



Building Real Estate Data Pipelines for Assette

MSBA Capstone 2025 – Shiv Nag, Courtney Vincent, Zicheng Wang



Problem Statement

Need: Assette’s goal is to expand their Snowflake database to house real estate data.

Challenge: Defining and sourcing complete, accurate data for real estate funds.

Goal: Deliver production-ready code with realistic real estate data.

Data Overview

Portfolio

Product Master – *Source: Synthetic*
Portfolio General Information – *Source: Nareit, REIT Notes, Synthetic*

Benchmark

Benchmark General Information – *Source: Yahoo Finance API*
Benchmark Characteristics – *Source: Yahoo Finance API*

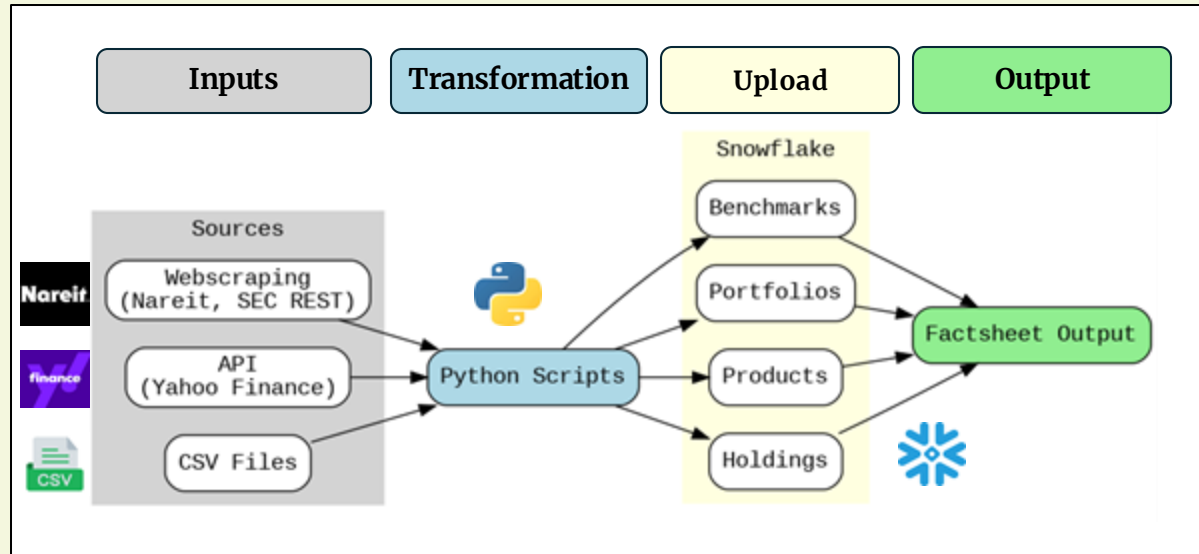
Security

Security Master – *Source: Yahoo Finance CSV*
Holdings Details – *Source: Yahoo Finance API, Synthetic*

Performance

Benchmark Performance – *Source: Yahoo Finance API*
Portfolio Performance – *Source: Yahoo Finance API*
Portfolio Benchmark Association – *Source: Benchmark General Information, Portfolio Performance*

Data Flow & Architecture



Challenges

Data availability & format: Some tickers had short histories or were delisted, requiring stronger validation.

Balancing real & synthetic data : Needed constant checks for realistic performance metrics.

Using Generative AI wisely: Helpful for speed, but all outputs were reviewed for accuracy.

Sample Output & Initial Results

9 Snowflake tables populated with real + synthetic REIT data

Modular Python scripts aligned with **Snowflake schemas** for easy updates

Standardized performance metrics across all portfolios & benchmarks

Scalable ingestion pipeline ready for new securities and asset classes

Data validation through cross-checks of **real** vs. **synthetic** outputs

Next Steps & References

Externalize Rules

Store logic in YAML/JSON for quick updates

Classify Holdings

Auto-assign by subclass (REIT, Core, Opportunistic)

Automate Updates

Schedule refreshes (Airflow, dbt, Snowflake Tasks)

Validate & Audit

Enhance accuracy with exception handling & anomaly detection

Expand Data

Broaden coverage to include more diverse Real Estate securities

Benchmark Integration

Link portfolios to performance benchmarks