

1. Record Nr.	UNINA9911004813803321
Autore	Campesato Oswald
Titolo	Data Wrangling Using Pandas, SQL, and Java
Pubbl/distr/stampa	Bloomfield : , : Mercury Learning & Information, , 2022 ©2022
ISBN	9781683929024 1683929020 9781683929031 1683929039
Edizione	[1st ed.]
Descrizione fisica	1 online resource (275 pages)
Disciplina	001.642
Soggetti	Computer programming Pandas
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover -- Title Page -- Copyright -- Dedication -- Contents -- Preface -- Chapter 1: Introduction to Python -- Tools for Python easy_install and pip virtualenv -- IPython -- Python Installation -- Setting the PATH Environment Variable (Windows Only) -- Launching Python on Your Machine -- The Python Interactive Interpreter -- Python Identifiers -- Lines, Indentation, and Multi-Lines -- Quotation and Comments -- Saving Your Code in a Module -- Some Standard Modules -- The help() and dir() Functions -- Compile Time and Runtime Code Checking -- Simple Data Types -- Working with Numbers -- Working with Other Bases -- The chr() Function -- The round() Function in Python -- Formatting Numbers in Python -- Working with Fractions -- Unicode and UTF-8 -- Working with Unicode -- Working with Strings -- Comparing Strings -- Formatting Strings in Python -- Uninitialized Variables and the Value None -- Slicing and Splicing Strings -- Testing for Digits and Alphabetic Characters -- Search and Replace a String in Other Strings -- Remove Leading and Trailing Characters -- Printing Text Without NewLine Characters -- Text Alignmet -- Working with Dates -- Converting Strings to Dates -- Exception Handling -- Handling User Input -- Command-Line Arguments -- Summary --

Chapter 2: Working with Data
 Dealing with Data: What Can Go Wrong?
 -- What is Data Drift? -- What are Datasets? -- Data Preprocessing --
 Data Types -- Preparing Datasets -- Discrete Data vs. Continuous Data
 -- "Binning" Continuous Data -- Scaling Numeric Data via
 Normalization -- Scaling Numeric Data via Standardization -- Scaling
 Numeric Data via Robust Standardization -- What to Look for in
 Categorical Data -- Mapping Categorical Data to Numeric Values --
 Working with Dates -- Working with Currency -- Working with Outliers
 and Anomalies -- Outlier Detection/Removal -- Finding Outliers with
 NumPy -- Finding Outliers with Pandas -- Calculating Z-Scores to Find
 Outliers -- Finding Outliers with SkLearn (Optional) -- Working with
 Missing Data -- Imputing Values: When is Zero a Valid Value? --
 Dealing with Imbalanced Datasets -- What is SMOTE? -- SMOTE
 Extensions -- The Bias-Variance Tradeoff -- Types of Bias in Data --
 Analyzing Classifiers (Optional) -- What is LIME? -- What is ANOVA? --
 Summary -- Chapter 3: Introduction to Pandas -- What is Pandas? --
 Pandas Data Frames -- Data Frames and Data Cleaning Tasks -- A
 Pandas Data Frame Example -- Describing a Pandas Data Frame --
 Pandas Boolean Data Frames -- Transposing a Pandas Data Frame --
 Pandas Data Frames and Random Numbers -- Converting Categorical
 Data to Numeric Data -- Merging and Splitting Columns in Pandas --
 Combining Pandas Data Frames -- Data Manipulation with Pandas Data
 Frames -- Pandas Data Frames and CSV Files -- Useful Options for the
 Pandas read_csv() Function -- Reading Selected Rows from CSV Files --
 Pandas Data Frames and Excel Spreadsheets -- Useful Options for
 Reading Excel Spreadsheets -- Select, Add, and Delete Columns in Data
 Frames -- Handling Outliers in Pandas -- Pandas Data Frames and
 Simple Statistics -- Finding Duplicate Rows in Pandas -- Finding
 Missing Values in Pandas -- Missing Values in an Iris-Based Dataset --
 Sorting Data Frames in Pandas -- Working with groupby() in Pandas --
 Aggregate Operations with the titanic.csv Dataset -- Working with
 apply() and mapapply() in Pandas -- Working with JSON-based Data --
 Python Dictionary and JSON -- Python, Pandas, and JSON -- Summary
 -- Chapter 4: RDBMS and SQL -- What is an RDBMS? -- What
 Relationships Do Tables Have in an RDBMS? -- Features of an RDBMS --
 What is ACID? -- When Do We Need an RDBMS? -- The Importance of
 Normalization -- A Four-Table RDBMS -- Detailed Table Descriptions
 -- The customers Table -- The purchase_orders Table -- The
 line_items Table -- The item_desc Table -- What is SQL? -- DCL, DDL,
 DQL, DML, and TCL -- SQL Privileges -- Properties of SQL Statements
 -- The CREATE Keyword -- What about MariaDB? -- Summary -- Index.

Sommario/riassunto

"This book contains a fast-paced introduction to as much relevant information about managing data that can be reasonably included in a book of this size. However, you will be exposed to a variety of features of NumPy and Pandas, how to create databases and tables in MySQL, and how to perform many data cleaning tasks and data wrangling. Some topics are presented in a cursory manner, which is for two main reasons. First, it's important that you be exposed to these concepts. In some cases, you will find topics that might pique your interest, and hence motivate you to learn more about them through self-study; in other cases, you will probably be satisfied with a brief introduction. In other words, you will decide whether or not to delve into more detail regarding the topics in this book. Second, a full treatment of all the topics that are covered in this book would significantly increase the size of this book, and few people have the time to read technical tomes"--