

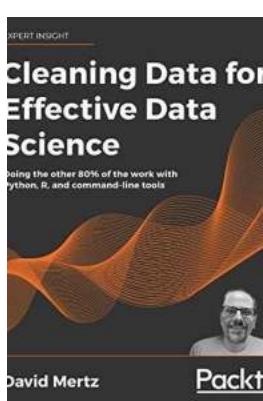
# Doing The Other 80 Of The Work With Python And Command Line Tools

In the world of coding, developers often face the challenge of managing the workload efficiently. While Python is a popular language for automation and large-scale data processing, it's important to explore the other 80% of the work to maximize productivity. This article will delve into how Python and command-line tools can help developers accomplish more.

## Why Python?

Python is a versatile programming language known for its simplicity and readability. With its extensive libraries and frameworks, Python provides developers with a wide range of tools and functionalities. It is particularly well-suited for tasks such as automating repetitive processes, manipulating data, and performing complex calculations.

By leveraging the power of Python, developers can write scripts and programs that significantly reduce the time and effort required to complete various tasks. However, Python shouldn't be seen as the sole solution for all problems. Complementing it with command-line tools can further enhance productivity and efficiency.



## Cleaning Data for Effective Data Science: Doing the other 80% of the work with Python, R, and command-line tools

by David Mertz (1st Edition, Kindle Edition)

4.8 out of 5

Language : English

Paperback : 25 pages

Item Weight : 4.2 ounces

Dimensions	: 8.5 x 0.06 x 11 inches
File size	: 8044 KB
Text-to-Speech	: Enabled
Screen Reader	: Supported
Enhanced typesetting	: Enabled
Print length	: 498 pages



## The Power of Command-Line Tools

Command-line tools are programs that can be executed directly from the command line or terminal. They provide developers with additional capabilities to interact with the operating system, perform system-level tasks, and access various functionalities. When combined with Python, command-line tools become a valuable asset in a developer's toolkit.

With command-line tools, developers can manipulate files, automate tasks, perform advanced data processing, manage version control systems, and much more. These tools offer flexibility and customizability, allowing developers to tailor their workflow according to specific requirements. Whether it's a small task or a complex operation, command-line tools provide the means to streamline the process.

## Practical Examples

To illustrate the potential of combining Python and command-line tools, let's explore a few practical examples:

### 1. File Operations

Python's `os` and `shutil` libraries provide functions for file operations. However, command-line tools like `grep`, `sed`, and `awk` can help developers efficiently

search, replace, and manipulate text within files. By utilizing these tools alongside Python, developers can perform complex file operations more effectively.

## 2. Task Automation

Python's subprocess module allows developers to run command-line tools from within a Python script. This feature enables seamless integration of command-line tools into automation workflows. Developers can create scripts that, for instance, automatically download files, interact with APIs, or deploy applications using command-line tools.

## 3. System Monitoring

Command-line tools like top, htop, and ps provide real-time information about system resources. By capturing the output of these tools using Python, developers can create custom monitoring dashboards, generate reports, or trigger specific actions based on system metrics. This integration enhances the ability to manage and optimize system performance.

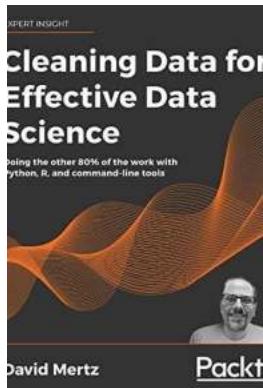
### The Importance of the Other 80%

While Python is indeed a powerful language, it cannot solve every problem on its own. By embracing the "other 80%"—command-line tools, developers can tap into a vast array of functionalities and optimize their workflow. This synergy allows for greater productivity, efficiency, and creativity.

Moreover, learning and incorporating command-line tools in your development process expands your skill set and makes you a more well-rounded developer. It equips you with additional problem-solving tools and helps you understand different aspects of system interaction.

Doing the "other 80%" of the work with Python and command-line tools is crucial if you want to maximize your efficiency as a developer. The combination of Python's versatility and command-line tools' power allows for more streamlined workflows, automation, and enhanced system management. By embracing both realms, you can unlock new possibilities and solve complex problems with greater ease.

So, don't limit yourself to Python alone—explore the world of command-line tools and harness their potential. It will undoubtedly take your coding skills to new heights.



## Cleaning Data for Effective Data Science: Doing the other 80% of the work with Python, R, and command-line tools

by David Mertz (1st Edition, Kindle Edition)

4.8 out of 5

Language	: English
Paperback	: 25 pages
Item Weight	: 4.2 ounces
Dimensions	: 8.5 x 0.06 x 11 inches
File size	: 8044 KB
Text-to-Speech	: Enabled
Screen Reader	: Supported
Enhanced typesetting	: Enabled
Print length	: 498 pages

DOWNLOAD E-BOOK

Think about your data intelligently and ask the right questions

## Key Features

- Master data cleaning techniques necessary to perform real-world data science and machine learning tasks
- Spot common problems with dirty data and develop flexible solutions from first principles
- Test and refine your newly acquired skills through detailed exercises at the end of each chapter

## **Book Description**

Data cleaning is the all-important first step to successful data science, data analysis, and machine learning. If you work with any kind of data, this book is your go-to resource, arming you with the insights and heuristics experienced data scientists had to learn the hard way.

In a light-hearted and engaging exploration of different tools, techniques, and datasets real and fictitious, Python veteran David Mertz teaches you the ins and outs of data preparation and the essential questions you should be asking of every piece of data you work with.

Using a mixture of Python, R, and common command-line tools, *Cleaning Data for Effective Data Science* follows the data cleaning pipeline from start to end, focusing on helping you understand the principles underlying each step of the process. You'll look at data ingestion of a vast range of tabular, hierarchical, and other data formats, impute missing values, detect unreliable data and statistical anomalies, and generate synthetic features. The long-form exercises at the end of each chapter let you get hands-on with the skills you've acquired along the way, also providing a valuable resource for academic courses.

## **What you will learn**

- Ingest and work with common data formats like JSON, CSV, SQL and NoSQL databases, PDF, and binary serialized data structures
- Understand how and why we use tools such as pandas, SciPy, scikit-learn, Tidyverse, and Bash
- Apply useful rules and heuristics for assessing data quality and detecting bias, like Benford's law and the 68-95-99.7 rule
- Identify and handle unreliable data and outliers, examining z-score and other statistical properties
- Impute sensible values into missing data and use sampling to fix imbalances
- Use dimensionality reduction, quantization, one-hot encoding, and other feature engineering techniques to draw out patterns in your data
- Work carefully with time series data, performing de-trending and interpolation

## **Who this book is for**

This book is designed to benefit software developers, data scientists, aspiring data scientists, teachers, and students who work with data. If you want to improve your rigor in data hygiene or are looking for a refresher, this book is for you.

Basic familiarity with statistics, general concepts in machine learning, knowledge of a programming language (Python or R), and some exposure to data science are helpful.

## **Table of Contents**

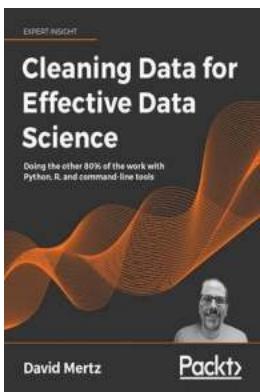
1. Data Ingestion – Tabular Formats
2. Data Ingestion - Hierarchical Formats
3. Data Ingestion - Repurposing Data Sources

4. The Vicissitudes of Error - Anomaly Detection
5. The Vicissitudes of Error - Data Quality
6. Rectification and Creation - Value Imputation
7. Rectification and Creation - Feature Engineering
8. Ancillary Matters - Closure/Glossary



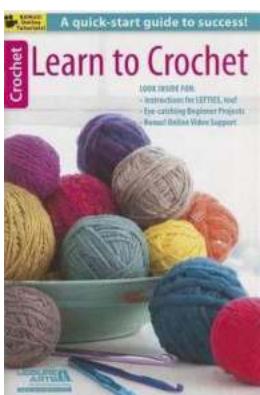
## The Ultimate Picture Rhyme Match Game: A Fun and Educational Activity for Preschool and Kindergarten Kids!

Are you looking for an engaging and interactive activity to help your preschool or kindergarten kids learn and have fun at the same time? Look no further! The Picture Rhyme...



## Doing The Other 80 Of The Work With Python And Command Line Tools

In the world of coding, developers often face the challenge of managing the workload efficiently. While Python is a popular language for automation and large-scale data...



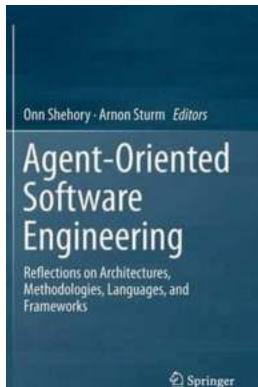
## Discover the Art of Crocheting: Learn to Crochet with Leisure Arts

Crocheting is an incredibly versatile and popular craft that allows you to create beautiful and functional pieces of art. Whether you are a beginner or have some experience,...



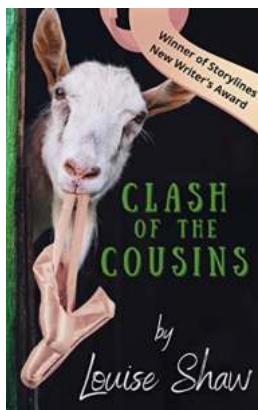
## Sculpting With Wool: Cute Easy Projects With Step By Step Tutorials

Have you ever heard of the fascinating art of sculpting with wool? If not, you are in for a treat! This unique and creative technique allows you to transform simple wool...



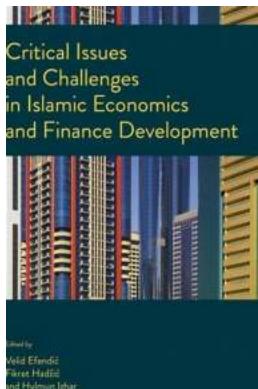
## Discovering the Power of Architectures: A Deep Dive into Methodologies, Languages, and Frameworks

Architectures form the backbone of our modern technological landscape. From software development to infrastructure design, they provide the necessary structure and...



## Clash of the Cousins: The Intriguing Story of Adventure and Mystery by Anna Ciddor

Are you ready for an epic adventure? Brace yourself, for Clash of the Cousins will take you on a journey like no other! Discover the captivating world created by Anna Ciddor,...



## Critical Issues And Challenges In Islamic Economics And Finance Development

The field of Islamic economics and finance has experienced tremendous growth in recent years. With over 1.8 billion Muslims around the world, there is a significant demand...



## The Captivating World of Butterfly 10 Cross Stitch Pattern - Mother Bee Designs

: Exploring the Beauty of Butterfly Cross Stitch Patterns Butterflies have long been regarded as symbols of transformation, beauty, and delicate grace....