)	Mention the right type of chart for the following visualizations: 1. Trends / Change over time 2. Distributions 3. Relationships 4. Fraction of whole Comparison of the values	**[05]**	**2**	**2**
Q - 3 (b)	Why the boxplot is known as 5- number summary plot. Explain the analytics obtained through boxplot.	**[05]**	**3**	**3**

Q - 4	Attempt any one/two.	**[05]**		
(i)	Create a 3x4 dimensional array of 12 values having some repeated values. Perform following operations on it: a) Find the median of array b) Change the shape of the array to 2x6 array c) Display 3rd and 4th element of 2nd row d) Find standard deviation of array Find variance of array		**3**	**4**
(ii)	Explain the merge command required to merge two dataframes. What are various parameters used in the command. What is the significance of inner, outer, left and right values of how parameter?		**3**	**4**

SECTION – II

Q - 1	**Answer the Following:**	**[05]**		
(i)	Which of the following thing can be data in Pandas? a) a python dict b) an ndarray c) a scalar value d) all of the mentioned		**1**	**2**
(ii)	Which of the following compute proportions from a contingency table? a) par() b) prop. Table() c) anova() d) mosaicplot()		**2**	**1**
(iii)	Which of the following finds the position of a quantile in a dataset? a) quantile() b) barplot() c) barchart() d) rep()		**1**	**2**
(iv)	Data can be visualized using? A. graphs B. charts C. maps D. All of the above		**1**	**2**
(v)	Which of the following groups find the correlation matrix? a) factor.model b) col.max(x) c) stem d) which.max(x)		**1**	**2**
Q - 2 (a)	Enlist and explain various applications of data science.	**[05]**	**1**	**1**
Q - 2 (b)	What are basic scales of measurement? Explain their types with example.	**[05]**	**2**	**1**

OR

Q - 2 (a)	**Answer in Short.** • What do you mean by data encoding? • Differentiate Probability vs. Statistics • What is the difference between Data Scientist and Data Analytics? • Enlist two python libraries.	**[05]**	**3**	**3**

- What do you mean by visual encoding?

Q - 2 (b)	Explain any one data analysis technique with an example.	**[05]**	**3**	**2**
Q - 3 (a)	Enlist and explain any one data encoding techniques.	**[05]**	**2**	**2**
Q - 3 (b)	Explain any two applications of data science.	**[05]**	**2**	**1**
	OR			
Q - 3 (b)	Explain retinal variables with suitable example.	**[05]**	**2**	**2**
Q - 4	**Answer the Following:**	**[05]**		
(i)	What do you mean by data visualization technology? Explain any one in detail.		**2**	**1**

****

CO : Course Outcome Number BTL : Blooms Taxonomy Level

Level of Bloom's Revised Taxonomy in Assessment

1: Remember	2: Understand	3: Apply
4: Analyze	5: Evaluate	6: Create