

Assignment 1: Calculating Descriptive Statistics and Understanding Market Structure

Overview

This project analyzes descriptive statistics of stocks and ETFs over time using data from SEC MIDAS. The analysis spans from Q1 2012 to Q1 2024.

Files Included

1. `Analysis-final.ipynb`: The Jupyter Notebook containing the final code used to compute and visualize the descriptive statistics for stocks and ETFs.
2. `Plots.pdf`: A PDF containing all the generated plots for individual metrics with all descriptives for 100 randomly sampled stocks and 100 ETFs. It includes individual metrics and comparisons between stocks and ETFs for each descriptive statistic (mean, p25, p50, p75, and standard deviation) across various metrics.
3. `Analysis and Reflection on Descriptive Statistics.pdf`: A PDF document that provides a detailed analysis and reflection on the descriptive statistics derived from individual metrics in the data. This document discusses the key insights gained from comparing stocks and ETFs.

Code Execution

Requirements

- Python 3.11
- Jupyter Notebook
- Pandas
- NumPy
- Plotly
- python-docx

How to Run

1. **Install Dependencies:** Make sure you have all the required libraries installed. You can install them using pip:

```
pip install pandas numpy plotly python-docx kaleido
```

2. **Open the Jupyter Notebook:** Open `Analysis-final.ipynb` in Jupyter Notebook.
3. **Run the Notebook:** Execute all cells in the notebook. The code will:
 - Load and preprocess the data for stocks and ETFs.
 - Calculate descriptive statistics for various metrics.

- Generate plots for the descriptive statistics over time.
 - Save the plots in the Images folder and compile them into a Word document (Plots2.docx).
4. Review the Outputs: The generated plots can be found in Plots.docx, and the analysis is detailed in Analysis and Reflection on Descriptive Statistics.pdf.

Citations:

- **OpenAI. (2024).** *Response from ChatGPT.* Online AI Model. Available at: <https://www.openai.com/chatgpt>.