Savitribai Phule Pune University
(Formerly University of Pune)

Two Year Degree Program in Data
Science   (Faculty of Science)

Syllabi for

M.Sc. (Data Science)
Part-II

(For Colleges Affiliated to Savitribai Phule Pune University)

Choice Based Credit System Syllabus under NEP

To be implemented from

Academic Year 2024-2025

Savitribai Phule Pune University
Board of Studies in Computer Science

M. Sc. (Data Science) Proposed Structure as per NEP 2020

SEMESTER I

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Schemeand Marks		
			TH	PR	TH	PR	CE	EE	Total
Major core	DS-501-MJ	Statistics for Data Science	4	-	4	--	30	70	100
	DS-502-MJ	Computational Mathematics	4	-	4	--	30	70	100
	DS-503-MJ	Fundamentals of DataScience	2	-	2	--	15	35	50
	DS-504-MJP	Lab course on Statistics for DataScience	-	2	--	4	15	35	50
	DS-505-MJP	Lab course on Computational Mathematics	-	2	--	4	15	35	50
Major Elective	DS-510-MJ	Data Mining and Datawarehousing	2	-	2	--	15	35	50
	DS-511-MJP	Lab course on DM	-	2	--	4	15	35	50
	OR								
	DS-512-MJ	Artificial Intelligence	2	-	2	--	15	35	50
	DS-513-MJP	Lab course on AI	-	2	--	4	15	35	50
RM	DS-531-RM	Research Methodology	4	-	4	--	30	70	100
		Total	16	6					

Semester - II

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-551-MJ	Database Technologies	4	-	4	--	30	70	100
	DS-552-MJ	Machine Learning	4	-	4	--	30	70	100
	DS-553-MJ	Python Programming for DataScience	2	-	2	--	15	35	50
	DS-554-MJP	Lab course on Database Technology	-	2	--	4	15	35	50
	DS-555-MJP	Lab course on Machine Learning using Python	-	2	--	4	15	35	50
Maj or Elect ive	DS-560-MJ	Big Data	2	-	2	--	15	35	50
	DS-561-MJP	Lab Course on Big Data	-	2	--	4	15	35	50
	OR								
	DS-562-MJ	Deep Learning	2	-	2	--	15	35	50
	DS-563-MJP	Lab Course on Deep Learning	-	2	--	4	15	35	50
OJT	DS-581-OJT	On Job Training in IT industry/Summer Project(120 Hours)	-	4	-	-	30	70	100
		Total	12	10					

Semester - III

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-601-MJ	Data Visualization and Analytics	4	-	4	--	30	70	100
	DS-602-MJ	Optimization Techniques	4	-	4	--	30	70	100
	DS-603-MJ	Predictive Analysis	2	-	2	--	15	35	50
	DS-604-MJP	Lab course Data Visualization and Analytics	-	2	--	4	15	35	50
	DS-605-MJP	Lab course Optimization Techniques and Predictive Analysis	-	2	--	4	15	35	50
Major Elective	DS-610-MJ	Exploratory Data Analysis	2	-	2	--	15	35	50
	DS-611-MJP	Lab Course on Exploratory Data Analysis	-	2	--	4	15	35	50
	OR								
	DS-612-MJ	Business Informatics	2	-	2	--	15	35	50
	DS-613-MJP	Lab Course on BI	-	2	--	4	15	35	50
RP	DS-631-RP	Research Project Work Research Paper in Conference	-	4	-	-	30	70	100
Total			12	10					

Semester - IV

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme And Marks		
			TH	PR	TH	PR	CE	EE	Total
MC	DS-651-MJP	Full Time Industrial Training (IT)	0	16	-	--	120	280	400
RP	DS-681-RP	Research Work/ Research paper in Journal	-	6			45	105	150
Total			4	22					

Abbreviations

TH	Theory		PR	Practical
CE	Continuous Evaluation		EE	End semester Evaluation
MJ	Major Core		ME	Major Elective
RM	Research Methodology		RP	Research Project
OJT	On Job Training			

NOTE: All Theory papers shall have lecture sessions of 48 hours, 12 hours for CE.

M.Sc. Data Science (2023)
Course Code : DS-601-MJ
Course Name : Data Visualization and Analytics

No. of Credits : 04 Credits

Continuous Evaluation : 30 Marks   External Exam : 70 Marks

Course Overview:

This course provides an in-depth exploration of data visualization techniques and their role in data analytics. Students will learn how to effectively visualize data using various tools and platforms, interpret visualizations, and gain insights for decision- making. The course will cover both theoretical concepts and practical applications,
equipping students with the skills needed to communicate complex data effectively.

Objectives:

By the end of the course, students should be able to:

1. Understand the principles of data visualization and its importance in data analytics.

2. Apply various data visualization techniques to represent different types of data.

3. Critically evaluate and interpret data visualizations.

4. Utilize tools and software for creating interactive and informative visualizations. Course

Outcome
CO1: Understand and create from the ground up business intelligence reports of expert calibre.

CO2: Create and employ the same instruments that experts in data science and analysis use.

CO3: Learn about the subtleties and uses of descriptive, predictive, and prescriptive analytics to improve your analytical abilities and make data-driven business choices in real time.

CO4: Analyse a dataset to learn new things and share what you find with impactful Data visualizations.

CO5:Apply data visualization techniques to real-world datasets for analysis and decision- making.

Unit No.	Name of Unit	CO	No.of Lect/ Hrs
1.	Introduction to Data Visualization and Analysis	CO3 & CO5	08

	1. Introduction to Data Visualization and Analysis
	  ○ Understanding the importance & Benefits of data visualization
	  ○ Principles of effective data visualization
	  ○ Challenges in data visualization
	  ○ Opportunities in data visualization
	  ○ Types of data visualizations (e.g., charts, graphs, maps)& their applications
	  ○ Data types and data structures
	  ○ Data cleaning and preprocessing
	  ○ Exploratory data analysis techniques		

2.	Data Visualization Tools and Software	CO 1, CO 3	09
	• Introduction to popular data visualization tools (e.g., Tableau, Power BI, matplotlib, ggplot2, Looker Studio, HighCharts, Adobe Illustrator) • Introduction to PowerBI, • Use cases and BI Tools , Power BI components, Power BI Desktop, workflows and reports , • Data Extraction with Power BI. • SaaS Connectors, Working with Azure SQL database, Python and R with Power BI • Power Query Editor, Advance Editor, Query Dependency Editor, Data Transformations, Shaping and Combining Data ,M Query and Hierarchies in Power BI. • Introduction to web-based visualization tools (e.g., D3.js, Plotly)		

3	Graphical Perception and Design Principles	CO4	09

	- Design Principles and Advanced Visualization Techniques:
	  ○ Color theory and color palettes
	  ○ Typography and layout
	  ○ Accessibility and inclusivity in visualization design
	  ○ Interactive visualizations
	  ○ Network visualization
	  ○ Human perception and cognition in visualization
	  ○ Color theory and visualization aesthetics		

4.	Static Data Visualization Techniques	CO3	04
	o What is Static data Visualization? o KeyComponents of Data Visualization o Common Types of Data Visualization o visualization types in big data o Examples of Data Visualization - Cinema,Art, Philosophy, Cartography etc.		
5	Geographic Data Visualization	CO1,CO2	10
	• Introduction to geographic data visualization • geovisualization • Visual variables • Properties of visual variables • The perception of visual variables • Types of information and their visualization • Map elements. Map composition • Types of thematic maps: proportional symbol maps, point density maps, isoline maps and choropleth maps. • Combining multiple layers • How do you visualize geographic data? • Mapping techniques and spatial analysis - Case studies in geographic data visualization		

6	Time Series Visualization	CO4	09
	- Characteristics of time series data - Applications and domains where time series data are commonly used - Challenges in visualizing time series data Decomposing time series data into trend, seasonal, and residual components - Visualizing seasonal patterns and trends using decomposition techniques - Analyzing and interpreting trends in time series data - Candlestick charts for visualizing financial time series data - Spectral plots and periodograms for analyzing periodicity in time series data - Heatmaps and calendar heatmaps for visualizing time series data over time and across categories - Techniques for analyzing temporal patterns - Hands-on exercises with time series data visualization		
7	Advanced Visualization Techniques	CO2	09
	Principles of Advanced Visualization: - Perception and cognition in visualization - Visual encoding techniques for multivariate data - Design principles for effective visualization communication Advanced Chart Types and Techniques: - sunburst charts, and tree diagrams - Sankey diagrams for flow visualization - Radar charts and parallel coordinate plots for multidimensional data Network and Graph Visualization: - Node-link diagrams and force-directed layouts - Community detection and network clustering algorithms Visualizing dynamic and large-scale networks Emerging Trends and Technologies:		

	- Augmented reality (AR) and virtual reality (VR) in data visualization - Immersive visualization environments and 3D visualization techniques - Data visualization in artificial intelligence and machine learning applications - Case studies and applications		
8	Ethical Considerations in Data Visualization	CO1,CO4	02
	- Ethical Considerations in Data Visualization - Bias, misrepresentation, and misleading visualizations - Privacy, consent, and data sensitivity - Cultural and societal implications		
	Recommended Readings		
	- "The Visual Display of Quantitative Information" by Edward Tufte - "Storytelling with Data: A Data Visualization Guide for Business Professionals" by Cole Nussbaumer Knaflic - "Interactive Data Visualization for the Web" by Scott Murray - "Data Points: Visualization That Means Something" by Nathan Yau - "Information Dashboard Design: Displaying Data for At-a- Glance Monitoring" by Stephen Few		

M.Sc. Data Science (2023)
Course Code: DS-602-MJ

Course Name : Optimization Techniques

No. of Credits :04 Credits

Continuous Evaluation : 30 Marks   External Exam : 70 Marks

Course Overview: Strong mathematical background (calculus, linear algebra), Analytical thinking, A basic programming skills (preferably), Familiarity with operations research concepts.

Objectives:

1. To learn history, methods, features and applications of OR
2. To learn structure of LP model, graphical solution of LPP.
3. To study advance methods like simplex, dual simplex method and transportation problem.
4. To learn Game theory and Project Scheduling by PERT-CPM. Course

Outcome

CO1: Understand and create from the ground up business intelligence reports of expert calibre.

CO2: Create and employ the same instruments that experts in data science and analysis use.

CO3: Learn about the subtleties and uses of descriptive, predictive, and prescriptive analytics to improve your analytical abilities and make data-driven business choices in real time.

CO4: Analyse a dataset to learn new things and share what you find with impactful data visualisations.

CO5:Apply data visualization techniques to real-world datasets for analysis and decision-making.

Unit No.	Name of Unit	Co	No.of Lect/ Hrs
1.	Unit 1: Introduction to OR	CO3 & CO5	05

	1.1 The Nature of OR 1.2 The history of OR 1.3 Definitions of OR 1.4 Principles of Problem solving using OR Models 1.5 Optimization and OR process 1.6 Applications of OR 1.7 Features of OR		
2.	Unit 2: Linear Programming	CO 1, CO 3	15

	2.1 Introduction
	2.2 Formulation of LP Models
	      2.2.1 General Mathematical Model of LPP
	      2.2.2 Guidelines on LPP Formulation
	      2.2.3 Examples of LPP Formulations
	2.3 Graphical Solution of LPP
	    2.3.1 Graphical Solution Methods of LP Problems
	    2.3.2 Special Cases in Linear Programming
	2.4 Simplex Method
	    2.4.1 Standard form, Canonical form of LP Problem
	    2.4.2 Simplex Algorithm (Maximization case)
	    2.4.3 Simplex Algorithm (Minimization case)
	    2.4.4 Revised Simplex Method
	    2.4.5 The Big-M Method (Use of Artificial variables)
	    2.4.6 Types of Linear Programming Solutions
	2.5 Alternative Introduction
	2.6 Formulation of LP Models
	    2.6.1 General Mathematical Model of LPP
	    2.6.2 Guidelines on LPP Formulations
	    2.6.3 Examples of LPP Formulations
	2.7 Graphical Solution of LPP
	    2.7.1 Graphical Solution Methods of LP Problems
	    2.7.2 Special Cases in Linear Programming
	2.8 Simplex Method
	    2.8.1 Standard form, Canonical form of LP Problem
	    2.8.2 Simplex Algorithm (Maximization case)
	    2.8.3 Simplex Algorithm (Minimization case)
	    2.8.4 Revised Simplex Method		

	2.8.5 The Big-M Method (Use of Artificial variables) 2.8.6 Types of Linear Programming Solutions 2.8.7 Alternative (Multiple) Optimal Solutions 2.8.8 Unbounded Solutions (Multiple) Optimal Solutions, Unbounded Solutions			
3	Unit 3: Duality theory and applications	CO4	09	
	3.1 Introductions 3.2 Formulation of Dual LPP 3.3 Standard Results on Duality 3.4 Advantage of Duality 3.5 Dual Simplex Method 3.5.1 Examples of Dual Simplex Method 3.6 Sensitivity analysis in LP			

4	Unit 4: Transportation Problem	CO3	12

	4.1 Mathematical Model of TP
	4.2 Unbalanced Supply and Demand
	4.3 Method for finding Initial basic feasible
	Solutions
	4.4 Degeneracy and it's Resolution
	4.5 North-west Corner Method
	4.6 Least Cost Method
	4.7 Vogel's Approximation Method
	4.8 Test of Optimality
	4.9 Steps of Modi Method
	4.10 Variations of the TP
	4.11 Assignment Problem
	    4.11.1 Mathematical Models of AP
	    4.11.2 Hungarian Method for Solving AP
	    4.11.3 Variations of AP
	    4.11.4 Alternate Optimal Solutions
	    4.11.5 Maximization case in AP
	    4.11.6 Unbalanced AP
		-	-

5	Unit 5: Game Theory	CO1,CO2	08
	5.1 Introductions 5.2 Two-Person Zero-Sum Games with saddle point, Two-Person Zero-Sum Games without Saddle 5.3 Algebraic Method for Solving 2 * 2 5.4 Graphical Method for Solving 2 * 2 5.5 Games Without Saddle Point		
6	Unit 6: Project Scheduling by PERT-CPM	CO4	12

Reference Books:

1. Operations Research: An Introduction, 7th Edn., H. Taha, Prenticeï¿½Hall, 2002

2. Operations Research: V.K.Kapoor

3. Operations Research: J.K.Sharma

4. Operations Research: Principles and Practice, A. Ravindran, D, Phillips, J Solberg, John Wiley Pub, 1987.

5. Linear Programming and Extensions, G Dantzig, Princeton University Press, 1963

6. Theory of Games and Economic Behaviour, J. von Neumann, O. Morgenstern, John Wiley Pub. 1967

M.Sc. Data Science (2023)

Course Code : DS-603-MJ
Course Name : Predictive Analysis

No. of Credits :02 Credits

Continuous Evaluation : 15 Marks   External Exam : 35 Marks

Pre-requisites:
• Students should familiar with terminologies of statistics, data science and data preparations.

Course Objectives:
• To provide an overview of an exciting growing field of predictive analytics.
• To give insights in the fundamentals of techniques and principles in predictive modeling.
• Compare the underlying predictive modeling techniques.
• To develop skills in students to solve complex real-world problems in decision support.

Course Outcomes: Students will be able to
• Use technical skills in predictive modelling to support business decision- making.
• Describe various techniques for predictive analysis.
• Fit predictive models for the sample data.

Chapter No.	Chapter Contents	No. of Lectures
1	Fundamentals of Predictive Analysis 1.1 Analysis 1.2 Data analysis 1.3 Prediction 1.4 Predictive Analysis 1.5 Why Predictive Analysis? 1.6 Steps/Lifecycle of Predictive Analysis 1.7 Applications of Predictive Analysis 1.8 Predictive modeling and its features 1.9 Types of Predictive models 1.10 Tools for Predictive Analysis	8

	1.11 Descriptive Vs Prescriptive Vs Predictive Analysis 1.12 AI Vs Predictive Analysis	

2	Predictive Modeling Techniques
	2.1 Factors for choosing Predictive Modeling Techniques
	2.2 Decision Trees –(CART Technique)
	2.3 Regression Techniques: Introduction
	      2.3.1 Linear Regression
	      2.3.2 Multiple Linear Regression,
	      2.3.3 Logistic Regression
	2.4 Neural Networks:
	      2.4.1 Introduction,
	      2.4.2 Architecture of ANN : Artificial vs Biological Neurons, Input Layer, Hidden Layers and Output Layer
	      2.4.3 Introduction to RNN,LSTM,CNN
		8

3	Introduction to Time Series Analysis and Forecasting
	3.1 Time Series Analysis
	      3.1.1 Time Series:Meaning and Utility
	      3.1.2 Components of Time Series
	      3.1.3 Time Series patterns
	      3.1.4 Randomness and Cyclical Pattern in a Time Series
	3.2 Forecasting
	      3.2.1 Forecast accuracy,
	      3.2.2 Auto-regressive and Moving average models
	      3.2.3 Casual models
	      3.2.4 Regression analysis for forecasting
		8

4	Predictive Model Evaluation and Deployment 4.1 Introduction 4.2 Model Validation 4.3 Rule Induction Using CHAID 4.4 Automating Models for Categorical and Continuous targets 4.5 Comparing and Combining Models 4.6 Evaluation Charts for Model Comparison 4.7 MetaLevel Modeling 4.8 Deploying Model 4.9 Assessing Model Performance 4.10 Updating a Model	6

Recommended Books:

1. Data Mining for Business Intelligence by Galit Shmueli, Nitin R. Patel, and Peter C. Bruce, (Wiley, 3rd ed., 2016
2. Predictive & Advanced Analytics (IBM ICE Publication)
3. Predictive Analytics and Data Mining By Vijay Kotu and Bala Deshpande
4. Predictive Analytics, Revised and Updated By Eric Siegel
5. Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd Edition By Dursun Del

Savitribai Phule Pune
University
M.Sc. Data Science
Course Code : DS-604-MJP
Lab Course On - Data Visualization and Analytics

No. of Credits	Teaching Scheme	Examination Scheme-02 Credits Continuous Evaluation -15 End Semester -35

Prerequisites

- Knowledge of Excel, Tableau, and programming languages like Python.
- Prior to starting a data visualization study, people may decide to learn data analytics, design.

Objectives
- Keep a journal of your experiments using tools and strategies for data visualisation.
- Demonstrate proficiency in using data visualization software and applying visualization principles.
- Write down your observations, understandings, and thoughts after each lab session.

Course Outcomes
CO1: Understand and use data visualisation techniques to effectively convey conclusions from an analysis of a dataset.
CO2: Use preprocessing, exploratory data analysis, and data cleaning techniques.
CO3: Demonstrate proficiency in using data visualization tools and software. CO4: Keep track of observations, evaluations, and conclusions from lab work. CO5: Develop your skills in exploratory data analysis, preprocessing, and data cleaning.

Assign No.	Practical Assignments
1.	Utilize a variety of visualization types such as bar charts, line charts, scatter plots, histograms, and heatmaps on weather dataset
2.	Design visually appealing and informative plots with appropriate labels, titles, and legends.on Employee Dataset.

3. Document experiments with advanced visualization techniques such as interactive visualizations

4. Document experiments with advanced visualization techniques such as geospatial mapping,

5. Document experiments with advanced visualization techniques such as network graphs.

6. Record challenges faced and solutions devised during the implementation of advanced visualizations.(use any case study such as cyber crime details , online attendnace , online buying behaviours ,etc)

7. Box Plots: Use box plots to visualize the distribution of data and identify outliers.

8. Heatmaps: Generate heatmaps to visualize correlations or patterns in 2D data.

9. 3D Plots: Experiment with creating 3D plots to visualize 3D data or to plot functions of two variables.

10. Interactive Plots: Explore interactive plotting libraries like Plotly.jl to create interactive visualizations that can be embedded in web pages or notebooks.

11. Geospatial Visualization: Plot data on maps using packages like GeoMaps.jl or Leaflet.jl to visualize geospatial data.

12. Network Visualization: Visualize network graphs using packages like LightGraphs.jl combined with plotting libraries to understand relationships in data.

13. Time Series Visualization: Plot time series data using specialized plots like line plots with time on the x-axis, candlestick plots, or spectral plots.

14. Customizing Plots: Experiment with customizing plots by changing colors, styles, labels, and adding annotations to make your visualizations more informative and appealing.

15. Statistical Plots: Create specialized statistical plots like QQ plots, violin plots, or empirical cumulative distribution function (ECDF) plots to analyze data distributions and relationships.

16. Dashboard Creation: Combine multiple plots into a dashboard using packages like Dashboard.jl to create interactive dashboards for exploring data.

17. Animated Visualizations: Generate animated visualizations using packages like Makie.jl or Plots.jl to show changes in data over time or in response to specific actions.

18. Practice importing data from various sources such as Excel spreadsheets, CSV files, databases (SQL Server, MySQL), online services (Google Analytics, Salesforce), and cloud storage (OneDrive, SharePoint) using POWER BI

19. Explore data transformation techniques such as data cleaning, filtering, merging, appending, grouping, and unpivoting using Power Query Editor using POWER BI

20. Create relationships between tables, define calculated columns, measures, and hierarchies to build a data model that supports your analytical requirements using POWER BI

21. Build different types of visualizations (e.g., bar charts, line charts, pie charts, scatter plots, maps, gauges) to represent your data and convey insights effectively.using POWER BI

22. Experiment with customizing visualizations by adjusting colors, fonts, labels, axes, legends, and formatting options to improve readability and visual appeal using POWER BI

23.Case study -

- Creating a dashboard to depict actionable insights in sales data.

- Marketing Campaign Analysis

- Financial Performance Analysis

- Sales Analysis

- Employee Performance Analysis

- Students Progression Analysis (Slow , Average & Advance Learners)

- E-commerce Analytics

- (Apart from above mentioned case studies internal teacher can consider

  any real time work of students as a case study and consider this case study

  for internal evaluation)

Savitribai Phule Pune University
M.Sc. Data Science (2023)
Course Code : DS-605-MJP

Lab Course On - Lab course Optimization Techniques and Predictive Analysis

DS-605- MJP
Lab course Optimization Techniques and Predictive Analysis

Software to be used:

1. R or R Studio
2. Weka Software

Course Outcomes:

CO1- To perform Analysis of data

CO2- To Performing operation over data

CO3- To Perform Regression and Classification CO4-
To implement Decision tree

CO5- To Understand Model Evaluation CO6-
To Implement Predictive analysis.

Practical Assignments:

Q.1) To Perform Data collection from online, local drive and .csv file.

Q.2) To Perform Data cleaning Operation over the data collected.

Q.3) Perform different data visualization for 2d and 3d visualization.

Q.4) Perform Regression over the dataset.

Q.5) Perform Classification of dataset.

Q.6) Perform decision tree operation over the dataset.

Q.7) Perform market basket analysis over the market data.

Q.8) Implement classical golf case for playing golf game or not.

Q.9) Create a small stock market analysis for bull or bear for a stock in NSE and BSE.

Q.10) Compare any three predictive models with model evaluation report

Savitribai Phule Pune University
M.Sc. Data Science (From 2024-25)

CS-610-MJ : EXPLORATORY DATA ANALYSIS

No. of Credits:02	Teaching Scheme Theory: 02 Hrs/Week	Examination Scheme Continuous Evaluation: 30 Marks End Semester : 70 Marks

Prerequisite
- Concepts of Basic Statistics & Data Mining
- Python Programming and Libraries

Objectives
- To outline an overview of exploratory data analysis.
- To implement data visualization using Python.
- To perform univariate data exploration and analysis.
- To apply bivariate data exploration and analysis.
- To use Data exploration and visualization techniques for multivariate and time series data.

Course Outcomes
On Completion of this course, student will be able to –
CO1: understand concepts of exploratory data analysis
CO2: implement EDS concepts using Python
CO3: aware the concepts of Univariate and bivariate data analysis
CO4: understand the time series concepts

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to exploratory data analysis	05	CO1

1.1 EDA fundamentals
1.2 Difference between Classical Data Analysis and EDA
1.3 Understanding Data Science
1.4 Significance of EDA
1.5 Types of Data, Types of EDA
1.6 Comparing EDA with classical and Bayesian analysis
1.7 Software tools for EDA - Visual Aids for EDA- Data transformation techniques merging database reshaping and pivoting, Transformation techniques

2	EDA using python	08	CO2

2.1 Data Manipulation using Pandas – Pandas Objects
2.2 Data Indexing and Selection – Operating on Data –Handling Missing Data –

Hierarchical Indexing – Combining datasets– Concat, Append, Merge and Join – Aggregation and grouping – Pivot Tables – Vectorized String Operations.			
3	Univariate Analysis	05	CO3

3.1 Introduction to Single variable

3.2 Distribution Variables - Numerical Summaries of Level and Spread- Scaling and Standardizing – Inequality.

4	Bivariate Analysis	05	CO3

Relationships between Two Variables - Percentage Tables - Analyzing Contingency Tables - Handling Several Batches - Scatter plots and Resistant Lines.

5	Multivariate and time series analysis	07	CO4

5.1 Introducing a Third Variable

5.2 Causal Explanations

5.3 Three-Variable Contingency Tables and Beyond

5.4 Fundamentals of Time Series Analysis

5.5 Characteristics of time series data

5.6 Data Cleaning – Time- based indexing – Visualizing – Grouping – Resampling.

Reference Books

- Exploratory Data Analysis Pearson Moden Class
- Python for Data Analysis by WES Mckinney
- Hands on Exploratory Data Analysis with Python Surekumar Mukhiya and Usman Ahmed
- Engineering statistics Handbook
- Introduction to Data Analysis Handbook Migrant and Seasonal head start Web

References:

Exploratory Data Analysis: Visualization | Lecture Notes: Introduction to Data Science (hcbravo.org) chapter4.pdf (cmu.edu)

Chapter 5 Exploratory Data Analysis | Introduction to Data Science (bookdown.org)

Lecture Notes

Lecture2_DescriptiveStats_EDA.ppt (washington.edu)

GitHub - PacktPublishing/Hands-on-Exploratory-Data-Analysis-with-Python: Hands-on Exploratory Data Analysis with Python, published by Packt

Swayam

Data Analysis For Social Science Teachers - Course (swayam2.ac.in) Python for Data Science - Course (nptel.ac.in)

Data Analysis with Python | Coursera

Savitribai Phule Pune University
M.Sc. Data Science (From 2024-25)

DS-611-MJP : Lab on Exploratory Data

Analysis

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

Prerequisite
- Statistics and Mathematics Concepts
- Programming with Python

Objectives
- To write Python programs and applications that make use of exploratory data analysis and data visualization.

Course Outcomes

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

CO1: Design Exploratory data analysis applications using Python.

1. Perform exploratory data analysis (EDA) with datasets like email data set. Export all your
emails as a dataset, import them inside a pandas data frame, visualize them and get
different insights from the data.
2. Explore various variable and row filters in Python for cleaning data. Apply various plot
features in Python on sample data sets and visualize.
3. Perform Time Series Analysis and apply the various visualization techniques.
4. Perform EDA on a Different Data Set.
5. Perform EDA to show outliers and anomalies from given data set.

Savitribai Phule Pune University
M.Sc. (Data Science)
SEMESTER - III
Course code: DS-612-MJ
Total Hours: 30

No. of Credits: 2   IA:15   UE: 35

Course Objectives:

1. Identify business opportunities for data-driven solutions.
2. Develop data-driven solutions to support decision-making in real-world business situations.
3. Understanding Business Data using efficient visualization Techniques

Course Outcomes:
On completion of the course, student will be able to–
CO1: Understand role of IT in Business
CO2: Explain Business Intelligence and its component framework
CO3: Perform different OLAP cube operation
CO4: Evaluate and assess risk in choosing a Business decision
CO5: Develop efficient Dashboard for discover insights and drill down into the data

Unit	Course Contents	ref	No. of Hours	CO/PO targeted
1	**Unit 1: Business view of Information Technology applications –** Business Enterprise Organization, Its functions and Core Business Processes IT in Business, E-Business, needs, challenges, Enterprise applications (CRM/ERP etc.), Information users and their requirements, BI component framework, BI users, BI applications, BI roles and responsibilities		4	CO1, CO2
2	**Unit 2: OLTP, OLAP and Enterprise reporting** Introduction OLTP, Queries scope of OLTP, Advantages and Challenges of OLTP, Introduction of OLAP, One, Two and Three Dimensional Data, Queries scope of OLAP Advantages of OLAP system OLAP Architectures-MOLAP, ROLAP, HOLAP OLTP Vs OLAP Data Models for OLTP and OLAP Role of OLAP tools in the BI architecture Multidimensional Data representation using Data Cube and its Operation-Slice, Dice, Rollup, Drill Down, Pivot, Drill-Across, Drill-Through.	1	6	CO3

	Enterprise Reporting (ER) perspectives, ER characteristics in OLAP, Balanced Scorecard, Dashboards, Scorecards Vs Dashboards			
3	**Unit 3: Decision Analysis [3]** Formulating Decision Problems Decision Strategies without Outcome Probabilities, Decision Strategies for a Minimize Objective • Decision Strategies for a Maximize Objective • Decisions with Conflicting Objectives , Decision Strategies with Outcome Probabilities , Average Payoff Strategy • Expected Value Strategy •Evaluating Risk, Decision Trees , Decision Trees and Monte Carlo Simulation • Decision Trees and Risk • Sensitivity Analysis in Decision Trees , The Value of Information Decisions with Sample Information • Bayes's Rule Utility and Decision Making Constructing a Utility Function • Exponential Utility Functions		8	CO4
4	**Unit 4: Business Analytics Visualisation Tool: Tableau** **Tableau Introduction and Architecture**: Tableau suite and system requirements, Tableau architecture, Data in Tableau, Connecting to data, Managing data source metadata, Working with extracts instead of live connections, Creating and using extracts, Joins and blends Joining tables, Cross database joins, Blending data sources, Filtering data, Filtering discrete and continuous fields, Filtering dates, Other filtering options, Sets in Tableau, Views in Tableau, Formatting in tableau: Font, Alignment, Shading, Border, Lines, Fields, Title, Caption, Workbook-level, Worksheet-level, Field-level formatting, Custom number and Date formatting **Creating charts**: Bar charts and its variations, Bullet chart, Bar-in-bar chart, Variations of date and time visualizations, Gantt Charts, Stacked bars, Treemaps, Area charts, Pie charts Circle charts ,Box and whisker plots, Histograms, Scatterplot, Dual axis and combination charts, Visualizing multiple axes to compare different measures, Visualizing distributions , Visualizing dates and times, Date parts, date values, and exact dates Introduction to calculations , An overview of Table Calculations, Creating and editing Table Calculations, Quick Table Calculations **Data Story with Dashboards:** Designing dashboards in Tableau, Building the views, manipulating objects on Dashboard, Creating the dashboard framework		12	CO5

References

1. Fundamentals of Business Analytics, 2nd Edition, R. N. Prasad, Seema Acharya, Wiley Publication
2. Data Science for Business, Foster Provost & Tom Fawcett, O'REILLY
3. Data Science Fundamentals and Practical Approaches, Gypsy Nandi, Rupam Sharma, BPB Publications, 2020
4. James, E.R. (2017). Business Analytics. UK: Pearson Education Limited.
5. Learning Tableau 2020: Fourth Edition by Joshua N. Milligan, published by Packt Publishing Lt

Savitribai Phule Pune University
M.Sc. (Data Science)
SEMESTER – III
Lab Course on BI

Course code: DS-613-MJP
No. of Credits: 2   IA:15   UE: 35

Course Objectives:

1. Understanding Business Data using efficient visualization Techniques.
2. Explore Tableau as a Visualization tool for Business Data Analytics.

Course Outcomes:
On completion of the course, student will be able to–
CO1: Perform OLAP cube operation.
CO2: Explore and understand business data and draw different charts
CO3: Develop efficient Dashboard for discover insights and drill down into the data

1. OLAP cube operations like slicing filters, drill down, drill up operations

2. Create dashboards by combining multiple visualizations into a single interactive dashboard. This can include filtering, highlighting, and tooltips.
3. Product Sales Data Analysis- sales data analysis project entails analyzing a company's sales data and indicating profit by product, sales, and other significant factors that might influence the company's performance.
4. Employee Performance Metrics- analyze employee performance metrics to assess productivity, attendance, and project completion rates. Create an Interactive Dashboard & Analyzed core underlying causes of poor employee performance. use Microsoft's/kaggle sample dataset.

5. Creating bar charts, line charts, pie charts, and scatter plots using sample datasets
6. creating calculated fields for advanced analysis, such as creating new metrics, performing date calculations, or using logical functions.
7. explore geographical data by creating maps in Tableau, including plotting data points on maps, customizing map layers, and using spatial files.
8. create a data story using Tableau, where they present insights derived from their visualizations in a coherent and engaging narrative format.
9. Create dynamic visualizations that allow users to interactively change aspects of the data being displayed.
10. Business analyst to explore historical stock market data.create compelling visualizations and craft an insightful report – building in-demand skills for communicating data-driven insights.
11. Explore customer data and build profiles of those likely to churn, applying statistics, PivotTables and charts.

Savitribai Phule Pune University,Pune
S.Y.M.Sc.(Data Science) Subject
Code: DS-631-RP

Subject: Research Project Work

No of credits 4

Course Objectives:-

1. Understand the research and research project.
2. Understand how to publish paper in conference.

Course Outcomes:-

After completion of this course the student will be able to-

CO1: Carry out a substantial research-based project

CO2: Demonstrate capacity to improve student achievement, engagement and retention

CO3: Demonstrate capacity to lead and manage change through collaboration with others

CO4: Demonstrate an understanding of the ethical issues associated with practitioner research

CO5: Analyze data and synthesize research findings

CO6: Report research findings in written and verbal forms

CO7: Publish research work in reputable journals, present at conferences or in recognized project competitions.

Course Contents

Sr. No. **Guidelines for Research Project Work**

1 Each student or group of students must submit a detailed project proposal outlining the research problem, objectives, methodology, and expected outcomes.

2 A mentor will be assigned by college to each group of students to provide guidance and support throughout the research process as well as to do internal assessment.

3 Students are required to conduct a thorough literature review to understand the current state of research in their chosen area.

4 Students should execute the research plan outlined in their proposal, adhering to ethical guidelines and academic standards.

5 Proper documentation of the research process, including experimental setup, data collection methods, and analysis techniques, should be maintained

6 Upon completion of the research work, students must prepare a project report and encouraged to publish their research work in reputed journals, present at conferences or in recognized project competitions to disseminate their findings.

8 Evaluation will be as per the University guidelines, based on the quality of the research work, adherence to the research plan, presentation skills, and contribution

Savitribai Phule Pune University
M.Sc. Data Science (2023)

CS-651-MJP : Full Time Industrial Training (IT)

No. of Credits: 16	**Total Duration** 480 Hours	**Examination Scheme** Continuous Evaluation: 120 MarksEnd Semester : 280 Marks

Objectives

- To provide students with an opportunity to apply theoretical knowledge gained throughout the program in a real-world industrial setting
- To foster professional skills such as teamwork, communication, time management, and problem-solving in an industrial environment.
- To expose students to the practices, technologies, and challenges prevalent in the IT industry or related sectors.
- To enable students to gain hands-on experience by working on projects or tasks relevant to their field of study.
- To facilitate networking opportunities with professionals in the industry, potentially leading to future career prospects..

Course Outcomes

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

CO1: Apply theoretical concepts learned in the classroom to solve practical problems encountered in an industrial setting.

CO2: Demonstrate proficiency in using industry-standard tools, technologies, and methodologies relevant to their area of specialization.

CO3: Apply analytical and problem-solving skills to address challenges encountered during the industrial training

CO4: Collaborate effectively with team members to achieve project goals and objectives.

CO5: Manage time and resources efficiently to complete assigned tasks and projects within the stipulated timeframe.

CO6: Prepare a comprehensive report documenting their experience, including project details, learnings, and reflections.

Sr. No.	Guidelines for Full Time Industrial Training (IT)
1	Students are required to secure an industrial/internship placement in any organization, institution, or IT industry relevant to their field of study.
2	Students must submit the offer letter from the organization within two weeks of starting the industrial training/internship, detailing the terms and duration of the internship.
3	Students must have to work full time in the organization as per their rules and regulations.

4	A mentor will be assigned to each group of students to provide guidance and support throughout the internship period.
5	The industrial training/ internship duration should span a minimum of 360 hours, equivalent to 12 credits.
6	Students may be assigned specific projects or tasks or assignments by the host organization, relevant to their area of specialization.
7	Students should provide regular updates to their mentor through progress reports time to time regarding their progress, challenges faced, and lessons learned during the industrial training.
8	Upon completion of the industrial training/ internship, students must submit a comprehensive report documenting their internship experience, including project/ assignment details, challenges and achievements as per the format specified.
9	Evaluation will be based on the quality content of the internship report, feedback from the host organization, and the overall performance during the internship/ industrial training period.

Evaluation Pattern

- Internal assessment will be carried by college guide/ mentor by continuous evaluationmethod.
- The final examination or presentation of the work carried during the training/internshipperiod will be in front the panel of examiners as per the schedule given by University.
- There will be a panel of three examiners for the final assessment
  1. Industry expert (Appointed by the college)
  2. Academic expert (Appointed by the University)
  3. College guide/Mentor of the student as an internal examiner

Parameters for Evaluation	Marks
Internal Assessment by Mentor Regular updates, timely report submission and deliverable (30 Marks) Professional Conduct, Learning and Skill Development (30 Marks) Work Undertaken and Learning Outcomes (30 Marks)	**120**
External Assessment by Industry expert and Academic expert	**280**
Relevance and significance of the project or tasks undertaken (40 Marks) Technical proficiency demonstrated during the internship (40 Marks) Communication skills and presentation of the internship experience (40 Marks) Work Undertaken and Quality of the internship report (40 Marks) Overall performance and contribution to the organization (50 Marks)	
Total Assessment	**400**

Savitribai Phule Pune University
M.Sc.(Data Science)
Progress Report for CS-651-MJP : Full Time Industrial Training (IT)
(This Progress report is to be submitted monthly to the college guide/Mentor)

Name of College	
Roll No./ID and Name of Student	
Date of Report Submission	
Duration of Report (From date – To date)	
Name of Organization	
Date of Joining in the organization	
Name of Industry Guide/Supervisor	
Name of College Guide/Mentor	

1. **Introduction** *(Mention brief overview of the internship objectives and the role of the studentwithin the organization)*

2. **Work Undertaken** *(Summary of the tasks or assignments or projects undertaken by the student during the reporting period with responsibilities assigned and progress made on eachtask)*

3. **Learning and Skill Development** *(Mention the summary of new skills, knowledge, andexperiences gained during the reporting period)*

4. **Challenges Faced** *(Put any challenges or obstacles encountered during the reporting periodand Strategies adopted to overcome these challenges and lessons learned from them)*

5. **Achievements and Contributions** *(Highlight of notable achievements, contributions, orsuccesses attained during the reporting period)*

6. **Future Plan** *(Specify future goals and objectives for the remaining period of the internship.Also put the plan of action to address any identified areas for improvement or skill enhancement)*

Signature: Signature:

Date: Date:

Name: Name:

Industry Guide/Mentor   **College Guide/Mentor**

SAVITRIBAI PHULE PUNE UNIVERSITY
SECOND YEAR M. Sc. (Data Science)
SEMESTER IV
CA 681 RP: Research Work – II

Teaching Scheme: Laboratory: 12 Hrs./Week	Credits 06	Examination Scheme: Continuous Evaluation: 45 Marks End-Semester : 105 Marks

Course Objectives

- To get first-hand experience to apply research methodology

Course Outcomes

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

- Apply research methodology to carry out research in a chosen problem domain
- Design and develop a novel methodology / framework etc
- Conduct experiments and analyze results

Guidelines for carry out Research Work

- Each student shall carry out the research work during semester IV under the guidance of the appointed faculty Advisor/Mentor. Preferably this work may be an extension of research work carried out by a student as a part of Research Work – I in Semester III.
- Students shall work on a research problem and publish a paper / file a copyright / patent based on the work carried out.
- The student shall prepare and submit a report based on the work carried out consisting of – Face Page, certificate, Acknowledgement, Abstract, Table of Contents, List of Tables, List of Figures, Abbreviations, and separate Chapters dealing with Introduction, Literature Review, Design details of Proposed System, Experimental Results and analysis, and a chapter providing Conclusions and future scope.
- List of Publications, Copyright/patent, references and appendix shall also be included in the report.

Guidelines for Assessment

The work carried out shall be evaluated on a continuous basis by the assigned faculty advisor / mentor for 45 marks and panel of examiners appointed shall evaluate the work based on the report for 105 marks.