Academic Tasks (22231)

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

Academic Task Number: 1

Course code: CAP447 (Section D2215 – G1)

Date of allotment: 30.09.2022

Course title: Data Warehousing and data mining Laboratory

Date of submission: 30.09.2022

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** **Link:** **https://archive.ics.uci.edu/ml/datasets/Secondary+Mushroom +Dataset** **To do the following on the dataset:** a) Create a new repository like DWDM22 b) Download a dataset from given the link and import into the	CO1	L1: Remember	15

Academic Tasks (22231)

	new repository (import only primary dataset) c) Analyze the dataset (instances, attributes) d) Show the missing attributes from the dataset using operator e) Show the no missing attributes from the dataset using operator f) Replace the single attribute missing value **(Cap-surface)** using mean function g) Replace only any five attribute missing values using mean function h) Replace the all the missing values using mean function i) Rename the **class** attribute into **target** j) Select the following attributes only from the dataset (cap-diameter, cap-shape, cap-color)			
Q2 **(SET 1)**	**Dataset: Titanic** **Link: Internal dataset** a) Analyze the dataset (instances, attributes)			

Academic Tasks (22231)

LOVELY
PROFESSIONAL
UNIVERSITY

Transforming Education Transforming India

	b) Show the missing attributes from the dataset using operator			
	c) Show the no missing attributes from the dataset using operator	CO2	L3: Apply	15
	d) Replace the single attribute missing value **(Cabin)** using mean function			
	e) Replace only any three attribute missing values using mean function			
	f) Replace the all the missing values using min function			
	g) Rename the Life Boat attribute into **Passenger**			
	h) Select the following attributes only from the dataset (Age, Cabin, and Passenger fare)			
	i) Use replacement functions for each column (use column entry)			
		CO3	L6:Creat	15

Academic Tasks (22231)

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

Q1 (SET 2)	**Dataset: Global superstore 2017** **Link: LPULive (Already shared)**		e	

Dataset: Global superstore 2017

Link: LPULive (Already shared)

a) Create a new repository like DWDM22

b) Download a dataset from given the link and import into the repository

c) Analyze the dataset (instances, attributes)

d) Show the missing attributes from the dataset using operator

e) Show the no missing attributes from the dataset using operator

f) Replace the missing value using max function

g) Rename the **Market** attribute into **Target**

h) Select the following attributes only from the dataset

(Market, Region, Category and Product name)

i) Visualize the dataset (any visualization)

Academic Tasks (22231)

Q2 **(SET 2)**	**Dataset: melb_data** **Link: https://www.kaggle.com/datasets/gunjanpathak/melb-data** a) Create a new repository like DWDMMCA22 b) Download a dataset from given the link and import into the 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 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) Use replacement functions for each column (use column			15

	entry)			