MANICALAND STATE UNIVERSITY OF APPLIED SCIENCES

FACULTY OF ENGINEERING, APPLIED SCIENCE AND TECHNOLOGY

DEPARTMENT: COMPUTER SCIENCE AND INFORMATION SYSTEMS

MODULE: DATA WAREHOUSING AND MINING

CODE: INSY412

SESSIONAL EXAMINATIONS
DECEMBER 2023

DURATION: 3 HOURS

EXAMINER: MS G MUTIPFORO

INSTRUCTIONS

1. *Answer **all** questions in **Section A***
2. *Answer **any three** questions in Section B*
3. *Start a new question on a fresh page*

Total marks 100

Section A, compulsory

Answer all questions in this section

Question 1

Consider the market basket transactions shown in the following table.

Transaction ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

Assume that min_support=40% and min_onfidence=70%. Further, assume that the Apriori algorithm is used to discover strong association rules among transaction items.

a) Show the candidate and the frequent itemsets or each database scan.

(20)

b) Generate all possible association rules from the frequent itemsets obtained in   the previous question. Calculate the confidence of each rule and identify all the strong association rules.

(10)

c) Describe two scenarios that would warranty the use of a data warehouse.

(10)

Section B
Answer any three questions from this section

Question 2

a) With the aid of a diagram, explain the three-tier architecture of a Data warehouse.   (10)

b) Illustrate the difference between a dependent and independent Data Mart. Indicate the ETL process on your diagram.   (10)

Question 3

a) Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Draw a start schema diagram for the data warehouse. (10)

b) Outline and explain the steps involved in data mining process. (10)

Question 4

a) Explore four reasons why organizations choose to use data marts? (12)

b) Explain the steps involved in the ETL process and how they contribute to preparing data for analysis. (8)

Question 5

a) Differentiate between operational systems and data warehousing systems. (10)

b) Explain the data warehouse model you are familiar with. (10)

END OF EXAMINATION