

B.COM – III

Course Code: CE 302 D

Subject: Computer Application

Paper Title: Data Warehousing & Data Mining

Objectives:

The objective of the course is

- To understand the need of Data Warehouses.
- To conceptualize the architecture of a Data Warehouse.
- To Learn Data Warehouse Technologies.
- To understand the need for Data Mining.
- To understand the concept of Analytical Processing (OLAP).
- To learn the algorithms used for various types of Data Mining problems.

Contents:

Unit – I Introduction to Data Warehousing and Data Mining:

The modern Data warehouse, Data Warehouse roles and structure, need of Data warehouse, The cost of Warehousing Data, Foundation of Data mining, The roots of Data Mining, The Approach to Data Exploration and Data Mining.

Unit –II The Data Warehouse:

Stores, Warehouses and Marts, the Data Warehouse Architecture, Metadata, Metadata Extraction, Implementing Data Warehouse, Data Warehouse technologies.

Unit – III Data Mining:

What is Data Mining, Online Analytical Processing, Techniques used to mine the data, Market Basket Analysis, Limitations and challenges to DM.

Unit – IV Data Analysis and Visualization:

Data Analysis: Correlation, Covariance, Rank and Percentile, Histogram and Moving Average. Data Visualization with advance Charts: Stock Chart, Surface Chart, Donut Chart, Bubble Chart and Radar Chart.

Main Reference Book:

1. George M. Marakas, “Modern Data Warehousing, Mining and Visualization: Core Concepts”, Pearson Education

Suggested Additional Reading:

1. Jiawei Han & Micheline Kamber, “Data Mining: Concepts & Techniques”, Second Edition, Morgan Kaufmann Publishers
2. Alex Berson & Stephen J. Smith, “Data Warehousing, Data Mining & OLAP”, Tata McGraw-Hill
3. Sam Anahory, Dennis Murray , “Data Warehousing in the real World”, Pearson Education

Chapter and Section-Wise Coverage from Main Reference Book:

Chapter	Articles
1	1.1 to 1.7
2	2.1 to 2.6
3	3.1 to 3.5

Accomplishment of Students after Completing the Course

At the end of the course the student will

- Learn and understand the usage, need and cost of Data Warehouse
- Learn various techniques for Data Warehouse and Data Mining.
- Understand Market Basket Analysis.

Practical

- 1 Find correlation between given two set of data.

Roll No	HSC Per%	Graduation Per%
1	77.86	60.67
2	76.00	77.67
3	84.43	77.89
4	67.57	65.89
5	75.00	57.56
6	52.00	58.78
7	82.86	71.44
8	67.29	63.56
9	88.86	72.89
10	77.43	61.56

- 2 Find Covariance for given set of data.

Day	Dollar Price	Nifty Close
1	46.62	5232.20
2	46.58	5277.90
3	46.31	5281.80
4	46.09	5263.10
5	45.83	5244.75
6	45.50	5249.40
7	45.60	5210.40
8	45.62	5233.95
9	45.74	5259.90
10	45.66	5252.20

- 3 Find rank and percentile for given student result.

SeatNo	Per%
9	98
3	92
7	89
10	87
1	79
5	78
6	73
8	67
2	64

4 Create histogram for student's marks.

RollNo	Marks
1	29
2	32
3	25
4	25
5	29
6	28
7	32
8	25
9	27
10	26
11	25
12	31
13	29
14	29
15	28
16	28
17	31
18	27
19	30
20	25

5 Find 3 day and 5 day Moving Average for given stock price.

Date	Close Price
01-04-2013	778.75
02-04-2013	793.90
03-04-2013	776.55
04-04-2013	767.55
05-04-2013	780.10
08-04-2013	789.40
09-04-2013	776.95
10-04-2013	779.40
11-04-2013	771.10
12-04-2013	775.35
15-04-2013	793.60
16-04-2013	804.95
17-04-2013	773.70
18-04-2013	781.60
22-04-2013	789.75
23-04-2013	803.70
25-04-2013	819.65
26-04-2013	792.40
29-04-2013	792.65
30-04-2013	788.20

6 Create Stock Chart for Reliance Stock price.

Date	Open Price	High Price	Low Price	Close Price
01-04-2013	781.00	790.80	773.80	778.75
02-04-2013	780.20	796.50	778.30	793.90
03-04-2013	797.80	799.00	771.00	776.55
04-04-2013	772.50	777.95	764.20	767.55
05-04-2013	769.00	782.40	767.70	780.10
08-04-2013	778.10	794.35	778.00	789.40
09-04-2013	792.10	795.60	771.10	776.95
10-04-2013	780.10	783.95	771.60	779.40
11-04-2013	784.00	785.10	764.15	771.10
12-04-2013	767.00	780.40	765.00	775.35
15-04-2013	773.05	799.95	773.05	793.60
16-04-2013	797.00	810.00	785.40	804.95
17-04-2013	805.85	809.60	769.20	773.70
18-04-2013	776.50	785.20	772.65	781.60
22-04-2013	786.00	794.70	778.30	789.75
23-04-2013	791.50	805.80	786.05	803.70
25-04-2013	807.60	826.60	807.60	819.65
26-04-2013	813.00	825.00	789.70	792.40
29-04-2013	793.10	798.90	788.45	792.65
30-04-2013	793.65	796.65	777.15	788.20

7 Demonstrate use of surface chart for given mathematical function $f(x,y) = x^2+y^2$.

	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0
-3.0	18.0	15.3	13.0	11.3	10.0	9.3	9.0	9.3	10.0	11.3	13.0	15.3	18.0
-2.5	15.3	12.5	10.3	8.5	7.3	6.5	6.3	6.5	7.3	8.5	10.3	12.5	15.3
-2.0	13.0	10.3	8.0	6.3	5.0	4.3	4.0	4.3	5.0	6.3	8.0	10.3	13.0
-1.5	11.3	8.5	6.3	4.5	3.3	2.5	2.3	2.5	3.3	4.5	6.3	8.5	11.3
-1.0	10.0	7.3	5.0	3.3	2.0	1.3	1.0	1.3	2.0	3.3	5.0	7.3	10.0
-0.5	9.3	6.5	4.3	2.5	1.3	0.5	0.3	0.5	1.3	2.5	4.3	6.5	9.3
0.0	9.0	6.3	4.0	2.3	1.0	0.3	0.0	0.3	1.0	2.3	4.0	6.3	9.0
0.5	9.3	6.5	4.3	2.5	1.3	0.5	0.3	0.5	1.3	2.5	4.3	6.5	9.3
1.0	10.0	7.3	5.0	3.3	2.0	1.3	1.0	1.3	2.0	3.3	5.0	7.3	10.0
1.5	11.3	8.5	6.3	4.5	3.3	2.5	2.3	2.5	3.3	4.5	6.3	8.5	11.3
2.0	13.0	10.3	8.0	6.3	5.0	4.3	4.0	4.3	5.0	6.3	8.0	10.3	13.0
2.5	15.3	12.5	10.3	8.5	7.3	6.5	6.3	6.5	7.3	8.5	10.3	12.5	15.3
3.0	18.0	15.3	13.0	11.3	10.0	9.3	9.0	9.3	10.0	11.3	13.0	15.3	18.0

- 8 Create donut chart of percentile for standard 10th students.

Percentile	Students
above 99	8091
above 98	15777
above97	23739
above 96	32135
above 94	48007
above 92	63979
above 90	81396

- 9 Demonstrate use of bubble chart for region wise sale for company.

Branch	Outlets	Sales	Per%
West	5	45789	21.66
East	7	65854	31.15
North	3	15241	7.21
South	5	84555	39.99
Total		211439	100.00

- 10 Create Radar chart displaying mobile sales in two cities for year 2012.

Mobile Sales Details		
Month-Year	Ahmedabad	Baroda
January-2012	8822	5766
February-2012	6110	7513
March-2012	4849	4947
April-2012	9331	2170
May-2012	3099	4585
June-2012	3249	1714
July-2012	5146	2168
August-2012	5166	1068
September-2012	3670	2314
October-2012	7941	2859
November-2012	2341	6671
December-2012	5556	7473