CS3353 FOUNDATIONS OF DATA SCIENCE L T P C 3 0 0 3

COURSE OBJECTIVES:

- · To understand the data science fundamentals and process.
- · To learn to describe the data for the data science process.
- To learn to describe the relationship between data.
- · To utilize the Python libraries for Data Wrangling.
- · To present and interpret data using visualization libraries in Python

UNIT I INTRODUCTION 9

Data Science: Benefits and uses – facets of data - Data Science Process: Overview – Defining research goals – Retrieving data – Data preparation - Exploratory Data analysis – build the model– presenting findings and building applications - Data Mining - Data Warehousing – Basic Statistical descriptions of Data

UNIT II DESCRIBING DATA 9

Types of Data - Types of Variables -Describing Data with Tables and Graphs -Describing Data with Averages - Describing Variability - Normal Distributions and Standard (z) Scores

UNIT III DESCRIBING RELATIONSHIPS 9

Correlation –Scatter plots –correlation coefficient for quantitative data –computational formula for correlation coefficient – Regression –regression line –least squares regression line – Standard error of estimate – interpretation of r2 –multiple regression equations –regression towards the mean

UNIT IV PYTHON LIBRARIES FOR DATA WRANGLING 9

Basics of Numpy arrays –aggregations –computations on arrays –comparisons, masks, boolean logic – fancy indexing – structured arrays – Data manipulation with Pandas – data indexing and selection – operating on data – missing data – Hierarchical indexing – combining datasets – aggregation and grouping – pivot tables

UNIT V DATA VISUALIZATION 9

Importing Matplotlib – Line plots – Scatter plots – visualizing errors – density and contour plots – Histograms – legends – colors – subplots – text and

annotation – customization – three dimensional plotting - Geographic Data with Basemap - Visualization with Seaborn.

COURSE OUTCOMES: At the end of this course, the students will be able to:

CO1: Define the data science process

CO2: Understand different types of data description for data science process

CO3: Gain knowledge on relationships between data

CO4: Use the Python Libraries for Data Wrangling

CO5: Apply visualization Libraries in Python to interpret and explore data

TOTAL:45 PERIODS

TEXTBOOKS:

- 1. David Cielen, Arno D. B. Meysman, and Mohamed Ali, "Introducing Data Science", Manning Publications, 2016. (Unit I)
- 2. Robert S. Witte and John S. Witte, "Statistics", Eleventh Edition, Wiley Publications, 2017. (Units II and III)
- 3. Jake VanderPlas, "Python Data Science Handbook", O'Reilly, 2016. (Units IV and V)

REFERENCE:

1. Allen B. Downey, "Think Stats: Exploratory Data Analysis in Python", Green Tea Press,2014