

1. Record Nr.	UNINA9911006690203321
Autore	Campesato Oswald
Titolo	Pandas Basics
Pubbl/distr/stampa	Bloomfield : , : Mercury Learning & Information, , 2022 ©2022
ISBN	9781683928249 1683928245 9781683928256 1683928253
Edizione	[1st ed.]
Descrizione fisica	1 online resource (215 pages)
Disciplina	005.133
Soggetti	COMPUTERS / Programming Languages / Python
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 -- Quotations 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 -- Formatting Numbers -- Working with Fractions -- Unicode and UTF-8 -- Working with Unicode -- Working with Strings -- Comparing Strings -- Formatting Strings -- 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 Alignment -- 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 Versus 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 Probability and Statistics -- What is a Probability? --
Calculating the Expected Value -- Random Variables -- Discrete versus
Continuous Random Variables -- Well-known Probability Distributions
-- Fundamental Concepts in Statistics -- The Mean -- The Median --
The Mode -- The Variance and Standard Deviation -- Population,
Sample, and Population Variance -- Chebyshev's Inequality -- What is a
p-value? -- The Moments of a Function (Optional) -- What is
Skewness? -- What is Kurtosis? -- Data and Statistics -- The Central
Limit Theorem -- Correlation versus Causation -- Statistical Inferences
-- Statistical Terms: RSS, TSS, R^2 , and F1 Score -- What is an F1
score? -- Gini Impurity, Entropy, and Perplexity -- What is the Gini
Impurity? -- What is Entropy? -- Calculating the Gini Impurity and
Entropy Values -- Multi-dimensional Gini Index -- What is Perplexity?
-- Cross-Entropy and KL Divergence -- What is Cross-Entropy? --
What is KL Divergence? -- What's Their Purpose? -- Covariance and
Correlation Matrices -- The Covariance Matrix -- Covariance Matrix: An
Example -- The Correlation Matrix -- Eigenvalues and Eigenvectors --
Calculating Eigenvectors: A Simple Example -- Gauss Jordan
Elimination (Optional) -- PCA (Principal Component Analysis) -- The
New Matrix of Eigenvectors -- Well-known Distance Metrics -- Pearson
Correlation Coefficient -- Jaccard Index (or Similarity) -- Local
Sensitivity Hashing (Optional).
Types of Distance Metrics -- What is Bayesian Inference? -- Bayes'
Theorem -- Some Bayesian Terminology -- What is MAP? -- Why Use
Bayes' Theorem? -- Summary -- Chapter 4: Introduction to Pandas (1)
-- What is Pandas? -- Pandas Options and Settings -- Pandas Data
Frames -- Data Frames and Data Cleaning Tasks -- Alternatives to
Pandas -- A Pandas Data Frame with a NumPy Example -- Describing a
Pandas Data Frame -- Pandas Boolean Data Frames -- Transposing a
Pandas Data Frame -- Pandas Data Frames and Random Numbers --
Reading CSV Files in Pandas -- Specifying a Separator and Column Sets
in Text Files -- Specifying an Index in Text Files -- The loc() and iloc()
Methods in Pandas -- Converting Categorical Data to Numeric Data --
Matching and Splitting Strings in Pandas -- Converting Strings to Dates
in Pandas -- Working with Date Ranges in Pandas -- Detecting Missing
Dates in Pandas -- Interpolating Missing Dates in Pandas -- Other
Operations with Dates in Pandas -- Merging and Splitting Columns in
Pandas -- Reading HTML Web Pages in Pandas -- Saving a Pandas Data
Frame as an HTML Web Page -- Summary -- Chapter 5: Introduction to
Pandas (2) -- Combining Pandas Data Frames -- Data Manipulation
with Pandas Data Frames (1) -- Data Manipulation with Pandas Data
Frames (2) -- Data Manipulation with Pandas Data Frames (3) --
Pandas Data Frames and CSV Files -- Managing Columns in Data
Frames -- Switching Columns -- Appending Columns -- Deleting

Columns -- Inserting Columns -- Scaling Numeric Columns --
Managing Rows in Pandas -- Selecting a Range of Rows in Pandas --
Finding Duplicate Rows in Pandas -- Inserting New Rows in Pandas --
Handling Missing Data in Pandas -- Multiple Types of Missing Values
-- Test for Numeric Values in a Column -- Replacing NaN Values in
Pandas -- Summary -- Chapter 6: Introduction to Pandas (3) --
Threshold Values and Outliers.
The Pandas Pipe Method -- Pandas query() Method for Filtering Data --
Sorting Data Frames in Pandas -- Working with groupby() in Pandas --
Working with apply() and mapapply() in Pandas -- Handling Outliers in
Pandas -- Pandas Data Frames and Scatterplots -- Pandas Data Frames
and Simple Statistics -- Aggregate Operations in Pandas Data Frames
-- Aggregate Operations with the titanic.csv Dataset -- Save Data
Frames as CSV Files and Zip Files -- Pandas Data Frames and Excel
Spreadsheets -- Working with JSON-based Data -- Python Dictionary
and JSON -- Python, Pandas, and JSON -- Window Functions in Pandas
-- Useful One-line Commands in Pandas -- What is pandasql? -- What
is Method Chaining? -- Pandas and Method Chaining -- Pandas
Profiling -- Alternatives to Pandas -- Summary -- Chapter 7: Data
Visualization -- What is Data Visualization? -- Types of Data
Visualization -- What is Matplotlib? -- Lines in a Grid in Matplotlib -- A
Colored Grid in Matplotlib -- Randomized Data Points in Matplotlib --
A Histogram in Matplotlib -- A Set of Line Segments in Matplotlib --
Plotting Multiple Lines in Matplotlib -- Trigonometric Functions in
Matplotlib -- Display IQ Scores in Matplotlib -- Plot a Best-Fitting Line
in Matplotlib -- The Iris Dataset in Sklearn -- Sklearn, Pandas, and the
Iris Dataset -- Working with Seaborn -- Features of Seaborn --
Seaborn Built-in Datasets -- The Iris Dataset in Seaborn -- The Titanic
Dataset in Seaborn -- Extracting Data from the Titanic Dataset in
Seaborn (1) -- Extracting Data from the Titanic Dataset in Seaborn (2)
-- Visualizing a Pandas Dataset in Seaborn -- Data Visualization in
Pandas -- What is Bokeh? -- Summary -- Index.

Sommario/riassunto

This book is intended for those who plan to become data scientists as well as anyone who needs to perform data cleaning tasks using Pandas and NumPy. It contains a variety of code samples and features of NumPy and Pandas, and how to write regular expressions. Chapter 3 includes fundamental statistical concepts and Chapter 7 covers data visualization with Matplotlib and Seaborn. Companion files with code are available for downloading from the publisher. FEATURES: Provides the reader with numerous code samples for Pandas and NumPy programming concepts, and an introduction to statistical concepts and data visualization. Includes an introductory chapter on Python. Companion files with code
