Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering

End Sem (Odd) Examination Dec-2019

CS3ED06 / IT3ED07 Data Science

Programme: B.Tech. Branch/Specialisation: CS/IT

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

Q.1 i. Point out the correct statement: **1**

(a) Raw data is original source of data

(b) Pre-processed data is original source of data

(c) Raw data is the data obtained after processing steps

(d) None of these

ii. Data Scientist should have the skills of: **1**

(a) Computer Science (b) Statistics

(c) Both (a) and (b) (d) None of these

iii. The joint probability is: **1**

(a) The likelihood of two events happening together

(b) The likelihood of an event happening given that another event has already happened

(c) Based on two mutually exclusive events

(d) Also called Prior probability

iv. A listing of the possible outcomes of an experiment and their corresponding probability is called **1**

(a) Random Variable (b) Contingency table

(c) Bayesian Table (d) Probability Distribution

v. A histogram: **1**

(a) Is a graphic representation of the frequency distribution of a continuous variable

(b) Is a graphic representation of the frequency distribution of a qualitative or categorical variable

(c) Is an alternative to a pie chart

(d) Is a bar chart

P.T.O.

vi. Which of the following information is not given by five-number summary? **1**
(a) Mean (b) Median (c) Mode (d) All of these

vii. Which plot might you use to show the relationship between attitudes towards exercise and physical fitness levels? **1**
(a) Box and whisker plot. (b) Stem and leaf plot.
(c) Scatter plot (d) Histogram

viii. Which of the following graph can be used for simple summarization of data? **1**
(a) Scatter plot (b) Word cloud
(c) Bar plot (d) All of these

ix. Which one is the python library? **1**
(a) Scikit-Learn (b) PyBrain
(c) NumPy (d) All of these

x. Which of these libraries contains a lot of efficient tools for machine learning and statistical modelling? **1**
(a) Scikit-Learn (b) SciPy
(c) NumPy (d) Matplotlib

Q.2 i. What are different skills required for data scientist? Give four points. **2**

ii. How data science is important in today's business world? **3**

iii. What are primary components of data science? Discuss working of every component. **5**

OR iv. Discuss about basic framework of data science in detail. **5**

Q.3 i. What is the conditional probability in statistics? Give example. **2**

ii. Explain descriptive and predictive analytics. Also give name of primary tools used in descriptive and predictive analytics. **8**

OR iii. Discuss about statistical modelling and statistical inference in detail. **8**

Q.4 i. What is the importance of exploratory data analysis in data science? Give three points. **3**

ii. Explain any four tools of exploratory data analysis. Describe box plot with its five-number summary. **7**

OR iii. Explain any case study using data science process. **7**

Q.5 i. What is data visualization? Why it is important in data science? Give two reasons. **4**

ii. What are the essential points of effective data visualization? Explain each in brief. **6**

OR iii. Explain any three comparison charts in data visualization with diagram. **6**

Q.6 Attempt any two:

i. Explain following python libraries with example: **5**
(a) Scikit-Learn (b) Matplotlib

ii. How python is useful in data science process? Give any four characteristics in detail. **5**

iii. Describe challenges of Data Science project management **5**

******

Marking Scheme
CS3ED06 / IT3ED07 Data Science

Q.1	i.	Point out the correct statement: (a) Raw data is original source of data	**1**
	ii.	Data Scientist should have the skills of: (c) Both (a) and (b)	**1**
	iii.	The joint probability is: (a) The likelihood of two events happening together	**1**
	iv.	A listing of the possible outcomes of an experiment and their corresponding probability is called (d) Probability Distribution	**1**
	v.	A histogram: (a) Is a graphic representation of the frequency distribution of a continuous variable	**1**
	vi.	Which of the following information is not given by five-number summary? (c) Mode	**1**
	vii.	Which plot might you use to show the relationship between attitudes towards exercise and physical fitness levels? (c) Scatter plot	**1**
	viii.	Which of the following graph can be used for simple summarization of data? (c) Bar plot	**1**
	ix.	Which one is the python library? (d) All of these	**1**
	x.	Which of these libraries contains a lot of efficient tools for machine learning and statistical modelling? (a) Scikit-Learn	**1**
Q.2	i.	Different skills required for data scientist(four points.) (0.5 mark*4)	**2**
	ii.	Each point has one mark which should describe importance (3 points) (1 mark*3)	**3**
	iii.	Minimum five Components (1 mark*5)	**5**
OR	iv.	Diagram of framework 1 mark Detail of framework 4 marks .	**5**
Q.3	i.	Description 1 mark Example 1 mark	**2**
	ii.	Description of descriptive analytics 3 marks Description of predictive analytics 3 marks. Name of tools of descriptive 1 mark Name of tools of predictive 1 mark	**8**
OR	iii.	Description of statistical modelling 4 marks Description of statistical inference 4 marks	**8**
Q.4	i.	Importance of exploratory data analysis in data science (Three points.) (1 mark*3)	**3**
	ii.	Four tools should be described in detail 4 marks Description of Box plot with diagram 3 marks	**7**
OR	iii.	Description of case study using data science process. (As per answer) 7 marks	**7**
Q.5	i.	Description of data visualization 2 marks Description of data visualization 2 marks	**4**
	ii.	Essential points of effective data visualization Each point in brief (1 mark*6)	**6**
OR	iii.	Three comparison charts in data visualization with diagram. (2 marks*3)	**6**
Q.6		Attempt any two:	
	i.	Explain following python libraries with example: (a) Scikit-Learn Description 1 mark Example 1.5 mark (b) Matplotlib Description 1 mark Example 1.5 mark	**5**
	ii.	Uses of python 1 mark Four characteristics in detail (1 mark*4)	**5**
	iii.	Description of 5 challenges (1 mark*5)	**5**

******