Time: 3 Hours   **Marks: 80**

Note: 1. Question 1 is compulsory

   2. Answer any three out of remaining questions.

Q1 A) What is dimensional modelling? Design the data warehouse for wholesale furniture [10] Company. The data warehouse has to allow analysing the company's situation at least with respect to the Furniture, Customer and Time. More ever, the company needs to analyse: The furniture with respect to its type, category and material. The customers with respect to their spatial location, by considering at least cities, regions and states. The company is interested in learning the quantity, income and discount of its sales.

B) Discuss different steps involved in Data Pre-processing.   [10]

Q2 A) The college wants to record the Marks for the courses completed by students using [10] the dimensions: i) Course, ii) Student, iii) Time & a measure Aggregate marks. Create a Cube and describe following OLAP operations:
(i) Slice (ii) Dice (iii) Roll up (iv) Drill down (v) Pivot

B) Apply the Naive Bayes classifier algorithm for buys computer classification and [10] classify the tuple X=(age="young", income="medium", student="yes" and credit-rating="fair")

Id	Age	Income	Student	Credit-rating	buys computer
1	young	high	no	fair	no
2	young	high	no	good	no
3	middle	high	no	fair	yes
4	old	medium	no	fair	yes
5	old	low	yes	fair	yes
6	old	low	yes	good	no
7	middle	low	yes	good	yes
8	young	medium	no	fair	no
9	young	low	yes	fair	yes
10	old	medium	yes	fair	yes
11	young	medium	yes	good	yes
12	middle	medium	no	good	yes
13	middle	high	yes	fair	yes
14	old	medium	no	good	no

Q3 A) Explain ETL of data warehousing in details?   [10]

   B) Explain types of attributes and data visualization for data exploration.   [10]

Q4 A) Illustrate the architecture of Data Warehouse system. Differentiate Data warehouse [10] and Data Mart

B) Explain K-Means clustering algorithm? Apply K-Means algorithms for the [10] following Data set with two clusters.
Data Set = { 15,15,16,19,19,20,20,21,22,28,35,40,41,42,43,44,60,61,65}

Q5 A) Explain Updates to dimension tables in detail. [10]

B) A database has ten transactions. Let minimum support = 30% and minimum [10] Confidence = 70%
i] Find all frequent patterns using Apriori Algorithm.
ii] List strong association rules.

Transaction _Id	Items
01	A,B,C,D
02	A,B,C,D,E,G
03	A,C,G,H,K
04	B,C,D,E,K
05	D,E,F,H,L
06	A,B,C,D,L
07	B,I,E,K,L
08	A,B,D,E,K
09	A,E,F,H,L
10	B,C,D,F

Q6 Write short note on the following (Answer any **FOUR**) [20]

a) Major issues in Data Mining

b) Metadata in Data Warehouse

c) FP Tree

d) DBSCAN

e) Hierarchical Clustering