HOME ASSIGNMENTS CO1 and CO2

Note: Attempt ANY 10 questions

1. Define ETL process?
2. Differentiate between Data Warehouse, Data Lake and Data Marts?
3. Differentiate between OLAP and OLTP
4. Differentiate between covariance and correlation.
5. Explain Schemas of data warehousing
6. Define KDD process?
7. Explain Datawarehouse Architecture.
8. List out a five-number summary
9. Explain any two partial materialization methods used in Data Cube computation.
10. Explain Simple Regression Analysis.
11. Short notes on OLAP operations a) Roll up operation b) Drill Down operation c) Slice operation d) Dice operation e) Pivot operation f)Star schema g) Snowflake, and h) Fact Constellation i) Symmetry, j) Positive k) Negative Skewed Distribution, l)Data Cleaning, m) Data Transformation n) Data Discretization
12. Calculate Principal Component Analysis without using pca() function. Matrix: ([3, 5], [4, 2], [1, 6])?
13. a) Location Cube
    b) Customer Cube

Dimensions	Count
{Item3, Vijayawada, Suresh}	9
{Item4, Vijayawada, Suresh}	2
{Item5, Vijayawada, Suresh}	9
{Item1, Guntur, Rohit}	2
{Item2, Guntur, Rohit}	1
{Item3, Guntur, Rohit}	4

    Design a Data Cube with the following data according to i) Location Cube ii) Customer Cube
14. **Given Data:**

Dimensions	Count
{ItemA, New York, Alice}	5
{ItemB, New York, Alice}	3
{ItemC, New York, Bob}	8
{ItemD, Los Angeles, Bob}	2
{ItemE, Los Angeles, Charlie}	6
{ItemF, San Francisco, Alice}	4
{ItemG, San Francisco, Charlie}	1
{ItemH, San Francisco, Charlie}	3

    Design a Data Cube according to:

    Location Cube and Customer Cube
15. Is there any relation between mean, median, and mode? Are uses of mean, median, and mode enough to represent a summary of data?
16. Explain various ways of handling redundancy in Data Integration. Briefly explain the Chi-Square test.
17. Sorted data for price (in dollars): 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34 given. Use the Binning method on given data.
18. Define normalization and its different types of normalization
19. Calculate the regression coefficient and obtain the lines of regression for the following data

X :	1	2	3	4	5	6	7
Y:	9	8	10	11	12	13	14

20. Mention Basic Statistical Descriptions of Data with necessary examples.