Academic Tasks (22231)

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

Academic Task Number: 2

Course code: CAP447 (Section D2215 – G1)

Date of allotment: 14.10.2022

Course title: Data Warehousing and data mining Laboratory

Date of submission: 14.10.2022

Maximum Marks: 50

Academic Task Type: Practical

Question Number	Question Statement	Course Outcome	Bloom's level	Marks per Question
Q1 **(SET 1)**	**Here we are looking to find the missing values and replace the missing values** **Dataset: Mushroom Dataset** **To do the following on the dataset:** a) Import the dataset using operator b) Analyze the dataset (instances, attributes) c) Show the missing attributes from the dataset using operator d) Show the no missing attributes from the dataset using operator	CO1	L1: Remember	15

Academic Tasks (22231)

	e) Replace the single attribute missing value **(Cap-surface)** using mean function f) Replace the all the missing values using mean function g) Rename any attribute using operator h) Select the following attributes only from the dataset (cap-diameter, cap-shape, cap-surface, cap-color) i) Change the role of class attribute into prediction role j) Remove duplicate values from class attribute k) Apply min-max normalization on single, subset and all attributes l) Apply sampling preprocessing (use only absolute and probability)			
Q2 **(SET 1)**	**Dataset: Titanic** **Link: Internal dataset** a) Analyze the dataset (instances, attributes) b) Show the missing attributes from the dataset using		L3:	

Academic Tasks (22231)

			CO2	Apply	15
	operator				

c) Replace the single attribute missing value **(Age)** using mean function

d) Replace the all the missing values using max function

e) Rename the **Age** attribute into **Passenger Age**

f) Select the following attributes only from the dataset

   (Age, Cabin, Passenger fare and Life Boat)

g) Use replacement functions for each column (use column entry)

h) Change the role of Passenger class attribute into prediction role

i) Apply min-max normalization (min:1; max: 10) on single, subset and all attribute

j) Remove duplicate values from Passenger class attribute

k) Sort the single attribute values (Ascending)

l) Apply sampling preprocessing (use only absolute and probability)

Academic Tasks (22231)

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

	m)			

Q1
 	
(SET 2)
	Dataset: Global superstore 2017
	
	Link: LPULive (Already shared)
	
	   a) import the dataset using operator
	
	   b) Analyze the dataset (instances, attributes)
	
	   c) Show the missing attributes from the dataset using operator
	
	   d) Show the no missing attributes from the dataset using operator
	
	   e) Replace the missing value using mean function
	
	   f) Rename the **Region** attribute into **decision class**
	
	   g) Select the following attributes only from the dataset
	
	(Market, Region, Category and Product name)
	
	   h) Change the role of decision class attribute into prediction role
		CO3	L6:Create	15

Academic Tasks (22231)

	i) Apply min-max normalization on single attribute, subset and all j) Sort the single attribute values (Descending order) k) Apply sampling preprocessing (use only absolute and probability) l) Visualize the dataset			
Q2 **(SET 2)**	**Dataset: melb_data** a) import the dataset using operator b) Show the missing attributes from the dataset using operator c) Show the no missing attributes from the dataset using operator d) Replace the single attribute missing value **(BuildingArea)** using mean function e) Replace only any two attribute missing values using min function			15

Academic Tasks (22231)

	f) Replace the all the missing values using max function g) Rename the **BuildingArea** attribute into **AreaBuilding** h) Select the following attributes only from the dataset (BuildingArea, YearBuilt and CouncilArea) i) Change the role of Building Area attribute into prediction role j) Apply min-max normalization on single, subset and all attribute k) Remove duplicate values from AreaBuilding attribute l) Apply sampling preprocessing (use only absolute and probability)			