



Data Analytics with Pandas

1. Data Analysis Intro

Data analysis fundamentals

- Collecting Data
- Data wrangling
- Exploratory data analysis (EDA)
- Drawing conclusions

Basic Statistics

- Sampling
- Descriptive statistics
- Prediction and forecasting
- Inferential statistics

Setting up a virtual environment

- Virtual environments
- Installing Python packages
- Why choose pandas?
- Jupyter

2. Pandas DataFrames

Pandas data structures

- Series
- DataFrame
- Index

Creating a DataFrame from...

- A Python object
- A file
- A database
- An API

Inspecting a DataFrame

- Examining the data
- Describing and summarizing the data

Grabbing subsets of the data

- Indexing
- Selecting columns
- Slicing
- Filtering

Adding and removing data

- Creating new data & adding data
- Removing/deleting unwanted data

3. Data Wrangling

What is data wrangling

- Cleaning data
- Transforming data
- Data enrichment

Cleaning data

- Renaming DataFrame columns
- Type conversion of DataFrame columns
- Reordering, reindexing, and sorting data of DataFrame

Reshaping data

- Transposing DataFrames
- Pivoting DataFrames
- Melting DataFrames

Handling duplicate, missing, or invalid data

- Finding the problematic/corrupt data
- Handling the issues in data

4. Aggregating DataFrames

Performing database-style operations

- Querying & Merging DataFrames

Using DataFrame operations to improve the quality of data

- Arithmetic and statistics
- Binning
- Using functions on DataFrames
- Window calculations
- Pipes

Aggregating data

- Summarizing DataFrames
- Aggregating by group
- Pivot tables and crosstabs

Time series data

- Time-based selection and filtering
- Shifting for lagged data
- Differenced data
- Resampling data
- Merging time series data

5. Pandas and Matplotlib for Data Visualization

Matplotlib

- Introduction
- Plot components
- More options

Pandas for plotting

- Chart to show evolution over time
- Chart to show relationships between variables
- Charts for different distributions

- Counts and frequencies

The pandas.plotting module

- Scatter matrices
- Lag plots
- Autocorrelation plots
- Bootstrap plots

6. Seaborn for plotting and Customization Techniques

Seaborn for advanced plotting

- Representing categorical data
- Representing correlations and heatmaps
- Regression plots
- Faceting

Formatting plots with matplotlib

- Chart titles and labels
- Chart legends
- Formatting chart axes

Customizing the visualizations

- Adding reference lines to the graph
- Shading chart regions
- Annotations
- Colors
- Textures

7. Rule-Based Anomaly Detection

Simulating login attempts

- Presuppositions
- The login_attempt_simulator package
- Simulating from the command line

Exploratory data analysis

Implementing rule-based anomaly detection

- Percent difference
- Tukey fence
- Z-score
- Evaluating performance

