Working with Descriptive Statistics Using Pandas

Janani Ravi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Getting started with Pandas

Calculating and visualizing mean, mode, and median

Calculating range, interquartile range, variance, and standard deviation

Understanding and calculating skewness and kurtosis

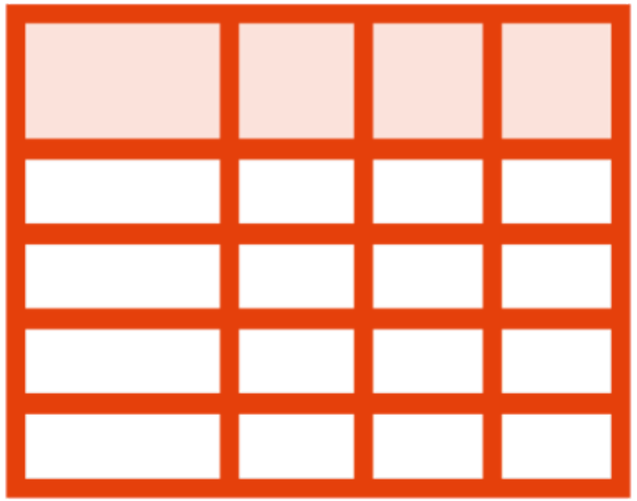
Exploring and visualizing correlations in data

Calculating and visualizing confidence intervals for population mean

Pandas

Extremely popular Python library for working with numerical tables and times series. Inspired by data frames in R.

Pandas



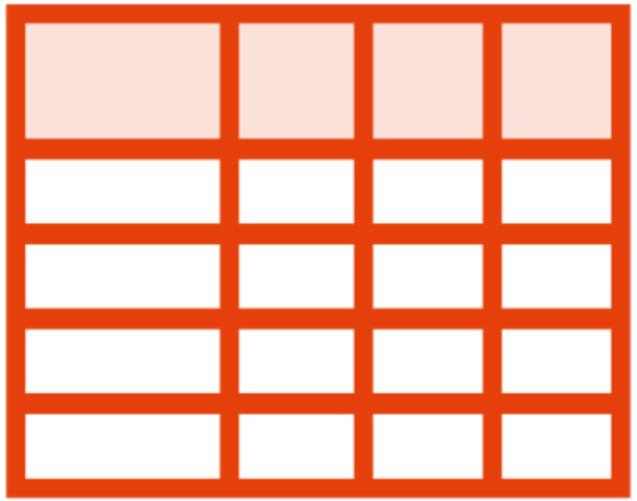
Tabular data with named rows and columns

Index value for each row

Indexable name for each column

Lookup, filter, pivot

Pandas



Easy to import and export data

Similar to functionality available in

- Excel
- R
- SQL querying

Demo

Getting started with Pandas

Demo

Calculating and interpreting mean, median, and mode

Demo

Calculating and interpreting interquartile range, variance, and standard deviation

Demo

Interpreting and visualizing summary statistics

Demo

Calculating and understanding skewness and kurtosis

Demo

Calculating and interpreting covariances and correlations

Demo

Calculating and visualizing confidence intervals for a measure

Summary

Getting started with Pandas

Calculating and visualizing mean, mode, and median

Calculating range, interquartile range, variance, and standard deviation

Understanding and calculating skewness and kurtosis

Exploring and visualizing correlations in data

Calculating and visualizing confidence intervals for population mean