# Project Progress Report

Project: Fama-French ML Universe – European Financial Sector

Date: October 30, 2025

Author: Christos Paseas

## 1. Objective

To construct a clean, modeling-ready dataset of European financial sector equities suitable for machine learning–based factor modeling and trading strategy development. The dataset must be investable (IBKR-tradable), stationary, and historically consistent for robust backtesting.

## 2. Achievements

### ✅ Data Universe Construction

• Collected raw price and return data for European financial sector stocks.  
• Excluded delisted or illiquid tickers.  
• Verified tradability and shortability on Interactive Brokers (IBKR).  
• Ensured sufficient history (≥10 years).

### ✅ Quality Control (QC)

• Conducted ADF and KPSS tests to assess stationarity.  
• Identified and removed non-stationary tickers (CLI.L, STAN.L, etc.).  
• Finalized list of 106 tickers passing all statistical and data integrity filters.

### ✅ Data Cleaning and Alignment

• Cleaned missing values and corporate actions.  
• Aligned price and return panels by date and ticker.  
• Created stable index spanning January 2013 – October 2024 (approx. 11+ years of data).

### ✅ Final Deliverables

|  |  |
| --- | --- |
| File | Description |
| Investment\_universe/modeling\_universe.csv | Final list of 106 financial sector tickers |
| Investment\_universe/modeling\_returns.csv | Cleaned daily return series |
| Investment\_universe/modeling\_prices.csv | Matching daily price series |
| Investment\_universe/modeling\_drop\_log.csv | Dropped tickers and reasons (stationarity or forced removal) |

## 3. Current Status

✅ Modeling-ready dataset completed  
• Universe: 106 European financial sector stocks  
• Period: 2013–2025  
• Frequency: Daily  
• Tradability: IBKR-confirmed  
• Stationarity: ADF/KPSS validated  
  
The dataset is now ready for feature engineering and predictive modeling.

## 4. Next Steps

|  |  |  |
| --- | --- | --- |
| Stage | Description | Deliverable |
| Feature Engineering | Add lagged returns, rolling vol, factor exposures (SMB, HML, Momentum, etc.) | features.parquet |
| Model Development | Train predictive ML models (Random Forest, XGBoost, Ridge, or Neural Nets) | Model artifacts |
| Backtesting | Simulate long/short portfolios using predicted signals | Performance report |
| Documentation | Write technical paper or report summarizing pipeline | modeling\_pipeline\_report.docx |

## 5. Summary

You have successfully completed the data engineering and QC phase of your machine learning pipeline. The resulting dataset is robust, reproducible, and fully suitable for factor modeling and algorithmic trading research.  
  
Milestone reached: Model-ready dataset of 106 IBKR-tradable European financial stocks (2013–2025) complete.  
Next milestone: Feature engineering and model design.