5. Download any dataset and do the following
a. Count number of categorical and numeric features
b. Remove one correlated attribute (if any)
c. Display five-number summary of each attribute and show it visually

DSE – 14
Data Science and Analytics using Python

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Data Science and Analytics using Python	**4**	**3**	**0**	**1**	**Class XII pass with Mathematics**	DSC-01(Programming using Python)

Learning Objectives:

1. *To introduce the students to real-world data analysis problems.*
2. *To introduce students with concepts of data wrangling and aggregation.*
3. *To give students hands-on knowledge of Pandas.*

Learning Outcomes:

1. Use data analysis tools in the pandas library.
2. Load, clean, transform, merge and reshape data.
3. Create informative visualization and summarize data sets.
4. Analyse and manipulate time series data.

UNIT-I **(10 Hours)**

Introduction: Introduction to Data Science, Exploratory Data Analysis and Data Science Process. Motivation for using Python for Data Analysis, Introduction of Python shell iPython and Jupyter Notebook.

Essential Python Libraries: NumPy, pandas, matplotlib, SciPy, scikit-learn, statsmodels

UNIT-II **(10 Hours)**

Getting Started with Pandas: Arrays and vectorized computation, Introduction to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. Data Loading, Storage and File Formats.

Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, Interacting with Web APIs, Interacting with Databases Data Cleaning and Preparation. Handling Missing Data, Data Transformation, String Manipulation

UNIT-III **(10 Hours)**

Data Wrangling: Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting.

Data Aggregation and Group operations: Group by Mechanics, Data aggregation, General split-apply-combine, Pivot tables and cross tabulation

UNIT-IV **(15 Hours)**

Time Series Data Analysis: Date and Time Data Types and Tools, Time series Basics, date Ranges, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion, Moving Window Functions.

Advanced Pandas: Categorical Data, Advanced Group by Use, Techniques for Method Chaining

References:

1. Dr. Anil Maheshwari, "Data Analytics", McGraw Hill Education (India) Private Limited.

Practicals :

Use data set of your choice from Open Data Portal (https://data.gov.in/) for the following exercises.

1. List of Practical based on NumPy ndarray
2. List of Practical based on Pandas Data Structures
3. List of Practical based on Data Loading, Storage and File Formats
4. List of Practical based on Interacting with Web APIs
5. List of Practical based on Data Cleaning and Preparation
6. List of Practical based on Data Wrangling
7. List of Practical based on Data Aggregation
8. List of Practical based on Time Series Data Analysis

DSE – 15
Organizational Behavior

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Organizational Behavior	4	3	1	0	Class XII pass with Mathematics	NIL

Learning Objectives:

1. *To deeply understand the role of individual, groups and structure in achieving organizational goals effectively and efficiently.*