Spring 4-15-2020

A tool for securities analysis and trading strategy back-testing

Qiaosheng Yu

A tool for securities analysis and trading strategy back-testing

by

Qiaosheng Yu

Applied Project report submitted to the Faculty of the Graduate School of the

in fulfillment of the requirements for CISC 699 of the

Computer and Information Sciences program

Supervised by: Abrar Qureshi, Ph.D.

Spring 2020

Problem Statement

The platform created in this project will help users to analyze stock/ETF performance, define and improve their asset allocation plans and trade ideas. Using real market data, user can do a deep dive into securities performance, historical correlation, risk and return. Furthermore, user can use the platform back test their trade ideas, modify the parameters, including percentage allocation, risk limits and rebalancing frequency, and create the best strategy which suits their risk/return criteria and long-term wealth management goals.

Requirements

Functional requirements

Below are the functional requirements of this application:

- The application shall allow user to select one or multiple securities to analyze from the database
- The application shall allow user to read and display historical price information of selected securities
- The application shall use basic regression metrics to analyze securities performances
- The application shall allow user to input their current asset allocation
- The application shall allow user to define a target percentage allocation
- The application shall calculate the cash flow needed for user's retirement plan

- The application shall have interactional buttons for users to change parameters and see results in real-time
- The application shall use charts and other diagrams to help visualize results

Non-functional requirements

Below are the non-functional requirements of this platform:

- The application shall be web-based and allow public access
- The application shall allow user to access via different operating systems, including Windows, MacOS and Linux
- This application shall allow data input and modification at real-time
- The application shall be reliable on returning results on edge cases
- The application shall be efficient and be able to handle large dataset in a quick fashion

Software Requirements Specification

- Programming language: R 3.6.1
- IDE: R Studio
- Libraries/packages:
    - shinydashboard
    - TTR
    - ggplot
    - rlang

- scripts:
    - ui.R
    - server.R
- data:
    - stocks.csv

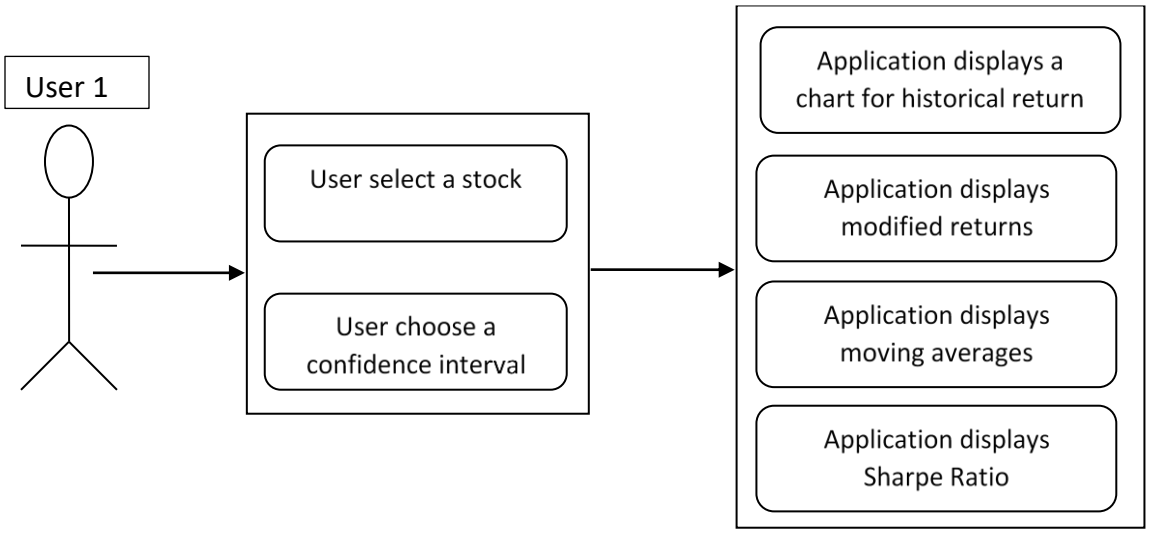
Use cases

Use case 1:

- Use case description:
    - analyze historical performance of a stock
- Use case requirements:
    - show historical return, confidence interval
    - show moving average and Sharpe Ratio
- Preconditions:
    - data is available
    - server is up and running
- Use case paths:
    - Main flow:
        - user select a stock
        - application displays a chart of historical return
        - user choose confidence interval

        - application displays a chart for modified returns
        - application displays a moving average
        - application displays the calculated Sharpe Ratio
    - Alternate flow:
        - none
    - Exception flow:
        - user select a stock with missing data
        - application removes missing data
        - application displays a chart of historical return with modified data
        - user choose confidence interval
        - application displays a chart for modified returns
        - application displays a moving average
        - application displays the calculated Sharpe Ratio

- Use case diagram:

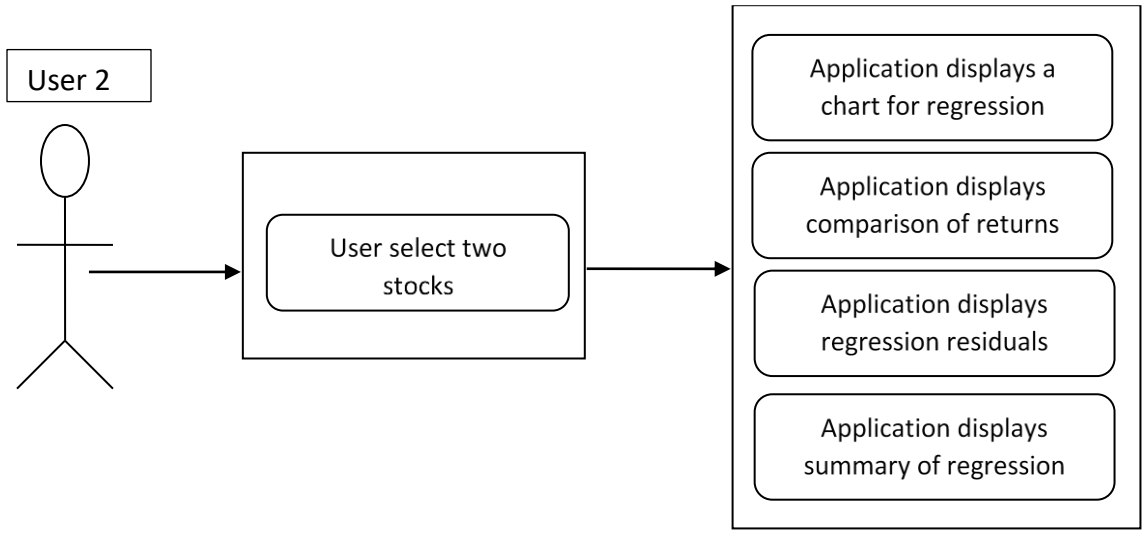


Use case 2:

- Use case description:
    - analyze historical correlation between two stocks
- Use case requirements:
    - show historical correlation of two stocks
    - show historical performance comparison of two stocks
- Preconditions:
    - data is available
    - server is up and running
- Use case path:
    - Main flow:
        - user select two stocks
        - application displays a chart for regression
        - application displays a comparison of returns
        - application displays regression residuals
        - application displays summary of regression
    - Alternate flow:
        - none
    - Exception flow:
        - user select two stocks with missing data
        - application removes missing data
        - application displays a chart for regression with modified data

- application displays a comparison of returns
- application displays regression residuals
- application displays summary of regression

- Use case diagram:

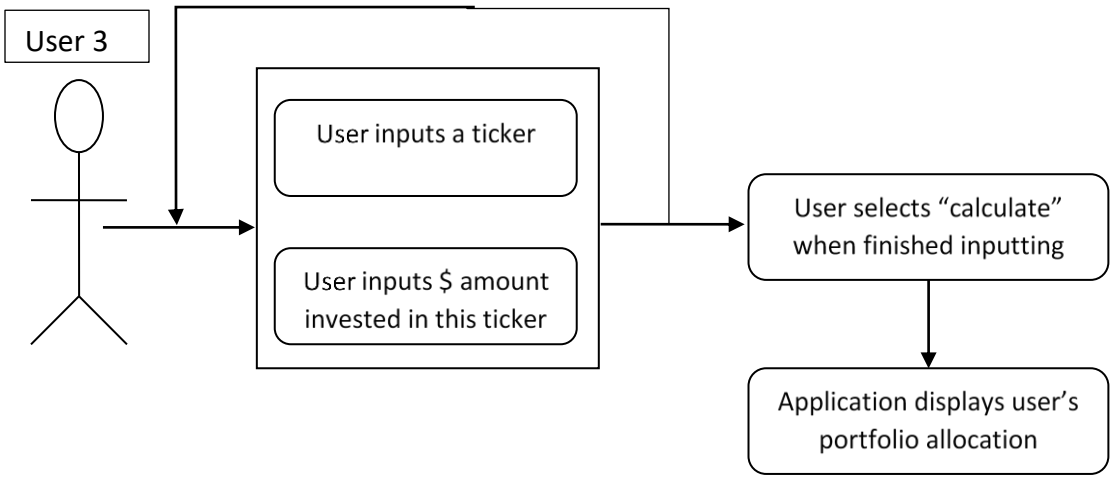


Use case 3:

- Use case description:
  - analyze user's portfolio allocation
- Use case requirements:
  - Allow user to input their current holdings
  - Calculate user's current portfolio allocation in different asset classes
- Preconditions:
  - data is available
  - server is up and running
- Use case paths:

- Main flow:
    - user inputs a ticker
    - user inputs $ amount invested in this ticker
    - application defaults to a category
    - repeat the above
    - user selects “calculate” when finishes inputting
    - application displays user’s portfolio allocation
- Alternate flow:
    - none
- Exception flow:
    - user inputs a ticker
    - user inputs $ amount invested in this ticker
    - the ticket user selects does not have a default category
    - user manually input the ticker’s category
    - repeat the above
    - user selects “calculate” when finishes inputting
    - application displays user’s portfolio allocation

- Use case diagram:

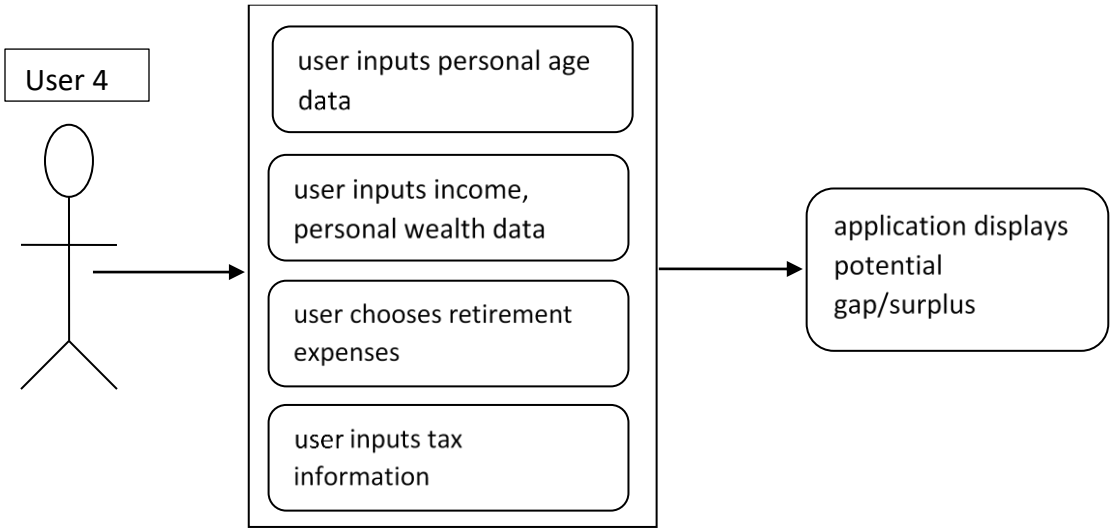


Use case 4:

- Use case description:
    - Retirement analysis
- Use case requirements:
    - Calculate user’s cashflow at retirement according to various inputs
    - Show user’s cashflow gap for retirement spendings
- Preconditions:
    - data is available
    - server is up and running
- Use case paths:
    - Main flow:
        - user inputs personal age data
        - user inputs income, current personal wealth data

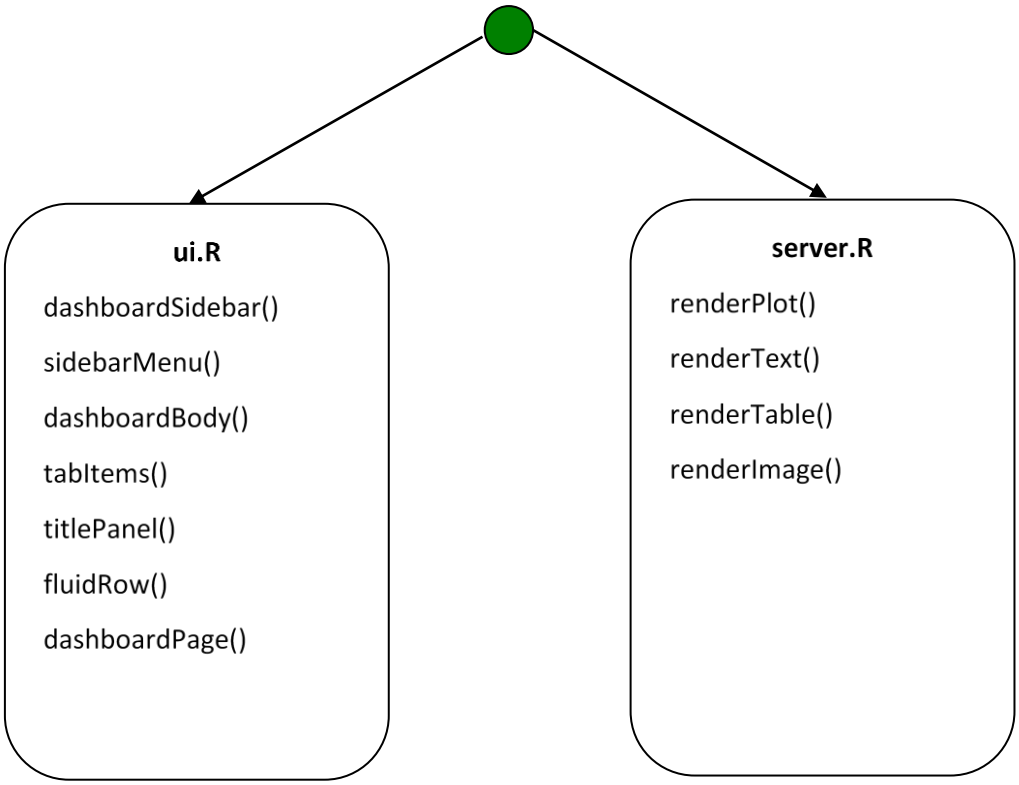
        - user chooses retirement expenses
        - user inputs tax information
        - application displays potential gap/surplus
    - Alternate flow:
        - none
    - Exception flow:
        - none

- Use case diagram:

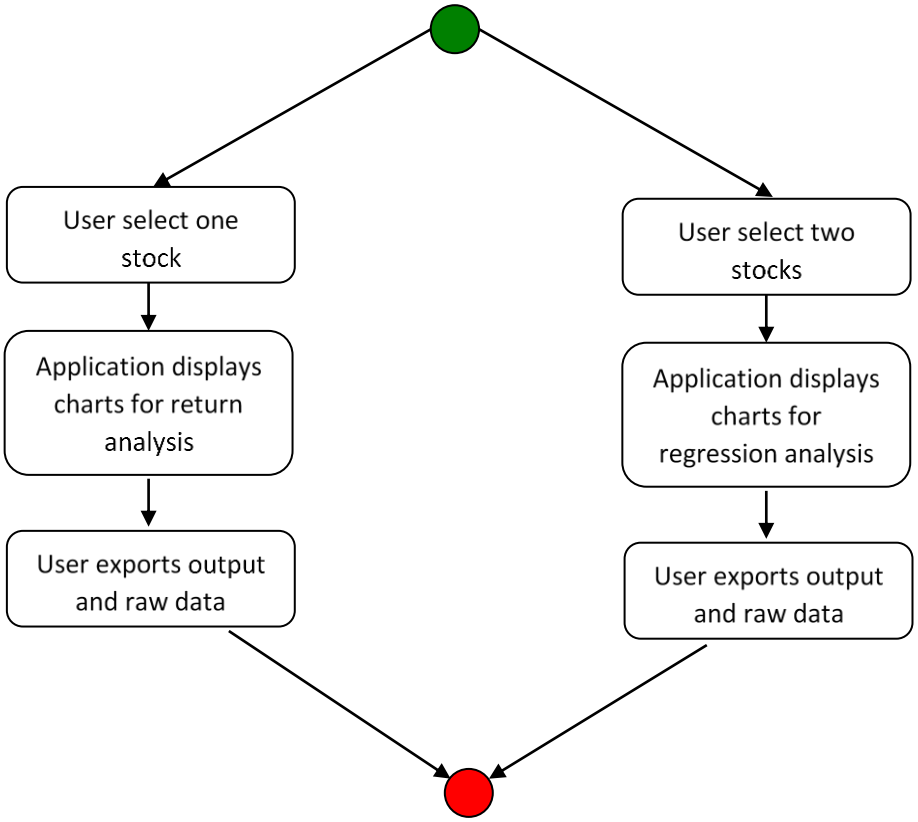


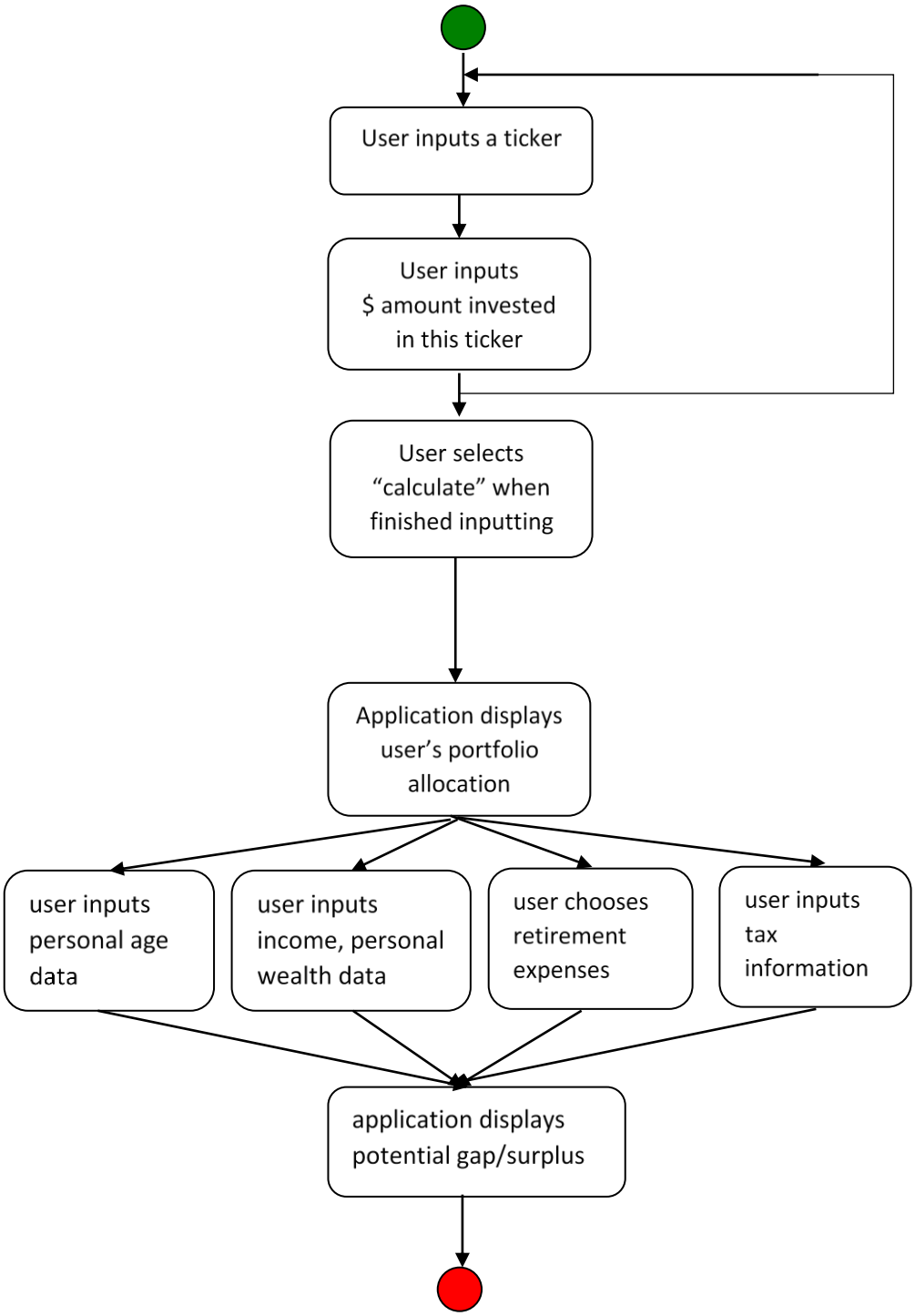
Class and Sequence diagrams

Class diagrams



Sequence diagrams




User inputs a ticker
User inputs $ amount invested in this ticker
User selects “calculate” when finished inputting
Application displays user’s portfolio allocation
user inputs personal age data
user inputs income, personal wealth data
user chooses retirement expenses
user inputs tax information
application displays potential gap/surplus

Design documents

Version History

Version	Author(s)	Description of version	Date completed
1.0.0	Qiaosheng Yu	Initial version	10/20/2019
1.0.1	Qiaosheng Yu	Second version	11/17/2019
2.0.0	Qiaosheng Yu	Third version	2/19/2020

Overview

This design document describes the operating environment, system requirements, full system and sub system architecture, database design, input format, output format, processing logic and user interfaces.

Operating environment

This platform is designed in Windows 10 platform. It can also be operated in Mac and Linux.

System requirements

System needs to have R (3.6.1 or above) and an integrated development environment (IDE), for example R studio in order to run the app.

The platform also requires the system to have access to internet browser.

Full system and sub system architecture

The platform utilizes two R scripts, including ui.R and server.R.

In ui.R, the interface of the platform is built. Components include a side bar, dropdown lists and dashboard layouts.

In server.R, the computational work and graph plotting are done. Main computations include log returns, regressions, confidence intervals, moving averages, and Sharpe Ratios, calculation of asset allocation, calculation and analysis of retirement portfolio.

Database design

The data will be downloaded from Yahoo finance and stored locally, in the same folder with ui.R and server.R.

Input format

In “dashboard” and “regression comparison” pages, the input will be in .csv format. First column (column A) will be the date, in DD-MM-YYYY format. The following columns will contain daily stock price data while the header will be the stock ticker, for example, BABA, COST, WMT, etc. The data for a specific stock ticker will be used once the user chooses stock(s) in the user interface.

In “my portfolio” and “retirement plan” pages, data input is done in the user interface. In “my portfolio” page, user will need to input the ticker, choose category of the asset, and input the amount of the respective asset. In “retirement plan” page, user will have to input their investment base from brokerage, 401K plan, traditional & Roth IRA and other Savings. Furthermore, user need to specify the parameters for further calculation, including Annual Investment Return Estimation in % (before retirement), Annual Investment Return Estimation in % (after retirement), Inflation Estimation in %, Your

Current Status, Your Age Now, Your age when you expect to retire, Your Life Expectancy, Your current Annually Expense and Planned Yearly Savings.

Output format

The output will be mainly in tables and graphs format, displayed in the shiny dashboard.

In “dashboard” and “regression comparison” pages, the log return charts will be bar charts with smoothed trend line. The regression charts will have the correlation function, R-squared, and p-value. User can also view the raw data directly from the dashboard by clicking “raw data” tab.

In “my portfolio” and “retirement plan” pages, the platform displays portfolio asset allocation pie chart given the data provided by the user, which also includes the asset class name and percentage currently allocated into this asset. Future Total Wealth Estimation, Future Expense Estimation and Annual Saving Required are displayed in tables.

Processing logic

In “dashboard” and “regression comparison” pages, the platform reads raw stock price data from data.csv. In “my portfolio” and “retirement plan” pages, the platform reads user input from the interface. After taking user inputs and instructions, it performs the calculation and displays results in table and graph formats.

User interfaces

The user interface will be browser based, relying on the shiny package. In the user interface, user can select/type in inputs and view outputs in table and graph formats.

Disclaimers

Test cases

Test case 1:

Test Case Number:	TC1
Revision:	Rev.1
Author:	Qiaosheng Yu
Date Conducted:	10/20/2019
Test Conductor:	Qiaosheng Yu
Customer Representative:	Haoting Chen
Description:	This Test Case will test the functionality of single stock analysis function which includes displaying a chart of historical return, a chart for modified returns, a moving average and Sharpe Ratio. The related use case for this Test case is UC01.
Pre-Test Setup:	ensure data is populated ensure server is up and running

Use Case: 01 Flow: Main Flow

No.	User action	Expected results	P/F	Comments
1	User runs the app in R	Home page is loaded	P	/
2	User selects “dashboard” from the side bar	Stock analysis dashboard is displayed	P	/
3	User clicks the dropdown menu on the top of the dashboard	A list of available stocks is displayed	P	/
4	User chooses a stock from the dropdown menu	A distribution for log return is displayed	P	/
5	User selects a confidence interval for mean	Range of log returns within the confidence interval is displayed	P	/
6	User selects a confidence interval for variance	Range of variance within the confidence interval is displayed	P	/
7	User selects a day range for moving average calculation	A moving average line is displayed in the chart	P	/
8	User scrolls to the end of page	A Sharpe Ratio is calculated and displayed	P	/
Post conditions: the platform performed as expected				

Test case 2:

Test Case Number:	TC2
Revision:	Rev.1
Author:	Qiaosheng Yu
Date Conducted:	10/20/2019
Test Conductor:	Qiaosheng Yu
Customer Representative:	Haoting Chen
Description:	This Test Case will test the functionality of two stocks correlation analysis function which includes displaying a chart for regression, a comparison of returns, regression residuals, summary of regression. The related use case for this Test case is UC02.
Pre-Test Setup:	ensure data is populated ensure server is up and running

Use Case: 02 Flow: Main Flow

No.	User action	Expected results	P/F	Comments
1	User runs the app in R	Home page is loaded	P	/
2	User selects “regression comparison” from the side bar	Stock regression analysis page is displayed	P	/
3	User chooses two stocks from the menu	A regression chart is displayed on the left hand side. On the right hand side, a comparison of returns, regression residuals, summary of regression are also displayed	P	/
Post conditions: the platform performed as expected				

Test case 3:

Test Case Number:	TC3
Revision:	Rev.1
Author:	Qiaosheng Yu
Date Conducted:	2/19/2020
Test Conductor:	Qiaosheng Yu
Customer Representative:	Haoting Chen
Description:	This Test Case will test the functionality of “my portfolio” analysis function which includes allowing user to input the assets they are currently holding and calculate the percentage of their asset allocated to each asset classes in a graph format. The related use case for this Test case is UC03.
Pre-Test Setup:	1. ensure server is up and running 2. ensure user has a keyboard to enter tickers and numbers

Use Case: 03 Flow: Main Flow

No.	User action	Expected results	P/F	Comments
1	User runs the app in R	Home page is loaded	P	/
2	User selects “my portfolio” from the side bar	“my portfolio” page is displayed	P	/
3	User input the ticker of the asset he/she is currently holding	Ticker is taken into calculation	P	/
4	User clicks the dropdown menu of “category” to choose a category for this ticker	The asset class user selected is taken into calculation	P	/
5	User input the amount of $ he/she has in this ticker	The amount user selected is taken into calculation	P	/
6	Repeat the above process until all tickers have been entered	All data entered by user is taken into calculation	P	/
7	User scrolls to the end of page	A pie chart is calculated and displayed	P	/
Post conditions: the platform performed as expected				

Test case 4:

Test Case Number:	TC4
Revision:	Rev.1
Author:	Qiaosheng Yu
Date Conducted:	2/19/2020
Test Conductor:	Qiaosheng Yu
Customer Representative:	Haoting Chen
Description:	This Test Case will test the functionality of "retirement plan" analysis function which includes calculation of Future Total Wealth Estimation and Future Expense Estimation, and annual saving required to breakeven based on the data user provides. The related use case for this Test case is UC04.
Pre-Test Setup:	1. ensure server is up and running 2. ensure user has a keyboard to enter numbers

Use Case: 04 Flow: Main Flow

No.	User action	Expected results	P/F	Comments
1	User runs the app in R	Home page is loaded	P	/
2	User selects “retirement plan” from the side bar	“retirement plan” analysis page is displayed	P	/
3	User input current holding in his/her brokerage account	Amount is taken into the calculation of “current total Current Asset” and been displayed	P	/
4	User input current holding in his/her 401k account	Amount is taken into the calculation of “current total Current Asset” and been displayed	P	/
5	User input current holding in his/her Traditional & Roth IRA account	Amount is taken into the calculation of “current total Current Asset” and been displayed	P	/
6	User input current holding in his/her other savings account	Amount is taken into the calculation of “current total Current Asset” and been displayed	P	/

7	User input Annual Investment Return Estimation in % (before retirement)	Data is taken into future total wealth calculation	P	Q
8	User input Annual Investment Return Estimation in % (after retirement)	Data is taken into future total wealth calculation	P	/
9	User input Inflation Estimation in %	Data is taken into future expense calculation	P	/
10	User input current age	Data is taken into future wealth and expense calculation	P	/
11	User input age expected to retire	Data is taken into future wealth and expense calculation	P	/
12	User input Life Expectancy	Data is taken into future wealth and expense calculation	P	/
13	User input current Annually Expense	Data is taken into future expense calculation	P	/
14	User input Planned Yearly Savings	Data is taken into future total wealth calculation	P	/

15	User scrolls to the end of page	Tables which include calculation of Future Total Wealth Estimation and Future Expense Estimation, and annual saving required to breakeven based on the data user provides can be viewed	P	/
Post conditions: the platform performed as expected				

Ethical and Societal effect

This project tries to help people who's not familiar with finance to better understand financial market and better prepare for their future retirement. Stock market is relevant to everyone's wealth. We all have our 401(k) invested in domestic and international stock/bond market or some target date funds. However, stock market won't going up forever, it's important for us to understand the correlation between different assets to help us outperform the market when the economy slows down or even when another financial market crash comes.

We also have different long-term/short-term plans in the future, including funding a house, children's education plan, retirement, etc. it's important to start putting some money aside for the retirement and let the money grow. This project provides people with a tool to quickly examine if they have saved enough for their retirement based on the numbers and assumptions they provide, if not, what's the gap between future asset and future consumption, and how much you need to save in order to be breakeven at your life expectancy. Hopefully by using this tool, people will be able to make a more informed decision in their retirement investments and after all help to create social wealth for the society.

Annotated Bibliography

Bai, R., Lu, X., & Tang, B. (2010). Business Performance Analysis of Listed Companies under the Stock Dividends Behavior: An Empirical Approach Based on Factor Analysis Model. *2010 International Conference on Management and Service Science.* Wuhan, China: IEEE.

This article introduces a few methods for investors to better measure and understand their portfolio performance. Specifically, an investor can use ratio analysis, benchmark analysis and trend analysis based on portfolio components and other investment parameters.

The analysis methodologies introduced in the article will be included in the asset allocation part.

Brasileiro, R. C., Souza, V. L., Fernandes, B. J., & Oliveira, A. L. (2013). Automatic method for stock trading combining technical analysis and the Artificial Bee Colony Algorithm. *2013 IEEE Congress on Evolutionary Computation.* Cancun, Mexico: IEEE.

This article introduces a few popular technical analysis trading softwares with their unique focus, features and highlights.

These information help to inspire me on designing the functionalities of the trading tool and determine which features are more important and relavant.

Chan, M.-C., Wong, C.-C., Cheung, B.-S., & Tang, G.-N. (6-9 Oct. 2002). Genetic algorithms in multi-stage asset allocation system. *IEEE International Conference on Systems, Man and Cybernetics.* Yasmine Hammamet, Tunisia, Tunisia: IEEE.

This article introduces the most common objectives of asset allocation models: preservation of capital, income, balanced, or growth. Furthermore, it explains in details how these four objectives should be achieved by allocating the portfolio into different asset classes given their unique risk and return charicteristics.

This article helps to connect user objectives with typical asset allocations to begin with. This will give user a quick guidance and they can modify the strategy based on their unique requirements.

Chang, K., & Tian, Z. (2016). Optimal asset allocation with mutual information. *2016 19th International Conference on Information Fusion (FUSION).* Heidelberg, Germany: IEEE.

This article gives examples of suggested asset allocation of three types of investors - aggressive, moderate and conservative. It further explains why each investor type would prefer different investment styles, given their age, investment horizon, income growth and risk tolerance.

This article will be used to provide default asset allocation suggestions in the asset allocation tab for user to begin with.

Fafuła, A., & Drelczuk, K. (2015). Buying stock market winners on Warsaw Stock Exchange - quantitative backtests of a short term trend following strategy. *2015 Federated Conference on Computer Science and Information Systems (FedCSIS).* Lodz, Poland: IEEE.

This article gives detailed breakdown and examples of back-testing. It gives a good foundation of knowledge before we start constructing our asset allocation themes and set up the back-testing model.

This article is crucial to the back-testing part, in which it provides the necessary background information for constructing the tool.

Guo, X., Zhang, H., Jiang, F., & Tian, T. (2018). Development of Stock Correlation Network Models Using Maximum Likelihood Method and Stock Big Data. *2018 IEEE International Conference on Big Data and Smart Computing (BigComp).* Shanghai, China: IEEE.

This article introduces positive and negative correlation in stock markets and further explains why correlation matters for investors and how to better utilize the information that correlation provides in stock investing. This article also gives detailed instructions on how to calculate stock correlations in the most efficient fashion.

This article provides very useful insights on how to calculate and interpret stock correlations. It will be used in the "investor tips" part to help user make a more informed investment decision.

Hung, K. K., Cheung, C. C., & Xu, L. (2000). New Sharpe-ratio-related methods for portfolio selection. *IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr) (Cat. No.00TH8520).* New York, NY: IEEE.

This article gives a detailed explanation on Sharpe Ratio, including the calculation, why is it important, what is considered a good Sharpe Ratio and why sometimes it is not a good measure of portfolio performance. It also gives examples step by step to help reader understand Sharpe Ratio.

This article will be used to provide background information on Sharpe Ratio and interpretation of calculated Sharpe Ratio after user perform the analysis.

Kossecki, P. (2009). Valuation and value creation of internet companies. *2009 International Multiconference on Computer Science and Information Technology.* Mragowo, Poland: IEEE.

This article provides a methodology of company performance valuation and how it is linked to stock performance. In this article, we also learn about the limitation

of the TRS (total returns to shareholders) method and how the company's market value may deviate from its value creation potential.

This article gives a unique view of company performance and stock price valuation. It can be used in advanced version of single stock analysis in addition to the traditional return calculation method.

Letchford, A., Gao, J., & Zheng, L. (2012). Optimizing the moving average. *The 2012 International Joint Conference on Neural Networks (IJCNN).* Brisbane, QLD, Australia: IEEE.

This article introduces three types of moving average calculations: Simple moving average (SMA), Exponential moving average (EMA) and Centered moving average and explained how moving averages with different time frames can provide a variety of information.

This article provides suggestions on how to choose the most common and most informative moving average time frames in the single stock performance analysis part.

Nuti, G., Mirghaemi, M., Treleaven, P., & Yingsaeree, C. (20 January 2011). Algorithmic Trading. *Computer*, IEEE.

This article introduces the top factors that user should pay most attention to when back-testing their strategies, including net profit or loss, volatility, average, capital exposure, wins-to-losses ratio, annualized return and risk-adjusted return, and a few tips that people should not ignore when conducting back-testing.

This article helps set up the parameters in the back-testing part and visualize the top factors based on their importance.

Bian, B., Yuan, Q., & Zhang, H. (2009). Financial valuation and optimal strategy for retirement benefits in a jump diffusion model. *2009 IEEE International Conference on Control and Automation.* Christchurch, New Zealand: IEEE.

This article discusses a defined benefit pension plan with the option of early retirement in a jump-diffusion model. The retirement plan valuation is characterized as the solution of an optimal stopping time problem.

This article provides a quantitative way to understand the retirement benefits and the optimal retirement strategy.

Christopher, L., Boler, W., Wieczorek, C., Crespo, J., Witcher, P., Hawkins, S. A., & Stewart, J. (2016). Asset allocation with swarm/human blended intelligence. *2016 Swarm/Human Blended Intelligence Workshop (SHBI).* Cleveland, OH, USA: IEEE.

This article discusses how multiple asset allocation plan can be exploited and changed real-time along with the situation. This blends computer simulation and human interaction.

The idea in this article will be blended in the project, helping to develop the user interaction part with scenario questions.

Consigli, G., Iaquinta, G., Moriggia, V., Tria, M. d., & Musitelli, D. (2012). Retirement planning in individual asset–liability management. *IMA Journal of Management Mathematics*, 365 - 396.

This article investigates the difference between traditional pension accumulation plans vs dynamic asset allocation plans from the perspective of individual asset–liability management.

This article discusses three different individuals' scenarios with different time horizons but common retirement goals. The results show the advantage of dynamic allocation strategies with different asset classes. This result will be examined in the project.

Jin, Z., Lan-jun, L., & Ming, S. (2010). Decomposition of Health Cost and Modeling of Asset Allocation. *2010 WASE International Conference on Information Engineering.* Beidaihe, Hebei, China: IEEE.

This article discusses the health cost, which is an important part to estimate individual's life expectancy and monthly expense. Furthermore, it also affects individual's volatility of earnings and asset allocation decisions.

This article provides some good thoughts on how to suggest an allocation ratio on risk assets based on individual's health condition and life expectancy.

Li, X. (2010). Accounting disposal of mutual transform between cost method and equity method in long-term equity investment. *2010 International Conference on Education and Management Technology.* Cairo, Egypt: IEEE.

This article explores the long-term equity investment options from an accounting perspective. It also adds complications when the invested fund is dynamically changing, meaning people may buy or sell assets, add or withdraw funds frequently during the investment horizon.

The accounting method discussed in this article will be applied to the project when calculating the fair value of investments.

Siddiaui, S. S., & Patil, V. A. (2018). Stock Market Valuation using Monte Carlo Simulation. *2018 International Conference on Current Trends towards Converging Technologies.* Coimbatore, India: IEEE.

This paper discusses Monte Carlo simulation and discounted cash flow analysis model which are used for stock valuation. The advantage of Monte Carlo simulation is that it generates random scenarios and has no human intervention therefore it prevents emotional bias.

The Monte Carlo simulation method discussed in this paper will be referenced when simulating the long term returns of value/growth stocks.

Wang, T., & Wang, J. (2008). Modeling and Simulation of Stock Prices by Contact Model and Statistical Analysis. *2008 IEEE Symposium on Advanced Management of Information for Globalized Enterprises.* Tianjin, China: IEEE.

This article models the stock market prices based on the Contact Model, which is a continuous time Markov process. The statistical properties of stock indices data and the simulation are analyzed, fat tails and the power-law distributions of returns are also mentioned.

The techniques presented in the study will be applied to the project to calculate stock price returns.

Yang, Z. (2010). Rural Public Pension and Endogenous Growth. *2010 Third International Conference on Information and Computing.* Wuxi, China: IEEE.

This article discusses the variation of public pension systems along with labor income growth rate and expense growth rate. It breaks down income and expenses to government allowance, subsidy, individual contribution, gift and education expenses.

The labor income growth rate and expense growth rate calculation presented in this article will be referenced in the project.

Yao, S., Zhang, Z., & Huang, W. (2011). A distribution phase investment model of defined contribution pensions and its monte carlo simulation. *2011 2nd IEEE International Conference on Emergency Management and Management Sciences.* Beijing, China: IEEE.

This paper introduces Dynamic plan theory and stochastic optimal control theory which are used to obtain the optimal investment strategies and the analytical solution of personal fund under optimal investment strategies.

These two theories discussed in the paper will be considered when establishing an investment model with risk and riskless assets.

Zhou, J., & Lao, L.-J. (2011). An Asset Allocation Model with Social Insurance. *2011 International Conference on Management and Service Science.* Wuhan, China: IEEE.

This article discusses different countries' popular asset allocation and precautionary savings plans. It also explores a two-period model framework and explains the significance of analytical solutions.

The article provides a good retirement spending model for this project, in which it assumes the expenses of individuals after retirement will come from both social insurance and precautionary savings.