Data Handling & Analysis using Python – Detailed Workflow

Document Owner: Indhumathi P
Team Members: Balaji P, Indhumathi P, Diljith R
Course: BCA

1. Overview

This document describes the detailed workflow of the Data Handling and Analysis project using Python. The goal is to transform raw, unstructured datasets into clean, organized, and visualized data ready for AI/ML applications. The project focuses on six main stages—Data Collection, Cleaning, Transformation, Statistical Analysis, Visualization, and Reporting.

Stage 1: Data Collection

Input: Raw datasets (CSV, Excel, JSON)
Process: Import data using Pandas (read_csv, read_excel) and inspect structure (.info(), .head())
Output: Data loaded into DataFrame for further processing
Tools: Pandas, Jupyter Notebook / VS Code

Stage 2: Data Cleaning

Input: Raw DataFrame
Process: Handle missing values (fillna(), dropna()), remove duplicates, correct formats, standardize data
Output: Cleaned DataFrame ready for transformation
Tools: Pandas, NumPy

Stage 3: Data Transformation

Input: Cleaned DataFrame
Process: Filter columns, group data, merge datasets (merge, concat), create calculated columns
Output: Transformed dataset ready for analysis
Tools: Pandas, NumPy

Stage 4: Statistical Analysis

Input: Transformed dataset
Process: Compute metrics (mean, median, sum, std), identify trends, perform correlation analysis
Output: Analytical insights and metrics
Tools: NumPy, Pandas

Stage 5: Data Visualization

Input: Analytical insights
Process: Create charts (bar, line, scatter, pie), generate heatmaps and trend graphs
Output: Visual dashboards and insights

Tools: Matplotlib, Seaborn

Stage 6: Export & Reporting

Input: Cleaned dataset and visualizations
Process: Export processed data (to_csv, to_excel), generate summary reports, integrate dashboards
Output: Final reports and datasets ready for use
Tools: Pandas, Matplotlib, Seaborn, Streamlit (optional)

3. Workflow Summary

Raw Data → Data Collection → Data Cleaning → Data Transformation → Statistical Analysis → Data Visualization → Export & Reporting

4. Future Enhancements

• Extend analysis to predictive ML models
• Automate dashboards using Streamlit or Power BI
• Integrate with CRM systems like Jewel Management CRM
• Build real-time AI pipelines for dynamic reporting