Exercise 1A: Build and Explore Price Data

BUSI 722: Data-Driven Finance II

Use Claude Code to query the Rice Data Portal directly. The relevant tables are:

- **SEP** (prices): `ticker`, `date`, `close` (split-adjusted), `closeadj` (adjusted for splits, dividends, and spinoffs), `volume`
- **DAILY** (daily metrics): `ticker`, `date`, `marketcap` (in USD millions)
- **TICKERS** (static info): `ticker`, `sector`, `industry`, `scalemarketcap`

Submission

Submit a **zip file** containing your **Jupyter notebook** (`.ipynb`) with all code, output, and charts, and at least one **screenshot** of your Claude Code session showing a database query. Use markdown cells for any written discussion.

Using Claude Code and the Rice Data Portal, build a monthly dataset for **all stocks** from January 2021 through January 2026.

1. From the **SEP** table, fetch end-of-month prices (`close` and `closeadj`). Compute monthly **returns** from `closeadj`: $r_t = \text{closeadj}_t / \text{closeadj}_{t-1} - 1$.
2. From the **DAILY** table, fetch end-of-month `marketcap`.
3. From the **TICKERS** table, fetch `sector` and `industry`.
4. Compute **momentum** for each stock: the cumulative return from month $t-13$ to month $t-2$ (skipping the most recent month to avoid short-term reversal).
5. Compute **lagged return**: the prior month's return.
6. Apply a penny-stock filter: drop rows where `close` $<$ \$5. Drop rows with missing `return`, `momentum`, or `marketcap`.
7. Report the number of rows, the number of unique tickers, and the date range. How many tickers have data in every month of the sample? Display summary statistics (mean, median, std, min, max) for `return`, `momentum`, and `marketcap`.