FACULTY OF COMPUTER APPLICATIONS
Bachelor of Science – Information Technology

- **Sem.** : 5
- **Subject Code** : 05BS0505
- **Subject** : Data Analytics
- **Course Objectives** :
    1. To introduce students to the fundamentals of data analytics, including types of data, data analysis techniques, and the data analytics life cycle.
    2. To develop proficiency in using Python for data analytics, including reading and writing data, working with data frames, and using data frame methods.
    3. To introduce statistical analysis techniques used in data analytics, including mean, median, mode, standard deviation, and variance and how to use the Scikit-learn library for data preprocessing.
    4. To teach students data visualization techniques using various plotting systems, including direct plotting, Seaborn plotting, and Matplotlib plotting.
    5. To provide an overview of various data analytics techniques, including quantitative and qualitative analysis, data mining, statistical analysis, machine learning, semantic analysis, and visual analysis.

- **Prerequisites :** Knowledge of Python programming language

Unit No.	Topics Covered	No. of lectures required
1	**Introduction of Data and Data Analytics:** Data, Dataset, Data Analysis, Data Analytics, Types of Digital Data: classification of Data (Structured, semi structured and unstructured), Characteristics of Data, Types of Data Analytics, Data Analytics Life Cycle Overview, Phases (Discovery, Data Preparation, Model Planning, Mode Building,	10

	Communicate Results, Operationalize), Data analytics real world applications	
2	**Python for data analytics:** Reading and writing data in Python, Data Frame and its methods Statistical analysis – Mean, Median, Mode, Q1, Q2 and Q3, standard deviation, variance	**08**
3	**Scikit-learn library for data preprocessing:** Filtering and Handling missing data, data scaling (StandardScaler, MinMaxScaler), categorical data to numerical encoding (LabelEncoder), feature subset selection, dimensionality reduction using PCA	**12**
4	**Data Visualization:** Direct Plotting – Line Plot, Bar Plot, Pie Chart, Box Plot, Histogram Plot, Scatter Plot Seaborn Plotting System - Strip Plot, Box Plot, Swarm Plot, Joint Plot Matplotlib Plot – Line Plot, Bar Chart, Histogram Plot, Scatter Plot, Stack Plot, Pie Chart	**10**
5	**Data Analytics Techniques:** Quantitative Analysis, Qualitative Analysis , Data Mining, Statistical Analysis (A/B Testing, Correlation, Regression), Machine Learning(Classification, Clustering, Outlier Detection, Filtering), Semantic Analysis, Visual Analysis	**10**

Course Outcomes : Students will able to

1. Understand the basics of data analytics, including types of data, data analysis techniques, and the data analytics life cycle.
2. Develop proficiency in using Python for data analytics, including reading and writing data, working with data frames, and using data frame methods.
3. Understand statistical analysis techniques and become proficient in using the Scikit-learn library for data preprocessing.
4. Develop skills in data visualization using various plotting systems, including direct plotting, Seaborn plotting, and Matplotlib plotting.
5. Understand various data analytics techniques, including quantitative and qualitative analysis, data mining, statistical analysis, machine learning, semantic analysis, and visual analysis.

FACULTY OF COMPUTER APPLICATIONS
Bachelor of Science – Information Technology

Course Outcomes – Program Outcomes Mapping Table :

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	L					H	H	L
CO2	L	L	M			L	H	H
CO3	L	M	M			L	H	H
CO4	L	M	M	H		L	H	H
CO5	L	H	H		M	L	H	H

Main References :

1. "Big Data and Analytics", Seema Acharyas, Subhashini Chellappan, Wiley, First Edition
2. "Python for Data Analysis" by Wes McKinney, O'REILLY, 3rd Edition
3. Thomas Erl,Wajid Khattak,and Paul Buhler: Big Data Fundamentals: Concepts,Dribers and techniques , Pearson, First Edition

Other Reference :

1. Data Science from Scratch: First Principles with Python" by Joel Grus, published by O'Reilly Media, 2nd edition.
2. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems" by Aurélien Géron, published by O'Reilly Media, 2nd edition.
3. "Data Analytics Made Accessible" by Anil Maheshwari, published by Apress, 1st edition.
4. "Data Analytics Handbook: A Guide for Beginners" by Samir Gupta, published by Independently Published, 1st edition.

Web Resources

1. Kaggle: https://www.kaggle.com/
2. DataCamp: https://www.datacamp.com/
3. GitHub: https://github.com/
4. Towards Data Science: https://towardsdatascience.com/
5. Analytics Vidhya: https://www.analyticsvidhya.com/

FACULTY OF COMPUTER APPLICATIONS
Bachelor of Science – Information Technology

Unit wise coverage from Text book(s) and Web Resources:

Unit 1	Book#	Topics
I	1	Chapter 1
II	2	Chapter 5
III	2	Chapter 6, 7
IV	2	Chapter 9
V	3	Chapter 7

FACULTY OF COMPUTER APPLICATIONS
Bachelor of Science – Information Technology

Lab Practical

Unit No	List of Practicals
1 (Theory u-2)	1. Write Python script to read and write excel 2. Write Python script to read and write csv file. 3. Write Python script to read and write XML file. 4. Write Python script to read and download Web data file. 5. Write Python script to workJSON file. 6. Write Python script to work with Database file. 7. Read iris.csv file and perform all the data frame operations. 8. Write python program to calculate Mean, Median, Mode, Q1, Q2 and Q3, standard deviation, and variance.
2 (Theory u-3)	Download Auto-mpg dataset and perform following operations : 1. read data 2. Performing Data Cleaning 3. Handling Missing Data 4. Removing Null data 5. Scaling Data 6. Encoding Data 7. Feature Selection 8. Implement Principle Component Analysis,
3 (Theory u-4)	1. Create 500 random temperature readings for sixcities over a season and then plot the generated datausing Matplotlib. 2. Load the well-known Iris data set, which lists measurements of petals and sepals of three iris species. Then plot the correlations between each pair using the .pairplot() method. 3. Load the well-known Tips data set, which shows the number of tips received by restaurant staff based on various indicator data; then plot the percentage of tips per bill according to staff gender. 4. Load the well-known Tips data set, which shows the number of tips received by restaurant staff based on various indicator data; then implement the factor plots to visualize the total bill per day according to staff gender. 5. Reimplement the above exercise using the Seaborn joint plot distributions.

	6. Python program of Barplot with all parameters of a sample data. 7. Python program of Pie-chart with all parameters of a sample data. 8. Python program of Histogram with all parameters of a sample data. 9. Python program of Line Plot with all parameters of a sample data. 10. Python program of Scatter plot with all parameters of a sample data.