**Preface (5 pages)**

This chapter gives a brief introduction to Pandas and data analysis.

* **What is the book about?**
* **Why use Python and Padas for Data Analysis?**

**Chapter 1: Getting Started with Pandas (10 Pages)**

This chapter will instruct the reader on obtain and install pandas and associated libraries.

* **Essential Python Libraries**
* **Setting up Pandas, IPython, NumPy, …**

By the end of this chapter the reader will have successfully installed pandas and be able to follow the examples in the rest of the book.

**Chapter 2: Examples of Manipulating Data with Pandas (30 pages)**

This chapter will demonstrate several important analytic capabilities easily accomplished by pandas through the use of several public data sets. I have yet to define which data sets, but I want to pull from some publicly available data from a data mart, and draw from three different domains.

* **Analysis of Public Data Set #1 (which one, TBD – likely financial in nature)**
* **Analysis of Public Data Set #2 (which one, TBD – likely economics related)**
* **Analysis of Public Data Set #3 (which one, TBD – likely social network related)**

By the end of the chapter the reader will have developed a sense of the power offered by pandas, and developed an appetite to learn more in the subsequent chapters.

**Chapter 3: Using IPython and NumPy: Foundations for Pandas (35 pages)**

Pandas relies heavily on the IPython and NumPy libraries. This chapter will give a brief introduction to both, explaining how both are instrumental to successful use of Pandas.

* **IPython**
  + **Using IPython**
  + **IPython HTML Notebooks**
  + **Coding techniques using IPython (debugging, profiling)**
* **NumPy**
  + **Multi-dimensional Arrays and Array functions**
  + **Data processing with Arrays**
  + **Linear Algebra**
  + **Random Number Generation**
  + **Example: Performing Random Walks**

By the end of this chapter the user will understand the value of using IPython to manage Pandas applications in workbooks (like in Mathematica), and have been refreshed on some of the mathematics of data science as offered by NumPy and useful in Pandas.

**Chapter 4: Using DataFrame and Indexes (35 pages)**

Pandas is built on top of NumPy and extends that library to provide tools for performing data analytics, not just mathematics. The chapter will introduce the reader to the concept of the DataFrame, and using indexes and series of data.

* **Pandas Data Structures: Series, Dataframe and Indexes**
* **Essential Capabilities: Indexing, sorting, data alignment, …**
* **Statistical Operations and Summarizations**

At the end of the chapter the reader will have a solid understanding of the concept of a DataFrame and how to start manipulating series data.

**Chapter 5: Accessing data with Pandas (30 pages)**

Pandas analyzes data. Data can come from a number of sources. In this chapter I will cover reading and/or writing data from text and binary files, as well as retrieving data from the web and database systems.

* **Reading and writing text data**
* **Reading and writing binary data**
* **Interacting with Web Data**
* **Interacting with Databases**

By the end of the chapter the reader will know how to bring data into Pandas from common sources, and write results to permanent storage.

**Chapter 6: Manipulating Data (25 pages)**

Once data is loaded into Pandas, it is then time to begin manipulating and deriving results from the data. This chapter will cover various concepts of data manipulation with pandas, as well how to handle common exceptions in data such as missing data.

* **Handling Missing Data**
* **Combining Data**
* **Pivoting Data**

**Chapter 7: Deriving results from manipulated data (25 pages)**

This chapter will show how to handle common exceptions in data such as missing data.

* **Transformations**
* **Grouping**
* **Aggregation**
* **Cross-tabulation**

At the end of the chapter the user will have learned how to reorder and derive results / analysis from data using the capabilities of pandas.

**Chapter 8: Time Series Operations (40 pages)**

Time series data is a critical form of data in many fields including finance, economics, social networks, physics, … Much of the interesting analysis from data comes from looking at how data has changed over time. This chapter will give the reader an understanding of time series data and how to include it in data analysis.

* **Time Series Basics**
* **Working with date ranges, frequencies and time shifts**
* **Working with Timezones**
* **Periods and arithmetic of periods**
* **Moving Windows**

At the end of the chapter the reader will have learned some of the basic concepts in time series analysis and be able to structure data and build applications to derived results from those data series.

**Chapter 9: Data Visualization (30 pages)**

Data visualization is one of the most important tasks in analyzing data as it helps to expose information that can otherwise be lost in the endless lists of numbers. This chapter will cover use of Matplotlib for visualizing data analyzed with Pandas.

* **Matplotlib overview**
* **Plotting Pandas Data**
* **Python Visualization Ecosystem**

At the end of the chapter readers will know how to use matplotlib to build line, bar, and several other types of charts from their pandas output.

**Chapter 10: Applications to Financial, Economic and Social Network Analysis (30 pages)**

This chapter will revisit the initial examples and drive into more detail of each, and providing a summary for the entire text. Whereas initial examples may show basic uses of pandas on the data sets, this will dive deeper into more complicated data analysis and visualization.

* **Financial example**
* **Economic example**
* **Social network analysis example**

By the finish of this chapter the reader will understand how all of the concepts given in the book apply to performing analysis on data in these domains, and gain a sense of how to use the tools to tackle their own data problems with pandas.\

**Summary**

Preface: (5 Pages)

Chapter 1: Getting Started with Pandas (10 Pages)

Chapter 2: Examples of Manipulating Data with Pandas (30 pages)

Chapter 3: Using IPython and NumPy: Foundations for Pandas (35 pages)

Chapter 4: Using DataFrame and Indexes (35 pages)

Chapter 5: Accessing data with Pandas (30 pages)

Chapter 6: Manipulating Data (25 pages)

Chapter 7: Deriving results from manipulated data (25 pages)

Chapter 8: Time Series Operations (40 pages)

Chapter 9: Data Visualization (30 pages)

Chapter 10: Applications to Financial, Economic and Social Network Analysis (30 pages)

**Total Pages: 295**