

Basavarajeswari Group of Institutions
BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT
 (Autonomous Institute under Visvesvaraya Technological University, Belagavi)

USN

--	--	--	--	--	--	--	--	--	--

Course Code

2	1	C	S	5	4	2
---	---	---	---	---	---	---

Fifth Semester B.E. Degree Examinations, April/May 2024

INTRODUCTION TO DATA ANALYTICS

(Common to CSE & AIML)

Duration: 3 hrs

Max. Marks: 100

*Note: 1. Answer any FIVE full questions choosing ONE full Question from each Module.
 2. Missing data, if any, may be suitably assumed*

<u>Q. No</u>	<u>Question</u>	<u>Marks</u>	<u>(RBTL:CO:PI)</u>
<u>Module-1</u>			
1.	a. Illustrate the data processing chain with a neat diagram.	07	(3 : 1 : 1.6.1)
	b. Interpret the requirement of Business Intelligence applications in education and hospitality industry and explain them?	07	(3 : 1 : 1.6.1)
	c. Outline business intelligence and data mining cycle.	06	(2 : 1 : 1.7.1)
(OR)			
2.	a. Discuss data visualization. What is the dashboard? How does it help?	07	(3 : 1 : 1.7.1)
	b. Interpret the different data mining techniques? Which of these would be relevant in your current work?	07	(3 : 1 : 1.7.1)
	c. Outline briefly on BI for better decisions and explain its types?	06	(2 : 1 : 1.6.1)
<u>Module-2</u>			
3.	a. Summarize an OLTP system? Discuss about its characteristics.	07	(2 : 2 : 1.6.1)
	b. Illustrate with a diagram the different design schemas along with its advantages and disadvantages. (i) Star Schema (ii) Snowflake Schema (iii) Fact Constellation Schema	07	(3 : 2 : 1.7.1)
	c. Explain the approaches of constructing a data warehouse. Illustrate with example.	06	(3 : 2 : 1.6.1)
(OR)			
4.	a. Explain the need of data warehouse and explain its characteristics.	07	(2 : 2 : 1.6.1)
	b. Illustrate BI and its main components with a neat diagram.	07	(3 : 2 : 1.7.1)
	c. Illustrate with a diagram the various steps involved in the data warehouse design.	06	(3 : 2 : 1.6.1)
<u>Module-3</u>			
5.	a. Identify advantages and disadvantages of data lake and explain its types.	07	(4 : 3 : 1.6.1)
	b. Outline the benefits and use cases of modern BI.	07	(2 : 3 : 1.6.1)
	c. Distinguish between structured and unstructured data.	06	(2 : 3 : 1.6.1)
(OR)			

Note: (RBTL - Revised Bloom's Taxonomy Level: CO - Course Outcome: PI- Performance Indicator)

6. a. Illustrate and explain the main components of data lake. **07** (3 : 3 : 1.6.1)
- b. Outline data quality and its best practices you see in today's industry related to data quality. **07** (2 : 3 : 1.6.1)
- c. Distinguish between data lake and data warehouse. **06** (2 : 3 : 1.6.1)

Module-4

7. a. Illustrate with a neat diagram the main types of hierarchical data visualization. **07** (3 : 4 : 2.8.4)
- b. Illustrate with a figure the following types of data visualization: **07** (3 : 4 : 2.8.4)
(i) Scatter plot (ii) Stacked area chart (iii) Gantt chart
- c. Summarize the aims of data visualization. **06** (2 : 4 : 2.8.2)

(OR)

8. a. List and explain network data visualization tools. **07** (3 : 4 : 2.8.4)
- b. Write note on importance of data visualization. **07** (2 : 4 : 2.8.2)
- c. Discuss the history of data visualization. **06** (2 : 4 : 2.8.2)

Module-5

9. a. Analyse and create a decision tree that helps make decisions about approving the play of an outdoor game? For the given dataset. **07** (4 : 5 : 2.8.4)

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	Normal	True	??

Dataset:

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

- b. Outline the various data visualization trends. **07** (2 : 5 : 2.8.2)
- c. Explain the following data visualization tools: **06** (2 : 5 : 2.8.2)
(i) Power BI (ii) Qlik View (iii) Tibco Spotfire

(OR)

- 10 a. Analyse and create a decision tree for the data given in data set. The objective is to predict the class category (Loan approved or not). **07** (4 : 5 : 2.8.4)

Age	Job	House	Credit	Loan Approved
Young	False	False	Good	??

Data Set:

Age	Job	House	Credit	Loan Approved
Young	False	No	Fair	No
Young	False	No	Good	No
Young	True	No	Good	Yes
Young	True	Yes	Fair	Yes
Young	False	No	Fair	No
Middle	False	No	Fair	No
Middle	False	No	Good	No
Middle	True	Yes	Good	Yes
Middle	False	Yes	Excellent	Yes
Middle	False	Yes	Excellent	Yes
Old	False	Yes	Excellent	Yes
Old	False	Yes	Good	Yes
Old	True	No	Good	Yes
Old	True	No	Excellent	Yes
Old	False	No	Fair	No

- b. Summarize the significance of data visualization best practices? **07** (2 : 5 : 2.8.2)
- c. Identify and explain the 3 key elements for decision tree algorithms. **06** (2 : 5 : 2.8.2)

** ** *