

# Contents

<b>Prerequisites: Tooling</b>	<b>5</b>
Python . . . . .	5
Editor . . . . .	6
Console (REPL) . . . . .	7
Using Spyder . . . . .	8
Anki . . . . .	9
Remembering What You Learn . . . . .	9
Installing Anki . . . . .	10
Using Anki with this Book . . . . .	11
<b>1. Introduction</b>	<b>12</b>
The Purpose of Data Analysis . . . . .	12
What is Data? . . . . .	12
Example Datasets . . . . .	14
Play-By-Play Data . . . . .	14
Player-Game Data . . . . .	14
ADP Data . . . . .	15
What is Analysis? . . . . .	15
Types of Data Analysis . . . . .	16
Summary Statistics . . . . .	17
Modeling . . . . .	17
High Level Data Analysis Process . . . . .	19
1. Collecting Data . . . . .	19
2. Storing Data . . . . .	19
3. Loading Data . . . . .	20
4. Manipulating Data . . . . .	20
5. Analyzing Data for Insights . . . . .	20
Connecting the High Level Analysis Process to the Rest of the Book . . . . .	21
<b>2. Python</b>	<b>22</b>
Introduction to Python Programming . . . . .	22
How to Read This Chapter . . . . .	22
Important Parts of the Python Standard Library . . . . .	23
Comments . . . . .	23

Variables . . . . .	24
Types . . . . .	25
Interlude: How to Figure Things Out in Python . . . . .	26
Bools . . . . .	28
if statements . . . . .	30
Container Types . . . . .	30
Unpacking . . . . .	32
Loops . . . . .	32
Comprehensions . . . . .	34
Functions . . . . .	36
Libraries are Functions and Types . . . . .	40
os Library and path . . . . .	40
<b>3. Pandas</b> . . . . .	<b>41</b>
Introduction to Pandas . . . . .	41
Types and Functions . . . . .	41
How to Read This Chapter . . . . .	42
Part 1. DataFrame Basics . . . . .	42
Loading Data . . . . .	42
DataFrame Methods and Attributes . . . . .	43
Working with Subsets of Columns . . . . .	44
Indexing . . . . .	46
Outputting Data . . . . .	49
Part 2. Things you can do with DataFrames . . . . .	49
Introduction . . . . .	49
1. Modify or create new columns of data . . . . .	50
2. Use built-in Pandas functions that work on DataFrames . . . . .	58
3. Filter Observations . . . . .	63
4. Change Granularity . . . . .	69
5. Combining two or more DataFrames . . . . .	74
<b>4. SQL</b> . . . . .	<b>83</b>
Introduction to SQL . . . . .	83
How to Read This Chapter . . . . .	83
Databases . . . . .	83
SQL Databases . . . . .	85
A Note on NoSQL . . . . .	85
SQL . . . . .	85
Pandas . . . . .	85
Creating Data . . . . .	86
Queries . . . . .	87
Filtering . . . . .	89
Joining . . . . .	90
LIMIT/TOP . . . . .	95
DISTINCT . . . . .	95
UNION . . . . .	96

Subqueries . . . . .	96
LEFT, RIGHT, OUTER JOINS . . . . .	96
Final SQL Example . . . . .	97
<b>5. Web Scraping and APIs</b>	<b>100</b>
Introduction to Web Scraping and APIs . . . . .	100
Web Scraping . . . . .	101
HTML . . . . .	101
BeautifulSoup . . . . .	103
Fantasy Football Calculator ADP - Web Scraping Example . . .	104
APIs . . . . .	109
Two Types of APIs . . . . .	109
Fantasy Football Calculator ADP - API Example . . . . .	110
HTTP . . . . .	112
JSON . . . . .	113
<b>6. Summary Stats and Data Visualization</b>	<b>115</b>
Introduction to Summary Stats . . . . .	115
Distributions . . . . .	115
Summary Stats . . . . .	119
Density Plots with Python . . . . .	120
Relationships Between Variables . . . . .	126
Scatter Plots with Python . . . . .	126
Correlation . . . . .	127
Line Plots with Python . . . . .	133
Composite Scores . . . . .	134
Plot Options . . . . .	137
Wrapping columns . . . . .	137
Adding a title . . . . .	138
Modifying the axes . . . . .	138
Legend . . . . .	139
Plot size . . . . .	139
Saving . . . . .	139
<b>7. Modeling</b>	<b>141</b>
Introduction to Modeling . . . . .	141
The Simplest Model . . . . .	141
Linear regression . . . . .	142
Statistical Significance . . . . .	146
Regressions hold things constant . . . . .	149
Other examples of holding things constant . . . . .	152
Fixed Effects . . . . .	152
Squaring Variables . . . . .	153
Logging Variables . . . . .	154
Interactions . . . . .	154
Logistic Regression . . . . .	156

Random Forest . . . . .	157
Classification and Regression Trees . . . . .	158
Random Forests are a Bunch of Trees . . . . .	158
Random Forest Example in Scikit-Learn . . . . .	159
<b>8. Conclusion</b>	<b>163</b>
<b>Appendix: Places to Get Data</b>	<b>164</b>
Ready-made Datasets . . . . .	164
RScraper . . . . .	164
Google Dataset Search . . . . .	164
Kaggle.com . . . . .	164
Data Available via Public APIs . . . . .	165
myfantasyleague.com . . . . .	165
fantasyfootballcalculator.com . . . . .	165
sportsdata.io . . . . .	165