

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Who this book is for . . . . .	3
1.2	Data in this Book . . . . .	4
1.3	Hints, Tables, and Images . . . . .	4
<b>2</b>	<b>Installation</b>	<b>5</b>
2.1	Anaconda . . . . .	5
2.2	Pip . . . . .	6
2.3	Jupyter Overview . . . . .	7
2.4	Summary . . . . .	9
2.5	Exercises . . . . .	9
<b>3</b>	<b>Data Structures</b>	<b>11</b>
3.1	Summary . . . . .	12
3.2	Exercises . . . . .	12
<b>4</b>	<b>Series Introduction</b>	<b>13</b>
4.1	The index abstraction . . . . .	14
4.2	The pandas Series . . . . .	14
4.3	The NaN value . . . . .	16
4.4	Optional Integer Support for NaN . . . . .	17
4.5	Similar to NumPy . . . . .	17
4.6	Categorical Data . . . . .	19
4.7	Summary . . . . .	21
4.8	Exercises . . . . .	21
<b>5</b>	<b>Series Deep Dive</b>	<b>23</b>
5.1	Loading the Data . . . . .	23
5.2	Series Attributes . . . . .	24
5.3	Summary . . . . .	25
5.4	Exercises . . . . .	25
<b>6</b>	<b>Operators (&amp; Dunder Methods)</b>	<b>27</b>
6.1	Introduction . . . . .	27
6.2	Dunder Methods . . . . .	27
6.3	Index Alignment . . . . .	28
6.4	Broadcasting . . . . .	28
6.5	Iteration . . . . .	30
6.6	Operator Methods . . . . .	30
6.7	Chaining . . . . .	30

## Contents

---

6.8	Summary . . . . .	32
6.9	Exercises . . . . .	32
<b>7</b>	<b>Aggregate Methods</b>	<b>33</b>
7.1	Aggregations . . . . .	33
7.2	Count and Mean of an Attribute . . . . .	34
7.3	.agg and Aggregation Strings . . . . .	35
7.4	Summary . . . . .	37
7.5	Exercises . . . . .	37
<b>8</b>	<b>Conversion Methods</b>	<b>39</b>
8.1	Automatic Conversion . . . . .	39
8.2	Memory Usage . . . . .	40
8.3	String and Category Types . . . . .	41
8.4	Ordered Categories . . . . .	41
8.5	Converting to Other Types . . . . .	42
8.6	Summary . . . . .	43
8.7	Exercises . . . . .	43
<b>9</b>	<b>Manipulation Methods</b>	<b>45</b>
9.1	.apply and .where . . . . .	45
9.2	If Else with Pandas . . . . .	48
9.3	Missing Data . . . . .	49
9.4	Filling In Missing Data . . . . .	50
9.5	Interpolating Data . . . . .	52
9.6	Clipping Data . . . . .	52
9.7	Sorting Values . . . . .	53
9.8	Sorting the Index . . . . .	54
9.9	Dropping Duplicates . . . . .	54
9.10	Ranking Data . . . . .	55
9.11	Replacing Data . . . . .	56
9.12	Binning Data . . . . .	58
9.13	Summary . . . . .	61
9.14	Exercises . . . . .	61
<b>10</b>	<b>Indexing Operations</b>	<b>63</b>
10.1	Prepping the Data and Renaming the Index . . . . .	63
10.2	Resetting the Index . . . . .	65
10.3	The .loc Attribute . . . . .	65
10.4	The .iloc Attribute . . . . .	72
10.5	Heads and Tails . . . . .	75
10.6	Sampling . . . . .	75
10.7	Filtering Index Values . . . . .	76
10.8	Reindexing . . . . .	76
10.9	Summary . . . . .	78
10.10	Exercises . . . . .	78
<b>11</b>	<b>String Manipulation</b>	<b>81</b>
11.1	Strings and Objects . . . . .	81
11.2	Categorical Strings . . . . .	82

---

11.3	The .str Accessor . . . . .	82
11.4	Searching . . . . .	84
11.5	Splitting . . . . .	85
11.6	Optimizing .apply with Cython . . . . .	87
11.7	Replacing Text . . . . .	88
11.8	Summary . . . . .	92
11.9	Exercises . . . . .	92
<b>12</b>	<b>Date and Time Manipulation</b>	<b>93</b>
12.1	Date Theory . . . . .	93
12.2	Loading UTC Time Data . . . . .	95
12.3	Loading Local Time Data . . . . .	96
12.4	Converting Local time to UTC . . . . .	97
12.5	Converting to Epochs . . . . .	98
12.6	Manipulating Dates . . . . .	98
12.7	Summary . . . . .	102
12.8	Exercises . . . . .	102
<b>13</b>	<b>Dates in the Index</b>	<b>103</b>
13.1	Finding Missing Data . . . . .	103
13.2	Filling In Missing Data . . . . .	104
13.3	Interpolation . . . . .	105
13.4	Dropping Missing Values . . . . .	106
13.5	Shifting Data . . . . .	107
13.6	Rolling Average . . . . .	107
13.7	Resampling . . . . .	110
13.8	Gathering Aggregate Values (But Keeping Index) . . . . .	112
13.9	Groupby Operations . . . . .	115
13.10	Cumulative Operations . . . . .	116
13.11	Summary . . . . .	118
13.12	Exercises . . . . .	118
<b>14</b>	<b>Plotting with a Series</b>	<b>119</b>
14.1	Plotting in Jupyter . . . . .	119
14.2	The .plot Attribute . . . . .	119
14.3	Histograms . . . . .	120
14.4	Box Plot . . . . .	121
14.5	Kernel Density Estimation Plot . . . . .	122
14.6	Line Plots . . . . .	122
14.7	Line Plots with Multiple Aggregations . . . . .	123
14.8	Bar Plots . . . . .	124
14.9	Pie Plots . . . . .	126
14.10	Styling . . . . .	128
14.11	Summary . . . . .	129
14.12	Exercises . . . . .	129
<b>15</b>	<b>Categorical Manipulation</b>	<b>131</b>
15.1	Categorical Data . . . . .	131
15.2	Frequency Counts . . . . .	131
15.3	Benefits of Categories . . . . .	132

## Contents

---

15.4	Conversion to Ordinal Categories . . . . .	132
15.5	The .cat Accessor . . . . .	133
15.6	Category Gotchas . . . . .	134
15.7	Generalization . . . . .	136
15.8	Summary . . . . .	138
15.9	Exercises . . . . .	138
<b>16</b>	<b>Dataframes</b>	<b>139</b>
16.1	Database and Spreadsheet Analogues . . . . .	139
16.2	A Simple Python Version . . . . .	139
16.3	Dataframes . . . . .	140
16.4	Construction . . . . .	141
16.5	Dataframe Axis . . . . .	143
16.6	Summary . . . . .	145
16.7	Exercises . . . . .	146
<b>17</b>	<b>Similarities with Series and DataFrame</b>	<b>147</b>
17.1	Getting the Data . . . . .	147
17.2	Viewing Data . . . . .	151
17.3	Summary . . . . .	152
17.4	Exercises . . . . .	152
<b>18</b>	<b>Math Methods in DataFrames</b>	<b>153</b>
18.1	Index Alignment . . . . .	153
18.2	Duplicate Index Entries . . . . .	154
18.3	Summary . . . . .	155
18.4	Exercises . . . . .	155
<b>19</b>	<b>Looping and Aggregation</b>	<b>157</b>
19.1	For Loops . . . . .	157
19.2	Aggregations . . . . .	157
19.3	The .apply Method . . . . .	159
19.4	Summary . . . . .	163
19.5	Exercises . . . . .	163
<b>20</b>	<b>Columns Types, .assign, and Memory Usage</b>	<b>165</b>
20.1	Conversion Methods . . . . .	165
20.2	Memory Usage . . . . .	165
20.3	Summary . . . . .	167
20.4	Exercises . . . . .	167
<b>21</b>	<b>Creating and Updating Columns</b>	<b>169</b>
21.1	Loading the Data . . . . .	169
21.2	More Column Cleanup . . . . .	173
21.3	Summary . . . . .	180
21.4	Exercises . . . . .	180
<b>22</b>	<b>Dealing with Missing and Duplicated Data</b>	<b>181</b>
22.1	Missing Data . . . . .	181
22.2	Duplicates . . . . .	183
22.3	Summary . . . . .	185

---

22.4 Exercises . . . . .	185
<b>23 Sorting Columns and Indexes</b>	<b>187</b>
23.1 Sorting Columns . . . . .	187
23.2 Sorting Column Order . . . . .	189
23.3 Setting and Sorting the Index . . . . .	189
23.4 Summary . . . . .	191
23.5 Exercises . . . . .	191
<b>24 Filtering and Indexing Operations</b>	<b>193</b>
24.1 Renaming an Index . . . . .	193
24.2 Resetting the Index . . . . .	193
24.3 Dataframe Indexing, Filtering, & Querying . . . . .	194
24.4 Indexing by Position . . . . .	196
24.5 Indexing by Name . . . . .	199
24.6 Filtering with Functions & <code>.loc</code> . . . . .	203
24.7 <code>.query</code> vs <code>.loc</code> . . . . .	204
24.8 Summary . . . . .	205
24.9 Exercises . . . . .	205
<b>25 Plotting with Dataframes</b>	<b>207</b>
25.1 Lines Plots . . . . .	207
25.2 Bar Plots . . . . .	211
25.3 Scatter Plots . . . . .	215
25.4 Area Plots and Stacked Bar Plots . . . . .	215
25.5 Column Distributions with KDEs, Histograms, and Boxplots . . . . .	217
25.6 Summary . . . . .	221
25.7 Exercises . . . . .	221
<b>26 Reshaping Dataframes with Dummies</b>	<b>223</b>
26.1 Dummy Columns . . . . .	223
26.2 Undoing Dummy Columns . . . . .	225
26.3 Summary . . . . .	226
26.4 Exercises . . . . .	226
<b>27 Reshaping By Pivoting and Grouping</b>	<b>229</b>
27.1 A Basic Example . . . . .	229
27.2 Using a Custom Aggregation Function . . . . .	232
27.3 Multiple Aggregations . . . . .	236
27.4 Per Column Aggregations . . . . .	238
27.5 Grouping by Hierarchy . . . . .	240
27.6 Grouping with Functions . . . . .	243
27.7 Summary . . . . .	247
27.8 Exercises . . . . .	247
<b>28 More Aggregations</b>	<b>249</b>
28.1 Aggregations while Keeping Rows . . . . .	249
28.2 Filtering Parts of Groups . . . . .	251
28.3 Summary . . . . .	254
28.4 Exercises . . . . .	254

<b>29</b>	<b>Cross-tabulation Deep Dive</b>	<b>255</b>
29.1	Cross-tabulation Summaries . . . . .	255
29.2	Adding Margins . . . . .	255
29.3	Normalizing Results . . . . .	256
29.4	Hierarchical Columns with Cross Tabulations . . . . .	257
29.5	Heatmaps . . . . .	257
29.6	Summary . . . . .	258
29.7	Exercises . . . . .	258
<b>30</b>	<b>Melting, Transposing, and Stacking Data</b>	<b>259</b>
30.1	Melting Data . . . . .	259
30.2	Un-melting Data . . . . .	262
30.3	Transposing Data . . . . .	263
30.4	Stacking & Unstacking . . . . .	264
30.5	Stacking . . . . .	267
30.6	Flattening Hierarchical Indexes and Columns . . . . .	269
30.7	Summary . . . . .	273
30.8	Exercises . . . . .	273
<b>31</b>	<b>Working with Time Series</b>	<b>275</b>
31.1	Loading the Data . . . . .	275
31.2	Adding Timezone Information . . . . .	276
31.3	Exploring the Data . . . . .	278
31.4	Slicing Time Series . . . . .	279
31.5	Missing Timeseries Data . . . . .	281
31.6	Exploring Seasonality . . . . .	283
31.7	Resampling Data . . . . .	286
31.8	Rules with Offset Aliases . . . . .	287
31.9	Combining Offset Aliases . . . . .	287
31.10	Anchored Offset Aliases . . . . .	288
31.11	Resampling to Finer-grain Frequency . . . . .	288
31.12	Grouping a Date Column with pd.Grouper . . . . .	289
31.13	Summary . . . . .	291
31.14	Exercises . . . . .	291
<b>32</b>	<b>Joining Dataframes</b>	<b>293</b>
32.1	Adding Rows to Dataframes . . . . .	293
32.2	Adding Columns to Dataframes . . . . .	294
32.3	Joins . . . . .	294
32.4	Join Indicators . . . . .	298
32.5	Merge Validation . . . . .	298
32.6	Joining Data Example . . . . .	299
32.7	Dirty Devil Flow and Weather Data . . . . .	299
32.8	Joining Data . . . . .	300
32.9	Validating Joined Data . . . . .	301
32.10	Visualization of Merged Data . . . . .	302
32.11	Summary . . . . .	304
32.12	Exercises . . . . .	304
<b>33</b>	<b>Exporting Data</b>	<b>305</b>

---

33.1	Dirty Devil Data . . . . .	305
33.2	Reading and Writing . . . . .	306
33.3	Creating CSV Files . . . . .	306
33.4	Exporting to Excel . . . . .	307
33.5	Feather . . . . .	308
33.6	SQL . . . . .	308
33.7	JSON . . . . .	310
33.8	Summary . . . . .	314
33.9	Exercises . . . . .	314
<b>34</b>	<b>Styling Dataframes</b>	<b>317</b>
34.1	Loading the Data . . . . .	317
34.2	Sparklines . . . . .	319
34.3	The .style Attribute . . . . .	319
34.4	Formatting . . . . .	320
34.5	Embedding Bar Plots . . . . .	320
34.6	Highlighting . . . . .	321
34.7	Heatmaps and Gradients . . . . .	321
34.8	Captions . . . . .	321
34.9	CSS Properties . . . . .	321
34.10	Stickiness and Hiding . . . . .	322
34.11	Hiding the Index . . . . .	322
34.12	Summary . . . . .	324
34.13	Exercises . . . . .	324
<b>35</b>	<b>Debugging Pandas</b>	<b>329</b>
35.1	Checking if Dataframes are Equal . . . . .	329
35.2	Debugging Chains . . . . .	332
35.3	Debugging Chains Part II . . . . .	336
35.4	Debugging Chains Part III . . . . .	337
35.5	Debugging Chains Part IV . . . . .	339
35.6	Debugging Apply (and Friends) . . . . .	341
35.7	Memory Usage . . . . .	343
35.8	Timing Information . . . . .	344
35.9	Summary . . . . .	345
35.10	Exercises . . . . .	346
<b>36</b>	<b>Summary</b>	<b>347</b>
	<b>About the Author</b>	<b>349</b>
	<b>Index</b>	<b>351</b>
	<b>Also Available</b>	<b>365</b>
	<b>One more thing</b>	<b>367</b>