CSV File Cleaner and Plotter Tool

December 13, 2025

The purpose of this project was to create a Python-based command-line tool that can clean, transform, plot and export data from CSV files. The tool should be useful with laboratory work results and datasets where the raw CSV files often contain inconsistent formatting, missing values, decimal commas, mixed units and other issues that make them difficult to use directly. The program aims to automate common cleaning steps, simplify data inspection, and provide a convenient way to export tables for reports.

The project consists of two Python files, `main.py` and `csv_actions.py`, both located in the same folder so that imports work correctly. The `main.py` file handles the user interface, such as selecting a CSV file, choosing operations and saving output. It also controls the execution sequence and manages the optional cleaning log. The `csv_actions.py` file contains all functions related to data cleaning, unit extraction, SI conversion, plotting and LaTeX export.

The project includes several cleaning features. It can remove rows or columns based on the percentage of missing values, strip unnecessary whitespace from column names and cell contents, normalize common missing-value patterns, convert decimal commas into dots to standardize numerical values, extract numeric values and units from cells or from column headers, which is useful when measurements are stored in combined formats, convert units into SI units, remove duplicate rows and allows the user to move rows or columns interactively within the table.

The plotting function allows the user to choose columns, select a plot type, customize labels, title, axis scales and grid visibility. The resulting plot is saved as a PNG file. The plotting process does not modify the CSV file.

The project also includes a LaTeX export feature. The user can select which columns and rows to include and it generates a standalone LaTeX document containing a formatted table. This file can be compiled directly with pdflatex and is useful for laboratory reports.

A cleaning log can be generated whenever the user selects at least one action that modifies the CSV file. The log describes what changes were made during each step. If the user selects only plotting or LaTeX export, the program automatically skips the log prompt since those actions do not modify the data.

I also added a config, for specific functions(extraction of numeric values and units, SI conversion, plotting and LaTeX export), since I think these are the most useful features to speed up report writing for my lab works.

The final project successfully cleans messy datasets, handles units consistently, visualizes data and exports tables. It is easy to extend with additional cleaning functions or new

unit types. The tool is useful for anyone working with experimental or measurement data stored in CSV format and helps reduce repetitive manual work while improving data reliability.