Contents

Prerequisites: Tooling **1**
- Files Included with this Book 1
- Python 2
  - Editor 4
  - Console (REPL) 4
  - Using Spyder and Keyboard Shortcuts 6
- Anki 6
  - Remembering What You Learn 6

1. Introduction **7**
- The Purpose of Data Analysis 7
- What is Data? 7
- Example Datasets 8
  - Play-By-Play Data 9
  - Player/Game Data 9
  - ADP Data 10
- What is Analysis? 11
- Types of Data Analysis 12
  - Summary Statistics 12
  - Modeling 13
- High Level Data Analysis Process 14
  - 1. Collecting Data 14
  - 2. Storing Data 15
  - 3. Loading Data 15
  - 4. Manipulating Data 15
  - 5. Analyzing Data for Insights 15
- Connecting the High Level Analysis Process to the Rest of the Book 16
- End of Chapter Exercises 17

2. Python **19**
- Introduction to Python Programming 19

How to Read This Chapter . . . 19
Important Parts of the Python Standard Library . . . 20
Comments . . . 20
Variables . . . 20
Types . . . 22
Interlude: How to Figure Things Out in Python . . . 23
Bools . . . 26
if statements . . . 27
Container Types . . . 27
Unpacking . . . 29
Loops . . . 30
Comprehensions . . . 31
Functions . . . 35
Libraries are Functions and Types . . . 42
os Library and path . . . 42
End of Chapter Exercises . . . 44

3. Pandas **47**
Introduction to Pandas . . . 47
Types and Functions . . . 47
How to Read This Chapter . . . 48
Part 1. DataFrame Basics . . . 48
Importing Pandas . . . 48
Loading Data . . . 49
DataFrame Methods and Attributes . . . 50
Working with Subsets of Columns . . . 51
Indexing . . . 52
Outputting Data . . . 56
Exercises . . . 57
Part 2. Things You Can Do With DataFrames . . . 58
Introduction . . . 58
1. Modify or Create New Columns of Data . . . 58
2. Use Built-In Pandas Functions That Work on DataFrames . . . 69
3. Filter Observations . . . 75
4. Change Granularity . . . 83
5. Combining Two or More DataFrames . . . 90

4. SQL 103
- Introduction to SQL . . . 103
- How to Read This Chapter . . . 103
- Databases . . . 103
  - SQL Databases . . . 105
  - A Note on NoSQL . . . 105
- SQL . . . 105
  - Pandas . . . 106
  - Creating Data . . . 106
  - Queries . . . 107
  - Filtering . . . 109
  - Joining, or Selecting From Multiple Tables . . . 111
  - Misc SQL . . . 115
  - SQL Example — LEFT JOIN, UNION, Subqueries . . . 117
- End of Chapter Exercises . . . 121

5. Web Scraping and APIs 122
- Introduction to Web Scraping and APIs . . . 122
- Web Scraping . . . 122
  - HTML . . . 123
  - BeautifulSoup . . . 125
  - Fantasy Football Calculator ADP - Web Scraping Example . . . 126
  - Exercises . . . 132
- APIs . . . 132
  - Two Types of APIs . . . 133
  - Fantasy Football Calculator ADP - API Example . . . 134
  - HTTP . . . 135
  - JSON . . . 136
  - Exercises . . . 137

6. Summary Stats and Data Visualization 138
- Introduction to Summary Stats . . . 138
- Distributions . . . 138
  - Summary Stats . . . 142
  - Density Plots with Python . . . 144
- Relationships Between Variables . . . 151
  - Scatter Plots with Python . . . 152
  - Correlation . . . 154

    - Line Plots with Python . . . 157
- Composite Scores . . . 161
- Plot Options . . . 161
    - Wrapping columns . . . 162
    - Adding a title . . . 162
    - Modifying the axes . . . 162
    - Legend . . . 163
    - Plot size . . . 163
    - Saving . . . 163
- End of Chapter Exercises . . . 165

7. Modeling **166**

- Introduction to Modeling . . . 166
    - The Simplest Model . . . 166
- Linear regression . . . 167
    - Statistical Significance . . . 172
    - Regressions hold things constant . . . 175
    - Other examples of holding things constant . . . 177
    - Fixed Effects . . . 178
    - Squaring Variables . . . 179
    - Logging Variables . . . 180
    - Interactions . . . 180
- Logistic Regression . . . 182
- Random Forest . . . 183
    - Classification and Regression Trees . . . 183
    - Random Forests are a Bunch of Trees . . . 184
    - Using a Trained Random Forest to Generate Predictions . . . 184
    - Random Forest Example in Scikit-Learn . . . 185
    - Random Forest Regressions . . . 189
- End of Chapter Exercises . . . 190

8. Intermediate Coding and Next Steps: High Level Strategies **192**

- Gall’s Law . . . 192
- Get Quick Feedback . . . 193
- Use Functions . . . 194
    - DRY: Don’t Repeat Yourself . . . 194
    - Functions Help You Think Less . . . 194
- Attitude . . . 195

Review 196

9. Conclusion **198**

Appendix A: Places to Get Data **199**
Ready-made Datasets 199
RScraper 199
Google Dataset Search 200
Kaggle.com 201
Data Available via Public APIs 201
myfantasyleague.com 201
fantasyfootballcalculator.com 201
sportsdata.io 201

Appendix B: Anki **202**
Remembering What You Learn 202
Installing Anki 203
Using Anki with this Book 204

Appendix C: Answers to End of Chapter Exercises **206**
1. Introduction 206
2. Python 208
3.0 Pandas Basics 212
3.1 Columns 214
3.2 Built-in Functions 218
3.3 Filtering 220
3.4 Granularity 222
3.5 Combining DataFrames 226
4. SQL 228
5.1 Scraping 230
5.2 APIs 233
6. Summary and Data Visualization 234
7. Modeling 240