

ST JOSEPH'S UNIVERSITY BENGALURU



A Public –Private-Partnership University under RUSA 2.0 of MHRD(Government of India), established by the
Karnataka Govt. Act No. 24 of 2021

ST. JOSEPH'S INSTITUTE OF INFORMATION TECHNOLOGY

DEPARTMENT OF ADVANCED COMPUTING

SYLLABUS FOR PG DIPLOMA IN DATA ANALYTICS

SUMMARY OF CREDITS IN PG DIPLOMA IN DATA ANALYTICS

DEPARTMENT OF ADVANCED COMPUTING (2026-2027)						
<u>Semester 1</u>	Code Number	Title	No. of Hours of Instructions	Number of teaching hours per week	Number of credits	Max Marks for SE - duration of examination
Theory	PGDDA1 126	Data Analytics and Visualization	45	3	3	50 -2 Hrs
Theory	PGDDA1 226	Database Management Systems	45	3	3	50 -2 Hrs
Theory	PGDDA1 326	Python for Data Analytics	45	3	3	50 -2 Hrs
Theory	PGDDA1 426	Basics of Statistics	45	3	3	50 -2 Hrs
Practical	PGDDA1 P126	Data Analytics and Visualization Lab	30	2	1	50 -2 Hrs
Practical	PGDDA1 P226	Python for Data Analytics Lab	30	2	1	50 -2 Hrs
Practical	PGDDA1 P326	Database Management Systems Lab	30	2	1	50 -2 Hrs
Practical	PGDDA1 P426	Mini Project	30	4	2	50 -2 Hrs
Total Number of credits:			17			
<u>Semester 2</u>	Code Number	Title	No. of Hours of Instructions	Number of teaching hrs /week	Number of credits	Max Marks for SE - duration of examination
Theory	PGDDA 2126	Machine Learning	45	3	3	50 -2 Hrs
Theory	PGDDA 2226	Linear Algebra	45	3	3	50 -2 Hrs
Theory	PGDDA 2326	Data visualization and Reporting Techniques	45	3	3	50 -2 Hrs
Practical	PGDDA 2P126	Machine Learning Lab	30	2	1	50 -2 Hrs

Practical	PGDDA 2P125	Data Visualization Lab	30	2	1	50 -2 Hrs
Practical	PGDDA 2P1	Internship/Project	100	24	12	50 -2 Hrs
Total Number of credits:			23			

COURSE OUTCOMES AND COURSE CONTENT

Semester	First
Paper Code	PGDDA1126
Paper Title	DATA ANALYTICS AND VISUALIZATION
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES

The objective of this course is to provide students with a strong foundation in data science concepts, the data science lifecycle, machine learning techniques, big data frameworks, and data visualization. The course also aims to develop the ability to analyze, process, and present large and complex datasets effectively for real-world decision making.

COURSE OUTCOMES

On successful completion of this course, students will be able to:

CO1: Explain the philosophies of data science, types of data (structured, unstructured, streaming, multimedia), and components of the big data ecosystem.

CO2: Apply the data science process to real-world problems, including data retrieval, cleaning, integration, transformation, exploratory data analysis, and automation.

CO3: Differentiate and implement machine learning approaches such as supervised, unsupervised, semi-supervised, and reinforcement learning for data science problems.

CO4: Analyze and utilize big data frameworks, NoSQL databases, graph databases, and text analytics techniques for large-scale and connected data applications.

CO5: Design and develop data visualizations and dashboards to present analytical results clearly and support data-driven decision making.

UNIT 1: PREPARING AND GATHERING DATA AND KNOWLEDGE

8 Hrs.

Philosophies of data science - Data science in a big data world - Benefits and uses of data science and big data - facts of data: Structured data , Unstructured data, Natural Language, Machine generated data, Audio, Image and video streaming data - The Big data Eco system: Distributed file system, Distributed Programming framework, Data Integration frame work, Machine learning Framework, NoSQL Databases, Scheduling tools, Benchmarking Tools, System Deployment, Service programming and Security.

UNIT 2: THE DATA SCIENCE PROCESS**8 Hrs.**

Overview of the data science process- Retrieving data –Data Preparation: Cleansing, integrating, and transforming data - Exploratory data analysis – Data Modeling: Model and variable selection, Model execution, Model diagnostic and model comparison - Presentation and automation: Presenting data, Automating data analysis

UNIT 3: INTRODUCTION TO MACHINE LEARNING**8 Hrs.**

Application for machine learning in data science- Tools used in machine learning- Modeling Process – Training model – Validating model – Predicting new observations –Types of machine learning Algorithm : Supervised learning algorithms, Unsupervised learning algorithms, Reinforcement Algorithm.- Semi supervised Learning

UNIT 4: BIG DATA, GRAPH DATABASES, TEXT ANALYTICS**8 Hrs.**

Distributing data storage and processing with frameworks - Case study: Assessing risk when loaning money - Join the NoSQL movement - Introduction to NoSQL - Case Study
Introducing connected data and graph databases - Text mining and text analytics - text mining in real world - text mining techniques - – Map Reduce – Dashboard development tools.

UNIT 5: DATA VISUALIZATION**8 Hrs.****SELF STUDY****5 Hrs.****TEXT BOOKS**

1. Introducing Data Science, Davy Cielen, Arno D. B. Meysman and Mohamed Ali, Manning Publications, 2016.
2. Think Like a Data Scientist, Brian Godsey, Manning Publications, 2017.

REFERENCE BOOKS

1. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O' Reilly, 1st edition, 2013.
2. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014
3. An Introduction to Statistical Learning: with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 1st edition, 2013

BLUE PRINT**Code number: PGDDA1126****Title of the paper: DATA ANALYTICS AND VISUALIZATION**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	8	10
Unit II	8	15
Unit III	8	15
Unit IV	8	15
Unit V	8	15
Self-Study	05	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	First
Paper Code	PGDDA1226
Paper Title	DATABASE MANAGEMENT SYSTEM
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES

The objective of this course is to introduce students to the fundamental concepts of database management systems, database models, relational theory, and SQL. The course aims to develop the ability to design, implement, normalize, and query databases using both relational and NoSQL approaches.

COURSE OUTCOMES

After successful completion of the course, students will be able to:

CO1: Explain the basic concepts of data, databases, DBMS architecture, database users, and different types of databases.

CO2: Design conceptual and logical database models using ER and EER diagrams, including specialization, generalization, and inheritance.

CO3: Apply relational database principles, functional dependencies, and normalization techniques to eliminate data anomalies.

CO4: Write and execute SQL queries for data definition, manipulation, and retrieval, and understand basic MongoDB concepts and data modeling.

UNIT 1: DATABASE MANAGEMENT SYSTEM INTRODUCTION

10 Hrs.

Data- Database- Database management system- Characteristics of the database approach- Role of Database administrators- Role of Database Designers- End Users- Advantages of Using a DBMS-Data models, Schema and Instances –Database design - Database Engine – 1 tier architecture – 2 tier architecture- 3 tier architecture – History of Database Management systems- Types of Databases.

UNIT 2: DATABASE MODELS AND IMPLEMENTATION

10 Hrs.

Data Model and Types of Data Model- Relational Data Model- Hierarchical Model- Network Data Model- Object/Relational Model- Object-Oriented Model- Entity-Relationship Model- Modeling using E-R Diagrams- Notation used in E-R Model- Relationships and Relationship Types- Cardinalities. Subclasses, Super classes and Inheritance – Specialization and Generalization – Characteristics of Specialization and Generalization – Modeling of UNION types with categories- An example University EER Schema.

UNIT 3: RELATIONAL DATABASES**10 Hrs.**

Structure of relational databases- Properties of relational databases and Tables –Structure of relational databases – Database Schema – Armstrong Axioms – Functional Dependency-Anomalies in a Database- Properties of Normalized Relations- First Normalization- Second Normal Form Relation- Third Normal Form- Boyce-Codd Normal Form (BCNF).

UNIT 4: SQL AND ADDITIONAL CONCEPTS**10 Hrs.**

Categories of SQL Commands; Data Definition; Data Manipulation Statements, SELECT - The Basic Form, Subqueries, Functions, GROUP BY Feature, Updating the Database, Data Definition Facilities. MongoDB Overview- MongoDB Data modeling.

SELF STUDY**5 Hrs.****REFERENCE BOOKS**

1. Elmasri Ramez and Navathe Shamkant B, Fundamentals of Database Systems, Addison-Wesley, 6th Edition, 2010.
2. Silberschatz, Korth, Sudarshan, Database System Concepts, 5 Edition, McGraw Hill, 2006.
3. O`neil Patricand, O`neil Elizabeth, Database Principles, Programming and Performance, 2nd Edition, Margon Kaufmann Publishers Inc, 2008.

BLUE PRINT**Code number: PGDDA1226****Title of the paper: DATABASE MANAGEMENT SYSTEM**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	10	15
Unit II	10	15
Unit III	10	20
Unit IV	10	20
Self-Study	05	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	FIRST
Paper Code	PGDDA1326
Paper Title	PYTHON FOR DATA ANALYTICS
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The objective of this course is to develop programming skills in Python with a focus on object-oriented concepts, data structures, and file handling. The course also aims to equip students with practical knowledge of data analysis, data wrangling, aggregation, and visualization using NumPy, pandas, and Python visualization libraries.

COURSE OUTCOMES:

On successful completion of the course, students will be able to:

CO1: Apply Python programming constructs, data structures, object-oriented principles, and file handling techniques to solve computational problems.

CO2: Use NumPy and pandas data structures and functions for efficient data manipulation, transformation, and analysis.

CO3: Perform data wrangling operations such as merging, reshaping, transformation, and string processing using regular expressions.

CO4: Implement data aggregation and group-wise operations, including pivot tables, cross-tabulations, and date-time analysis.

CO5: Create and customize data visualizations using Matplotlib and Seaborn to effectively communicate analytical insights.

UNIT 1: DATA STRUCTURES AND OOP

8 Hrs.

Python Program Execution Procedure – Statements – Expressions – Flow of Controls – Functions – Numeric Data Types – Sequences – Strings – Tuples – Lists – Dictionaries. Class – Constructors – Object Creation – Inheritance – Overloading. Text Files and Binary Files – Reading and Writing.

UNIT 2: NUMPY AND PANDAS PACKAGES

8 Hrs.

NumPy ndarray - Vectorization Operation - Array Indexing and Slicing - Transposing Array and Swapping Axes - Saving and Loading Array - Universal Functions - Mathematical and Statistical Functions in NumPy .

Series and DataFrame data structures in pandas - Creation of Data Frames – Accessing the columns in a DataFrame - Accessing the rows in a DataFrame - Panda's Index Objects - Reindexing Series and DataFrames - Dropping entries from Series and Data Frames - Indexing, Selection and Filtering in Series and Data Frames - Arithmetic Operations between Data Frames and Series - Function Application and Mapping.

UNIT 3: DATA WRANGLING**8 Hrs.**

Combining and Merging Data Sets – Reshaping and Pivoting – Data Transformation – String manipulations – Regular Expressions.

UNIT 4: DATA AGGREGATION AND GROUP OPERATIONS**8 Hrs**

Group By Mechanics – Data Aggregation – GroupWise Operations – Transformations – Pivot Tables – Cross Tabulations – Date and Time data types.

UNIT 5: VISUALIZATION IN PYTHON**8 Hrs.**

Matplotlib and Seaborn Packages – Plotting Graph - Controlling Graphs – Adding Text – More Graph Types – Getting and Setting Values – Patches.

SELF STUDY**5 Hrs.****REFERENCE BOOKS:**

1. Gowrishanker and Veena, “Introduction to Python Programming”, CRC Press, 2019.
2. Python Crash Course, 2nd Edition, By Eric Matthes, May 2019
3. NumPy Essentials, By Leo Chin and Tanmay Dutta, April 2016
4. Joel Grus, “Data Science from scratch”, O'Reilly, 2015.
5. Wes Mc Kinney, “Python for Data Analysis”, O'Reilly Media, 2012.
6. Kenneth A. Lambert, (2011), “The Fundamentals of Python: First Programs”, Cengage Learning
7. Jake Vanderplas. Python Data Science Handbook: Essential Tools for Working with Data 1st Edition.

BLUE PRINT**Code number: PGDDA1326****Title of the paper: PYTHON PROGRAMMING**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	8	10
Unit II	8	15
Unit III	8	15
Unit IV	8	15
Unit V	8	15
Self-Study	05	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	FIRST
Paper Code	PGDDA1426
Paper Title	BASICS OF STATISTICS
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The objective of this course is to provide students with a strong foundation in data collection methods, data visualization principles, and basic statistical techniques used in business analytics. The course aims to develop the ability to analyze data using descriptive statistics, correlation, and regression techniques for informed decision making.

COURSE OUTCOMES:

CO1: Explain concepts of measurement, scales, sampling techniques, data quality issues, and methods for cleaning and handling missing data.

CO2: Apply appropriate data visualization principles and techniques to present business data effectively.

CO3: Compute and interpret descriptive statistical measures such as measures of central tendency, dispersion, skewness, kurtosis, and frequency distributions.

CO4: Analyze relationships between variables using correlation, rank correlation, and regression techniques, including multiple and partial correlation.

UNIT 1: DATA COLLECTION

10 Hrs.

Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with data collection systems with examples from business, cleaning and treatment of missing data, Sampling techniques.

UNIT 2: DATA VISUALIZATION

10 Hrs.

Principles of data visualization and different methods of presenting data in business analytics

UNIT 3: BASIC STATISTICS

10 Hrs.

Frequency table, histogram, measures of location, measures of spread, skewness, kurtosis, percentiles, box plot, relative frequency distribution as a statistics model

UNIT 4: CORRELATION AND REGRESSION

10 Hrs.

Covariance, Correlation coefficient, properties of Correlation coefficient, Rank correlation, linear regression (two variables), Multiple correlation and partial correlation.

SELF STUDY

5 Hrs.

REFERENCE BOOKS:

1. Statistical Inference : P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall.
2. Introduction to Linear Regression Analysis: Douglas C. Montgomery

BLUEPRINT

Code number: PGDDA1426

Title of the paper: BASIC STATISTICAL METHODS

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	10	15
Unit II	10	15
Unit III	10	20
Unit IV	10	20
Self-Study	5	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	FIRST
Paper Code	PGDDA2P126
Paper Title	DATA ANALYTICS AND VISUALIZATION LAB
Number of teaching hrs per week	2 Hrs
Total number of teaching hrs per semester	30
Number of credits	1

1. Creating and manipulating vector in R
2. Creating matrix and manipulating matrix in R
3. Operations on Data Frames in R
4. Operations on Lists in R.
5. Programs on If – else statements in R
6. Programs on For Loops in R.
7. Customizing and Saving to Graphs in R.
8. PLOT Function in R to customize graphs
9. 3D PLOT in R to customize graphs
10. Implement in R Programming the concept to find Sum, Mean and Product of a Vector, ignore element like NA or NaN.
11. Implement in R Programming the concept to find missing values.

12. Implement the concept to create a list of data frames and access each of those data frames from the list using R.
13. Implement the concept of matrix multiplication and addition using R.
14. Implement linear regression model and compare predicted value with actual value using Visualization.
15. Implement logistic regression model and compare predicted value with actual value using Visualization.
16. Implement k-means clustering.
17. Data Visualization

Semester	FIRST
Paper Code	PGDDA2P226
Paper Title	PYTHON FOR DATA ANALYTICS LAB
Number of teaching hrs per week	2 Hrs
Total number of teaching hrs per semester	30
Number of credits	1

List of programs -

1. Introduction to Python interpreter
2. Control statements
3. functions, I/O, File handling, Packages/Libraries
4. Exception Handling, OO Programming.
5. Use of different packages for Data analytics and visualization

Semester	FIRST
Paper Code	PGDDA2P326
Paper Title	DATABASE MANAGEMENT SYSTEMS LAB
Number of teaching hrs per week	2 Hrs
Total number of teaching hrs per semester	30
Number of credits	1

1. Creation of relational tables with appropriate data types and constraints
2. Demonstration of DDL commands (CREATE, ALTER, DROP, TRUNCATE)
3. Demonstration of DML commands (INSERT, UPDATE, DELETE)
4. Execution of basic and advanced SELECT queries
5. Demonstration of JOIN operations (INNER, LEFT, RIGHT, FULL)
6. Use of subqueries and nested queries
7. Implementation of aggregate functions and GROUP BY / HAVING clauses
8. Demonstration of views and constraints

Semester	SECOND
Paper Code	PGDDA2126
Paper Title	MACHINE LEARNING
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The objective of this course is to introduce students to the fundamental concepts, mathematical foundations, and processes of machine learning. The course aims to develop the ability to apply supervised, unsupervised, and dimensionality reduction techniques to analyze and model real-world data.

COURSE OUTCOMES:

After successful completion of the course, students will be able to:

CO1: Explain the fundamentals of machine learning, probability theory, probability distributions, and decision theory used in machine learning models.

CO2: Apply **supervised learning algorithms** such as linear models, decision trees, Bayesian methods, ensemble techniques, and support vector machines.

CO3: Implement **unsupervised learning techniques** including clustering and mixture models to discover patterns in data.

CO4: Analyze and apply **dimensionality reduction techniques** such as PCA, LDA, ICA, and t-SNE for feature reduction and visualization.

UNIT 1: MACHINE LEARNING INTRODUCTION

10 Hrs.

Machine Learning–Types of Machine Learning –Machine Learning process- preliminaries, testing Machine Learning algorithms, turning data into Probabilities, and Statistics for Machine Learning Probability theory – Probability Distributions – Decision Theory.

UNIT 2: SUPERVISED LEARNING

10 Hrs.

Linear Models for Regression, Linear Models for Classification, Discriminant Functions, Probabilistic Generative Models, Probabilistic Discriminative Models, Decision Tree Learning, Bayesian Learning, Naïve Bayes, Ensemble Methods – Bagging and Boosting, Mixture of experts, Support Vector Machines.

UNIT 3: UNSUPERVISED LEARNING

10 Hrs.

Clustering- K-means, EM Algorithm, Mixtures of Gaussians, Estimating means of K Gaussians

UNIT 4: DIMENSIONALITY REDUCTION

10 Hrs.

Dimensionality Reduction, Linear Discriminant Analysis, Factor Analysis, Principal Components Analysis, Independent Components Analysis, TSNE.

SELF STUDY

5 Hrs.

REFERENCE BOOKS:

1. Tom Mitchell, "Machine Learning", McGraw-Hill, 1997.
2. Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2007. \
3. Stephen Marsland, "Machine Learning – An Algorithmic Perspective", Chapman andHall, CRC Press, Second Edition, 2014.
4. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.
5. Ethem Alpaydin, "Introduction to Machine Learning", MIT Press, Third Edition, 2014.

BLUEPRINT**Code number: PGDDA2126****Title of the paper: Machine Learning**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	10	15
Unit II	10	20
Unit III	10	20
Unit IV	10	15
Self-Study	5	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	SECOND
Paper Code	PGDDA2226
Paper Title	LINEAR ALGEBRA
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

To help students understand the 'intuition' behind the concepts of Linear Algebra and which in turn will help them to see its applications in later courses.

COURSE OUTCOMES:

CO1: Understand the most fundamental concept 'vector' that constructs Linear Algebra.

CO2: Able to gain knowledge of two Fundamental topics of Linear Algebra and Vector Space

CO3: Understanding two Fundamentals topics of Linear Algebra and Linear Transformation

CO4: Building the Basics of Linear Programming

UNIT 1: VECTORS

10 Hrs.

Introduction to Linear Algebra, Difference Between Linear Algebra & Matrix Analysis, Revision of Basic Geometry, Definition of Vectors - Examples, Two Fundamental Vectors – Geometric Vectors and R_n Vectors, Properties of Vectors, Linear Combination of Vectors, Decomposition of Vectors, Linear Independent & Linearly Dependent Vectors and Span of Vectors.

UNIT 2: VECTOR SPACE

10 Hrs.

Definition of Vector Space – Examples, Definition of Subspaces – Examples, Union & Intersection of Subspaces, Definition of Basis Vectors – Standard Basis and Dimension of Vector Space

UNIT 3: LINEAR TRANSFORMATION

10 Hrs.

Definition of Linear Transformation – Examples, Introduction to Matrix, Matrix as Linear Transformation, Matrix Multiplication (Composition of Linear Transformations) – Three Perspectives: 1. Column, 2. Row & 3. Dot Product, Concept of Determinant – Area, Volume, Hyper-plane, etc., System of Linear Equations – Column & Null Space, Gaussian Elimination, Row Reduced Echelon Form, Eigenvalues & Eigenvectors, Inverse Matrix and Positive Definite & Semi-Definite Matrix.

UNIT 4: LINEAR PROGRAMMING

10 Hrs.

Introduction to Linear Programming – Examples, Problems in LP, Convex Sets, Corner Points, Feasibility, Basic Feasible Solutions and Simplex Method

SELF STUDY

5 Hrs.

REFERENCE BOOKS :

1. Introduction to Linear Algebra, Gilbert Strang 5th Edition.
2. Linear Programming, G. Hadley.

BLUEPRINT

CODE NUMBER: PGDDA2226

TITLE OF THE PAPER: LINEAR ALGEBRA

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	10	15
Unit II	10	20
Unit III	10	20
Unit IV	10	15
Self-Study	5	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	SECOND
Paper Code	PGDDA2326
Paper Title	DATA VISUALIZATION AND REPORTING TECHNIQUES
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVE:

The objective of this course is to provide students with a comprehensive understanding of data visualization principles, tools, and design practices for effective visual communication. The course aims to develop the ability to design, build, and manage interactive visualizations and dashboards that support data-driven decision making while adhering to ethical and governance standards.

Course Outcomes (COs)

After successful completion of the course, students will be able to:

CO1: Explain the principles, fundamentals, and perceptual aspects of data visualization and apply best practices for effective visual storytelling.

CO2: Compare and use appropriate visualization tools and platforms and prepare, format, and structure data for visualization workflows.

CO3: Design and develop a wide range of charts, maps, and visual reports using effective design principles and interactive features.

CO4: Apply data modeling concepts and calculations such as measures, dimensions, relationships, and calculated fields to support accurate visual analysis.

CO5: Build, publish, and manage interactive dashboards, considering user experience, collaboration, ethical visualization, data privacy, and governance best practices.

UNIT 1: INTRODUCTION TO VISUALIZATION

9 HRS

Principles of data visualization, fundamentals of visualization, importance of visualization, advantages of visualization, limitations of visualization, human perception aspects in visualization, visual encoding techniques, best-practice guidelines for effective visualizations, storytelling with data, narrative techniques in visualization, visual communication strategies, role of visualization in analytics, role of visualization in decision-making.

UNIT 2: VISUALIZATION TOOLS, TECHNIQUES, AND DATA PREPARATION

9 HRS

Overview of visualization tools (Tableau, Power BI, Python libraries, R, Excel), comparison of visualization platforms, visualization workflow (data import, cleaning, transformation), understanding visualization components, preparing datasets for visualization, importance of data types, formatting and structuring data for visualization.

UNIT 3: DESIGNING AND DEVELOPING DATA VISUALIZATIONS

12 HRS

Building charts and visual reports, bar charts, line charts, pie charts, histograms, density plots, scatter plots, bubble plots, geo-visualizations and maps, using filters and slicers, design principles for effective visualization.

UNIT 4: DATA MODELING AND CALCULATIONS FOR VISUALIZATION

8 HRS

Basics of data modelling in visualization workflows, creating relationships between datasets, understanding measures, understanding dimensions, understanding metadata, introduction to calculations for visualization, data Calculation and calculated fields.

UNIT 5: BUILDING AND MANAGING INTERACTIVE DASHBOARDS

7 HRS

Principles of dashboard design, creating interactive dashboards, user experience in dashboards, advantages of dashboards, limitations of dashboards, publishing dashboards, sharing dashboards, collaboration in dashboards, ethical visualization considerations, data privacy considerations, best practices for visualization governance.

SUGGESTED BOOKS:

1. Storytelling with Data: A Data Visualization Guide for Business Professionals, ColeNussbaumer Knaflic
2. Fundamentals of Data Visualization , Claus O. Wilke
3. Mastering Microsoft Power BI By Brett Powell
4. Beginning Power BI by Dan Clark
5. Pro Power BI Architecture by Reza Rad Meadowbank

BLUEPRINT

Code number: PGDDA2326

Title of the paper: Data Visualization and Reporting Techniques

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions)
Unit I	10	15
Unit II	10	20
Unit III	10	20
Unit IV	10	15
Self-Study	5	
TOTAL	45	70
Maximum marks for the paper (Excluding bonus question)= 50		

Semester	SECOND
Paper Code	PGDDA2P126
Paper Title	MACHINE LEARNING LAB
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

1. Introduction to Python environment for machine learning (NumPy, pandas, scikit-learn)
2. Data loading and preprocessing (handling missing values, normalization)
3. Exploratory data analysis for machine learning datasets
4. Implementation of linear regression and model evaluation
5. Implementation of classification algorithms (Logistic Regression, Naïve Bayes)
6. Decision tree construction and visualization
7. Implementation of ensemble methods (Bagging and Boosting)
8. Support Vector Machine (SVM) implementation
9. Implementation of K-means clustering
10. Implementation of Gaussian Mixture Models / EM algorithm
11. Dimensionality reduction using PCA
12. Visualization using t-SNE
13. Model evaluation using accuracy, confusion matrix, and cross-validation
14. Mini project: end-to-end machine learning model on a real dataset

Semester	SECOND
Paper Code	PGDDA2P226
Paper Title	MACHINE LEARNING LAB
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

1. Introduction to data visualization tools (Power BI / Tableau / Excel / Python)
2. Importing data from different sources (CSV, Excel, databases)
3. Data cleaning and preparation for visualization
4. Understanding data types and formatting data
5. Creating basic charts: bar chart, line chart, pie chart
6. Creating statistical charts: histogram, box plot, scatter plot
7. Designing advanced charts: bubble chart and geographic map
8. Applying colors, labels, legends, and formatting to charts
9. Using filters, slicers, and interactive controls
10. Creating calculated fields and basic measures

11. Designing interactive dashboards
12. Publishing and sharing dashboards
13. Understanding ethical visualization and data privacy
14. Mini project: end-to-end visualization and dashboard creation using a real dataset