



**Computer Science paper-IV  
Foundations of Data Science  
[Discipline Specific Course]**

<b>Semester: V</b>	<b>Credits: 02</b>	<b>Subject Code: BS52204</b>	<b>Lectures: 36</b>
--------------------	--------------------	------------------------------	---------------------

<b>Course Outcomes:</b>
<b>At the end of this course, the learner will be able to:</b>
<ul style="list-style-type: none"> <li>● Apply the fundamental concepts and techniques in data science.</li> <li>● Learn how to obtain data from different sources, clean/process, and transform data</li> <li>● Develop the ability to build and assess data-based models using statistical concepts</li> <li>● Represent data using various data visualization tools and techniques.</li> </ul>

<b>Unit 1: Introduction to Data science</b>	<b>06</b>
<ul style="list-style-type: none"> <li>● What is Data Science?</li> <li>● Applications of Data science</li> <li>● Big Data and Data Science</li> <li>● The Data Science Lifecycle</li> <li>● Types of Data             <ul style="list-style-type: none"> <li>○ Structured, semi-structured, Unstructured Data, Problems with unstructured data</li> <li>○ Data sources</li> <li>○ Open Data, Social Media Data, Multimodal Data, standard datasets Data Formats Integers, Floats, Text Data, Text Files, Dense Numerical Arrays, Compressed or Archived Data, CSV Files, JSON Files, XML Files, HTML Files , Tar Files, GZip Files, Zip Files, Image Files: Rasterized, Vectorized, and/or Compressed</li> </ul> </li> </ul>	

<b>Unit 2: Statistical Data Analysis</b>	<b>10</b>
<ul style="list-style-type: none"> <li>● Descriptive Statistics: Measuring the Frequency.</li> <li>● Measuring the Central Tendency: Mean, Median, and Mode</li> <li>● Measuring the Dispersion: Range, Standard deviation, Variance, Interquartile Range Statistical Inference: Populations and samples</li> <li>● Statistical modeling - probability distributions - fitting a model - Hypothesis Testing.</li> <li>● Measuring Data Similarity and Dissimilarity             <ul style="list-style-type: none"> <li>○ Data Matrix versus Dissimilarity Matrix, Proximity Measures for Nominal Attributes, Proximity Measures for Binary Attributes, Dissimilarity of Numeric Data: Euclidean, Manhattan, and Minkowski distances, Proximity Measures for Ordinal Attributes</li> </ul> </li> <li>● Concept of Outlier, types of outliers, outlier detection methods</li> </ul>	

<b>Unit 3: Data pre-processing</b>	
<ul style="list-style-type: none"> <li>● Data Objects and Attribute Types: What Is an Attribute?, Nominal , Binary, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes</li> <li>● Data cleaning –</li> </ul>	<b>10</b>

Board of Studies	Name	Signature
Chairperson	Ashwini Kulkarni	



<ul style="list-style-type: none"> <li>○ Missing Values, Noisy Data (Duplicate Entries, Multiple Entries for a Single Entity, Missing Entries, NULLs, Huge Outliers, Out-of- Date Data, Artificial Entries, Irregular Spacings,</li> <li>○ Formatting Issues - Irregular between Different Tables/Columns, Extra Whitespace, Irregular Capitalization, Inconsistent Delimiters, Irregular NULL Format, Invalid Characters, Incompatible Datetimes)</li> <li>● Data Transformation–             <ul style="list-style-type: none"> <li>○ Rescaling, Normalizing, Binarizing, Standardizing, Label and One Hot Encoding</li> </ul> </li> <li>● Data reduction and Data Discretization</li> </ul>	
--	--

<b>Unit 4: Data Visualization</b>	<b>10</b>
<ul style="list-style-type: none"> <li>● What is Data Visualization?</li> <li>● Benefits of Good Data Visualization</li> <li>● Purpose of Data visualization</li> <li>● Different Types of Analysis for Data Visualization             <ul style="list-style-type: none"> <li>○ Univariate Analysis Techniques for Data Visualization, Bivariate Analysis , Multivariate analysis</li> </ul> </li> <li>● Techniques for Data Visualization</li> <li>● Tools for data visualization.(PRACTICAL IMPLEMENTATIONS)             <ul style="list-style-type: none"> <li>○ Histograms, Bar charts/graphs, Scatter plots, Line charts, Area plots, Pie charts, Donut charts</li> <li>○ Specialized data visualization tools Boxplots, Bubble plots, Heat map, Dendrogram, Venn diagram, Treemap, 3D scatter plots</li> <li>○ Advanced data visualization tools- Word clouds</li> <li>○ Visualization of geospatial data open source libraries</li> <li>○ Introduction to Tableau.</li> </ul> </li> <li>● Data Visualization types</li> </ul>	

<b>Recommended Basic Reading:</b>
<ul style="list-style-type: none"> <li>● <b>Peter Bruce, Andrew Bruce &amp; Peter Gedeck, <i>Practical statistics for data scientists</i>, O'reilly</b></li> </ul>

<b>Recommended Reference Books:</b>
<ul style="list-style-type: none"> <li>● Cathy O'Neil and Rachel Schutt , "<i>Doing Data Science</i>", O'Reilly, 2015.</li> <li>● Chirag Shah , <i>A Hands-On Introduction to Data Science</i>, Washington Cambridge University Press</li> <li>● Field Cady, John Wiley &amp; Sons , <i>The Data Science Handbook</i>, Inc, 2017</li> <li>● Gypsy Nandi, Rupam Sharma, <i>Data Science Fundamentals and Practical Approaches</i>, BPB Publications, 2020</li> </ul>

<b>Journals:</b>
<ul style="list-style-type: none"> <li>● Springer.com</li> </ul>
<b>Websites:</b>
<ul style="list-style-type: none"> <li>● Github</li> <li>● Kaggle.com, Google colab</li> </ul>

Board of Studies	Name	Signature
Chairperson	Ashwini Kulkarni	



<b>E-resources:</b>
• MOOCS -Data camp, Swayam etc..

Board of Studies	Name	Signature(in white cell)	
Chairperson (HoD)	Ms. Ashwini Kulkarni		
Faculty	Ms. Ashwini Kulkarni		
Faculty	Ms. Alka Kalhapure		
Subject Expert (Outside SPPU)	Prof. Mr. Aniket Nagane		
Subject Expert (Outside SPPU)	Dr. Manisha Divate		
VC Nominee	Dr. Manisha Bharambe		
Industry Expert	Ms. Snehal Biyala		
Alumni	Ms. Mamta Choudhary		

Board of Studies	Name	Signature
Chairperson	Ashwini Kulkarni	