

FOUNDATIONS OF DATA SCIENCE USING PYTHON

CONTENTS

UNIT - I PYTHON LANGUAGE BASICS AND DATA STRUCTURES

1.1	Introduction to Python	1.1
1.1.1	Why Python for Engineering?	1.2
1.1.2	Features and Advantages of Python	1.2
1.1.3	Python Installation and Environment Setup	1.3
1.1.4	Running Python Programs: Interactive vs Script Mode	1.5
1.1.5	Basic Syntax and Indentation Rules	1.7
1.1.6	Comments and Documentation Strings	1.8
1.1.7	Variables and Identifiers	1.10
1.1.8	Input and Output Operations	1.13
1.1.9	Data Types in Python	1.15
1.1.10	Type Conversion and Casting	1.17
1.1.11	Operators in Python	1.20
1.1.12	Expressions and Statements	1.24
1.2	Scalar Types and Control Flow	1.26
1.2.1	Understanding Scalar Types: int, float, bool, str	1.26
1.2.2	Numeric Operations and Math Module	1.31
1.2.3	Strings and String Operations	1.33
1.2.4	String Formatting and Methods	1.36
1.2.5	Control Flow: if, elif, else Statements	1.38
1.2.6	Looping Constructs: for and while	1.41
1.2.7	Loop Control Statements	1.43
1.2.8	Iterators and Range Function	1.45
1.2.9	Comprehensions Overview	1.47
1.3	Introduction to Data Structures	1.49
1.3.1	Tuples – Definition, Indexing, Slicing, and Immutability	1.50
1.3.2	Lists – Creation, Operations, and Mutability	1.55
1.3.3	List Methods and Functions	1.58
1.3.4	Nested Lists and Matrix Representation	1.62
1.3.5	Built-in Sequence Functions	1.64
1.3.6	Dictionary (DICT) – Concept and Operations	1.68
1.3.7	Dictionary Methods and Iteration	1.71
1.3.8	Sets – Definition, Operations, and Applications	1.75
1.4	Introduction to Functions	1.80
1.4.1	Defining and Calling Functions	1.80
1.4.2	Arguments and Return Values	1.82
1.4.3	Default, Keyword, and Variable-Length Arguments	1.86

1.4.4	Namespaces and Scope (Local, Global, Built-in)	1.89
1.4.5	Local Functions (Nested Functions and Closures)	1.95
1.4.6	Returning Multiple Values	1.98
1.4.7	Functions as Objects – Higher-Order Functions	1.101
1.4.7.1	Higher-Order Functions (map, filter, reduce)	1.102
1.4.7.2	Lambda Functions	1.104
1.5	Files and the Operating System in Python	1.106
1.5.1	File Handling Basics	1.106
1.5.1.1	Using Context Managers	1.107
1.5.2	Reading Files	1.107
1.5.2.1	Reading Methods	1.107
1.5.2.2	Reading Binary Files	1.107
1.5.3	Writing Files	1.107
1.5.3.1	Write and Append Operations	1.107
1.5.3.2	Writing Binary Data	1.108
1.5.4	Working with File Paths	1.108
1.5.4.1	The pathlib Module (Modern Approach)	1.108
1.5.4.2	The os.path Module (Legacy Approach)	1.108
1.5.5	Directory Operations	1.108
1.5.5.1	Creating and Removing Directories	1.108
1.5.5.2	Listing Directory Contents	1.108
1.5.6	The os Module	1.109
1.5.6.1	Environment Variables	1.109
1.5.6.2	System Information	1.109
1.5.6.3	File Operations	1.109
1.5.7	Error Handling	1.109
Short Question and Answers		1.110
Big Questions		1.115
Practical		1.117

UNIT - II NUMPY BASICS

2.1	NumPy-Basics	2.1
2.1.1	Core Features of NumPy	2.1
2.1.2	Role of NumPy in Data Analysis and Scientific Computing	2.2
2.1.3	Historical Background	2.2
2.1.4	Efficiency and Performance Considerations	2.3
2.1.5	Performance Comparison	2.3
2.2	The NumPy ndarray: A Multidimensional Array Object	2.3

2.2.1	Creating ndarrays	2.5
2.2.1.1	Other Array Creation Functions	2.6
2.2.2	Data Types for <i>ndarray</i>	2.8
2.2.3	Arithmetic with NumPy Arrays	2.11
2.2.4	Basic Indexing and Slicing	2.12
2.2.4.1	Indexing in Multidimensional Arrays	2.13
2.2.4.2	Indexing with Slices	2.15
2.2.5	Boolean Indexing	2.18
2.2.6	Fancy Indexing	2.21
2.2.7	Transposing Arrays and Swapping Axes	2.23
2.3	Universal Functions(ufuncs): Fast Element-Wise Array Functions	2.25
2.3.1	Returning Multiple Outputs	2.27
2.3.2	Using the out Argument	2.27
2.3.3	Categories of Universal Functions	2.28
2.4	Array-Oriented Programming with Arrays	2.29
2.4.1	Expressing Conditional Logic as Array Operations	2.32
2.4.2	Mathematical and Statistical Methods	2.34
2.4.3	Methods for Boolean Arrays	2.37
2.4.4	Sorting	2.39
2.4.5	Unique and Other Set Logic	2.41
2.5	File Input and Output with Arrays	2.42
2.6	Linear Algebra	2.4
2.7	Pseudorandom Number Generation	2.47
Short Question and Answers		2.50
Big Questions		2.56
Practical		2.58
 UNIT - III PANDAS BASICS		
3.1	Introduction to Pandas Data Structure	3.1
3.1.1	Series	3.1
3.1.2	Dataframe	6.3
3.2	Loading And Understanding Data	3.5
3.2.2	Understanding Your Data	3.9
3.2.2.1	Initial Data Exploration	3.9
3.3	Data Aggregation For Computing Descriptive Statistics	3.10
3.4	Data Cleaning and Preprocessing	3.12
Short Question and Answers		3.45
Big Questions		3.51
Practical		3.53

UNIT - IV DATA LOADING, STORAGE, AND FILE FORMATS

4	Data Loading, Storage, and File Formats	4.1
4.1	Reading and Writing Data in Text Format	4.1
4.1.1	Reading Text Files in Pieces	4.12
4.1.2	Writing Data to Text Format	4.14
4.1.3	Working with Other Delimited Formats	4.15
4.1.4	JSON Data	4.18
4.1.5	XML and HTML: Web Scraping	4.21
4.1.5.1	Parsing XML with lxml.objectify	4.23
4.2	Binary Data Formats	4.27
4.2.1	Reading Microsoft Excel Files	4.28
4.2.2	Using HDF5 Format	4.30
4.3	Interacting with Web APIs	4.33
4.4	Interacting with Databases	4.36
	Short Question and Answers	4.40
	Big Questions	4.46
	Practical	4.49

UNIT - V DATA EXPLORATION

5.1	Data Transformation	5.1
5.1.1	Removing Duplicates	5.1
5.1.2	Transforming Data Using a Function or Mapping	5.4
5.1.3	Replacing Values	5.7
5.1.4	Renaming Axis Indexes	5.9
5.1.5	Discretization and Binning	5.10
5.1.6	Detecting and Filtering Outliers	5.14
5.1.7	Permutation and Random Sampling	5.17
5.1.8	Computing Indicator/Dummy Variables	5.19
5.2	String Manipulation	5.24
5.2.1	Python Built-In String Object Methods	5.24
5.2.2	Regular Expressions	5.27
5.2.3	String Functions in pandas	5.32
5.3	Data Wrangling	5.38
5.3.1	Reordering and Sorting Levels	5.43
5.3.3	Indexing with a DataFrame's columns	5.45
5.4	Combining and Merging Datasets	5.47
5.4.1	Database-Style DataFrame Joins	5.48
5.4.2	Merging on Index	5.56
5.4.3	Concatenating Along an Axis	5.63

5.4.4	Combining Data with Overlap	5.72
5.5	Reshaping and Pivoting	5.75
5.5.1	Reshaping with Hierarchical Indexing	5.75
5.5.2	Pivoting “Long” to “Wide” Format	5.80
5.5.3	Pivoting “Wide” to “Long” Format	5.86
Short Question and Answers		5.90
Big Questions		5.96
Practical		5.99
 UNIT - VI DATA WRANGLING		
6.1	Data Aggregation	6.1
6.1.1	Group Operations	6.2
6.1.2	Iterating over Groups	6.8
6.1.3	Selecting a Column or Subset of Columns	6.11
6.1.4	Grouping with Dictionaries and Series	6.13
6.1.5	Grouping with Functions	6.15
6.1.6	Grouping by Index Levels	6.16
6.2	Data Aggregation	6.17
6.2.1	Column-Wise and Multiple Function Application	6.21
6.2.2	Returning Aggregated Data Without Row Indexes	6.26
6.3	Apply: General split-apply-combine	6.27
6.3.1	Suppressing the Group Keys	6.30
6.3.2	Quantile and Bucket Analysis	6.31
6.3.3	Example: Filling Missing Values with Group-Specific Values	6.34
6.3.4	Example: Random Sampling and Permutation	6.38
6.3.5	Example: Group Weighted Average and Correlation	6.40
6.3.6	Example: Group-Wise Linear Regression	6.43
6.4	Pivot Tables and Cross-Tabulation	6.44
6.4.1	Cross-Tabulations: Crosstab	6.48
6.5	Date and Time Data Types	6.50
6.5.1	The datetime Object	6.50
6.5.2	The timedelta Object	6.51
6.5.3	Data Types in the datetime Module	6.51
6.5.4	Converting Between String and Datetime	6.51
6.5.5	Common strftime Format Codes	6.52
6.5.6	Parsing Strings into Datetime Objects	6.52
6.5.7	Date Parsing with pandas	6.52
6.5.8	Locale-Specific Date Formatting	6.53

Short Question and Answers	6.54
Big Questions	6.61
Practical	6.64

UNIT - VII DATA VISUALIZATION

7.1	Introduction to Data Visualization	7.1
7.1.1	Visualization in Data Analysis and Storytelling	7.2
7.1.2	Principles of Effective Visualization (Clarity, Accuracy, Design)	7.2
7.1.3	Visualization Tools and Python Libraries	7.3
7.2	Visualizing categorical data.....	7.5
7.2.1	Understanding Categorical Variables	7.5
7.2.2	Bar Charts and Column Charts	7.7
7.2.3	Pie Charts and Donut Charts	7.12
7.2.4	Count Plots Using Seaborn.....	7.16
7.2.5	Stacked and Grouped Bar Charts	7.17
7.3	Visualizing Time-Series Data	7.19
7.3.1	Understanding Time Series	7.19
7.3.2	Visualization Techniques for Time-Series Data	7.22
7.4	Visualizing Multiple Variables	7.29
7.4.1	Concept of Multivariate Visualization	7.29
7.4.2	Scatter Plots and Color/Size Encoding.....	7.31
7.4.3	Pair Plots (Seaborn pairplot).....	7.33
7.4.4	Heatmaps (Correlation Matrix Visualization).....	7.34
7.4.5	Grouped Box Plots and Violin Plots	7.36
7.4.6	Faceted Plots (Subplots by Category).....	7.37
7.5	Visualizing Distributions and Relationships	7.39
7.5.1	Understanding Data Distributions	7.39
7.5.3	Visualizing Relationships Between Variables.....	7.53
7.6	Multivariate and Time-Series Visualization Exploration.....	7.54
7.6.1	Conceptual Foundation.....	7.55
7.6.2	Combining Temporal and Categorical Dimensions	7.55
7.6.3	Heatmaps for Time-Series Analysis.....	7.55
7.6.4	FacetGrid and Subplot Layouts	7.56
7.6.5	3-D and Animated Visualizations	7.56
7.6.6	Interactive Exploration and Dashboards	7.57
7.6.7	Integration of Multivariate and Temporal Dimensions	7.58
Short Question and Answers		7.59
Big Questions		7.65
Practical		7.69