

Course Code: 602
Course Title: Data Analytics using Python

Course Code	602																																																						
Course Title	Data Analytics using Python																																																						
Credits	4																																																						
Course Category	Major Course																																																						
Level of Course	400-499 (Advance Level)																																																						
Teaching Hours	30 Hours of class-room teaching + 60 Hours of Applied work(Project)																																																						
Minimum Hours/ Semester	30 hours of Theory + 60 Hours of Applied work (Project) (Including class work, examination, preparation etc.)																																																						
Review / Revision	-																																																						
Implementation Year:	Only during A.Y. 2025-2026																																																						
Purpose of Course	This course aims to introduce students to data analytics techniques using Python, with a focus on Exploratory Data Analysis (EDA), regression, and supervised learning. It equips learners with practical skills in handling data, automating EDA, and applying machine learning concepts in real-world scenarios.																																																						
Course Objective	1) Explain the concepts and techniques of exploratory data analysis including handling missing data, outliers, and distribution types. 2) Apply Python libraries like Pandas and NumPy to automate and perform efficient data analysis tasks. 3) Describe the basics of regression and correlation, and understand their role in data analysis and machine learning. 4) Understand key concepts of machine learning including types, benefits, and real-world applications. 5) Evaluate supervised learning models using loss functions such as MSE and MAE, and explain concepts like overfitting and underfitting.																																																						
Pre-requisite	Students should have a basic understanding of Python programming and core concepts of statistics and probability. Familiarity with libraries like Pandas and NumPy will be helpful but not mandatory.																																																						
Course Outcomes	CO1: Understanding: Explain key concepts of exploratory data analysis, including univariate, bivariate, and multivariate techniques, and describe data distributions. CO2: Applying: Use Python libraries like Pandas and NumPy to perform and automate exploratory data analysis on various datasets. CO3: Analyzing: Analyze relationships between variables using regression techniques, and interpret covariance and correlation in datasets. CO4: Understanding & Applying: Describe the types and benefits of machine learning, and apply basic ML lifecycle steps to real-world problems. CO5: Evaluating: Evaluate supervised learning models using loss functions such as MSE and MAE, and assess issues like overfitting and underfitting.																																																						
Mapping between Course Outcomes(CO) with Program Specific Outcomes(PSO)	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th>PSO1</th> <th>PSO2</th> <th>PSO3</th> <th>PSO4</th> <th>PSO5</th> <th>PSO6</th> <th>PSO7</th> <th>PSO8</th> </tr> </thead> <tbody> <tr> <td>CO1</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CO2</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CO3</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CO4</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CO5</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>		PSO1	PSO2	PSO3	PSO4	PSO5	PSO6	PSO7	PSO8	CO1									CO2									CO3									CO4									CO5								
	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6	PSO7	PSO8																																															
CO1																																																							
CO2																																																							
CO3																																																							
CO4																																																							
CO5																																																							
Course Content	Unit-1 : Fundamentals of Data Analytics : 1.1 Exploratory Data Analysis (EDA) 1.1.1 Types of Exploratory Data Analysis: 1.1.2 Univariate Analysis 1.1.3 Bivariate Analysis 1.1.4 Multivariate Analysis 1.1.5 Handling Missing Data and Outliers 1.2 Understanding the Data: 1.2.1 Quantitative Data : Discrete and Continuous																																																						

- 1.2.2 Qualitative Data : Non-numerical (Normal and Ordinal)
- 1.3 Spread of Data
 - 1.3.1 Normal Distribution
 - 1.3.2 Skewed Distribution
 - 1.3.3 Skewness and Kurtosis

Unit-2 Automate EDA (Exploratory Data Analysis)

- 2.1 Python Libraries to Automate Exploratory Data Analysis
 - 2.1.1 Pandas and Numpy
- 2.2 Regression
 - 2.2.1 Characteristics of Regression
 - 2.2.2 Dependent and Independent variables
 - 2.2.3 Covariance and Correlation
- 2.3 Machine Learning Basics:
 - 2.3.1 Concepts of Machine learning
 - 2.3.1.1 Understanding machine learning
 - 2.3.1.2 Benefit of machine learning
 - 2.3.1.2 Machine learning life cycle
- 2.4 Types of Machine Learning:
 - 2.4.1 Supervised and Unsupervised Learning
 - 2.4.2 Applications of ML in real-world scenarios

Unit-3: Understanding Supervised Learning

- 3.1 Overview of Supervised Learning
 - 3.1.1 Concepts of Supervised Learning
 - 3.1.2 Difference between Classification and Regression
- 3.2 Basic Terminologies
 - 3.2.1 Dataset, Features, Labels
 - 3.2.2 Training Data, Test Data, validation Data
 - 3.3.3 Overfitting, Underfitting
- 3.3 Loss functions :
 - 3.3.1 Mean Squared Error (MSE)
 - 3.3.2 Definition of MSE
 - 3.3.2.1 Computing MSE and its properties
 - 3.3.3 Mean Absolute Error (MAE)
 - 3.3.3.1 Definition of MAE
 - 3.3.3.2 Computing MAE and its properties
 - 3.3.3.3 Understanding Regression and R².

Unit-4 : Vedic Mathematics Sutras :

- 4.1 Nikhilam Navatashtcaramam Dashataha : "All from 9 and the last from 10."
- 4.2 Ekadhikena Purvena : "By one more than the previous one."
- 4.3 Udharan : "The extraction."
- 4.4 Paraavartya : "Transposition and cancellation."
- 4.5 Shunyam Saamyasamuccaye : "When the sum is the same that sum is zero."
- 4.6 Anurupyena : "Proportionately."
- 4.7 Sankalana-Vyavakalanabhyam : "By addition and by subtraction."
- 4.8 Puranapuranabhyam : "By the completion or non-completion."
- 4.9 Chalana-Kalana : "By motion or by applying a shift."
- 4.10 Yavadunam : "Whatever is the deficiency."
- 4.11 Vyastisamanstih : "The parts and the whole."
- 4.12 Sesanyan : "The remainder."
- 4.13 Gunitasamuchyah : "The product of the sum."
- 4.14 Vistaran : "Expansion."
- 4.15 Rupan : "Form."
- 4.16 Chidana : "By splitting."

[Implementation of all sutras of Unit-4 in computer Lab. Using C / Python / Any Prog. Language]

[Students will submit E-Document for Project report.

One internal guide will be allocated for every ten groups

	<p>All groups are required to contact their internal guides once a week to endorse their project progress work.] [Students need to complete their project by 10th of February/15th of August(Winter session)]</p> <p>[All Units carry Equal Weightage]</p>
Reference Books	<p>1) Data Analytics Using Python, Bharti Motwani, Wiley India Pvt. Ltd., ISBN: 978-93-87034-42-0</p> <p>2) Data Science and Analytics, V.K. Jain, Khanna Publishing House, ISBN: 978-93-86173-66-3</p> <p>3) Data Analytics with Python, R. N. Prasad, Seema Acharya, Wiley India Pvt. Ltd., ISBN: 978-81-265-9337-4</p> <p>4) Foundations of Data Science, Avrim Blum, John Hopcroft, Ravindran Kannan, Universities Press (India) Pvt. Ltd., ISBN: 978-93-86279-47-1</p> <p>5) Python for Data Analysis, Reema Thareja, Oxford University Press India, ISBN: 978-0-19-948017-1</p> <p>6) Python for Data Analysis, Wes McKinney, O'Reilly Media, ISBN: 978-1-491-95766-0</p> <p>7) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, Aurélien Géron, O'Reilly Media, ISBN: 978-1-492-03264-1</p> <p>8) Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media, ISBN: 978-1-491-91205-8</p> <p>9) An Introduction to Statistical Learning, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, ISBN: 978-1-4614-7138-7</p> <p>10) Think Stats: Exploratory Data Analysis in Python, Allen B. Downey, O'Reilly Media, ISBN: 978-1-449-39416-6</p>
Teaching Methodology	Class Work, Discussion, Presentation, Self-Study, Seminars and/or Assignments
Evaluation Method	<p>50% Internal assessment.</p> <ul style="list-style-type: none"> - Attendance, Class and home Assignment, Unit tests. - Internal project presentation and demonstration, project documentation. <p>50% External assessment.</p> <ul style="list-style-type: none"> - Theory / written examination - project presentation and demonstration, viva-voce and e-project report.