elite Analytical Solutions®
Business Analytics + Data Consultants

Python™ Data Prep

Length: 3 Days / 6 Half Days

Overview: This course is designed for students who are new to Python or have limited experience with programming. Students will learn the fundamentals of Python with a focus on descriptive analytics and statistics. You will explore data manipulation and transformation techniques, including how to work with lists, dictionaries, loops, and functions, as well as utilizing libraries like Pandas and NumPy. By the end of the course, you will be equipped to clean and analyze data effectively, laying a strong foundation for data-driven insights.

Objectives: This course aims to teach the basics of Python and data preparation.Topics include:

- Python basics
- Lists
- Dictionaries
- Loops
- Functions
- Packages and libraries
- Pandas
- NumPy
- Querying databases
- Data cleaning
- Filtering
- Joining
- Aggregations
- Outputs

Python Basics

- **Hello Python**
  - Install software (anaconda)
  - The Jupyter notebook environment
  - Print(Hello World!)
    - Adding Comments
  - Variable types
  - Variable assignment
  - Calculations with variables
  - Type conversions
- **Python Lists – learn to store, access, and manipulate data in lists (helps working with big data)**
  - Create a list
  - Mutability
  - Sub-setting
- **Dictionaries**
  - Create Dictionaries
  - Access Dictionaries
  - Dictionary Manipulation
- **Logic, Control Flow, and Filtering**
  - Comparison operators
  - Boolean operators
  - Conditional Statements
  - Loops
- **Packages & Functions**
  - Packages
  - What are Functions & Methods?
    - Built-in Functions
    - String methods
    - List methods

Intro to Pandas DataFrame Library

- **Pandas & DataFrames**
  - Creating Table Data using Pandas DataFrame
  - Adding Row and/or Column Data to Your Table
  - Adding an Aggregate Column and/or Row
  - Optional: Multilevel DataFrame

continued

- **Data Exploration**
  - Importing and Exporting Data
  - Excel, CSV, and Txt files
  - Loading in Chunks of Data
  - Reading and Moving through your Data
  - Viewing your Data
  - Slicing and DataFrame vs Series
- **Data Filtering**
  - Basics
  - Dates
  - Filtering Based on Letters in a String
  - Optional: Multi-Level Filtering
  - Sorting

Advanced Pandas Usage

- **Descriptive Analytics**
  - Grouping Basics
  - Advanced Grouping
  - Optional: Multi-Level Grouping and Aggregating
  - Plotting or Charting to describe your Data
  - Basic Charting
  - Plotting Against v. Plotting Together
  - Using Parameters to Add Features and Change Format
  - Practicing describing trends
- **Data Cleaning**
  - Making Changes to Your Data
  - Warning: Redefining Columns and Rows
  - Renaming Columns and Rows
  - Changing Specific Entries
  - Removing Data
  - Replace and Where Function
  - Using a Filter to Change Columns and Rows
  - Cleaning: Getting Rid of Empty Values
  - Using Fillna() to Combine Rows and Columns
  - Removing Data from Multi-Level Data
  - Lambda Functions
- **Data Joining**
  - Last Call for Combining Data
  - Concat and Append
  - Join
  - Left and right
  - Inner and Outer
  - Multiple Join
  - Merge
  - Merging Multiple Tables

Prerequisites
No experience with Python is required, or other coding languages. However, some experience in programming will be helpful in picking concepts up more quickly. Some material will be harder and/or take longer for classes who have no background or understanding of data architecture (joins, pivots, etc.)

Materials
- All students will receive slides with lecture material and data and labs.
- Software needed on each student PC:
  - Microsoft Excel 2010 or later (2013 or later recommended)
  - Internet access
- Related data and lab files will be provided

Other Python Courses Available: Python Diagnostic; Python Predictive