VIT-AP UNIVERSITY

Apply Knowledge. Improve Life!

QUESTION PAPER

Name of the Examination: FALL 2023-2024 – CAT-2

Course Code: CSE1006 **Course Title: Foundations for Data Analytics**

Set number: 1 **Date of Exam:** 19/10/2023 (FN) (E1)

Duration: 90 Minutes **Total Marks: 50**

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. Create the following **dataframe (df)** and answer the following question. **(15M)**

Roll.no	Name	Percentage	color	maths	phy	chem
1	hemant	86.5	black	50	60	70
2	prashant	91.5	red	45	67	89
3	eshwar	93.27	orange	58	NA	39
4	vishvesh	99.5	red	47	98	88
5	rohit	NA	black	87	48	90
6	soham	67.666	red	56	67	49
7	srijay	98.9	black	100	95	99
8	ajay	NA	orange	88	49	85

(a) Write a command to print non-duplicated records of dataframe.

(b) Write command to fill the missing values in the column **"phy"** with the mean of remaining values in the column **"phy"**.

(c) Write a command to print the number of **"NA"** in **"percentage"** variable of a dataframe.

(d) What is the output of the following command: **df[complete.cases(df),]**

(e) What is the output of the following command: **is.na(df$Percentage)**.

Q2. Write a R program to create the data frame given below and perform the following join operation. **(10M)**

"PERSON"

Id	First_name	Last_name	M_id
1	Adam	Smith	1
2	Ravi	Kumar	2
3	Susan	Davidson	5
4	Jenny	Adrianna	8
6	Lee	Pong	10

"MOVIE"

M_id	Title	category
1	Real Steel	Animations
2	The adventures of tintin	Animations
3	Jungle Book	Animations
4	Inside Out	Animations
5	Unforgiven	Drama
6	Superman	Action
7	Terminator	Action
8	Insidious	Horror
9	Cyrus	Comedy
10	Pi	Horror

(a) Write a command to perform inner join and write the output.
(b) Write a command to perform left outer join and write the output.
(c) Write a command to perform cross join and write the output.

Q3. Create a dataframe for the following observations and write as Emp.txt file. Read the data from Emp.txt file and answer the following questions using the functions available in the package "dplyr". **(15M)**

Name	Age	Address	Salary	Position
Ram	25	Delhi	20000	Programmer
Hardik	29	Kota	35000	Manager
Krishna	36	Amaravati	19000	Analyst
Shiva	54	Bhopal	34000	Programmer
Pavani	43	Amaravati	52000	Analyst
Suresh	34	Mumbai	60000	Manager
Alice	53	Pune	65000	Analyst

(a) Write a command to create a new column "Salary_year" which represents the salaryamount for a year. The existing column "Salary" indicates the salary amount in rupees for one month.

(b) Write a command to display the rows of employee dataset where employee's **Age** is greater than 30 and not living in Amaravati.

(c) Write a command to display the all information of employees whose **Position** is Analyst and salary is greater than 30000.

(d) Write a command to display the all information of employees whose **Salary** is greater than 30000 and less than 60000.

Q4. Write R program to do exploratory analysis using plots for given scenario. There are **98** students in a class, following table shows the obtained **Grade** by the number of student. Create a Data frame name as **Result.** Calculate **Percentage** of students whom got the respective grades and plot a Pie chart with **Percentage** of student for **Grade** with different colours. **(10M)**

Grade	No. of students
S	11
A	20
B	35
C	15
D	10
F	7

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	3	2	1, 2, 3	1, 2, 3, 4	1,3	15
Q2	3	2	1, 2, 3	1, 2, 3, 4	1, 3	10
Q3	4	3	1, 2, 3	1, 2, 3, 4	2,3	15
Q4	4	3	1, 2, 3	1, 2, 3, 4	2,3	10