1st Year Syllabus

Course Code	CSL109	Course Title	Introduction to Data and Analytics			
Category	Core	Credit Assigned	L	T	P	C
			3	0	2	4
Pre-requisite (If any)	-	Type of Course	CSE with DSA			

Course Outcomes:

1. Identify and describe the methods and techniques commonly used in data science.
2. Demonstrate proficiency with the methods and techniques for obtaining, organizing, exploring, and analyzing data.
3. Recognize how data analysis, inferential statistics, modeling, machine learning, and statistical computing can be utilized in an integrated capacity.
4. Create and modify customizable tools for data analysis and visualization per the evaluation of characteristics of the data and the nature of the analysis.
5. Demonstrate the ability to clean and prepare data for analysis and assemble data from a variety of sources.

Course Contents:

Module – I: Introduction

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

Module – II: Data Collection and Data Pre-Processing

Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Module – III: Exploratory Data Analytics

Descriptive Statistics – Mean, Standard Deviation, Skewness, and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA.

Module -IV: **Data Definitions and Analysis Techniques**

Elements, Variables, and Data categorization, Levels of Measurement, Data management and indexing, Introduction to statistical learning

Module -V: **Descriptive Statistics**

Measures of central tendency, Measures of the location of dispersions, Practice, and analysis Statistical hypothesis generation, and testing, Chi-Square test, t-Test, Analysis of variance, Correlation analysis, Maximum likelihood test, Practice, and analysis

Text Books:

1. An introduction to Data Science by Jeffrey Stanton
2. The Elements of Data Analytic Style by Jeff Leek
3. Exploratory Data Analysis with R, by Roger Peng
4. OpenIntro Statistics, by Diez, Barr, and Centinkaya-Rundel
5. R Programming for Data Science, by Roger Peng

References:

1. UC Irvine Machine Learning Repository https://archive.ics.uci.edu/ml/index.php