

Comprehensive Data Science with Python

Code:	ACCEL-PYTH-CDS
Length:	5 days
URL:	View Online

This Comprehensive Data Science with Python training course teaches engineers, data scientists, statisticians, and other quantitative professionals the Python programming skills they need to chart, visualize, and apply inferential statistics. Our experienced instructors guide you through the full range of topics, starting with the basics, and equip you for advanced data science work.

Skills Gained

- Understand the difference between Python basic data types
- Know when to use different Python collections
- Implement python functions
- Understand control flow constructs in Python
- Handle errors via exception handling constructs
- Quantitatively define an answerable, actionable question
- Import both structured and unstructured data into Python
- Parse unstructured data into structured formats
- Understand the differences between NumPy arrays and pandas dataframes
- Simulate data through random number generation
- Understand mechanisms for missing data and analytic implications
- Explore and clean data
- Create compelling graphics to reveal analytic results
- Reshape and merge data to prepare for advanced analytics
- Find test for group differences using inferential statistics
- Implement linear regression from a frequentist perspective
- Understand non-linear terms, confounding, and interaction in linear regression
- Extend to logistic regression to model binary outcomes

Prerequisites

All attendees should have prior programming experience and an understanding of basic statistics.

Course Details

Software Requirements

- Anaconda Python 3.6 or later
- Spyder IDE and Jupyter notebook (Comes with Anaconda)

Outline

An Accelerated Introduction and Overview to Python for Data Science Foundations

- Introduction to course and computing environment
- Up and running with Jupyter notebooks
- Fundamental Python types: String literals, numeric, Boolean, and dates
- Understanding Python 'variables' (reference assignment)
- Slicing syntax
- Fundamental collections: tuples, lists, dictionaries, and sets
- Control flow iteration in Python (if/then, for, while, list comprehension)
- Writing your own functions
- Handling exceptions

Matrix Computing with NumPy

- Introduction to the ndarray
- Dtypes in NumPy
- NumPy operations, uFuncs
- Broadcasting
- Missing data in NumPy (masked array)
- Random number generation

Managing, Exploring, and Cleaning Data with Pandas

- Fundamental Pandas: Series and DataFrames
- Exploring objects with attributes/methods
- Importing data from different structured sources
- Basic DataFrame summaries
- Creating new variables (columns)
- Scaling and standardizing data elements
- Discretizing continuous data
- Mapping categorical data to new values
- Establishing dummy codes (one hot encoding)
- Filtering rows and selecting columns
- Managing the indices
- Identifying duplicate rows
- Quantifying and managing missing data
- Combining datasets
- Merging datasets
- Transposing datasets

- Changing data from long to wide formats and back

Exploratory Data Analysis with Pandas (including visualization with Seaborn)

- Univariate Statistical Summaries and Detecting Outliers, visually with graphical approaches and numerically.
- Multivariate Statistical Summaries and Outlier Detection, visually with graphical approaches and numerically.
- Groupwise calculations
- Pivot Table type operations to aggregate by group
- Pandas DataFrame plotting methods

Data Pseudo-Coding Process, Extension to Data-Centric Problems

- Identifying data verbs
- Answering a question using a well-formatted analytic dataframe
- Understanding the unit of analysis
- Identifying the unit of analysis for a given question - is my dataframe organized this way?
- Leveraging normalized data to create the analytic dataframe through combinations of data verbs
- Identify the question and unit of analysis
- Define the desired analytic dataframe
- Examine the normalized source data
- Create data pseudo-code to map source data to the final analytic dataframe
- Implement with Python

Focus on Graphics with Python: Seaborn, Matplotlib, and Plotly

- Using seaborn for 1 and 2 variable summaries
- Advanced statistical plots with Seaborn
- Controlling plot details through Seaborn
- Making graphs interactive with Plotly
- Introduction to Matplotlib for full control of parameters

Overview of Descriptive versus Inferential Analytics

- Identifying the null hypothesis
- P-value interpretation
- The idea of statistical power and type 1/2 errors

Implementing Inferential Statistics in Python

- Analyzing an A/B randomized test:
- T-tests/ANOVA
- Chi-square tests
- Correlation methods

Multivariate Models: Linear Regression

- Estimating the mean
- Identifying p-values of interest
- Adding a categorical predictor and the link to t-tests
- Nonlinear trends: Polynomial regression and spline modeling
- Interaction terms
- Confounding
- Model building approaches (choosing the best model)
- Scoring new data from the model (making predictions)

Multivariate Models: Logistic Regression

- GLMs and the link function
- Understanding the logit function
- The binomial distribution and
- Recovering the average event probability from the model
- Interpreting the coefficient - the odds ratio
- Categorical predictors and the connection to the chi-square test
- Expansion to more complex models (non-linear trends, multiple predictors)
- Confounding
- Interaction terms
- Making predictions
- Comparing models and picking the 'best' model

Conclusion

Optional modules depending on student interest and timing:

Analyzing unstructured data with Python

- Analyzing unstructured data with Python
- Overview of structure versus unstructured data
- Implementing regular expressions in Python
- Converting unstructured data to structured data for analysis
- Missing Data
- Exploring and understanding patterns in missing data
- Missing at Random
- Missing Not at Random
- Missing Completely at Random
- Data imputation methods

Download Whitepaper: Accelerate Your Modernization Efforts with a Cloud-Native Strategy

Get Your Free Copy Now

ExitCertified® Corporation and iMVP® are registered trademarks of ExitCertified ULC and ExitCertified Corporation, respectively
Copyright ©2024 ExitCertified ULC & ExitCertified Corporation. All Rights Reserved.

Generated 4