

PYTHON for Data Science



C MUTHU
M C PRAKASH

PYTHON

FOR DATA SCIENCE

PYTHON FOR DATA SCIENCE

Dr. C. Muthu

Head, Department of Data Science
Director, Robotics Lab
Loyola College (Autonomous),
Chennai

M. C. Prakash

Industry Expert



Vijay Nicole Imprints Private Limited
CHENNAI

Published by



Vijay Nicole Imprints Private Limited

No. 1B, Second Main Road,
Ram Nagar South, Chennai – 600 091
Phone: +91-44-4283 7178, 72002 58368, 97898 33533
Email: vni@vijaynicole.co.in Website: www.vijaynicole.co.in

Python for Data Science

ISBN (Print): 978-81-974025-5-5

ISBN (ePub): 978-81-974025-0-0

ISBN (ePdf): 978-81-974025-1-7

Copyright © 2024, Vijay Nicole Imprints Private Limited

No part of this publication may be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a database or retrieval system without the prior written permission of the publishers and copyright holders. The program listings (if any) may be entered, stored and executed in a computer system, but they may not be reproduced for publication.

Information contained in this work has been obtained by publishers, from sources believed to be reliable. However, neither publishers nor copyright holders guarantee the accuracy or completeness of any information published herein, and neither publishers nor copyright holders shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that publishers and copyright holders are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Contents

Preface

xi

Chapter-1 Python Basics

1.1	Introduction	1
1.2	Computational Problem Solving	1
1.3	Computer Hardware, Software and Programming Languages	3
1.4	Installing Python	4
1.5	Tokens	11
1.6	Operators	13
1.7	Data Types	18
1.8	Expressions	19
1.9	Reading Input	20
1.10	Displaying Output	21
1.11	Comments	24
1.12	<i>Summary</i>	27
1.13	<i>Try Yourself</i>	28

Chapter-2 Control Statements

2.1	Introduction	31
2.2	Selection Control Statements	31
2.3	The 'while' Statement	39
2.4	Infinite Loops	43
2.5	The 'for' Statement	43
2.6	Definite vs Indefinite Loops	48
2.7	Jump Statements - 'break' and 'continue'	48
2.8	The 'pass' Statement	49
2.9	'Loop Else' Clause	50
2.10	<i>Summary</i>	51
2.11	<i>Try Yourself</i>	52

Chapter-3 Functions

3.1	Introduction	55
3.2	Built-in Functions	55
3.3	Modules in Python	56
3.4	Defining a Function	60
3.5	Calling a Function and Passing Parameters	61
3.6	Value-Returning Functions	63
3.7	Non-Value-Returning Functions	63
3.8	Scope and Lifetime of Variables	70
3.9	Default Arguments	72
3.10	Keyword Arguments	73
3.11	Variable-length Arguments	74
3.12	Functional Programming	75
3.13	Turtle Graphics	77
3.14	<i>Summary</i>	81
3.15	<i>Try Yourself</i>	82

Chapter-4 Lists

4.1	Introduction	85
4.2	Creating Lists	85
4.3	Reading the Items of a List through the Keyboard	85
4.4	Basic List Operations	86
4.5	List Functions and Methods	90
4.6	List Comprehension	101
4.7	<i>Summary</i>	103
4.8	<i>Try Yourself</i>	104

Chapter-5 Dictionaries

5.1	Introduction	107
5.2	Basic Dictionary Operations	108
5.3	Dictionary Functions and Methods	109
5.4	Default Dictionary	112
5.5	<i>Summary</i>	119
5.6	<i>Try Yourself</i>	119

Chapter-6 Tuples and Sets

6.1	Introduction	123
6.2	Creating Tuples	123
6.3	Basic Tuple Operations	123
6.4	Tuple Methods	126
6.5	Sets	130
6.6	Creating a Set	131
6.7	Basic Set Operations	131
6.8	<i>Summary</i>	137
6.9	<i>Try Yourself</i>	138

Chapter-7 File Processing

7.1	Introduction	141
7.2	Types of Files	141
7.3	File Paths	142
7.4	Creating and Using Text Files	144
7.5	Reading and Writing Binary Files	151
7.6	The Pickle Module	153
7.7	Reading and Writing CSV Files	154
7.8	<i>Summary</i>	158
7.9	<i>Try Yourself</i>	159

Chapter-8 String Processing and Exception Handling

8.1	Introduction	163
8.2	Traversing a String	163
8.3	String Operators	163
8.4	Slicing the Strings	165
8.5	String Methods	167
8.6	Exception Handling	170
8.7	Built-in Exceptions	170
8.8	User-defined Exception-handling Mechanism	172
8.9	The 'else' block	175
8.10	'Finally' block	175
8.11	<i>Summary</i>	176
8.12	<i>Try Yourself</i>	177

Chapter-9 Object-Oriental Programming

9.1	Introduction	179
9.2	Classes and Objects	179
9.3	Defining Classes	179
9.4	Creating Objects	180
9.5	Encapsulation	186
9.6	Inheritance	188
9.7	Polymorphism	191
9.8	Multiple Inheritance	193
9.9	Private Data Members and Methods	195
9.10	Recursive Functions	196
9.11	<i>Summary</i>	196
9.12	<i>Try Yourself</i>	197

Chapter-10 NumPy and Pandas Libraries

10.1	NumPy Library	201
10.2	NumPy Array Creation	201
10.3	NumPy Array Initialization	202
10.4	NumPy Array Attributes	204
10.5	Indexing in NumPy Arrays	204
10.6	Basic Arithmetic Operations on NumPy Arrays	205
10.7	Mathematical Functions in NumPy Library	208
10.8	Changing the Shape of a NumPy Array	209
10.9	Stacking and Splitting NumPy Arrays	209
10.10	Broadcasting in NumPy Arrays	210
10.11	Pandas Library and Series Data Structure	212
10.12	<i>Summary</i>	216
10.13	<i>Try Yourself</i>	217

Chapter-11 Data Visualization

11.1	Introduction	219
11.2	The Pyplot Module	219
11.3	Functions in Pyplot Module	220
11.4	Line Plot	221
11.5	Colours, Markers and Line Styles	222
11.6	Labels, Ticks and Legends	223
11.7	Bar Chart	224

11.8	Pie Chart	225
11.9	Histogram	227
11.10	Scatter Plot	228
11.11	Seaborn Library	230
11.12	<i>Summary</i>	243
11.13	<i>Try Yourself</i>	244

Chapter-12 Exploring and Plotting Data

12.1	Introduction	247
12.2	Creation of a Data Frame	247
12.3	Accessing the Columns in a Data Frame	248
12.4	Row Index and Column Names	249
12.5	Data Frame Indexing and Selecting Data	251
12.6	Reading Data from a CSV File	253
12.7	Sorting the Data	254
12.8	Basic Attributes and Methods of DataFrame Class	257
12.9	Plotting in Pandas	263
12.10	DateTime Column	271
12.11	The 'apply()' Function	279
12.12	<i>Summary</i>	281
12.13	<i>Try Yourself</i>	281

Chapter-13 Data Aggregation

13.1	Introduction	285
13.2	Basic Aggregation Operation	285
13.3	Aggregating Functions	294
13.4	Calculations Based on Multiple Columns	297
13.5	Grouping Data Using Dates and Times	299
13.6	<i>Summary</i>	302
13.7	<i>Try Yourself</i>	302

Chapter-14 Combining Datasets

14.1	Introduction	305
14.2	Concatenate and Append Operations	305
14.3	Merge Operation	307
14.4	Inner Merge Operation	308
14.5	Left Merge Operation	309

14.6	Right Merge Operation	310
14.7	Outer Merge Operation	312
14.8	<i>Summary</i>	321
14.9	<i>Try Yourself</i>	322
Chapter-15 Machine Learning		
15.1	Introduction	325
15.2	Supervised Learning	325
15.3	Unsupervised Learning	326
15.4	<i>Summary</i>	329
15.5	<i>Try Yourself</i>	330
Chapter-16 Regression Models		
16.1	Introduction	333
16.2	Implementation in Python	334
16.3	Hyperparameters	355
16.4	<i>Summary</i>	355
16.5	<i>Try Yourself</i>	356
Chapter-17 Classification Models		
17.1	Introduction	361
17.2	Implementation in Python	361
17.3	Goodness of Fit and Accuracy of Logistic Regression Model	370
17.4	<i>Summary</i>	372
17.5	<i>Try Yourself</i>	372
Chapter-18 Clustering Models		
18.1	Introduction	377
18.2	K-Means Clustering Method	378
18.3	Hierarchical Clustering Method	388
18.4	<i>Summary</i>	390
18.5	<i>Try Yourself</i>	391