**General Project overview**

This project is designed to handle stock market data for different groups of stocks, including American (S&P 500) and European (Euro Stoxx 50) indices. The project involves collecting, preprocessing, and analyzing stock data to assist in manual financial analysis performed in Excel. This document outlines the folder structure, workflows, and preprocessing steps used in the project.

**Folder Structure**

The project is organized as follows:

/stock\_project/

│

├── /assets/ # Main folder for different stock groups or asset classes

│ ├── /crypto/ # Stores data related to cryptocurrency assets

│ ├── /raw\_data/ # Stores raw S&P 500 data files

│ ├── /processed\_data/ # Stores preprocessed S&P 500 data

│ ├── /metadata/ # Stores Crypto metadata

│ ├── /europe/ # Stores data related to European stocks (could include Euro Stoxx 50)

│ ├── /raw\_data/ # Stores raw S&P 500 data files

│ ├── /processed\_data/ # Stores preprocessed S&P 500 data

│ ├── /metadata/ # Stores EuroStoxx50 metadata

│ ├── /sp500/ # Stores data related to S&P 500

│ ├── /raw\_data/ # Stores raw S&P 500 data files

│ ├── /processed\_data/ # Stores preprocessed S&P 500 data

│ ├── /metadata/ # Stores S&P 500 metadata

│

├── /deprecated/ # Archive for outdated files and scripts

│

├── /docs/ # Stores project documentation and references

│ ├── general\_overview.docx # General project description file

│

├── /scripts/ # Python scripts for data preprocessing

**Workflow**

1. **Data Collection**

**Raw Data**: The raw data for each stock group (e.g., S&P 500, Euro Stoxx 50) is stored in their respective /raw\_data/ folders. Files are typically in .csv or .xlsx format and are imported into the project directly from external sources.

**Metadata**: Any additional information regarding the datasets, such as descriptions or special considerations, are stored in the /metadata/ folders. These files help document the data's source, update frequency, or any important notes.

1. **Data Preprocessing**

**Python Scripts**: Preprocessing is done using Python scripts stored in the /scripts/preprocessing\_scripts/ folder. These scripts are responsible for tasks such as:

Cleaning the raw data (e.g., handling missing values, normalizing formats)

Applying basic transformations (e.g., calculating stock returns, adjusting for splits)

Saving the preprocessed data in a format ready for Excel (e.g., cleaned .csv or .xlsx files).

**Processed Data**: The output of the preprocessing scripts is saved in the /processed\_data/ folders for each stock group. These files will be used for manual calculations and analysis in Excel.

1. **Analysis in Excel**

**Manual Analysis**: Once the data is preprocessed, it can be imported into Excel for manual financial analysis. This step involves using Excel functions to calculate metrics like price-to-earnings ratios, moving averages, or portfolio returns.

**Excel Templates**: If needed, you can create a standard Excel template to streamline the analysis. This template can be updated periodically with the latest preprocessed data.

1. **Archiving**

**Deprecated Folder**: Outdated files or scripts that are no longer used in the project are moved to the /deprecated/ folder to keep the workspace clean and organized. This ensures that historical data or legacy scripts are not lost but do not clutter the main project directories.