

# CONTENTS IN DETAIL

<b>ACKNOWLEDGMENTS</b>	<b>xix</b>
------------------------	------------

<b>INTRODUCTION</b>	<b>xxi</b>
---------------------	------------

Why This Book? . . . . .	xxii
Why Python? . . . . .	xxiii
How This Book Is Organized . . . . .	xxiv
Dataframes, Your New Best Friends . . . . .	xxvii

## **PART I: FROM SPREADSHEETS TO DATAFRAMES** **1**

<b>1</b>	
<b>GETTING STARTED WITH PYTHON</b>	<b>3</b>

Technical Considerations for Working with Python . . . . .	4
Python Version . . . . .	4
Distribution Platform . . . . .	4
Integrated Development Environment . . . . .	4
The IDLE Shell: A Simple Interface . . . . .	5
Spyder: A More Robust Working Environment . . . . .	6
Summary . . . . .	8

<b>2</b>	
<b>DISPLAYING DATA AND UNDERSTANDING DATA TYPES</b>	<b>11</b>

Printing String Variables and Strings of Strings . . . . .	12
Data Types . . . . .	13
Printing Numbers and Numeric Variables . . . . .	14
Concatenating Different Data Types . . . . .	15
Learning from Your Mistakes . . . . .	16
Customizing Functionality with Parameters . . . . .	18
Printing an Empty Line . . . . .	18
Escape Characters and Escape Sequences . . . . .	19
Potential Pitfalls for Python Novices . . . . .	20
Using Quotation Marks Inconsistently . . . . .	20
Ignoring Case Sensitivity . . . . .	22
Not Knowing Your Data . . . . .	22
Summary . . . . .	23

### 3

## **CREATING AND MANIPULATING DATAFRAMES AND LISTS 25**

What Exactly Is a Dataframe? . . . . .	26
How to Create a Dataframe . . . . .	26
Importing the pandas Module and Manually Creating a Dataframe . . . . .	26
Copying a Dataframe . . . . .	31
Subsetting a Dataframe by Column . . . . .	31
Common Dataframe Operations . . . . .	33
Counting Rows Using the len() Function . . . . .	33
Counting Rows and Columns with the shape Attribute . . . . .	34
Deleting Rows with a Specific Value . . . . .	34
Identifying and Dropping Duplicated Rows . . . . .	35
Concatenating Dataframes . . . . .	35
Lists: The DNA of Dataframes . . . . .	40
Creating a List . . . . .	42
Isolating Unique List Values . . . . .	42
Appending a Single Item to a List . . . . .	43
Adding a List to a List . . . . .	43
Sorting a List . . . . .	43
Identifying Minimum, Maximum, and Mean List Values . . . . .	44
Removing a List Element . . . . .	44
Comparing Lists . . . . .	44
Creating an Empty Dataframe with a List . . . . .	45
Converting a Column into a List . . . . .	46
Another Fragment of Dataframe DNA: The Series Object . . . . .	46
Summary . . . . .	47

### 4

## **ADDING, MODIFYING, AND CALCULATING COLUMN DATA 49**

Defining a Column . . . . .	50
Changing How Columns Appear in a Dataframe . . . . .	50
Reordering Columns in a Dataframe . . . . .	50
Dropping Columns . . . . .	51
Renaming Columns . . . . .	52
Sorting Dataframes by Column Values . . . . .	53
Printing Select Columns . . . . .	54
Changing Values in Columns . . . . .	55
Overwriting All Column Values . . . . .	55
Replacing Particular Values in Dataframes and Target Columns . . . . .	55
Replacing Substrings . . . . .	57
Creating New Columns . . . . .	58
Adding a New Column with a Single Value . . . . .	58
Duplicating a Column . . . . .	58
Concatenating Two or More Columns into a New One . . . . .	59
Math Methods and Operators . . . . .	60
Applying the sum() Method to a Dataframe Column . . . . .	60
Returning Average, Maximum, or Minimum Values . . . . .	61
Calculating Median Values . . . . .	61

Rounding a Single Column or Full Dataframe . . . . .	63
Converting to Absolute Values . . . . .	64
Stringing Together Multiple Methods . . . . .	65
Storing Calculated Values in a New Column . . . . .	66
Comparing Values. . . . .	67
Conditional Logic. . . . .	68
Controlling Execution Flow with if Logic . . . . .	68
Iterating Through a List. . . . .	69
Repeating a Process with the while Loop . . . . .	69
Iterating Through a Dataframe . . . . .	70
Replicating SUMIF with the where() Conditional Statement . . . . .	71
Storing Values Returned from iterrows() in a New Column . . . . .	72
Storing Conditional Results in a New Column with List Comprehensions. . . . .	72
Handling Exceptions with a try-except Block. . . . .	73
Summary . . . . .	74

## **5 ACCESSING AND TRANSFORMING INDIVIDUAL CELL VALUES 77**

An Overview of Values and Variables. . . . .	78
Converting Integers, Floats, and Strings . . . . .	79
Converting Individual Variables . . . . .	79
Converting an Entire Dataframe Column . . . . .	81
Converting Objects to Strings . . . . .	82
Manually Inserting Values with the input() Function . . . . .	82
Answering a Question and Saving the Answer . . . . .	83
Selecting a Menu Option . . . . .	84
Pausing Execution to Review Output . . . . .	85
Working with NaN Objects. . . . .	86
Manually Creating NaN Objects . . . . .	86
Manually Entering NaN Objects into a Dataframe . . . . .	86
Replacing NaN Objects. . . . .	87
Slicing Techniques for Strings and Lists . . . . .	88
Slicing a Single String . . . . .	88
Slicing Within an Entire Column . . . . .	90
Slicing a List from a List . . . . .	91
Pulling a Select Element from a List . . . . .	92
Indexing Techniques for Dataframes and Series Objects . . . . .	93
Pulling Values by Index Position . . . . .	93
Pulling Values by Unique Index Label . . . . .	94
Targeting Single Values . . . . .	95
Pulling a Value from a Series . . . . .	95
Modifying Existing Cells. . . . .	96
Splitting Techniques . . . . .	98
Splitting a Single String . . . . .	98
Handling Inconsistent Delimiters . . . . .	99
Splitting Columns in a Dataframe . . . . .	100
Summary . . . . .	102

## **6 FILTERING AND DISPLAYING DATAFRAMES 105**

A Closer Look at the DataFrame() Method . . . . .	106
Using Optional Arguments . . . . .	107
Creating an Empty Dataframe with a Column List . . . . .	108
Working with Dataframe Indexes . . . . .	109
Naming and Renaming an Index . . . . .	109
Renumbering or Resetting an Index . . . . .	110
Sorting by the Index . . . . .	111
Moving Values Between the Index and a Column . . . . .	112
Subsetting Dataframes . . . . .	113
By Column List . . . . .	113
By Relative Index Location . . . . .	116
By Index Label . . . . .	118
By Matching Values in Columns . . . . .	120
By Excluding Values in Columns . . . . .	121
By a Substring Value . . . . .	121
By Column Labels in a List . . . . .	122
By Inclusion . . . . .	122
By Exclusion . . . . .	123
By Mathematical Condition . . . . .	123
By NaN Objects . . . . .	124
Controlling the Appearance of Dataframe Output . . . . .	125
Printing the First or Last Few Rows . . . . .	125
Printing Specific Rows . . . . .	126
Printing the Rightmost Columns . . . . .	126
Customizing Global Display Settings . . . . .	127
Working with Dictionary Objects . . . . .	130
Declaring a Single Dictionary . . . . .	131
Declaring a List of Dictionaries . . . . .	132
Accessing and Modifying Dictionary Contents . . . . .	132
Creating New Dictionary Key-Value Pairs on the Fly . . . . .	133
Storing Dictionaries in Dataframes . . . . .	134
Summary . . . . .	134

## **PART II: TOOLS TO REPLICATE EXCEL FUNCTIONALITY 137**

### **7 COUNTING AND SUMMING VALUES 139**

The value_counts() Method . . . . .	140
Counting Every Value in a Column . . . . .	140
Counting Specific Values in a Column . . . . .	141
Counting a List of Specific Values in a Column . . . . .	141
Normalizing value_counts() Results . . . . .	143

The <code>crosstab()</code> Method . . . . .	144
Creating Basic Cross-Tabulations . . . . .	144
Performing Math Operations . . . . .	147
Adding Row and Column Totals . . . . .	152
Handling Missing Values . . . . .	153
Normalizing <code>crosstab()</code> Data . . . . .	155
The <code>pivot_table()</code> Method . . . . .	157
Breaking Down the Basic Form of <code>pivot_table()</code> . . . . .	158
Grouping Unique Values . . . . .	161
Calculating Math Values . . . . .	161
Handling NaN Objects . . . . .	163
Organizing Dataframe Values . . . . .	164
Summary . . . . .	165

## **8 MERGING DATAFRAMES 167**

The Basics of VLOOKUP and <code>merge()</code> . . . . .	168
How <code>merge()</code> Handles Orphaned Keys . . . . .	170
The Full <code>merge()</code> Method Syntax . . . . .	172
Specifying Join Type . . . . .	174
Analyzing Merge Results . . . . .	177
Defining Keys . . . . .	179
Handling Different Key Column Labels . . . . .	181
Checking Your Match Expectations . . . . .	183
Quality Control with the <code>shape</code> Attribute . . . . .	185
Summary . . . . .	186

## **9 FORMATTING AND CALCULATING DATES AND TIMES 189**

Introducing the Datetime Module and <code>datetime.now()</code> . . . . .	190
Creating Datetime Objects . . . . .	190
Isolating Units of Time as Integers . . . . .	191
Converting Datetime Objects to Strings . . . . .	193
Transforming Timestamps . . . . .	193
Converting a Single Datetime Object to a Custom-Formatted String . . . . .	196
Isolating Units of Time as Strings . . . . .	197
Removing Leading Zeros from Single-Digit Time Elements . . . . .	197
Working with Time Durations: The <code>Timedelta</code> Object . . . . .	198
Comparing and Calculating Dates and Times . . . . .	199
Datetime Objects in Dataframes . . . . .	201
Dataframe Datetime Operations . . . . .	202
Using Directives to Customize to <code>_datetime()</code> Results . . . . .	203
Calculating <code>Timedeltas</code> in a New Column . . . . .	203
Subsetting a Dataframe Using Datetime Objects . . . . .	204
Summary . . . . .	205

**10****READING EXCEL FILES INTO DATAFRAMES 209**

Creating or Downloading Your Excel Spreadsheet . . . . .	210
Introducing the read_excel() Method. . . . .	210
Importing a Specific Tab from a Workbook . . . . .	212
Importing All Tabs at Once. . . . .	213
Filtering Source Data . . . . .	214
Parsing Input Spreadsheets . . . . .	214
Dealing with More Complex Spreadsheets . . . . .	216
Setting the Column Labels for Your Dataframe. . . . .	217
Setting an Excel Column as the Dataframe Index. . . . .	218
Handling Hard Returns in Excel Data. . . . .	219
Reading in a CSV File . . . . .	220
Summary . . . . .	221

**11****SAVING DATAFRAMES TO EXCEL 223**

Simple Single-Tab Export. . . . .	224
Exporting a More Complex Dataframe . . . . .	225
Designating the Tab Name. . . . .	226
Excluding the Dataframe Index . . . . .	227
Freezing the View . . . . .	228
The Six Steps to Exporting and Formatting Excel Files . . . . .	228
Step 1: Creating a Writer Object and Excel File . . . . .	230
Step 2: Adding Multiple Dataframes to One Excel Workbook. . . . .	231
Step 3: Closing the Writer Object and Excel File. . . . .	232
Step 4: Creating and Populating the Workbook Object and Excel File. . . . .	233
Step 5: Formatting the Excel File . . . . .	234
Step 6: Closing the Workbook Object and Excel File. . . . .	243
Emailing from Python . . . . .	244
Sending a Basic Email . . . . .	244
Converting Dataframes to HTML Code. . . . .	245
Sending an Email Containing HTML . . . . .	246
Summary . . . . .	247

**12****THERE AND BACK AGAIN: AN EXCEL-PYTHON-EXCEL WORKFLOW 249**

The Scenario . . . . .	250
Analyzing the Vet's Workbook . . . . .	250
Flowcharting the Manual Process . . . . .	251
Coding with Modularity . . . . .	252
Writing and Calling UDFs . . . . .	252
Defining Required and Optional Parameters. . . . .	253
Returning One or More Values . . . . .	254
Saving UDFs in a Separate Script . . . . .	256

Writing Your First Script: Rolling Over a File . . . . .	257
Importing Your Favorite Modules . . . . .	257
Importing the Ducks UDFs . . . . .	258
Printing a Header . . . . .	258
Setting Your Dataframe Display Preferences . . . . .	258
Setting the Date . . . . .	259
Creating a New File for Today’s Date . . . . .	260
Automating Exception Reports and Data Management Tasks . . . . .	262
Importing the Vet’s Workbook Tabs . . . . .	262
Generating and Sending an Email . . . . .	263
Automating the Overdue Invoice Process . . . . .	264
Filtering Today’s Appointments . . . . .	266
Pausing the Program’s Execution . . . . .	266
Automating Daily Appointment Email . . . . .	267
Recording New Appointments . . . . .	269
Exporting Dataframes to Excel Tabs . . . . .	270
Updating and Formatting the Excel File in Spyder . . . . .	271
Automating Updates in Excel Files . . . . .	276
Analyzing Trends with Dynamic Pivot Tables . . . . .	280
Using Spyder as Your Daily Workflow GUI . . . . .	282
Summary . . . . .	282

## A

### **WORKING WITH FOLDERS, FILES, AND PATHNAMES 285**

Pathnames in File Explorer vs. Python . . . . .	286
Viewing and Changing Your Working Directory . . . . .	287
Listing the Contents of a Folder . . . . .	288
Creating a New Directory . . . . .	289
Checking Whether a Pathname Exists . . . . .	289
Copying a File . . . . .	290
Renaming or Moving a File or Folder . . . . .	291
Deleting a File . . . . .	292
Deleting Folders . . . . .	293
File Management Quick Reference . . . . .	294

## B

### **CLEANING UP A MESSY SPREADSHEET 295**

Reading a Messy Spreadsheet into a Dataframe . . . . .	296
Customizing Column Labels . . . . .	297
Specifying a Column as the Index . . . . .	297
Dropping a Record by Index Label . . . . .	298
Sorting by Index Label . . . . .	298
Splitting Columns: Converting Full Names to First and Last . . . . .	299
Renaming Columns . . . . .	299
Reordering Columns . . . . .	300
Changing a Specific Cell Value . . . . .	300
Converting Strings to Datetime Objects . . . . .	301
Exporting the Cleaned-Up Dataframe Back to Excel . . . . .	301
Formatting the Clean Excel File . . . . .	302
The Final Product . . . . .	303

<b>C</b>	
<b>THE DUCKS MODULE</b>	<b>305</b>
<b>PYTHON QUICK REFERENCE</b>	<b>311</b>
<b>INDEX</b>	<b>313</b>