Savitribai Phule Pune University

(Formerly University of Pune)

Bachelors Degree in Data Science

(Faculty of Science and Technology)

Syllabi for
B.Sc. (Data Science)-Third Year
Sem-V and VI

(For Colleges Affiliated to Savitribai Phule Pune University)

Choice Based Credit System (CBCS) Syllabus
Under National Education Policy (NEP)

To be implemented from Academic Year 2026-27

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2026-27

TY (Level 5.5) SEMESTER V

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS-301-MJ-T	Data Visualization and Modelling	4	-	4	-	30	70	100
	DS-302-MJ-T	R Programming	2	-	2	-	15	35	50
	DS-303-MJ-T	Foundations of Artificial Intelligence	2	-	2	-	15	35	50
	DS-304-MJ-P	Lab Course on DS-301-MJT (Data Visualization and Modelling)	-	2	-	4	15	35	50
	DS-305-MJ-P	Lab Course on DS-302-MJT (R Programming)	-	2	-	4	15	35	50
Major Elective	DS-310-MJ-T	Business Analytics	2	-	2	-	15	35	50
	DS-311-MJ-P	Lab Course on DS-310-MJT (Business Analytics)	-	2	-	4	15	35	50
	OR					-			
	DS-312-MJ-T	Social Media Analytics	2	-	2	-	15	35	50
	DS-313-MJ-P	Lab Course on DS-312-MJT (Social Media Analytics)	-	2	-	4	15	35	50
VSC	DS-321-VSC-P	Business Intelligence and Dashboarding Tools		2	-	4	15	35	50
FP/CEP	DS-331-FP	Project on Data Analytics & Visualization	-	2	-	4	15	35	50
Minor	DS-341-MN-T	Categorical and Multivariate Data Analysis	2	-	2	-	15	35	50
		Total	**12**	**10**	**12**	**20**			**550**

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2026-27

TY (Level 5.5) SEMESTER VI

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS-351-MJ-T	NoSQL databases	4	-	4	-	30	70	100
	DS-352-MJ-T	Artificial Intelligence in Data Science	2	-	2	-	15	35	50
	DS-353-MJ-T	Data Privacy and Security	2	-	2	-	15	35	50
	DS-354-MJ-P	Lab Course on DS-351-MJ-T (NoSQL databases)	-	2	-	4	15	35	50
	DS-355-MJ-P	Lab Course on DS-352-MJ-T (Artificial Intelligence in Data Science)	-	2	-	4	15	35	50
Major Elective	DS-360-MJ-T	HR Analytics	2	-	2	-	15	35	50
	DS-361-MJ-P	Lab Course on DS-360-MJ-T (HR Analytics)	-	2	-	4	15	35	50
	OR								
	DS-362-MJ-T	Financial Analytics	2	-	2	-	15	35	50
	DS-363-MJ-P	Lab Course on DS-362-MJ-T (Financial Analytics)	-	2	-	4	15	35	50
VSC	DS-371-VSC-P	Lab Course on Statistical Analysis using PSPP	-	2	-	4	15	35	50
OJT	DS-381-OJT	On Job Training	-	4	-	8	30	70	100
		Total	**10**	**12**	**10**	**24**			**550**

Detail Syllabus
B.Sc. (Data Science)
Semester-V

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-301-MJ-T : Data Visualization and Modelling

No. of Credits: 4	Teaching Scheme Theory: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 30 Marks End Semester : 70 Marks

Prerequisites

- Fundamental understanding of programming concepts using Python
- Basic knowledge of statistics

Objectives

- To introduce fundamental concepts of data visualization and data handling
- To develop skills in loading, cleaning, and preprocessing data using Python libraries
- To provide knowledge of exploratory data analysis techniques for understanding data
- To build a foundation in predictive modelling and machine learning algorithms
- To enable students to apply data science techniques to real-world datasets

Course Outcomes

On Completion of this course, student will be able to –
CO1: Explain the need, importance, and applications of data visualization techniques
CO2: Load, clean, preprocess, and visualize data using Python-based data science libraries
CO3: Perform exploratory data analysis and interpret data patterns and correlations
CO4: Build, evaluate, and interpret regression and classification models for prediction
CO5: Apply unsupervised learning and association rule mining techniques to analyze datasets.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to Data Visualization and Data Handling**	11	CO1, CO2

1.1 Introduction to Data Visualization

Need for Data Visualization, Importance of Data Visualization, Applications of Data, Visualization, Challenges of Data Visualization.

1.2 Data Loading and Preprocessing

Loading datasets from CSV and Excel files, Data loading using Pandas, Data cleaning techniques, Data preprocessing methods.

1.3 Introduction to Data Science Libraries

NumPy, Pandas, Matplotlib, Seaborn, Plotly, SciPy, Statsmodels, Scikit –learn, XGBoost

1.4 Basic Charts and Graphs

Bar Chart ,Line Chart, Histogram, Pie Chart, Scatter Plot, Spike Plot

2	**Advanced Data Visualization and Exploratory Data Analysis**	**10**	**CO1, CO2, CO3**

2.1 Advanced Visualization Techniques

Box Plot / Box Whisker Plot, Violin Plot, Joint Plot, Pair Plot, Heatmap, Word Cloud, Bubble Chart, Area Chart, Subplots

2.2 Exploratory Data Analysis (EDA)

Concept and importance of EDA, Identifying data patterns and trends, Correlation analysis using visualization, Visual techniques for exploratory analysis.

3	**Predictive Modelling and Machine Learning**	**20**	**CO3, CO4**

1.1 Introduction to Machine Learning

Definition of Machine Learning, Types of Machine Learning (Supervised Learning, Unsupervised Learning, Reinforcement Learning), Applications of Machine Learning.

1.2 Introduction to Statistical Modelling

Model building process, Training and testing data concepts.

1.3 Simple Linear Regression

Model formulation, Assumptions of Linear Regression, Interpretation of regression coefficients.

1.4 Multiple Linear Regression

Model formulation, Interpretation of regression coefficients, Multicollinearity

1.5 **Model Evaluation for Regression Models**

R-square, Adjusted R-square, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Residual analysis, Interpretation and comparison of regression models.

1.6 Regularization Techniques

Ridge Regression, Lasso Regression.

1.7 Practical Applications of Regression Models

4	**Classification and Ensemble Learning**	**10**	**CO4**

1.1 Classification Algorithms

Logistic Regression, K -Nearest Neighbors (KNN), Decision Tree Classifier, Naïve Bayes Classifier.

4.2 Ensemble Learning Techniques

Concept of Bagging and Boosting, Random Forest, Gradient Boosting, Ada Boost.

4.3 Model Evaluation Techniques for Classification

Confusion Matrix, Accuracy, Precision, Recall, F1 Score, ROC Curve and AUC, Interpretation and comparison of classification models.

4.4 Model Performance Issues

Bias–Variance Trade-off, Overfitting, Underfitting.

5	**Unsupervised Learning and Association**	**9**	**CO5**

	Analysis		

5.1 Introduction to Unsupervised Learning

5.2 Clustering Techniques

Concept of clustering, Distance and similarity measures, K-Means Clustering, Applications of clustering, Hierarchical Clustering, DBSCAN.

5.3 Dimensionality Reduction

Principal Component Analysis (PCA).

5.4 Association Rule Mining

Concept of Association Rules, Apriori Algorithm, Market Basket Analysis.

Reference Books

1. Book1 : Han, Kamber & Pei, *Data Mining: Concepts and Techniques*
2. Book2 : Learning Python, 4th Edition by Mark Lutz
3. Book3: Data Science Essentials in Python: Collect, Organize, Explore, Predict, Value. Dmitry Zinoriev, The Pragmatic Programmers LLC, 2016
4. Book4: James et al., *An Introduction to Statistical Learning*
5. Book5: Géron, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-302-MJ-T: R Programming

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks

Prerequisites

- Basic knowledge of Statistics
- Basic knowledge of Computer Software

Objectives

- Understand R basics, set up R Studio, and customize the environment
- Develop proficiency in using R data structures: vectors, matrices, lists, and data frames
- Use R – software to find the summary statistics.
- Compute probabilities of the distributions using R – software.
- Represent statistical data by diagrammatically and graphically using R-software

Course Outcomes

On Completion of this course, student will be able to -

CO1: Apply basic R programming commands and data management techniques using R/RStudio.

CO2:Create and interpret diagrammatic and graphical representations for data visualization in R

CO3: Analyze measures of central tendency, dispersion, and perform matrix computations in R

CO4: Summarize datasets using frequency distributions and advanced statistical summaries.

CO5: Perform correlation and regression analysis to study relationships between variables.

CO6: Simulate and analyze discrete and continuous probability distributions and conduct hypothesis testing using R.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to R Programming**	**5**	**CO1**

1.1 Introduction to R, History, Features and Applications

1.2 Installing R and RStudio, Running Simple Programs

1.3 Data Types, Variables, Operators and Type Conversion

1.4 Meaning of package, base (standard package) package, to install a package, to load a package, to delete a package, learning about a package, getting help.

1.5 Operators & expressions in R – Arithmetic, logical, relational operators & expressions, precedence rule.

2	**Working with Vectors**	**5**	**CO1, CO3**

2.1 Vectors: creating a vector, modifying a vector, deleting a vector.

2.2 Working with vectors: the functions to be discussed-c(), rep(), rev(), sort(), diff(), max(), min(), colon operator(:), seq(), scan(), cut(), cat(), table(), which(), unique(), is.vector(), as.vector(), print()

2.3 Mathematical functions: abs(), sqrt(), ceiling(), floor(), trunc(), log(), log10(), exp(), sin(),

cos(), tan(), atan(), sign(), round(), range(), length(), prod(), sum(),.

2.4 Character functions: nchar(), substr(), grep(), sub(), paste(), strsplit(), toupper(), tolower(). Meaning of data frame, creation of data frame, modifying a data frame, deleting data frame, extracting elements from a data frame, use of $ sign. Functions to be discussed: subset(), transform(), attach(), detach(), with(), data.entry(), edit(), is.data.frame(), as.data.frame()

3	**Matrix Operations**	4	**CO1, CO3**

3.1 Creating a matrix, scalar multiplication of matrix, matrix addition, subtraction and multiplication, transpose of a matrix, inverse of a matrix, solving system of linear equations, finding row and column sums and means.

3.2 The functions to be discussed: matrix(), colMeans(), colSums(), rowMeans(), rowSums(), solve(), t(), diag(), is.matrix(), as.matrix()

4	**Data Handling and Visualization**	5	**CO2, CO4**

4.1 Reading and Writing Data Files, Data Preprocessing and Cleaning

4.2 Packages used for Data Visualization

4.3 Diagrams - Simple bar diagram, Subdivided bar diagram, multiple bar diagram, Pie diagram, Stem and leaf diagram

4.4 Graphs - Boxplot for one and more than one variables, spike plot, histogram for both equal and unequal class intervals, frequency polygon, ogive curves

5	**Statistical Analysis and Predictive Modelling**	6	**CO3, CO5**

5.1 Measures of Central tendency - mean, mode, median, Partition values - quartiles, deciles, percentiles, geometric mean and harmonic mean.

5.2 Dispersion: variance, standard deviation, coefficient of variation, mean deviation.

5.3 Summary Statistics using fivenum() and summary() for quick data summaries.

5.4 Skewness: Bowley's coefficient and Karl Pearson's coefficient of skewness.

5.5 Moments: Computations of raw and central moments, measures of skewness and kurtosis.

5.6 Correlation and Regression Analysis

6	**Probability Distributions and Hypothesis Testing**	5	**CO6**

6.1 Discrete Probability Distributions

a) Simulating data from Bernoulli, Binomial, and Poisson distributions.

b) Calculating probabilities and percentiles using dbinom, pbinom, dpois, etc.

6.2 Continuous Probability Distributions

a) Visualization of normal, Chi square, t-distribution, f-distribution using R

b) Applications of standard normal distribution (Z-scores)

6.3 Hypothesis Testing : t test, Chi square test, f Test for testing variances, One Way ANOVA

Reference Books

1. Book1 : Richard Cotton, "Learning R", SPD O'Reilly Publications
2. Book2 : Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, Third Edition, Sultan Chand and Sons Publishers, New Delhi.

Important Links

https://www.w3schools.com/r/r_math.asp

https://www.stats4stem.org/pdfjs/web/viewer.php?file=using_r_for_introductory_statistics.pdf

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V
DS-303-MJ-T: Foundations of Artificial Intelligence

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks

Prerequisites

- Foundational knowledge of programming, data structures, discrete mathematics, probability and statistics, and data science.

Objectives

- To introduce students to the fundamental concepts, scope, and evolution of Artificial Intelligence.
- To develop understanding of problem solving, knowledge representation, reasoning, and learning in AI.
- To bridge Data Science foundations with Artificial Intelligence concepts.
- To create awareness of ethical, social, and practical implications of AI.

Course Outcomes

On Completion of this course, student will be able to –

- **CO1:** Understand the basic concepts, goals, and applications of Artificial Intelligence.
- **CO2:** Formulate problems and analyze AI-based problem-solving techniques.
- **CO3:** Describe knowledge representation and logical reasoning mechanisms.
- **CO4:** Understand learning paradigms and the role of expert systems in Artificial Intelligence.
- **CO5:** Analyze ethical issues and future trends in Artificial Intelligence.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to Artificial Intelligence**	5	CO1

- What is Artificial Intelligence?
- Scope and importance of Artificial Intelligence
- Brief Histroy and Evolution of AI
- Types of Artificial Intelligence
- AI Problems and Basic Concepts
- Application of Artificial Intelligence
    - Game Playing
    - Natural Language Processing
    - Computer Vision
    - Speech Processing

2	**Problem Solving Techniques in Artificial Intelligence**	5	CO2

- Problem definition and problem space
- Characteristics of AI Problem
- State space representation
- Production systems
- Problem reduction

- Issues in the Design of Search Programs

3	**Knowledge Representation**	**6**	**CO3**

- Role of knowledge in Artificial Intelligence
- Knowledge representation and mapping
- Approaches to knowledge representation
- Issues in knowledge representation
- Rule-based systems
- Procedural versus Declarative knowledge

4	**Logic-Based Representation and Reasoning under Uncertainty**	**8**	**CO3**

- Propositional Logic
- Predicate Logic:
- Representing facts and relationships
- Modus Ponens
- Inference techniques:
- Resolution
- Natural Deduction
- Forward chaining and Backward chaining
- Structured knowledge representation:
- Semantic Networks
- Frames
- Reasoning under uncertainty:
- Probabilistic reasoning (conceptual overview)
- Certainty factors
- Comparison of logical and probabilistic reasoning

5	**Learning and Expert Systems**	**4**	**CO4**

Learning in Artificial Intelligence

- Concept of learning in AI
- Types of learning:
  - Rote learning
  - Learning by taking advice
  - Learning in problem solving
  - Learning by induction
  - Explanation-based learning

Expert Systems

- Need and justification of expert systems
- Components of an expert system
- Limitations of expert systems

6	**Ethical Issues in AI**	**2**	**CO5**

- Ethical challenges in AI
- Bias and fairness in AI systems
- Transparency and Eexplainability
- Privacy and data protection
- Social impact of Artificial Intelligence
- Responsible and trustworthy AI

Reference Books

Book1 : Artificial Intelligence by Elaine Rich & Kevin Knight
Book2 : Artificial Intelligence: A Modern Approach by Stuart Russell & Peter Norvig
Book 3: Artificial Intelligence: Foundations of Computational Agents by David Poole & Alan Mackworth
Book 4: Ethics of Artificial Intelligence and Robotics by Edited by Patrick Lin, Keith Abney & Ryan Jenkins

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-304-MJ-P: Lab Course on DS-301-MJ-T (Data Visualization and Modelling)

No. of Credits: 2	Teaching Scheme Practical : 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic knowledge of Statistics
- Basic knowledge of Python

Objectives

- To introduce essential libraries for data analysis and visualization.
- To develop practical skills in data loading, data cleaning, preprocessing, and exploratory data analysis using Python.
- To understand and apply various data visualization techniques for effective graphical representation of data.
- To introduce basic data modelling techniques such as regression, classification, and clustering using Python.
- To enable students to build, evaluate, and interpret simple data models for real-world datasets.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Implement Python programs for data loading, preprocessing, visualization and exploratory data analysis.

CO2: Create effective data visualizations using Python libraries for graphical representation of data.

CO3: Apply statistical techniques and basic data modelling methods to real-world datasets.

CO4: Build, evaluate, and interpret simple data models for analysis and prediction.

CO5: Integrate visualization and modelling results to support data-driven decision making.

Operating Environment:

- Windows (Windows 10/11)
- Linux (Ubuntu, Fedora, Debian, etc.)
- macOS (Latest versions)

Python Programs can be executed using:

- VS Code
- Jupyter Notebook

Sr No.	Name of Practical	Hours	CO Targeted
1	**Overview of Python Programming Environment and Data Visualization Libraries**	12	CO1
Overview of Python environment, installation and use of Jupyter Notebook. Introduction to data visualization libraries: NumPy, Pandas, Matplotlib, ggplot2 and Seaborn			

2	**Data Cleaning, and Preprocessing using Pandas**	**4**	**CO1**
Loading datasets from CSV/Excel files. Handling missing values, data type conversion, filtering, sorting, and data preprocessing			
3	**Basic and Advanced Data Visualization**	**8**	**CO2**
Creating basic plots such as line plot, bar chart, histogram, and scatter plot using data visualization libraries such as Matplotlib. Advanced visualizations including boxplot, joint plot, violin plot, pair plot, word cloud, bubble chart, area chart, pie chart, and heatmap.			
4	**Exploratory Data Analysis (EDA) and Correlation Analysis with Visualization**	**12**	**CO1, CO2**
Performing Exploratory Data Analysis to understand data patterns and trends. Correlation analysis using correlation matrix and visualization using heatmaps.			
5	**Simple Linear Regression**	**8**	**CO3, CO4**
Implementation of Simple Linear Regression model. Interpretation of regression coefficients, analysis of residual plots and visualization of regression line.			
6	**Multiple Linear Regression and Model Evaluation**	**8**	**CO4, CO5**
Building Multiple Linear Regression model using Python. Evaluation of model performance using metrics such as R^2, MAE, MSE, and RMSE.			
7	**Logistic Regression and K-Means Clustering**	**8**	**CO4, CO5**
Building Logistic Regression model using Python for classification. Application of K-Means Clustering technique for grouping unlabeled data. Selection of appropriate number of clusters and visualization of clusters using Python.			

Reference Books

1. Book1: Python Programming: An introduction to computer science, John Zelle, 3rd Edition.
2. Book2: Statistics and Machine Learning in Python by Edouard Duchesnay, Tommy Löfstedt, Feki Younes
3. Book3: Practical Statistics for Data Scientists, Second Edition by Peter Bruce, Andrew Bruce, and Peter Gedeck, O'Reilly, 2020

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-305-MJ-P : Lab Course on DS-302-MJ-T (R Programming)

No. of Credits: 2	**Teaching Scheme** Practical : 04 Hrs/Week	**Examination Scheme** Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic knowledge of Statistics
- Basic knowledge of computers and data handling

Objectives

- To familiarize students with the R programming environment and RStudio
- To develop practical skills in handling data using R data structures
- To perform statistical computations using R software
- To visualize data using diagrams and graphical techniques in R
- To apply R programming for statistical analysis, probability distributions, and hypothesis testing

Course Outcomes

On Completion of this course, student will be able to -

CO1: Use basic R commands, functions, and packages for data handling and computation

CO2: Create and interpret diagrammatic and graphical representations of data using R

CO3: Compute and analyze measures of central tendency, dispersion, and matrix operations using R

CO4: Perform summary statistics, skewness, kurtosis, and correlation analysis using R

CO5: Apply R programming to probability distributions, hypothesis testing, regression, and ANOVA

List of Practical Assignments:

Practical No.	Title of Practical	Hours
1	Introduction to R and RStudio: Writing and executing basic R programs, understanding R environment and packages	4
2	Working with R Data Types and Operators: Variables, type conversion, arithmetic, logical and relational operations	4
3	Creating and Manipulating Vectors in R using built-in functions	4
4	Mathematical and Character Functions in R	4
5	Creating and Managing Data Frames: Data extraction, modification, and summary using R	8

6	Matrix Operations in R: Matrix creation, arithmetic operations, transpose, inverse, and solving linear equations	8
7	Diagrammatic Representation of Data using R (Bar diagrams, Pie charts, Stem-and-Leaf)	4
8	Graphical Representation of Data using R (Histogram, Boxplot, Frequency Polygon, Ogive Curves)	4
9	Measures of Central Tendency and Dispersion using R	4
10	Summary Statistics, Skewness, and Kurtosis Computation using R	4
11	Correlation and Regression Analysis using R	4
12	Probability Distributions and Hypothesis Testing using R (Binomial, Normal, t-test, Chi-square test, ANOVA)	8

Reference Books

2. Book1 : Introduction to Probability and Statistics Using R by G. Jay Kerns
3. Book2 : An R Companion to Linear Statistical Models by Christopher Hay-Jahans
4. Book3 : Introductory Statistics with R, Second Edition by Peter Dalgaard, Springer

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-310-MJ-T: Business Analytics

No. of Credits: 02	Teaching Scheme Theory: 02 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic Mathematical operations
- Fundamental understanding of Data Science workflows
- Knowledge of problem solving tools like algorithms, flowcharts and pseudo codes will be an added advantage

Objectives

- To acquaint students with the process of translating business problems into analytical queries
- To introduce data mining processes (CRISP/SEMMA) and data preparation techniques
- To identify the nature of business data across different organizational domains
- To apply advanced statistical tools for trend analysis and prediction.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Identify types of analytics (Descriptive to Prescriptive) and translate vague business problems into data queries.

CO2: Prepare data for analysis using cleaning tools while following ethical and privacy laws (GDPR/DPDP).

CO3: Perform business benchmarking and value-based profiling using Pareto (80/20) and ABC analysis.

CO4: Measure relationships between business drivers using Correlation and Regression models.

CO5: Predict future trends using Time Series forecasting and evaluate accuracy with error metrics (MAE/MSE).

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**INTRODUCTION**	**05**	**CO 1**

1.1 Definition of analytics, Evolution of analytics, Concept of analytics, Types of Analytics.

1.2 The Growing Role of Business Analytics, the distinction between Business Analysis (Process), BI (Reporting), and Data Science (Algorithms).

1.3 Applications of Business Analytics: Marketing and Sales, Human Resource, Healthcare, Product Design, Service Design, finance, operation, Customer Service and Support.

1.2 Categorization of Analytical methods and models: - Descriptive (What happened?), Diagnostic (Why?), Predictive (What will happen?), and Prescriptive (How to make it happen?).

1.3 The process of translating a vague business problem into a structured data query.

1.4 The analytical decision-making process vs. intuitive decision-making vs. evidence-based decision making.

2	**DATA BASES, DATA WAREHOUSING**	**05**	**CO 2**
2.1 Importance of data in business analytics, Differences between data, information and knowledge, Quality of data. **2.2** Types of Data Sources- Structured Vs. Semi structured Vs. Unstructured data, **2.3** Data Warehouse Vs. Databases, Relational Database Vs. Non-Relational Database, RDBMS Data structures, and Columnar Data structures. **2.4** 5Vs (Volume, Velocity, Variety, Veracity, Value) of Big Data, Big Data Collection and Ethics. **2.5** Ethical collection methods, and data privacy laws (GDPR/Digital Personal Data Protection basics). **2.6** Data preparation for knowledge discovery: Data understanding and data cleaning tools, Data transformation, Data Discretization, Data Visualization. **2.7** Data Mining Process: CRISP and SEEMA, Supervised and unsupervised learning techniques.			
3	**DESCRIPTIVE AND DIAGNOSTIC BUSINESS ANALYTICS (focus must on "how to interpret and apply" using Python.)**	**10**	**CO 3**
3.1 Comparative Analysis: Benchmarking performance across categories (Region, Product, and Department) or fixed time periods. **Statistical Tools:** Mean, Median, Variance, Standard Deviation, and Coefficient of Variation (for stability analysis). Visual Tools: Side-by-Side Bar Charts and Box-and-Whisker plots for distribution comparison. **3.2** Value-Based Analysis (Impact Analysis): Identifying high-impact business segments and resource optimization. **Statistical Tools:** Pareto Analysis (80/20 Rule), ABC Analysis (Inventory/Customer classification), and Cumulative Frequency Distribution. **3.3** Trend & Growth Analysis (Historical): Identifying secular, seasonal, and cyclical shifts; distinguishing "Noise" from "Signal." **Statistical Tools:** Moving Averages (Simple & Exponential), Percentage Change, YoY Growth, and CAGR (Compound Annual Growth Rate).			
4	**PREDICTIVE AND RELATIONSHIP ANALYTICS (focus must on "how to interpret and apply" using Python.)**	**10**	**CO 4, CO 5**
4.1 Correlation & Relationship Analysis: Measuring the strength and direction between business drivers (e.g., Marketing Spend vs. Revenue). Statistical Tools: Pearson's Correlation (Linear), Spearman's Rank (Non-linear), and Scatter Plot Matrices. **4.2** Regression Analysis (Causal Modeling): Predicting continuous business outcomes based on one or more predictors. Statistical Tools: Simple and Multiple Linear Regression. Diagnostics: R^2 (Goodness of fit), P-values (Significance), and Residual Analysis for error checking. **4.3** Time Series Forecasting (Predictive): Analyzing interval-based data to project future business metrics. Components: Trend, Seasonality, Cyclicity, and Irregularity. Models: Decomposition models, Autoregression (AR), and Holt-Winters Exponential Smoothing. Accuracy Metrics: Mean Absolute Error (MAE) and Mean Squared Error (MSE).			

Reference Books

1. Business Analytics: The Science of Data-Driven Decision Making by U. Dinesh Kumar (Wiley India). **(Unit Coverage: Excellent for Unit 1 (Foundations)), and Unit 4 (Regression & Time Series))**
2. Business Analytics: Data Analysis & Decision Making by Albright & Winston (Cengage Learning).**(Unit Coverage: The good standard for Unit 4 (Methodologies))**
3. Business Intelligence, Analytics, and Data Science: A Managerial Perspective by Ramesh Sharda, Dursun Delen, & Efraim Turban (Pearson). **(Unit Coverage: Best for Unit 1 and Unit 2).**
4. Data Science for Business by Foster Provost & Tom Fawcett. **(Unit Coverage: Unit 2).**
5. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends by Michael Minelli. **(Unit Coverage: Unit 2).**
6. Python for Data Analysis by Wes McKinney (O'Reilly). **(Unit Coverage : Unit 3)**
7. Business Analytics by Sahil Raj (Pearson Education). **(Unit Coverage : Unit 3)**
8. Forecasting: Principles and Practice by Hyndman & Athanasopoulos (O'Reilly/Online). **(Unit Coverage : Unit 4)**
9. Introduction to Linear Regression Analysis by Montgomery & Peck. **(Unit Coverage : Unit 4)**
10. Hands-on Machine Learning with Scikit-Learn by Aurélien Géron. **(Unit Coverage : Unit 4)**

OTHERS

1. **Competing on Analytics: The New Science of Winning** by *Thomas H. Davenport*.
2. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling** by *Ralph Kimball*.
3. **Big Data: A Revolution That Will Transform How We Live, Work, and Think** by *Viktor Mayer-Schönberger*.
4. **Data Mining for Business Analytics: Concepts, Techniques, and Applications** by *Galit Shmueli et al. (Wiley)*.
5. **Introduction to Data Mining** by *Pang-Ning Tan, Michael Steinbach, & Vipin Kumar*.
6. **Forecasting: Principles and Practice** by *Rob J. Hyndman & George Athanasopoulos*.
7. **Naked Statistics: Stripping the Dread from the Data** by *Charles Wheelan*.

VIDEO REFERENCES LINK

1. **Analytics Overview & Types:** Business Analytics Definition, Types, and Applications
2. **BI vs. BA vs. Data Science:** Business Intelligence vs. Business Analytics
3. **Problem Framing:** Translating Business Problems into Analytical Questions
4. **Data Architecture:** Database vs. Data Warehouse vs. Data Lake Explained
5. **Big Data Fundamentals:** The 5 Vs of Big Data
6. **Mining Processes:** CRISP-DM & SEMMA Methodologies
7. **Supervised vs. Unsupervised Learning:** Machine Learning Categories Explained
8. **Forecasting:** Time Series Analysis and Forecasting for Beginners
9. **Full Course Reference:** Business Analyst Full Course 2026
10. Business Analyst Full Course

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-311-MJ-P : Lab Course on DS-310-MJ-T (Business Analytics)

No. of Credits: 02	**Teaching Scheme** Practical: 04 Hrs/Week	**Examination Scheme** Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic Mathematical operations
- Fundamental understanding of Data Science workflows
- Knowledge of problem solving tools like algorithms, flowcharts and pseudo codes will be an added advantage

Objectives

- To acquaint students with the process of translating business problems into analytical queries.
- To introduce data mining processes (CRISP/SEMMA) and data preparation techniques.
- To identify the nature of business data across different organizational domains.
- To apply advanced statistical and machine learning tools for trend analysis and prediction

Course Outcomes

On Completion of this course, student will be able to -

CO1: Execute classification, clustering, and association rule mining to solve business use cases.
CO2: Perform time series forecasting and trend analysis to predict future business outcomes.
CO3: Apply comparative and value-based analysis to optimize business performance.
CO4: Evaluate model performance using professional metrics like Confusion Matrix and Accuracy.

Sr. No.	**List of Practical Assignments**	**Teaching Hours**
1	Write a program to identify missing values, duplicate records, and outliers in a business dataset using pandas and seaborn.	**04**
2	Create a script to convert continuous variables (e.g., Age) into categorical bins (e.g., Age Groups) and apply One-Hot Encoding to unstructured text categories.	**04**
3	Use Python to generate histograms and density plots to analyze the **Volume** and **Variety** of a dataset.	**04**
4	Write a program to compare sales across different regions using **Groupby** operations and visualize the consistency using **Side-by-Side Bar Charts** and **Box-and-Whisker plots**.	**04**
5	Implement a script to calculate cumulative contribution of products to total revenue and identify the top 20% "Vital Few." By using **Pareto (80/20) analysis.**	**04**
6	**ABC Inventory Classification:** Create a program to categorize inventory/customers based on usage value into A, B, and C classes.	**04**

7	**Trend & Growth Calculator:** Write a script to calculate **Year-over-Year (YoY) growth** and **Compound Annual Growth Rate (CAGR)** for a 5-year financial dataset.	**04**
8	Apply Simple and Exponential Moving Averages to smooth out daily fluctuations in stock or sales data	**04**
9	**Correlation Matrix & Scatter Plot Matrix:** Write a program using seaborn.pairplot() to measure the strength of relationships between multiple business drivers like "Marketing Spend," "Social Media Reach," and "Sales."	**04**
10	**Simple & Multiple Linear Regression:** Implement a regression model using statsmodels to predict sales and interpret the **R-squared** and **P-values** to check model significance.	**04**
11	**Time Series Decomposition:** Create a script to decompose a business time-series into its **Trend, Seasonality, and Residual** components.	**04**
12	**Holt-Winters Forecasting:** Use the ExponentialSmoothing library to forecast future demand and calculate accuracy using **Mean Absolute Error (MAE)**.	**04**

Reference Books

1. "Business Analytics" by Sahil Raj (Pearson Education).

2. "Storytelling with Data: A Data Visualization Guide for Business Professionals" by Cole Nussbaumer Knaflic.

3. "Data Mining for Business Analytics: Concepts, Techniques, and Applications in R/Python" by Galit Shmueli, Peter C. Bruce, et al. (Wiley).

4. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.

5. "Introduction to Data Mining" by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar.

6. "Forecasting: Principles and Practice" by Rob J. Hyndman and George Athanasopoulos.

7. "Business Analytics: Data Analysis & Decision Making" by S. Christian Albright and Wayne L. Winston.

8. Forecasting: Principles and Practice by Hyndman & Athanasopoulos.

9. Hands-On Machine Learning (Scikit-Learn) by Aurélien Géron.

10. An Introduction to Statistical Learning by James, Witten, Hastie, Tibshirani.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-312-MJ-T: Social Media Analytics

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Graph Theory, Data Mining, Python programming

Objectives

- Familiarize the learners with the concept of social media.
- Familiarize the learners with the concept of social media analytics and understand its significance.
- Familiarize the learners with different tools of social media analytics.
- Familiarize the learner with different visualization techniques for Social media analytics.
- Examine the ethical and legal implications of leveraging social media data.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Understand the concept of Social media

CO2: Understand the concept of social media Analytics and its significance.

CO3: Learners will be able to use different Social media analytics tools effectively and efficiently.

CO4: Learners will be able to use different effective Visualization techniques to represent social media analytics.

CO5: Acquire the fundamental perspectives and hands-on skills needed to work with social media data

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Social Media Analytics: An Overview**	**6**	**CO1, CO2**

1.1 Core Characteristics of Social Media, Types of Social Media.

1.2 Need for Social Media Analytics (SMA), SMA in small & large organizations.

1.3 Social Media vs. Traditional Business Analytics.

1.4 Seven Layers of Social Media Analytics.

1.5 Types of Social Media Analytics,

1.6 Social Media Analytics Cycle

1.6 Challenges to Social Media Analytics

1.7 Social Media Analytics Tools

2	**Social Network Structure, Measures & Visualization**	**8**	**CO3, CO4, CO5**
2.1 Basics of Social Network Structure - Nodes, Edges & Tie 2.2 Describing the Networks Measures - Degree Distribution, Density, Connectivity, Centralization 2.3 Centrality: Degree Centrality, Eigenvector Centrality, Katz Centrality, PageRank, Betweenness, Centrality, Closeness Centrality, Group Centrality 2.4 Network Visualization - Graph Layout, Visualizing Network features, Scale Issues. 2.5 Social Media Network Analytics - Common Network Terms, Common Social Media Network Types 2.6 Types of Networks, Common Network Terminologies, Network Analytics Tools.			
3	**Social Media Text, Action and Hyperlink Analytics**	**10**	**CO3,CO5**
3.1 Social Media Text Analytics- Types of Social Media Text, Purpose of Text Analytics, Steps in Text Analytics, Social Media Text Analysis Tools 3.2 Social Media Action Analytics -What Is Actions Analytics? Common Social Media Actions, Actions Analytics Tools 3.3 Social Media Hyperlink Analytics-Types of Hyperlinks, Types of Hyperlink Analytics, Hyperlink Analytics Tool			
4	**Social Media Location & Search Engine Analytics**	**6**	**CO4, CO5**
4.1 Location Analytics - Sources of Location Data, Categories of Location Analytics, Location Analytics and Privacy Concerns, Location Analytics Tools 4.2 Search Engine Analytics - Types of Search Engines, Search Engine Analytics, Search Engine Analytics Tools			
Reference Books			
1. Gohar F. Khan. Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine and Location Data. ISBN: 1507823207, 2015. 2. Analyzing the Social Web 1st Edition by Jennifer Golbeck			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V
DS-313-MJ-P: Lab Course on DS-312-MJ-T (Social Media Analytics)

No. of Credits: 02	**Teaching Scheme** Practical: 4 Hrs/Week	**Examination Scheme** Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Types of Graphs, Data Mining, Data Analytics

Objectives

- To understand the fundamental concepts of social media networks.
- To learn various social media analytics tools and evaluation matrices.
- To collect and store social media data.
- To analyze and visualize social media data
- To design and develop social media analytics models.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Understand characteristics and types of social media networks.

CO2: Use social media analytics tools for business.

CO3: Collect, monitor, store and track social media data.

CO4: Analyze and visualize social media data from multiple platforms.

CO5: Design and develop content and structure based social media analytics models.

Unit No.	**Practical Assignment**	**Teaching Hours**	**CO Targeted**
1	Study various - i) Social Media platforms (Facebook, X, YouTube etc) ii) Social Media analytics tools (Facebook insights, google analytics etc) iii) Applications of Social media analytics for business. e.g. Google Analytics: https://marketingplatform.google.com/about/analytics/	**4**	**CO1**
2	Select **two social media platforms** (e.g., Facebook & Instagram OR Twitter/X & LinkedIn). Study their **network characteristics**, types of users, and interaction patterns. • Identify network type (ego network, follower network, community network) • Identify nodes, edges, and ties • Compare user engagement mechanisms	**4**	**CO2, CO4**
3	Data Cleaning and Storage- Preprocess, filter and store social media data for business (Using Python, MongoDB, etc)	**4**	**CO3**

4	Analyze engagement metrics of a selected social media page/account. **Tasks:** □ Compute likes, shares, comments, reach	**4**	**CO2, CO4**
	• Identify best-performing content • Suggest business insights		
5	Plot the given graph in python and calculate its Degree centrality, density, connectivity	**4**	**CO1**
6	Analyze **public sentiment** on a trending topic (e.g., budget, movie release, product launch). **Tasks:** • Clean text data • Perform sentiment analysis • Generate word cloud	**4**	**CO3, CO4**
7	Data Collection-Select the social media platforms of your choice (X, Facebook, LinkedIn, YouTube, Web blogs etc), connect to and capture social media data for business (scraping, crawling, parsing).	**4**	**CO3, CO4**
8	Use **Google Trends** to study interest patterns for a topic/product. **Tasks:** • Compare keyword trends • Analyze regional popularity • Correlate with social media trends	**4**	**CO5**

Reference Books

1. Gohar F. Khan. Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine and Location Data. ISBN: 1507823207, 2015.
2. Analyzing the Social Web 1st Edition by Jennifer Golbeck
3. Python Social Media Analytics: Analyze and visualize data from Twitter, YouTube, GitHub, and more Kindle Edition by Siddhartha Chatterjee , Michal Krystyanczuk
4. Matthew A. Russell. Mining the Social Web: Data Mining Facebook, Twitter, Linkedin, Google+, Github, and More, 2nd Edition, O'Reilly Media, 2013

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-321-VSC-P : Business Intelligence and Dashboarding Tools

No. of Credits: 2	**Teaching Scheme** Practical : 04 Hrs/Week	**Examination Scheme** Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic understanding of statistics
- Fundamental knowledge of data concepts (rows, columns, datasets)
- Familiarity with spreadsheets (Excel/CSV files)
- Basic exposure to programming concepts is desirable

Objectives

- To introduce students to advanced data visualization and business intelligence tools
- To develop practical skills in Tableau and Power BI for data analysis
- To enable students to perform statistical analysis using BI tools
- To create interactive dashboards for decision making
- To expose students to integration of Python/R scripts with BI tools.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Understand concepts of business intelligence and data visualization

CO2: Create interactive visualizations and dashboards using Tableau Public

CO3: Perform data transformation and modelling using Power BI Desktop

CO4: Apply DAX for basic statistical calculations and analytical measures

CO5: Integrate Python/R scripts in Power BI for advanced analytics

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to Tableau and Data Visualization**	**12**	**CO1, CO2**

- Introduction to Business Intelligence (BI) and Data Visualization
- Importance of visualization in data science and analytics
- Overview of Tableau: features, architecture, and applications
- Installation of Tableau Public / Tableau Desktop
- Tableau interface: worksheets, dashboards, stories
- Understanding data types, measures, and dimensions
- Connecting Tableau to Excel, CSV, and text files for data visualization.

2	**Data Analysis and Dashboarding using Tableau**	**12**	**CO2**

- Data preparation and data cleaning in Tableau
- Creating basic and advanced visualizations
- Formatting and customization of charts
- Filters, sorting, and grouping
- Calculated fields and table calculations
- Parameters for dynamic analysis
- Designing interactive dashboards

- Dashboard actions and story creation
- Best practices for effective dashboard design

3	**Introduction to Power BI and Data Modeling**	**16**	**CO3, CO4**

- Introduction to Power BI and its components
- Installation and overview of Power BI Desktop
- Connecting Power BI to various data sources
- Introduction to Power Query Editor
- Data cleaning and transformation
- Data modelling concepts
- Creating and managing relationships
- Understanding cardinality and star schema

4	**Data Analysis and Reporting using Power BI**	**12**	**CO4**

- Introduction to DAX (Data Analysis Expressions)
- DAX syntax and structure
- Calculated columns and measures
- Aggregation and statistical DAX functions
- Basic statistical calculations using DAX
    - Aggregation functions: SUM(), AVERAGE(), MIN(), MAX(), COUNT()
    - Logical functions: IF(), SWITCH()
    - Statistical functions: MEDIAN(), STDEV.P(), VAR.P()
- Filter context and row context
- Using CALCULATE() function
- Time intelligence functions
    - TOTALYTD(), TOTALMTD(), TOTALQTD()
- Creating interactive reports using filters and slicers

5	**Advanced Power BI and Integration**	**8**	**CO5**

- Introduction to Python and R integration in Power BI
- Installing and configuring Python and R environments for Power BI
- Using Python scripts in Power BI for data preprocessing and visualization
- Executing R scripts in Power BI for statistical analysis and modelling
- Creating Python/R visuals within Power BI reports
- Advantages and limitations of using Python and R scripts in Power BI Desktop

Reference Books

1. Book1: Microsoft Power BI Cookbook by Brett Powell
2. Book2: Tableau Your Data by Daniel G. Murray, ISBN: 978-1-118-61204-0, John Wiley & Sons, Inc., Indianapolis, Indiana
3. Book3: Mastering Microsoft Power BI by Brett Powell, ISBN 978-1-78829-723-3, Packt Publishing Ltd.
4. Book4: Learning Tableau 2019, Third Edition by Joshua N Milligan, ISBN 978-1-78883-952-5
5. Book5: Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-331-FP: Project on Data Analytics & Visualization

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Knowledge of programming concepts
- Knowledge of statistics fundamentals
- Knowledge of database concepts
- Knowledge of spreadsheet handling
- Knowledge of Python/R Programming and Power BI/ Tableau

Objectives

- Understand the complete data science lifecycle from data collection to visualization.
- Apply programming concepts using Python or R for data analysis and pre-processing.
- Perform exploratory data analysis and statistical analysis on real-world datasets.
- Develop analytical and problem-solving skills using data science techniques.
- Interpret data patterns, trends, and predictive insights for decision-making.
- Enhance project documentation

Course Outcomes

On Completion of this course, student will be able to -

CO1: Identify and formulate real-world data science problems.

CO2 : Collect, clean, pre-process, and organize datasets using Python or R.

CO3 : Perform exploratory data analysis and statistical interpretation.

CO4 : Apply suitable machine learning or analytical techniques for problem-solving.

CO5 : Design interactive dashboards and visualizations using Power BI or Tableau.

CO6 : Present analytical findings through reports, dashboards, and presentations.

Software & Tools

Programming Tools

- Python
- RStudio

Visualization Tools

- Microsoft Power BI
- Tableau

Supporting Tools

- Jupyter Notebook
- Microsoft Excel
- CSV / JSON datasets

Sr. No.	Guidelines for Project
1	**Problem Statement** • Define the project problem • Identify objectives and scope

	• Determine expected outputs Students should choose a real-world problem from areas such as: • Healthcare • Banking • Education • Agriculture • Retail • Finance • Social Media • Sports Analytics • Transportation and many more
2	**Data Collection** • Collect data from reliable sources • Understand dataset structure and attributes
3	**Data Preprocessing** Using Python/R perfor**m:** • Missing value handling • Duplicate removal • Data cleaning • Data transformation • Feature selection
4	**Exploratory Data Analysis (EDA)** Students should analyze: • Trends • Relationships • Correlations • Outliers • Statistical summaries Visualization Techniques • Bar Charts • Histograms • Pie Charts • Scatter Plots • Heat Maps • Line Graphs
5	**Model Building (Optional)** • Using Regression, Classification or Clustering techniques of statistics
6	**Dashboard Development** Using Power BI or Tableau: • Create interactive dashboards • Add filters and slicers • Include KPI indicators • Display summary reports and charts

7	**Result Interpretation** Students should: • Explain analytical findings • Compare outcomes • Draw conclusions from data
8	**Documentation and Presentation** • Students should prepare: • Project report (Hardcopy & Softcopy) • PPT presentation • Dashboard demonstration

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-341-MN-T: Categorical and Multivariate Data Analysis

No. of Credits: 02	**Teaching Scheme** Theory: 02 Hrs/Week	**Examination Scheme** Continuous Evaluation: Marks 20 End Semester: 30 Marks

Prerequisites

- Fundamental knowledge of Probability Theory
- Basic Statistical Inference
- Introductory Linear Algebra (matrices, eigenvalues, eigenvectors)

Objectives

- Introduce statistical techniques for analyzing categorical (qualitative) data**.**
- Develop understanding of logistic and log-linear modeling frameworks**.**
- Provide foundations of multivariate statistical methods**.**
- Enable students to apply dimension reduction and classification techniques.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Explain the structure and properties of categorical and multivariate data.

CO2: Apply measures of association, and trend tests to analyze contingency tables.

CO3: Fit and interpret logistic regression and log-linear models for categorical outcomes.

CO4: Compute and interpret multivariate summary measures.

CO5: Perform and interpret Principal Component Analysis and Factor Analysis.

CO6: Apply classification and clustering techniques.

Unit No.	**Name of Unit**	**Teaching Hours**	**CO Targeted**
1	**Introduction to Categorical Data Analysis**	**08**	**CO1, CO2**

Types of Categorical Data: Nominal vs Ordinal variables, Binary data, multinomial data, Contingency tables (2-way and r×c tables)

Measures of Association: Odds and Odds Ratio, Relative Risk, Yule's Q, Phi coefficient, Cramér's V

Chi-Square Methods: Pearson's Chi-square test, Likelihood ratio test, Test of independence, Test of homogeneity

Exact & Trend Tests: Fisher's Exact Test**,** Mantel–Haenszel Test**,** Test for linear trend (ordinal data)

2	**Log-linear Models for Categorical Data**	**07**	**CO3**

Log-linear Models: Poisson regression model**,** Two-way log-linear model**,** Interaction terms**,** Model selection (Backward/Forward approach)

Logistic Regression: Binary logistic regression**,** Interpretation of coefficients**,** Odds ratio interpretation**,** Goodness-of-fit measures

Multinomial & Ordinal Logistic Models: Baseline category logit model**,** Proportional odds model (overview)

3	**Introduction to Multivariate Data Analysis**	**07**	**CO4, CO5**

Multivariate Concepts: Random vectors**,** Mean vector & covariance matrix**,** Correlation matrix**,** Mahalanobis distance

Multivariate Normal Distribution: Definition and properties (without proof)

Principal Component Analysis (PCA): Eigenvalues and eigenvectors**,** Variance explained**,** Dimensionality reduction

Factor Analysis: Model formulation**,** Factor loadings**,** Communality**,** Varimax rotation (conceptual overview)

4	**Dimension Reduction and Classification**	**08**	**CO6**

Discriminant Analysis: Linear Discriminant Analysis (LDA)**,** Classification rule Misclassification probability

Cluster Analysis: Hierarchical clustering**,** K-means clustering**,** Distance measures

Introduction to Canonical Correlation Analysis: Concept**,** Interpretation**,** Applications in data science

Reference Books

1. Agresti, A. (2018). *Categorical data analysis* (3rd ed.). Wiley.
2. Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). Wiley.
3. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
4. Johnson, R. A., & Wichern, D. W. (2018). *Applied multivariate statistical analysis* (6th ed.). Pearson.
5. McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall.
6. Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists: 50+ essential concepts using R and Python* (2nd ed.). O'Reilly Media.
7. Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.
8. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Detail Syllabus
B.Sc. (Data Science)
Semester-VI

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-351-MJ-T: NoSQL Databases

No. of Credits: 4	Teaching Scheme Theory:4 Hrs./Week	Examination Scheme Continuous Evaluation: 30 Marks End Semester: 70 Marks

Prerequisites

- Knowledge of DBMS concepts
- Basics of SQL- data retrieval and aggregation
- Knowledge of file system concepts
- A firm foundation of any RDBMS package
- Fundamentals data structures for efficient data storage and analysis

Objectives

- Gives a basic understanding of NoSQL technology.
- Provide an insight into different types of NoSQL databases.
- Provide hands-on experience with popular NoSQL databases
- Helps students choose the right database based on application needs.
- Enables students to use NoSQL databases through suitable frameworks.

Course Outcomes

On Completion of this course, student will be able to -
CO1: To know principles, and architecture of NoSQL databases
CO2: Analyze and design appropriate aggregation models
CO3: Apply distribution models for scalable data systems
CO4: Basics of Key-Value and Column-Family databases
CO5: CRUD Operation- Document and Graph databases
CO6: Apply Polyglot Persistence and MapReduce techniques
CO7: Use NoSQL databases in real-world Data Science applications
CO8: Select suitable NoSQL databases for data science and big data applications

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to NOSQL**	**8**	**CO1**

1.1 Evolution of data management systems
1.2 Limitations of RDBMS for Big Data
1.3 Characteristics of Big Data (5Vs)
1.4 Introduction to NoSQL databases
1.5 CAP Theorem and BASE properties
1.6 ACID vs BASE
1.7 Types of NoSQL databases
1.8 Schema-agnostic nature of NoSQL databases
1.9 Use cases in Data Science and Big Data ecosystems

2	**Aggregation Models**	**8**	**CO2**
2.1 Concept of aggregation in NoSQL systems 2.2 Aggregate-oriented databases 2.3 Key-Value aggregation model 2.4 Document aggregation model 2.5 Column-family aggregation model 2.6 Aggregate design principles 2.7 Query-driven data modelling 2.8 Comparison with relational normalization 2.9 Schema flexibility and schema evolution in aggregate models			
3	**Distribution Models**	**10**	**CO3**
3.1 Need for distributed data storage 3.2 Data partitioning strategies 3.3 Sharding techniques: Range-Based Sharding, Hashed Sharding, Zone-Based Sharding (Tag-Aware Sharding) 3.4 Replication models: Master–Slave Replication, Multi-Master Replication, Peer-to-Peer architectures, Log-Based/Incremental Replication 3.6 Consistency models: strong consistency, weak consistency (including eventual consistency), and causal consistency 3.8 Fault tolerance and availability 3.9 Version stamps and conflict resolution in distributed systems			
4	**NoSQL Databases : Key Value & Column Family**	**8**	**CO4**
4.1 Key-Value databases: concepts and architecture 4.2 Operations and data access patterns 4.3 In-memory databases 4.4 Redis architecture and commands 4.5 Riak overview 4.6 Column-family databases 4.7 Cassandra architecture and data model 4.8 HBase architecture 4.9 Data modelling and querying 4.10 Use cases in analytics, caching, and logging			
5	**NoSQL Databases: Document Store & Graph Database**	**10**	**CO5**
5.1 Document-oriented databases 5.2 JSON, BSON, and XML document formats 5.3 MongoDB architecture 5.4 CRUD operations and indexing 5.5 Aggregation framework in MongoDB 5.6 Graph database fundamentals 5.7 Property graph model 5.8 Neo4j architecture 5.9 Graph traversal and queries 5.10 Use cases: Social Networks, Recommendation Engines			

6	**Polyglot Persistence, Map Reduce**	**8**	**CO6,CO8**

6.1 Introduction to Polyglot Persistence
6.2 Benefits and challenges
6.3 Database selection strategies
6.4 Integrating multiple data stores
6.6 Introduction to MapReduce
6.7 Map and Reduce functions
6.8 Hadoop ecosystem overview
6.9 Database Performance Monitoring tools
6.10 Introduction to cloud databases: Amazon RDS, Google Cloud Firestore, Microsoft Azure SQL Database
6.11 Introduction to AI Databases: Pinecone, Milvus, Weaviate, Qdrant

7	**NoSQL in Data Science Applications**	**8**	**CO7,CO8**

7.1 Role of NoSQL in Data Science lifecycle
7.2 Data ingestion and preprocessing
7.3 NoSQL for real-time analytics
7.4 Handling unstructured and semi-structured data
7.5 NoSQL integration with Python and R
7.6 Case studies:

- Recommendation systems
- Social media analytics
- IoT and sensor data analysis

7.7 Mini case study discussion

Reference Books

1 NoSQL Distilled: Pramod J. Sadalage and Martin Fowler, Addison-Wesley, Pearson
2 NoSQL for Mere Mortals, Dan Sullivan, Addison-Wesley
3 Professional NoSQL, Shashank Tiwari, Wrox Press
4 MongoDB: The Definitive Guide, Kristina Chodorow, O'Reilly Media.
5 Seven Databases in Seven Weeks, Eric Redmond and Jim R. Wilson, Pragmatic Bookshelf.
6 Ian Robinson, Jim Webber, and Emil Eifrem, Graph Databases, O'Reilly Media.

Online Learning Resources

2 MongoDB University – Practical labs and case studies
3 Apache Cassandra Documentation – Industry-grade reference
4 Redis University – Key-Value store fundamentals
5 Neo4j GraphAcademy – Hands-on graph database learning

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)

Sem-VI

DS-352-MJ-T : Artificial Intelligence in Data Science

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks

Prerequisites

- Fundamental concepts, scope, and evolution of Artificial Intelligence.
- Knowledge of Programming Language
- Data Analytics Skill

Objectives

- Introduce students to AI problem-solving techniques relevant to data-driven applications.
- Enable understanding and application of search and heuristic algorithms for decision making.
- Familiarize students with game theory, constraint satisfaction, and planning techniques used in AI systems.
- Develop the ability to analyse AI strategies and evaluate their performance in data science problems.
- Provide exposure to modern AI trends**,** including Generative AI and Explainable AI.
- Prepare students to integrate AI techniques with data science workflows through practical implementation.

Course Outcomes

On Completion of this course, students will be able to –

CO1: Apply uninformed and informed search techniques to solve data-driven AI problems.
CO2: Represent problems and knowledge using suitable AI representations.
CO3: Design AI-based decision-making solutions using game search and CSP techniques.
CO4: Analyse and evaluate AI model behaviour and search performance.
CO5: Analyse and implement AI-based planning and scheduling techniques for data science applications under constraints and uncertainty.
CO6: Understand recent trends and applications of AI in Data Science

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Problem Solving Using Search Techniques**	**9**	**CO1,CO2**
1.1 Problems Solving methods 1.2 Problem-Solving Agents 1.3 Example Problems 1.4 Search Algorithms			

1.5 Blind Search Techniques:
- Breadth-first search(BFS)
- Depth-first search(DFS)
- Uniform cost Search.

1.6 Heuristic search techniques: -
- Hill Climbing
- Best First search
- Constraint Satisfaction
- A*

2	**Game Theory and Decision Making**	**9**	**CO3,CO4**

2.1 Introduction Game Theory
- Types of Game

2.2 Optimal Decisions in Games
- Application of Optimal Decisions in Games

2.3 Heuristic Alpha–Beta Tree Search
- Application of Heuristic Alpha–Beta Tree Search

2.4 Monte Carlo Tree Search
- Application of Monte Carlo Tree Search

2.5 Constraint Satisfaction Problems (CSP).
- Application of Constraint Satisfaction Problem

2.6 Limitations of Game Search Algorithms

3	**Planning and Scheduling for Data Science Applications**	**7**	**CO5**

3.1 Introduction to Classical and Automated Planning
3.2 Planning Algorithms and State-Space Models
3.3 Heuristic-Based Planning Techniques
3.4 Hierarchical Task Planning
3.5 Planning under Uncertainty and Resource Constraints
3.6 Comparative Analysis of Planning Approaches

4	**Recent Trends and Role of AI in Data Science**	**5**	**CO6**

4.1 Role of AI in Data Science
4.2 Information Retrieval and Information Extraction
4.3 Introduction to Generative AI
4.4 Explainable AI and Ethical Considerations
4.5 Computer Vision Applications
4.6 AI Applications in Social Media and Finance
4.7 AI Application in Business and HR Analytics

Reference Books

Book1 : Artificial Intelligence: A Modern Approach by Stuart Russell & Peter Norvig
Book2 : Artificial Intelligence by Elaine Rich & Kevin Knight
Book3: Artificial Intelligence: Foundations of Computational Agents by David Poole & Alan Mackworth
Book4: Artificial Intelligence Basics: A Non-Technical Introduction, Apress by Tom Taulli

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-353-MJ-T : Data Privacy and Security

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic knowledge of **computer networks** and operating systems
- Introductory knowledge of **data science workflows**
- Understanding of **database management systems**

Objectives

- To introduce students to the principles of data security and privacy in digital systems
- To explain cryptographic techniques and their applications in securing data
- To familiarize students with privacy laws, frameworks, and ethical considerations
- To develop skills for designing secure data science systems
- To expose students to emerging trends like blockchain and differential privacy.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Understands the concepts of confidentiality, integrity, and availability in data security.

CO2: Apply basic cryptographic methods (encryption, hashing, digital signatures) to secure data.

CO3: Understand privacy principles and compliance with data protection laws.

CO4: Design secure mechanisms for data storage, transmission, and access control in data Science systems.

CO5: Understand role of emerging technologies: (Blockchain, differential privacy, AI-driven security) in data protection.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to Data Security**	**4**	**CO1**
1.1 Introduction: CIA triad: confidentiality, integrity, availability 1.2 Types of threats: malware, phishing, insider threats 1.3 Vulnerabilities: software bugs, weak passwords, misconfigurations 1.4 Risk assessment and mitigation strategies 1.5 Security policies: organizational rules, compliance requirements			
2	**Cryptography Basics**	**8**	**CO2**
2.1 Introduction: Plain text and cipher Text 2.2 Substitution techniques 2.3 Transposition techniques			

2.4 Symmetric and Asymmetric key cryptography (Only introduction of DES,AES,RSA,SHA,MD-5) 2.4 Steganography 2.5 Digital signatures and certificates			
3	**Privacy Concepts**	**6**	**CO3**
3.1 Definitions: privacy vs. security 3.2 Legal frameworks: GDPR, Indian IT Act, HIPAA basics 3.3 Data anonymization: k-anonymity, l-diversity 3.4 Pseudonymization and masking techniques 3.5 Privacy in data publishing and sharing			
4	**Secure Systems in Data Science**	**6**	**CO4**
4.1 Access control models: RBAC (Role Based), ABAC (Attribute Based) 4.2 Authentication methods: passwords, biometrics, multifactor authentication 4.3 Secure storage: encryption at rest, database security 4.4 Secure transmission: Introduction to SSL/TLS, VPNs 4.5 Cloud security: shared responsibility model, data isolation 4.6 Big data security: Hadoop/NoSQL vulnerabilities, secure pipelines			
5	**Emerging Topics**	**6**	**CO5**
5.1 Differential privacy: noise addition, privacy budget 5.2 Blockchain: distributed ledger, immutability, smart contracts 5.3 AI-driven threat detection: anomaly detection, intrusion detection systems 5.4 Security challenges in IoT 5.5 Ethical considerations in data security and privacy			
Reference Books			
Book1: William Stallings, Cryptography and Network Security Book2: Daniel J. Solove, Understanding Privacy Book3: Arvind Narayanan et al., A Primer on Differential Privacy Book4: Atul Kahate, Cryptography and Network Security			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-354-MJ-P: Lab Course on DS-351-MJ-T (NoSQL Databases)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic knowledge of database concepts (tables, records, keys)
- Understanding of programming fundamentals
- Familiarity with JSON data format
- Basic knowledge of data structures and relationships

Objectives

- Introduce NoSQL database technologies using MongoDB and Neo4j
- Perform CRUD operations and manage data efficiently in MongoDB
- Design and query complex document structures using JSON/BSON formats
- Understand indexing and aggregation techniques for performance optimization
- Model real-world applications using document-oriented and graph databases
- Implement graph data models and traversal queries using Neo4j
- Integrate document and graph databases for analytical applications

Course Outcomes

On Completion of this course, student will be able to -
CO1: Perform CRUD operations in MongoDB using standard queries.
CO2: Design and manage nested MongoDB documents using JSON/BSON.
CO3: Optimize MongoDB queries using indexing and aggregation.
CO4: Design real-world applications using MongoDB document schemas.
CO5: Create and manage nodes, relationships, graph traversal and constraints in Neo4j.
CO6: Develop graph-based applications such as social networks and recommendation systems.

Assignment 1 : Perform CRUD operations using MongoDB **CO1**

Perform CRUD operations in MongoDB: create a collection, insert documents, read data with queries, update records, and delete documents.

Assignment 2 : Create MongoDB documents using JSON/BSON formats **CO2**

Insert and manipulate nested documents in MongoDB using JSON/BSON formats. Query the nested fields effectively.

Assignment 3 : Create indexes on MongoDB collections **CO3**

Create indexes on MongoDB collections and perform queries using filters, projections, and sorting. Measure query performance before and after indexing.

Assignment 4 : Perform MongoDB aggregation **CO3**

Implement MongoDB aggregation pipelines: filter, group, and summarize data from a collection. Demonstrate aggregation using multiple stages

Assignment 6 : Design a small application using MongoDB	**CO4**
Design and implement a document schema in MongoDB for an application (e.g., e-commerce or blogging). Populate with sample data and perform queries.	
Assignment 7 : Create nodes and relationships	**CO5**
In Neo4j, create nodes and relationships representing a domain like **users, products, posts**. Use Cypher queries to insert data.	
Assignment 8 : Graph traversal queries	**CO5**
Perform graph traversal queries in Neo4j: find all connected nodes, shortest paths, and neighbors up to N levels.	
Assignment 9 : Design a small application using Neo4j	**CO6**
• Model a social network in Neo4j with users, friendships, posts, and likes. Write queries to find friends-of-friends and user activity. • Build a recommendation system in Neo4j using graph algorithms or traversal queries to suggest products or connections	
Assignment 10 : Integrate MongoDB and Neo4j	**CO6**
Integrate MongoDB and Neo4j: Store transactional data in MongoDB and relationships in Neo4j. Perform cross-database queries for analytics.	
Assignment 11 : Perform advanced Neo4j operations	**CO6**
Perform advanced Neo4j operations: create labels, constraints, indexes, and execute complex Cypher queries to extract insights**.**	

Reference Books:

1. MongoDB: The Definitive Guide – Kristina Chodorow – O'Reilly Media
2. MongoDB Applied Design Patterns – Shannon Bradshaw, Eoin Brazil, Kristina Chodorow – O'Reilly Media
3. Graph Databases – Ian Robinson, Jim Webber, Emil Eifrem – O'Reilly Media
4. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence – Pramod J. Sadalage, Martin Fowler – Addison-Wesley
5. Learning Neo4j – Rik Van Bruggen – Packt Publishing
6. Graph Algorithms: Practical Examples in Apache Spark and Neo4j – Mark Needham, Amy E. Hodler – O'Reilly Media
7. NoSQL Distilled – Jesse Anderson – Addison-Wesley

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-355-MJ-P : Lab Course on DS-352-MJ-T (Artificial Intelligence in Data Science)

No. of Credits: 02	Teaching Scheme Practical: 4 Hrs./Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks

Prerequisites

- Fundamentals of Artificial Intelligence
- Data Structures and Algorithms
- Python Programming
- Basic Discrete Mathematics

Objectives

- To implement AI algorithms using Python
- To analyse search, planning, and decision-making techniques
- To apply AI methods to real-world data science problems

Course Outcomes

On Completion of this course, students will be able to –

CO1: Implement and analyse uninformed and informed (heuristic) search techniques to solve state-space and graph-based problems relevant to data science

CO2: Formulate and solve constraint satisfaction problems such as scheduling and resource allocation using appropriate AI techniques.

CO3: Implement game-theoretic decision-making algorithms to achieve optimal strategies in competitive and adversarial environments.

CO4: Apply heuristic planning techniques for intelligent task scheduling and resource optimization in data-driven applications.

CO5: Design, implement, and evaluate rule-based intelligent systems, and analyse the performance of AI algorithms from a data science perspective**.**

Operating Environment:

- **Operating Systems:**
  - Microsoft Windows (Windows 10 or higher)
  - Linux (Ubuntu 20.04 or higher)
  - macOS (macOS Catalina or higher)
- **Programming Language:**
  - Python 3.x
- **Hardware Requirements:**
  - Minimum 4 GB RAM
  - Processor: Intel i3 / AMD equivalent or higher
  - Minimum 10 GB free disk space

Python Development Environments

To write, execute, and debug Python programs efficiently, students can use any of the following tools:

- Anaconda Distribution (Recommended for Data Science students)
- Python IDLE
- PyCharm Community Edition
- Visual Studio Code (VS Code)
- Jupyter Notebook / JupyterLab

Required Python Libraries (As applicable to experiments)

- numpy
- pandas
- matplotlib
- scikit-learn
- networkx (for graph-based experiments – optional)

Execution Environment

Programs may be executed using:

- Command Line / Terminal
- Integrated Development Environment (IDE)
- Jupyter Notebook environment

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Assignment on Uninformed Search Techniques**	**6**	**CO1, CO5**

- Implementation of Breadth First Search (BFS)
- Implementation of Depth First Search (DFS)
- Implementation of UCS for a weighted graph
- Application Oriented Implementation:
    - Cycle Detection in data pipelines(DFS)
    - Network traversal (BFS)
    - Supply Chain Analytics using UCS

2	**Assignment on Informed (Heuristic)Search Techniques**	**8**	**CO1, CO4,CO5**

- Implementation of A* Search Algorithm
- Implementation of Hill Climbing
- Implementation of Best First Search
- Application Oriented Implementation:
    - Delivery Route Optimization using A* Search
    - Hill Climbing Algorithm for Feature Selection
    - Best First Search for Delivery Prioritization
    - Heuristic Planning for Task Scheduling using Best First Search

3	**Assignment on Constraint Satisfaction Problem**	**6**	**CO2, CO4**
• Application-Oriented Implementation o Resource Allocation using Constraint Satisfaction Techniques o Scheduling using Constraint Satisfaction Techniques			
4	**Assignment on Game Theory**	**4**	**CO3, CO5**
• Implement of Alpha–Beta Pruning Algorithm • Application Oriented Implementation: o Tic-Tac-Toe game using Alpha–Beta Pruning			
5	**Assignment on Rule based system**	**6**	**CO5**
• Implementation of Simple Chabot • Implementation of Rule-based Spam Filter • Implementation of Rule-based Fraud Detection			

Reference Books
Book1: Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, Pearson Education.
Book2: Elaine Rich, Kevin Knight, Artificial Intelligence, McGraw-Hill.
Book3: Ivan Vasilev et al., Python Deep Learning, Packt Publishing.
Book4:Python Data Science Handbook *O'Reilly Media*

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-360-MJ-T : HR Analytics

No. of Credits: 02	**Teaching Scheme** Theory: 2 Hrs/Week	**Examination Scheme** Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Knowledge of Descriptive Statistics.
- Basic Knowledge of Data Science.
- Practical knowledge of R Programming.

Objectives

- To explore evolving nature of HR analytics
- To understand analysis strategies required for HR analytics.
- To have hands on practice for Predictive HR analytics using R programming language.

Course Outcomes

On Completion of this course, student will be able to - CO1:
Defined HR analytics and its metrics
CO2: Preform HR data importing and merging in R environment
CO3: Do statistical test on HR data using R language
CO4: Perform analytics like diversity analytics, employee engagement and performance analytics, predictive analytics for employee turnover using R programming language.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**HR analytics :Introduction**	**5**	**CO1**

1.1. Analytics in HR, Defined Predictive HR analytics
1.2. Introduction of HCM (Human Capital Management) Model, Metrics and Analytics
1.3. Understanding Descriptive Analytics and Perspective Analytics,
1.4. Understanding the need (and business case) for mastering and utilizing predictive HR analytic techniques
1.5. Human capital data storage and 'big (HR) data' manipulation
1.6. Predictors, prediction and predictive modelling
1.7. Current state of HR analytic capabilities and professional or academic training
1.8. Business applications of modelling
1.9. HR analytics and HR people strategy
1.10. Becoming a persuasive HR function

2	**HR information systems and data**	**5**	**CO2**

1.1. Information sources
1.2. Analysis software options
1.3. Using artificial intelligence and large language models to aid analyses Using R
1.4. Importing and merging data
1.5. Big data

3	**Analysis strategies**	**12**	**CO3**

1.1. From descriptive reports to predictive analytics
1.2. Statistical Significance
1.3. HR analytic metrics/measures
1.4. Data integrity
1.5. Types of data, Categorical and Continuous variables
1.6. Group level and individual level data
1.7. Dependent and Independent variables
1.8. Types of statistical tests
1.9. Statistical test for categorical data
1.10. Statistical test for continuous data/interval level data
1.11. Factor analysis and reliability analysis

4	**Case Studies in HR analytics**	**8**	**CO4**

1.1. Diversity analytics: Equality, diversity and inclusion, Approaches to measuring and managing D&I, Testing the impact of diversity.
1.2. Employee attitude surveys: employee engagement and workforce perceptions, measure employee engagement, Interrogating the measures,
1.3. Predicting employee turnover: Descriptive turnover analysis, measuring turnover
1.4. Predicting employee performance: measures for performance, predict performance using multiple linear regression .

Reference Books

1. *Using R in HR Analytics: A Practical Guide to Analysing People Data*. Edwards, Martin R., Kirsten Edwards, and Daisung Jang.London, Kogan Page, 2024

2. *The Practical Guide to HR Analytics: Using Data to Inform, Transform, and Empower HR Decisions*, Waters, Shonna D., Valerie Streets, Lindsay McFarlane, and Rachael Johnson-Murray. Alexandria, VA: Society for Human Resource Management, 2018

3. *The New HR Analytics: Predicting the Economic Value of Your Company's Human Capital Investments*. New York. Fitz-Enz, Jac, AMACOM, 2010

4. *Leveraging Data for Competitive Advantage*. New Delhi, Soundararajan, Ramesh, and Kuldeep Singh. *Winning on HR Analytics.*Sage Publications / Atlantic Publishers & Distributors, 2019

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-361-MJ-P : Lab course on DS-360-MJ-T (HR Analytics)

No. of Credits: 02	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- R Programing
- Knowledge of HR Analytics

Objectives

- To introduce students to the R programming environment for creating, importing, managing, and pre-processing HR datasets.
- To develop students' ability to apply statistical techniques in R for analysing employee's diversity and inclusion data within organizations.
- To enable students to analyse employee attitude and engagement data using appropriate analytical tools and techniques in R.
- To equip students with the skills to build and interpret predictive models for employee turnover using R.
- To provide hands-on experience in performing predictive analytics for evaluating and forecasting employee performance outcomes.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Create or import datasets in R environment

CO2: Perform employee diversity analytics using R

CO3: Perform Employee attitude analytics using R language CO4:
Do Predictive analytics for employee turnover.

CO5: Do Predictive analytics for employee performance.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**R Studio Environment**	**4**	**CO1**
• Explore R Studio Environment, Revision Data types, Data structures (vectors, lists, matrices, arrays), create Data set and Data frame, Make plots and visualize data (line, scatter, pie, bar), Importing and Merging Data.			
2	**Diversity analytics**	**8**	**CO2**

- Prepare and import data set with useful attributes for Gender and grade analysis
- Write Program using R for Gender and job grade analysis using frequency tables and Chi square
- Write Program using R for exploring ethnic diversity across teams descriptive statistics
- Write Program using R for Multiple linear regression to model and predict ethnic diversity variation across teams.

3	**Employee Attitude Survey**	**16**	**CO3**

- Prepare datasets for Employee survey to assess employee engagement.
- Write Program using R for Exploratory and factor analysis
- Write Program using R for Regression and descriptive analytics for survey data

4	**Predicting Employee Turnover**	**16**	**CO4**

- Prepare datasets suitable for do the analytics regarding employee turnover including attributes like Gender,Age,LengthOf Service, AppraisalRating, Country, LeaverStatus etc.
- Write Program in R language using frequency tables to explore regional differences in staff turnover
- Write Program in R language using Chi-square analysis to explore regional differences in individual staff turnover
- Write Program in R language for predicting individual turnover
- Write Program in R language for comparing expected length of service for men vs women.

5	**Predicting Employee Performance**	**16**	**CO5**

- Prepare datasets suitable for do the analytics regarding employee turnover
- Write Program in R language uses multiple regression to predict sickness absence using employee attitude scores.
- Write Program in R language uses multiple regression to explore patterns of performance with employee profile data in an organisation.

Reference Books

1. *Using R in HR Analytics: A Practical Guide to Analysing People Data*. Edwards, Martin R., Kirsten Edwards, and Daisung Jang.London, Kogan Page, 2024
2. *R for HR: An Introduction to Human Resource Analytics Using R*. Caughlin, David E. 2022. Online book. https://rforhr.com/

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-362-MJ-T : Financial Analytics

No. of Credits: 02	Teaching Scheme Theory: 02 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic Statistics knowledge.
- Python or R proficiency.
- Understanding of Linear Algebra fundamentals and finance.

Objectives

- To acquaint students with corporate financial modeling and insolvency prediction.
- To introduce portfolio optimization and risk-adjusted performance metrics.
- To identify market volatility patterns using advanced time-series frameworks.
- To apply technical indicators and visual analytics for trend detection.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Evaluate corporate solvency and project viability using financial modeling and bankruptcy analytics.

CO2: Quantify market risk and price fluctuations using volatility models and Value at Risk (VaR).

CO3: Construct optimized asset portfolios and assess performance using risk-adjusted metrics.

CO4: Identify market trends and signals through technical indicators and visual candlestick analysis.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Corporate Finance, Financial Modeling & & Portfolio Management**	15	CO 1, CO 3

1.1 Introduction to Finance & Accounting: Understanding financial statements (Balance Sheet, P&L, Cash Flow), data sources, and data cleaning for accounting analytics.

1.2 Project Analysis & Modeling: Cash flow forecasting, Cost of Capital (WACC) using sensitivity analysis, and Financial Break-even modeling.

1.3 Capital Budgeting Models: Payback Period, NPV, IRR, and MIRR.

1.4 Bankruptcy & Fraud Analytics: Modeling corporate distress using Beaver's t-test, Ohlson's Logistic Regression, and the Altman Z-score. Introduction to Fraud detection in accounting data.

1.5 Modern Portfolio Theory: Markowitz's Mean-Variance Optimization and the Capital Asset Pricing Model (CAPM).

1.6 Performance Metrics: Sharpe Ratio, Treynor Ratio, and Cluster Analysis for portfolio categorization.

2	**Financial Market & Time Series Analysis**	**15**	**CO 2, CO 4**

2.1 Data Acquisition: Importing financial data from web portals (APIs), adjusting for stock splits and mergers, and handling time-series gaps.

2.2 Risk and Return: Estimation and prediction of returns for bonds and stocks.

2.3 Advanced Time Series Modeling: Examining data stationarity, EWMA (Exponentially Weighted Moving Average), and Value at Risk (VaR).

2.4 Volatility Modeling: Implementing ARMA, ARCH, and GARCH models to predict market fluctuations.

2.5 Technical Indicators: RSI (Relative Strength Index), ROC, MACD, and Moving Averages.

2.6 Visual Analytics: Candlestick chart patterns and volume analysis.

Reference Books

For (UNIT 1)

1. Financial Modeling by **Simon Benninga**.
2. Corporate Financial Analysis by **Blaine Robertson**.
3. Corporate Financial Distress and Bankruptcy by **Edward Altman**.
4. *Modern Portfolio Theory and Investment Analysis* by **Edwin J. Elton & Martin J. Gruber**.

For (UNIT 2)

5. Analysis of Financial Time Series by **Ruey S. Tsay**.
6. Introductory Econometrics for Finance by **Chris Brooks**.
7. Forecasting: Principles and Practice by **Hyndman & Athanasopoulos**.
8. *Python for Finance: Mastering Data-Driven Finance* by **Yves Hilpisch**.
9. *Technical Analysis of the Financial Markets* by **John J. Murphy**.

OTHERS

10. Options, Futures, and Other Derivatives by **John C. Hull**.
11. Investment Analysis and Portfolio Management by **Reilly & Brown**.
12. Quantitative Trading by **Ernest P. Chan**.
13. Python for Algorithmic Trading by **Yves Hilpisch**.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-363-MJ-P : Lab Course on DS-362-MJ-T (Financial Analytics)

No. of Credits: 02	Teaching Scheme Practical: 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic Statistics knowledge.
- R Programming proficiency.
- Financial Math fundamentals.

Objectives

- To acquaint students with R-based financial modeling and insolvency prediction.
- To introduce portfolio optimization and risk-adjusted performance evaluation.
- To identify market volatility patterns using statistical time-series frameworks.
- To apply technical indicators and visual analytics for trend detection.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Execute corporate finance and project feasibility models using R.

CO2: Quantify market risk and volatility using GARCH and VaR.

CO3: Construct optimized portfolios using the Markowitz Efficient Frontier.

CO4: Build automated technical dashboards for market signal interpretation.

Sr. No.	List of Practical Assignments	Teaching Hours
1	Import raw CSV/Excel accounting data and use dplyr to calculate key liquidity and profitability ratios.	04
2	Write an R function to calculate **NPV and IRR** for uneven cash flows and perform sensitivity analysis on the discount rate.	06
3	Implement a script to compare multiple projects using **Payback Period and MIRR** (Modified Internal Rate of Return).	06
4	Apply the **Altman Z-Score** formula to real-time company data to classify financial distress zones.	04
5	Use the Portfolio Analytics package to plot the **Markowitz Efficient Frontier** and identify the minimum variance portfolio.	06
6	Calculate Beta (β) for a stock by regressing its returns against a market index (e.g., Nifty 50) to determine the Cost of Equity.	06

7	Compute and visualize the **Sharpe and Treynor Ratios** for a multi-asset portfolio using the Performance Analytics library.	**04**
8	Use quantmod to fetch live stock data via Yahoo Finance API and adjust for **stock splits and dividends**.	**04**
9	Perform the **Augmented Dickey-Fuller (ADF) test** on stock prices and transform them into log-returns for analysis.	**04**
10	Calculate the **Value at Risk (VaR)** using the Historical Simulation and Parametric (Gaussian) methods in R.	**04**
11	Fit **ARCH and GARCH (1,1)** models to capture volatility clustering in financial returns using the rugarch package.	**04**
12	Generate and plot **RSI, MACD, and Moving Averages** overlaid on interactive Candlestick charts using tidyquant.	**04**
13	Analyze the correlation between trading volume and price breakouts using visual scatter plots and ROC (Rate of Change) indicators.	**04**

Reference Books

1. **Reproducible Finance with R** by **Jonathan K. Regenstein** (**Covers:** Practicals 1, 5, 6, 7, and 8.)
2. **Analysis of Integrated Financial Statistics with R** by **Ronald D. Fricker & N.M. Lewis.** (**Covers:** Practicals 9, 10, and 11).
3. **Technical Analysis with R** by **Christopher Hare.** (**Covers:** Practicals 12 and 13).
4. **Financial Analytics with R** by **Mark J. Bennett & Dirk L. Hugen.** (**Covers:** Practicals 2, 3, and 4).
5. **Statistics and Data Analysis for Financial Engineering** by **David Ruppert & David S. Matteson.** (**Covers:** Practicals 5, 6, and 11).

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-371-VSC-P : Lab Course on Statistical Analysis using PSPP

No. of Credits: 02	Teaching Scheme Practical: 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks

Prerequisites

- Basic Statistics knowledge.

Objectives

- To understand the fundamentals of statistical data analysis using PSPP.
- To perform data entry, coding, and management of research data effectively.
- To apply descriptive and inferential statistical techniques for data analysis.
- To interpret statistical results and prepare meaningful research reports.
- To develop practical skills in using open-source statistical software

Course Outcomes

At the end of this course, students will be able to:

1) **CO1:** Demonstrate competence in data management using PSPP, including data entry, variable coding, transformation, and preparation of datasets for analysis.
2) **CO2:** Apply descriptive statistics, correlation, regression, and hypothesis testing techniques using PSPP and accurately interpret the statistical output.
3) **CO3:** Analyze real-world datasets using appropriate statistical procedures in PSPP and present clear, evidence-based conclusions in report format.

Sr. No.	List of Practical Assignments	Teaching Hours
1	**Introduction to PSPP Interface and Data Entry:** Understanding Data View and Variable View, Defining variables (numeric, string, labels, value labels), Saving and opening data files (.sav format)	06
2	**Data Import and Coding:** Importing data from Excel/CSV, Recoding variables (Recode into Same/Different Variables)	06
3	**Data Visualization:** Frequency tables, Diagram and graphs	06

4	**Descriptive Statistics – Univariate Analysis:** Computation of Mean, Median, Mode, Standard deviation, Variance	**06**
5	**Cross Tabulation and Chi-Square Test:** Creating contingency tables, Performing Chi-square test of independence, Interpretation of output	**06**
6	**Correlation Analysis:** Computation of Pearson's correlation coefficient, Spearman's rank correlation, Testing significance, Interpretation of correlation matrix	**06**
7	**Fitting of Simple, Multiple and Logistic Linear Regression Model**	**06**
8	**Testing of hypothesis:** Testing difference between two means, Checking homogeneity of variance, Interpretation of output	**06**
9	**Parametric and Non-parametric Tests**	**06**
10	**Reliability Analysis:** Cronbach's Alpha, Scale reliability, Interpretation for survey data	**06**

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-381-OJT: On Job Training

No. of Credits: 4	Duration: 120 Hrs	Examination Scheme Continuous Evaluation: 30 Marks End Semester: 70 Marks

Objectives

- To apply data science concepts to real-world industrial problems.
- To gain practical exposure to data collection, preprocessing, and analysis techniques.
- To work with data science tools, programming languages, and visualization platforms.
- To develop analytical thinking and problem-solving abilities using data-driven approaches.
- To understand business intelligence and decision-making processes.
- To improve teamwork, communication, and professional ethics.
- To gain industry exposure in analytics, and dashboard development.

Course Outcomes

On Completion of this course, student will be able to -

CO1: Understand real-world applications of data science in industry.

CO2 : Perform data preprocessing, cleaning, and transformation using suitable tools.

CO3 : Apply Python/R programming for data analysis and basic machine learning tasks.

CO4 : Create interactive dashboards and reports using visualization tools

CO5 : Analyze datasets and derive meaningful insights for decision-making

CO6 : Prepare professional reports and presentations based on job training experience.

Sr. No.	Guidelines For On Job Training (OJT)
1	Student must undergo for OJT/Internship during semester-VI after teaching and practical hours
2	Student are expected to complete the related work/project within 120 hours assigned by organization (company/ industry/ consultancy/ institution)
3	The internship work may involve the Data Analytics or IT related assignment(s) OR the maintenance of existing project OR the design/development of new project OR equivalent work
4	College should assign the mentors/guides for students to monitor the progress throughout the OJT
5	Students have to submit the weekly progress report duly signed by the concern authorities of organization to the assigned mentor
6	At the end of OJT, students should prepare the documentation and submit a report to the college in prescribed format given below
7	After completion, the final presentation and documentation will be evaluated by the examination panel as per the University norms
	OJT Report Documentation Format
	• Front Page • College Certificate • Company Certificate • Index

	1) Learning objectives Every OJT comes with learning objectives, something you want to gain and acquire from that company. Hence, listing down the knowledge and skills will allow the employers/directors to know if you've accomplished your learning objectives upon finishing the OJT. 2) Company information [Company name] was founded in [year] with the vision to [details of the vision]. The company's mission is to [mission statement]. [Company name] plays an important role in the [name of the industry] industry. I chose to OJT in this organization because I want to contribute to their mission which is aligned and relevant to my career goals. 3) OJT description My role at [Company name] was to [details of your job description]. I worked directly with the [department] and [other relevant departments (if any)] in ensuring [OJT responsibilities]. 4) Position and OJT responsibilities / Assignments done 5) Learning, challenges, and skills while doing OJT 6) Overview of OJT experience During my OJT with [Company name], I was able to experience working first-hand with [the tasks you've completed], and through that, I'm able to develop my [skills you've acquired]. I found out that my [experiences] are useful in solving issues related to [problems you've solved]. Although I am considered to be less experienced with [the tasks/issues you've faced], I'm able to solve the problems with the help of my team by [describing how you apply the skills you've acquired]. 7) Outgoing consideration While I've gained many experiences and skills at the [Company name], I feel that I'm still lacking the confidence to solve [tasks you wished you were better at doing]. If I'm allowed more time to learn and participate at the [the aspects of the job you wanted to join], I would be able to polish my skills more in solving the [tasks/issues you've faced]. 8) Observation & Findings a. Front end tool used b. Back-end tools used c. Tools used other than curriculum d. Code of conduct e. Employment process f. Appraisal process g. Areas where you need improvement h. What is your learning / outcome of OJT?