

Python Analytics: Pandas, Seaborn & Power BI

Learn to clean, analyse and visualise data with pandas, Seaborn and Plotly, then connect Python to SQL Server and Tableau.

20 Days

DURATION

20

MODULES

134

TOPICS

11

HANDS-ON ACTIVITIES

4 weeks (20 working days)

COURSE OVERVIEW

This comprehensive 4-week training program provides an intensive deep dive into data analytics and visualization using Python, with a primary focus on Excel as the data source. Designed for professionals with existing Python programming experience, this course bridges the gap between programming skills and practical data analysis capabilities. Participants will master essential data manipulation techniques using pandas and NumPy, create compelling visualizations with Matplotlib, Seaborn, and Plotly, and gain foundational knowledge of data warehousing concepts. The training includes hands-on experience with extracting data from MS SQL Server and creating interactive dashboards in Tableau, ensuring participants can work effectively across the modern analytics technology stack.

WHO THIS COURSE IS FOR

Data Analysts

Business Intelligence Developers

Python Developers

BI Developers

Data Engineers

Reporting Analysts

PREREQUISITES

- Proficiency in Python programming (variables, data types, control structures, functions, and object-oriented programming)
- Basic understanding of data structures (lists, dictionaries, tuples)
- Familiarity with file handling in Python
- Basic knowledge of databases and SQL concepts (beneficial but not mandatory)
- Understanding of basic statistical concepts (mean, median, standard deviation)

What You Will Learn

Upon successful completion of this training, participants will be able to:

- Import, clean, transform, and manipulate data from Excel files using pandas and openpyxl libraries
- Perform comprehensive exploratory data analysis (EDA) to uncover patterns, trends, and insights from datasets
- Apply statistical analysis techniques including descriptive statistics, correlation analysis, and hypothesis testing
- Create professional static and interactive visualizations using Matplotlib, Seaborn, and Plotly
- Understand fundamental data warehousing concepts including ETL processes, dimensional modeling, and OLAP
- Connect to and extract data from MS SQL Server databases using pyodbc
- Integrate Python analytics with Tableau for advanced dashboard development
- Implement data quality checks and handle missing values, duplicates, and outliers effectively
- Automate data analysis workflows and generate reports programmatically
- Apply best practices for reproducible data analysis and documentation

COURSE SYLLABUS

20 Modules, 134 Topics

01

Environment Setup and Excel Data Import

TOPICS

- Setting up Python environment (Anaconda, Jupyter Notebook, VS Code)
- Installing and managing essential libraries (pandas, NumPy, Matplotlib, Seaborn, openpyxl)
- Reading Excel files using pandas (read_excel, ExcelFile, sheet navigation)
- Working with openpyxl for advanced Excel operations (workbook manipulation, cell operations)
- Handling multiple sheets and workbooks programmatically
- Excel file formats and compatibility considerations (xlsx, xls, xlsxm)
- Writing data back to Excel with formatting preservation
- Performance optimization techniques for large Excel files

02

Pandas Data Structures - Part 1

TOPICS

- Understanding pandas Series and DataFrame architecture
- Data indexing, selection, and filtering techniques (loc, iloc, boolean indexing)
- Handling different data types and type conversions
- Working with categorical data and encoding techniques
- Practical exercises on data selection and filtering
- Creating DataFrames from various sources
- Understanding DataFrame properties and methods

03

Pandas Data Structures - Part 2

TOPICS

- Data aggregation and grouping operations (groupby, pivot tables, crosstab)
- Merging, joining, and concatenating datasets from multiple sources
- Reshaping data (melt, pivot, stack, unstack)
- Time series data handling and datetime operations
- Advanced groupby operations and multi-level indexing
- Combining multiple datasets for analysis
- Performance considerations for large-scale operations

04

Data Cleaning and Quality Assessment

TOPICS

- Identifying and handling missing values (isna, fillna, dropna, interpolation)
- Detecting and removing duplicate records
- Data validation and quality assessment techniques
- String manipulation and text processing using pandas string methods
- Outlier detection and treatment strategies
- Data profiling and quality metrics
- Creating data quality reports

05

Hands-on project

Data Transformation and Feature Engineering

TOPICS

- Data normalization and standardization techniques
- Feature engineering and creating derived columns
- Using apply, map, and lambda functions for custom transformations
- Type conversions and data formatting
- Binning and discretization techniques
- Creating categorical variables from continuous data

06

Hands-on exercise

Exploratory Data Analysis Fundamentals

TOPICS

- Understanding the EDA process and its importance in analytics
- Univariate analysis techniques (distribution analysis, frequency tables)
- Summary statistics using describe, info, and value_counts
- Creating data profiling reports
- Identifying data patterns and initial insights
- Documenting EDA findings effectively

07

Case study

Bivariate and Multivariate Analysis

TOPICS

- Bivariate analysis (scatter plots, correlation matrices, cross-tabulations)
- Multivariate analysis and relationships between multiple variables
- Identifying patterns, trends, and anomalies in data
- Correlation vs causation understanding
- Creating comprehensive EDA reports
- Analyzing relationships across multiple dimensions

08

Statistical Analysis - Part 1

TOPICS

- Descriptive statistics (mean, median, mode, variance, standard deviation)
- Probability distributions and their applications
- Understanding normal distribution and other common distributions
- Measures of central tendency and dispersion
- Using Python statistics module for basic computations
- Handling skewness and kurtosis in distributions
- Practical statistical analysis with real datasets

09

Activity

Statistical Analysis - Part 2

TOPICS

- Correlation analysis (Pearson, Spearman, Kendall coefficients)
- Hypothesis testing fundamentals (t-tests, chi-square tests, ANOVA)
- Confidence intervals and p-value interpretation
- Statistical significance and practical significance
- Using SciPy for advanced statistical computations
- Type I and Type II errors in hypothesis testing

10

Data Visualization with Matplotlib

TOPICS

- Understanding Matplotlib architecture (Figure, Axes, pyplot interface)
- Creating basic plots (line charts, bar charts, scatter plots, histograms)
- Customizing plot elements (colors, markers, line styles, legends)
- Working with subplots and figure layouts
- Annotating and labeling visualizations effectively
- Saving figures in various formats (PNG, PDF, SVG)
- Creating publication-quality static visualizations
- Best practices for choosing appropriate chart types

11

Practical exercise

Advanced Visualization with Seaborn

TOPICS

- Understanding Seaborn's statistical plotting capabilities
- Distribution plots (histograms, KDE plots, rug plots, dist plots)
- Categorical plots (bar plots, count plots, box plots, violin plots, swarm plots)
- Relationship plots (scatter plots, line plots, regression plots)
- Customizing Seaborn themes and color palettes
- Creating visually appealing statistical graphics

12

Case study

Seaborn Advanced Features and Multi-dimensional Plots

TOPICS

- Matrix plots (heatmaps, cluster maps, correlation matrices)
- Facet grids and pair plots for multi-dimensional analysis
- Combining Seaborn with Matplotlib for enhanced visualizations
- Creating complex multi-panel visualizations
- Color palette strategies for effective communication
- Styling and theming best practices

13

Hands-on

Interactive Visualizations with Plotly

TOPICS

- Introduction to Plotly and interactive visualization concepts
- Creating interactive charts (scatter, line, bar, bubble, pie charts)
- 3D visualizations and surface plots
- Adding interactivity (hover tooltips, zooming, panning, filtering)
- Creating dashboards with Plotly Express
- Exporting interactive visualizations as HTML files
- Integration considerations for web applications

14

Data Warehousing Fundamentals

TOPICS

- Understanding data warehouse architecture and components
- Core data warehousing concepts (subject-oriented, integrated, time-variant, non-volatile)
- Dimensional modeling principles (Star schema, Snowflake schema)
- Fact tables and dimension tables design patterns
- ETL vs ELT processes and workflows
- OLAP vs OLTP systems and their use cases
- Data marts and their relationship to enterprise data warehouses
- Modern data warehousing trends (cloud data warehouses, data lakehouses)

15

Activity

Geographic and Specialized Visualizations

TOPICS

- Geographic visualizations and map-based analytics with Plotly
- Time series visualization techniques
- Financial and business-specific chart types
- Creating dynamic and animated visualizations
- Dashboard design principles and best practices
- Choosing the right visualization for your data story

16

Hands-on

Connecting Python to MS SQL Server - Part 1

TOPICS

- Installing and configuring pyodbc driver for SQL Server connectivity
- Creating connection strings and establishing database connections
- Executing SQL queries from Python and retrieving results
- Parameterized queries and SQL injection prevention
- Reading SQL query results into pandas DataFrames
- Error handling and connection management best practices

17

Case study

Connecting Python to MS SQL Server - Part 2

TOPICS

- Writing DataFrame data to SQL Server tables
- Managing database transactions and connection pooling
- Bulk insert operations for performance optimization
- Creating automated data extraction pipelines
- Combining SQL queries with pandas transformations
- Working with stored procedures from Python

18

Python and Tableau Integration - Part 1

TOPICS

- Understanding TabPy (Tableau Python Server) architecture
- Installing and configuring TabPy server
- Connecting Tableau Desktop to TabPy
- Creating Python-calculated fields in Tableau
- Passing data between Tableau and Python functions
- Basic analytics integration examples
- Setting up the integration environment

Python and Tableau Integration - Part 2

TOPICS

- Implementing custom analytics using Python scripts in Tableau
- Use cases for Python-Tableau integration (advanced statistics, ML predictions)
- Deploying TabPy solutions to Tableau Server
- Performance considerations for Python scripts in Tableau
- Debugging and troubleshooting integration issues
- Creating advanced calculated fields with Python

Automation, Best Practices, and Final Project

TOPICS

- Automating repetitive data analysis tasks with Python scripts
- Creating reusable functions and modules for analytics workflows
- Generating automated reports with formatted Excel outputs
- Scheduling Python scripts for regular execution
- Version control for analytics code using Git
- Code documentation and commenting best practices
- Creating reproducible analysis pipelines
- Performance optimization and memory management techniques

HANDS-ON ACTIVITIES

Practise What You Learn

#	MODULE	ACTIVITY
05	Data Transformation and Feature Engineering	Hands-on project: Complete data cleaning pipeline
06	Exploratory Data Analysis Fundamentals	Hands-on exercises: Hands-on exercises with real-world datasets
07	Bivariate and Multivariate Analysis	Case study: Complete EDA workflow
09	Statistical Analysis - Part 2	Hands-on statistical testing exercises: Hands-on statistical testing exercises
11	Advanced Visualization with Seaborn	Practical exercises: Practical exercises with business datasets
12	Seaborn Advanced Features and Multi-dimensional Plots	Case study: Creating comprehensive analytical dashboards

#	MODULE	ACTIVITY
13	Interactive Visualizations with Plotly	Hands-on: Building interactive analytical reports
15	Geographic and Specialized Visualizations	Comprehensive visualization project: Comprehensive visualization project combining all techniques
16	Connecting Python to MS SQL Server - Part 1	Hands-on: Extracting data from SQL Server databases
17	Connecting Python to MS SQL Server - Part 2	Case study: Building ETL pipeline from SQL Server to Excel
19	Python and Tableau Integration - Part 2	Hands-on project: Building Python-enhanced Tableau dashboard

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Hands-on exercises throughout the course
- Delivery by certified instructors and subject matter experts
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training

Live Online Classes (ILO)

Fly-Me-A-Trainer (FMAT)

Flexi (Self-Paced)

1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com

✉ info@koenig-solutions.com