

P P SAVANI UNIVERSITY

Third Semester of M.Sc.DS Examination

December 2022

SSDM8021 Data Analytics

05.01.2023, Thursday

Time: 10:00 a.m. To 12:30 p.m.

Maximum Marks: 60

Instructions:

1. The question paper comprises of two sections.
2. Make suitable assumptions and draw neat figures wherever required.
3. Use of scientific calculator is allowed.

SECTION - I

- Q - 1** MCQ/Short Question/Fill in the Blanks [05] CO BTL
- (i) Data Analytics uses _____ to get insights from data. CO1 2
- (ii) Linear regression analysis is used to predict the value of a variable based on the value of another variable. (True/False) CO3 2
- (iii) _____ are used when we want to visually examine the relationship between two quantitative variables. CO1 2
- (iv) What is a hypothesis? CO2 1
- (v) Tableau is a _____ tool. CO4 2
- a) Visualization
 - b) Analytical
 - c) Data Exploration
 - d) All of the mentioned above
- Q - 2 (a)** Explain different types of parameters of hypothesis testing. [05] CO2 2
- Q - 2 (b)** Elaborate the following terms with example, if needed [05] CO2 3
- a) Paired T test
 - b) One-tailed T test
 - c) T test formula
 - d) Multivariate ANOVA
 - e) p-value

OR

- Q - 2 (a)** Short Note on Analysis of Variance, explain with example. [05] CO2 4
- Q - 2 (b)** Explain Types of Statistics in detail. [05] CO2 3
- Q - 3 (a)** Elaborate Process of Data Analytics in your own words. [05] CO1 3
- Q - 3 (b)** Why is big data analytics important? [05] CO1 3
- Q - 4** Attempt anyone. [05]
- (i) Explain Elements of Data Analytics. CO1 2
- (ii) Explain Level of Measurement. CO1 2

SECTION - II

- Q - 1** MCQ/Short Question/Fill in the Blanks (Any Five) [05]
- (i) Amongst which of the following is / are the measures of central tendency - CO3 2
- a) Mean
 - b) Median
 - c) Mode
 - d) All of the mentioned above
- (ii) A data set is a distribution of n number of scores or values. (True/False) CO3 2
- (iii) What is F-test? CO3 1

- (iv) If the null hypothesis is false then which of the following is accepted? C04 3
- Null Hypothesis
 - Positive Hypothesis
 - Negative Hypothesis
 - Alternative Hypothesis
- (v) Movie Recommendation systems are an example of C04 4
- Classification
 - Clustering
 - Reinforcement Learning
 - Regression
- 2 only
 - 1 and 3
 - 1 and 2
 - 2 and 3
- Q - 2 (a) Explain in detail Multiple Linear Equation. [05] C03 1
- Q - 2 (b) State down Requirements of clustering in data mining in detail. [05] C03 3
- Q - 3 (a) How Hierarchical Method is classified and useful in clustering? [05] C04 2
- Q - 3 (b) Draw and explain Experiment Design Process. [05] C04 4
- OR**
- Q - 3 (a) Write down the steps to implement Fishbone Diagram. [05] C04 3
- Q - 3 (b) What is the usage of Box and Whisker Plot, explain with figure or example. [05] C04 3
- Q - 4 Attempt any one [05]
- (i) Find the missing value of Weight using KNN for the following given data sample. C03 5

ID	Height	Age	Weight
1	5	45	77
2	5.11	26	47
3	5.6	30	55
4	5.9	34	59
5	4.8	40	72
6	5.8	36	60
7	5.3	19	40
8	5.8	28	60
9	5.5	23	45
10	5.6	32	58
11	5.5	38	?

- (ii) Find the frequent item sets and generate association rules on following data. C03 5
 Assume that minimum support threshold as 2 and minimum confident threshold is 60%

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

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