Training Day 10 Report

Date: 9 July 2025

Topic: Data Cleaning, Exploratory Data Analysis (EDA), and Pivot Tables using Pandas

Session Summary:
The tenth day of the training focused primarily on data cleaning techniques, exploratory data analysis (EDA), and pivot table creation using Pandas in Python. The session also revisited previous topics on handling missing data, visualizing data insights, and understanding the concepts of covariance and correlation.

Key Concepts Covered:

- **1. Recap of Previous Sessions:** Reviewed NumPy and Pandas basics: creating, reading, and writing CSVs, using info(), shape, describe(), head(), tail(), and plotting basics using matplotlib.
- **2. Exploratory Data Analysis (EDA):** Used Titanic dataset for initial data inspection with info(), describe(), head(), and tail(). Identified data patterns and anomalies.
- **3. Handling Missing Values:** Explored four methods: dropping rows, dropping columns (threshold), filling missing values with mean/median/mode, and forward/backward filling using fillna().
- **4. Handling Special Null Representations:** Used replace() to convert placeholders like 'None' or empty strings into NaN values.
- **5. Removing Duplicates:** Used duplicated() and drop_duplicates() to identify and remove redundant rows.
- **6. Pivot Tables in Pandas:** Created pivot tables with pivot_table() to analyze survival rate, gender, and passenger class relationships. Learned parameters like values, index, columns, and aggfunc.
- **7. Data Type Conversion (Typecasting):** Checked column types using dtypes and converted types using astype().
- **8. String Operations in Pandas:** Applied string functions: upper(), lower(), len(), and strip() for cleaning textual data.
- **9. Covariance, Correlation, and Causation:** Understood covariance as direction, correlation as strength (-1 to 1), and causation as direct cause-effect relationship.
- **10. Correlation Matrix and Heatmaps:** Introduced correlation visualization techniques like heatmaps to interpret variable relationships.

Skills Practiced:

- Data cleaning and preprocessing
- Handling missing values and duplicates
- Using Pandas functions effectively
- Creating and interpreting pivot tables
- Understanding covariance and correlation • Applying string operations for data formatting

Reflection & Learning Outcome:

Day 10 provided deep insights into real-world data handling and EDA workflows. Participants learned to clean data, prepare it for analysis, and derive insights through pivot tables and correlation analysis. This session effectively bridged theoretical understanding with hands-on implementation.