SYLLABUS

CS3352 - FOUNDATIONS OF DATA SCIENCE

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 r^2 – 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 – Hiearchical indexing – combining datasets – aggregation and grouping – pivot tables.

UNIT – V: DATA VISUALIZATION

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 – Visuali with Seaborn. 9

CONTENTS

UNIT – I: INTRODUCTION

UNIT – II: DESCRIBING DATA

UNIT – V: DATA VISUALIZATION