

Branch Name:	MCA
Program Code:	CS201
Course Name:	Data Analytics using Python Practical
Course Code:	4CS2010204P
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	
-	-	4	2	-	-	25	25	50

Tools and Libraries Used for Practical:

- Python (Jupyter Notebook)
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- Statsmodels
- Scipy
- Other required Tools & Library

Suggestive Practical List:

1	Python for Data Analysis • Setting up Python environment (Jupyter, Anaconda) • Basic Python exercises (loops, functions, lists, dictionaries) • Data import and export (CSV, Excel) • Data cleaning using Pandas • Descriptive statistics and visualizations using Matplotlib and Seaborn
2	Probability and Sampling • Simulating probability experiments using Python • Visualizing probability distributions (normal, binomial, etc.) • Random sampling from datasets using Pandas/NumPy • Demonstration of Central Limit Theorem using simulations

3	Hypothesis Testing - Performing one-sample and two-sample t-tests using scipy.stats - Applying z-test for proportions - Conducting chi-square test for independence and goodness-of-fit - Visualizing test statistics and p-values
4	ANOVA and Regression Analysis - One-way and Two-way ANOVA using Python (e.g., statsmodels) - Fitting simple linear regression models - Multiple linear regression with feature selection - Residual analysis and checking assumptions (normality, multicollinearity)
5	Logistic Regression and ROC Analysis - Building logistic regression models with categorical and numerical variables - Confusion matrix, ROC curve, AUC calculation - Model evaluation metrics: accuracy, precision, recall, F1-score - Practical case study on binary classification (e.g., loan approval, disease prediction)
6	Clustering and CART - K-Means and Hierarchical Clustering on real datasets - Dendrograms and cluster visualization - Building Classification and Regression Trees (CART) - Feature importance and pruning techniques

Textbooks:

- McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc."
- Swaroop, C. H. (2003). A Byte of Python. Python Tutorial.
- Ken Black, sixth Editing. Business Statistics for Contemporary Decision Making. "John Wiley & Sons, Inc"
- Anderson Sweeney Williams (2011). Statistics for Business and Economics. "Cengage Learning".
- Douglas C. Montgomery, George C. Runger (2002). Applied Statistics & Probability for Engineering. "John Wiley & Sons, Inc"
- Jay L. Devore (2011). Probability and Statistics for Engineering and the Sciences. "Cengage Learning".
- David W. Hosmer, Stanley Lemeshow (2000). Applied logistic regression (Wiley Series in probability and statistics). "Wiley-Interscience Publication".
- Jiawei Han and Micheline Kamber (2006). Data Mining: Concepts and Techniques. "
- Leonard Kaufman, Peter J. Rousseeuw (1990). Finding Groups in Data: An Introduction to Cluster Analysis. "John Wiley & Sons, Inc"

List of Useful websites / MOOCs

- 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	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
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