Certified Entry-Level Data Analyst with Python (PCED)

Duration: 3 days/ 24 hours
Prerequisites: Knowledge of Python Programming.

1. Data Acquisition and Preprocessing

- Types and sources of data (CSV, Excel, JSON, APIs, scraping)
- Ethical and legal considerations (data privacy, GDPR)
- Web scraping basics using requests and BeautifulSoup
- Reading and writing files with pandas
- Handling missing values and duplicates
- Data type conversion and formatting
- Normalization and standardization
- Encoding categorical data (Label, One-Hot)
- Feature extraction basics

Labs:

- Read and clean a CSV file with pandas
- Scrape a public dataset (e.g., quotes or book listings)
- Handle missing and duplicated values in a messy dataset
- Normalize and encode a categorical dataset

---

2. Python Programming Essentials

- Python syntax, variables, and data types
- Lists, dictionaries, tuples, and sets
- Control structures: conditionals and loops
- Functions and scope
- File handling (read/write)
- Importing modules and working with packages

Labs:

- Write a Python script to summarize a dataset
- Parse and transform structured data (e.g., JSON to CSV)
- Create a utility function for data cleaning

---

3. SQL and Database Integration

- SQL basics: SELECT, WHERE, ORDER BY, GROUP BY, HAVING
- Joins: INNER JOIN and LEFT JOIN
- Using SQLite: creating tables, inserting and retrieving data
- Connecting Python to SQLite using sqlite3
- Querying and transforming database results in pandas

Labs:

- Create and query a SQLite database
- Join two tables and filter results
- Connect to SQLite from Python and perform queries
- Export SQL results to a pandas DataFrame

---

4. Statistical Analysis

- Descriptive statistics: mean, median, mode, std dev, variance
- Visualizing distributions
- Correlation and covariance
- Introduction to inferential statistics
- Simple linear regression and interpretation

Labs:

- Compute summary statistics using numpy and scipy
- Visualize distributions with histograms
- Perform and interpret simple linear regression using scikit-learn

---

5. Data Analysis with Pandas

- Loading and exploring datasets
- Filtering, grouping, and sorting data
- Merging and joining datasets
- Aggregation functions and pivot tables
- Exploratory Data Analysis (EDA)

Labs:

- Perform filtering, sorting, and grouping operations on a real dataset
- Merge two datasets and generate summary statistics
- Conduct EDA on a sample dataset (e.g., Titanic, sales data)

---

6. Data Modeling

- Introduction to machine learning workflow
- Train-test split using scikit-learn
- Building a linear regression model
- Evaluating model performance (MSE, R^2)

Labs:

- Build and evaluate a regression model
- Interpret coefficients and prediction errors
- Test model performance using a train/test split

---

7. Data Visualization

Subtopics:

- Plotting with matplotlib and seaborn
    - Bar chart, histogram, scatter plot, boxplot
    - Correlation heatmap
- Visual storytelling and choosing the right chart
- Optional: Intro to streamlit for dashboard creation

Labs:

- Create visual summaries for numerical and categorical data
- Visualize correlations between features
- Build a simple interactive dashboard using streamlit (if time permits)

---

8. Capstone Project & PCED Exam Preparation

Subtopics:

- End-to-end workflow: data acquisition → cleaning → analysis → modeling → visualization
- Presenting insights
- PCED certification exam overview and readiness tips
- Mock quiz and discussion

Labs:

- Complete a mini-project using a real-world dataset (e.g., sales, HR, or public health data)