

Branch Name:	MCA
Program Code:	CS201
Course Name:	Data Analytics using Python
Course Code:	4CS2010204T
Pre-requisite Course:	Basic Statistics and Mathematics, Fundamentals of Programming

Course Objective:

1. The course provides a comprehensive foundation in data analytics using Python.
2. It covers key topics such as statistical analysis, hypothesis testing, regression, classification, and clustering techniques.
3. Students will develop practical skills in data-driven decision-making.
4. Real-world datasets and analytics tools are used to enhance hands-on learning.
5. The course prepares students for careers in data science, analytics, and applied research.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture	Tutorial	Practical	Credit	Theory		Practical		Total
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
3	-	-	3	60	40	-	-	100

Course Contents:

Unit No	Topics	Total Hours	Weightage (%)
1	Introduction to Data Analytics and Python Fundamentals • Basics of Data Analytics: Definition, Applications, and Process • Introduction to Python for Data Analysis ◦ Python Environment Setup ◦ Data Types, Variables, and Operators ◦ Control Flow and Functions • Working with Libraries: NumPy, Pandas, Matplotlib • Importing, Cleaning, and Exploring Data • Exploratory Data Analysis (EDA) Basics	11	25

2	Foundations of Statistical Inference • Introduction to Probability ◦ Basic Concepts and Rules ◦ Conditional Probability and Bayes' Theorem • Sampling and Sampling Distributions ◦ Central Limit Theorem ◦ Estimators and Confidence Intervals • Hypothesis Testing: ◦ Null and Alternative Hypotheses ◦ Type I & II Errors, p-value, z-test, t-test	05	10
3	Comparative Tests and ANOVA • Two-Sample Testing: Independent and Paired Samples • One-Way ANOVA: Concepts and Implementation • Two-Way ANOVA: Interaction Effects • Introduction to Linear Regression ◦ Least Squares Method ◦ Interpretation of Coefficients	05	10
4	Advanced Regression Techniques • Simple and Multiple Linear Regression • Assumptions and Diagnostics • Introduction to Maximum Likelihood Estimation (MLE) • Logistic Regression: Concept and Implementation • Model Evaluation: ◦ ROC Curve ◦ Accuracy, Precision, Recall, F1-score	12	30
5	Advanced Analytics – Clustering and Decision Trees • Chi-Square (χ^2) Test: Goodness of Fit and Independence • Introduction to Cluster Analysis ◦ K-Means Clustering ◦ Hierarchical Clustering • Classification and Regression Trees (CART) ◦ Tree Construction and Pruning ◦ Use Cases and Performance Metrics	12	25

Textbooks:

1. McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc."
2. Swaroop, C. H. (2003). A Byte of Python. Python Tutorial.
3. Ken Black, sixth Editing. Business Statistics for Contemporary Decision Making. "John Wiley & Sons, Inc".
4. Anderson Sweeney Williams (2011). Statistics for Business and Economics. "Cengage Learning".

5. Douglas C. Montgomery, George C. Runger (2002). Applied Statistics & Probability for Engineering. "John Wiley & Sons, Inc"
6. Jay L. Devore (2011). Probability and Statistics for Engineering and the Sciences. "Cengage Learning".
7. David W. Hosmer, Stanley Lemeshow (2000). Applied logistic regression (Wiley Series in probability and statistics). "Wiley-Interscience Publication".
8. Jiawei Han and Micheline Kamber (2006). Data Mining: Concepts and Techniques. "
9. Leonard Kaufman, Peter J. Rousseeuw (1990). Finding Groups in Data: An Introduction to Cluster Analysis. "John Wiley & Sons, Inc".

List of Useful websites / MOOCs

1. Learners are advised to opt for NPTEL and SWAYAM courses that are relevant to this course

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Apply fundamental Python programming and libraries (Pandas, NumPy, Matplotlib) for data exploration and visualization.	Applying
CLO2	Understand and implement basic probability concepts, sampling techniques, and hypothesis testing for data-driven inference.	Understanding Analyze
CLO3	Analyze and interpret results from ANOVA and linear regression techniques, both simple and multiple.	Analyze
CLO4	Implement logistic regression and evaluate classification models using ROC and other performance metrics.	Evaluate
CLO5	Conduct clustering analysis and build classification trees (CART) for predictive modeling.	Create
CLO6	Develop complete end-to-end analytics workflows, including model building, evaluation, and interpretation of results.	Create Evaluate

Mapping of CLOs with Pos & PSOs

Course Learning Outcomes	Program Out comes (POs)												Program Specific Outcomes (PSOs)	
	PO1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2
CLO1	H	M	M	M	H	L	M	L	L	L	L	L	H	L
CLO2	H	H	M	H	M	L	M	L	L	M	L	L	M	M
CLO3	H	H	H	H	M	L	M	L	M	M	M	L	H	M
CLO4	H	H	H	H	H	M	M	L	M	M	M	M	H	H
CLO5	H	H	H	H	H	L	M	L	M	M	M	M	H	H
CLO6	H	H	H	H	H	M	H	M	H	M	H	H	H	H

H: High, M: Medium, L: Low