Exercise 2C: Data Quality Checks

BUSI 722: Data-Driven Finance II

Using the returns data from Exercise 2A and fundamentals from Exercise 2B, merge and inspect the combined dataset.

Submission

Submit a **zip file** containing your **Jupyter notebook** (`.ipynb`) with all code, output, and charts, the **`merged.parquet`** file produced in this exercise, and at least one **screenshot** of your Claude Code session showing Claude using the `merge-data` skill. Use markdown cells for any written discussion.

Merge the monthly returns with fundamentals and save as `merged.parquet`. Using the merged dataset:

1. For each fundamental variable, report the fraction of stock-months where the variable is non-missing.
2. Compute the cross-sectional mean and standard deviation of `roe`, `gp_to_assets`, and `asset_growth` for each month. Plot these over time. In a markdown cell, discuss whether the means are stable or show trends.
3. Check for extreme outliers: what fraction of `roe` values are above 1.0 or below -1.0? What fraction of `asset_growth` values are above 5.0?
4. Report the final dataset's shape and column list.