

sdmay19-06: Quantitative Research Modeling Library

Week 3 Report

September 22 - September 28

Client: Joseph Byrum

Advisor: Srikanta Tirthapura

Team MembersJosiah Anderson — *Meeting Facilitator*Doh Yun Kim — *Scribe*Gabriel Klein — *Report Manager*Drake Mossman — *Communication Manager*Nathan Schaffer — *Overseer*Jacob Richards — *Quality Assurance Manager*

Summary of Progress this Report

This week was mostly spent on examining example scripts provided by Principal once again. We delved deeper into many packages and mostly focused on Python which we feel will likely be the direction we want to take the project. We also drafted plans for several aspects of the project including requirements, schedule, resources, solutions, and testing processes. The results from this week have prepared us to begin discussing concrete implementation details.

Past Week Accomplishments

Josiah Anderson

- Studied the advantages of Python v. R in data science
 - R
 - has extensive packages available
 - very active data science community that helps with maintenance and advances
 - exceptional visualization capabilities
 - Python
 - Very flexible
 - much easier learning curve
 - more accepted on the engineering side of things
 - growing availability of packages
- Began researching Python and R packages that would be useful for our project
 - panda
 - data.table
 - lattice
 - caret
 - NumPy
- Began establishing a concrete set of deliverables for our project

Doh Yun Kim

- Begin to try and plan basic functions that could be used for the Python library
 - For now trying to see if any preprocessing work will help save redundant writing of code
- Worked on developing the risk management chart for the Project Plan

- Created the tentative project schedule for the Project plan
- Researched estimate_model_coefficients.r
 - data.table, glmnet, splines libraries researched from said code
 - Looked for links between this code and the previous rolling window Python code (unfortunately different enough for no easy connections to be found)

Gabriel Klein

- Continued to study Principal scripts
 - Looked further into numpy, pandas, and sklearn
 - StandardScaler
 - LabelEncoder
 - mutual_info_classif
 - mutual_info_regression
 - numpy.corrcoef
 - make_scorer
 - Reviewed intent of each section in rolling window script
 - splitting testing/training data
 - preventing contamination
 - engineering non-linear features
 - removing correlated columns
 - cross validation
 - feature selection
- Compared and contrasted several solutions for the project
 - R/Python Library
 - Standalone Application
 - Web Application

Drake Mossman

- Drafted out initial plan for testing deliverables
 - Requirements testing for each functional and non-functional requirement
 - Usability testing
 - Have client and potential users engage with software at various iterations stages
 - Elicit feedback through follow-up interviews
- Planned out short-term and long term vision / goals for the project
 - Our idea for what we will deliver based on current knowledge
 - Timeline and process to get there
- Installed and began familiarizing with Jupyter

Jacob Richards

- Furthered Python knowledge
 - Ran through online Python examples
- Looked at Principal example scripts
 - Signal Performance and Hedging Strategy Performance
- Furthered R knowledge
 - Went through online R examples

Nathan Schaffer:

- Studied Principal example scripts
 - Looked at Signal Generation and Portfolio Simulation
 - Looked into

- Researched Python
 - Experimented with reading input files
- Researched R
 - Got to know the basics

Group:

- Drafted a cohesive project plan for lifecycle of the project
 - Defined requirements
 - Analyzed constraints
 - Brainstormed solutions
 - Estimated resource costs
 - Rough schedule
-

Pending Issues

The final interviews with Principal employees are planned for next week on the first and third of October. Once we have completed them and compiled the results we will be prepared to finalize our project scope and begin moving from research and ground work to working on concrete implementation.

Plans for Upcoming Reporting Period

Josiah Anderson

- Look at Principal scripts and see what packages they use and how
- Interview Principal data science employees
 - Ask about R v. Python
 - Determine what types of functions/visualization/collaboration they need for their projects
 - research which packages would be most efficient for provided those needs
- Go over deliverables with Principal
- Begin writing test scripts in R to evaluate the learning curve and the usefulness of the packages I studied last week.

Doh Yun Kim

- Study the R scripts of Principal further
 - Especially look into if somehow refactoring R scripts into a similar format of Python will help (doubt it but worth a try)
- Compare how the data is processed in R and Python to set up a standard
 - See if examples of rolling window in R is available at Principal or get an example from the internet
- Try and understand what the estimate model is doing during the processing of data step and the weird(?) output currently (probably something I'm doing wrong)

Gabriel Klein

- Finish up looking at script examples
 - Portfolio simulation script
 - Stock selection model
- Conduct final interviews at Principal
- Finalize scope for project
- Begin discussing concrete implementation
 - Draft up concepts for libraries
 - Define interfaces to get feedback on

Drake Mossman

- Look into new scripts and data shared with us
 - Focus on scripts in R, learning idiosyncrasies with the language
- Collect and process input from data scientists
 - Interviews in DSM with members of data science team
 - Summarize results of interviews with team
- Flush out plan going forward
 - Details technical tasks, planned deliverables, and goal deadlines

Jacob Richards

- Further Python and R knowledge and experience
- Look more into Principal scripts
 - Research what libraries they use
 - Try to better understand what the scripts are doing
- Interview Principal employees

Nathan Schaffer:

- Continue to study scripts from Principal
 - Look at Factor Portfolio Data Aggregation
 - Research packages used
- Conduct more interviews on site at Principal
 - Interview team lead Krisoye and 3 others on data science team on Wednesday.
- Study and get to know R better

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Josiah Anderson	Python v. R research, project plan, deliverables, problem statement, packages	6	21
Doh Yun Kim	Script research, Python, risk management, tentative schedule	6	20.5
Gabriel Klein	In-depth rolling window study, project approach	9	25
Drake Mossman	Jupyter, short/long term goals, testing processes	6.5	20
Nathan Schaffer	Script research, Python/R learning	6	17
Jacob Richards	Python and R learning, Principal script research	7	16

Gitlab Activity Summary

Nothing to report.
