Matrance

(Matrix Finance portfolio manager)

"Financially powering your future through financial analytics."

Report and code written by:
Matthew Huang

Matrance

(Matrix Finance Portfolio Manager)

Product Description

Matrix Finance portfolio manager (Matrance) is an automated advanced portfolio management product that helps clients and users in equity asset management using historical and real-time data. Matrance uses a wide variety of sources and datasets to make the most informed decisions possible when managing assets. Matrance is fully customizable, allowing clients to input which stocks they have, current portfolio and risk levels in order to determine the best matching portfolio. It is a continually evolving application being created using modular functions and having a maintainable infrastructure such that new information and financial equations are continuously added as soon as they appear in the market. Matrance uses highly sophisticated algorithms and artificial intelligence in order to additionally gauge market sentiments and market trends to better calculate future cash flows and profitability. Datasets are based on a broad spectrum of market and sentiment indicators, news and heterogeneous data streams.

Basic Use and interface

Matrance will be an online web-based platform that connects to a central server to protect intellectual property as well as allow users to access and manage their portfolios wherever and whenever they want.

Matrix Finance portfolio manager uses:

Create a new portfolio with a customizable risk level.

Clients can select from a range of one to ten their desired risk level. Then using our financial models that incorporate risk such as modern portfolio theory we can generate a new optimal portfolio that suits their risk level.

Check whether buying or selling a stock an existing portfolio is worth it.

Using Matrance it can also be used as a validity checker for their own chosen stock by inputting the stock they wish to buy and writing and then determining if their stock predictions align with ours. This grants the client peace of mind when buying new stocks as they have evidence driven second opinion.

This is particular useful in checking for last minute changes in news or consumer sentiments as our predictive analytics delves into unstructured data that may be otherwise hard for clients to com across

---

Create or edit existing portfolio using client-selected stocks to find appropriate weighting that maximizes return.

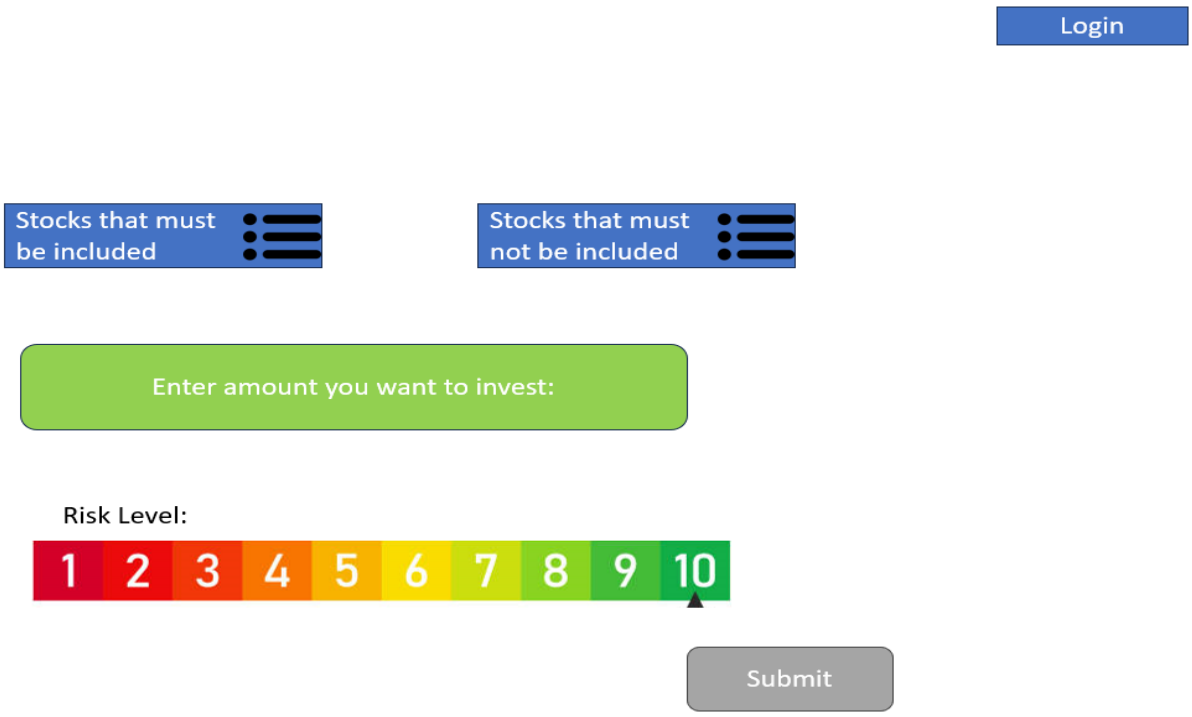
Matrance is fully customizable with what stocks the client wants, if they happen to love a particular company, they can make sure it is included in their portfolio and vice versa. This allows clients to have a portfolio that aligns with their beliefs such as a portfolio consisting of only green sustainable companies.

Use case scenario:

Background: Client has no current assets and has no preference for what stocks must be included. However, he wants a low-risk portfolio with a level of 5.

1. Login
2. Select new portfolio.
3. Webpage will show up with input boxes:
4. (optional) Choose stocks the client wants included and stocks the client doesn't want included.
5. Enter the amount that the client is willing to invest.
6. Selection scale for the amount of risk the client wants from 1-10
7. Submit
8. Portfolio will then be generated and will be displayed on a new screen, with the option of allowing us to create an actual portfolio for them at an added cost.

Draft user Interface:



Input Data

Fundamental Data:

Government data through quarterly economic indicators microeconomic indicators and Government fiscal policies. This comes from the economic indicators excel file.

Market Data:

Stock price history of the top 10 ASX stocks. This comes from the asx200top10 excel file. Additional market data can be attained through using the yahoo finance Api which has been configured and ready for use in the python code.

Alternative Data:

Relevant Company News from Twitter and other news sources in the form of a Json news dump. This comes from the news dump Json file provided.

User Data:

User inputs their desired risk level, the amount they're willing to invest and any stocks that they want to be included/excluded in the portfolio calculation.

Data calculated, stored and cleaned based on ASX excel data.

All data files have been loaded and stored such that they are ready for use in station 2, however some files have been additionally cleaned to ensure the data does not contain errors or any unnecessary information. These changes are:

- Client Details Excel: Removed 'AT Equity' in all column names as it is apparent and extra wording only seeks to create confusion and unnecessary clutter in data manipulations.
- Client Details Excel: Calculated the sum of equities to check if they equal to 100 (when rounded to 4 d.p.), else it indicates that the data might be incomplete, so raise an IO exception.
- Economic Indicators Excel: Got rid of the first two months (first two columns) as they contained incomplete data for some rows, which could cause errors and inconsistencies in calculations.
- ASX200Top10 Excel: Removed the first row and appended the stock name to the second row as the first row had lots of blank spaces making it hard to do calculations as it generates fake column names.

Output Data

Once the data has been cleaned it is then output as data frames ready to be used in station 2 and station 3, where it is used to generate models and predictive analytics. For stock prices the output format is a data frame with

Extracting, transforming and loading data

Station 1 is all about Extracting, transforming and loading data from different sources and formats and making ready for computation in code. This station is split into three different stages which each serve a core function before it is ready to be ready and interpreted by data analysts and predictive models.

Extracting

Financial prediction is all about understanding the market and every key aspect to ensure the right decision is made. Not all information is good, and it is imperative to extract the information that serves to improve our financial model and discard information that seeks to hurt it. Information can firstly be harmful when it is from an untrustworthy source, this can cause the data to be speculative and dramatized to garner attention, this can occur in things such as media sources. Information that is diluted or derived from too many sources at once can also seek to impede models as the data gets translated and transformed so much that it begins to deviate from its original data. Lastly it is important to be mindful of biased data as data from the company tends to emphasize its strengths and omit shortfalls.

Hence extraction is a core part of station 1 as we need to extrapolate correct data as well as relevant data. Data can either come in structured such as a csv file or it can come from an unstructured source such as a news website or text file from a twitter API. In station one we will take in data from both in order to gauge a more holistic understanding. Unstructured data is useful for more opinionated data to gauge market and consumer sentiments for equities for future predictions. On the other hand, structured data tends to be in the form of more historical and numerical data creating historical evidence to back up future predictions. Combining both structured and unstructured data will thus allow us to better predict future asset prices and make informed decisions.

Transforming

Data also needs to be transformed to ensure there are no errors in the data and it is clean relevant data for station 2. This can take on many forms such as duplicate data, unnecessary information and errors.

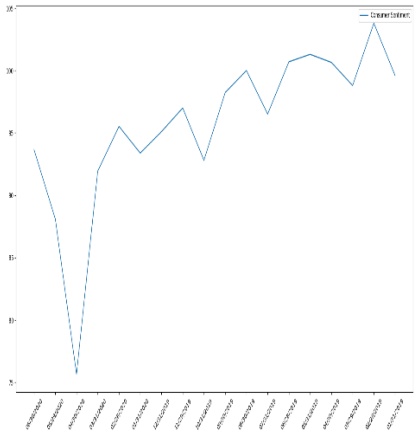
Loading

Once the data has been transformed into clean data it can be loaded and saved as a Data frame.

Prompt and Features Engineering

Inputs

Core inputs for this station will be the ASX200top10 data frame and economic indicators. Additional inputs such as the Json news dump have been processed and are ready for model creations. Client details have been also feature engineered to only have the different companies as the columns, removing risk profile and age group so it can perform correlations between stocks and optimal portfolio calculations.



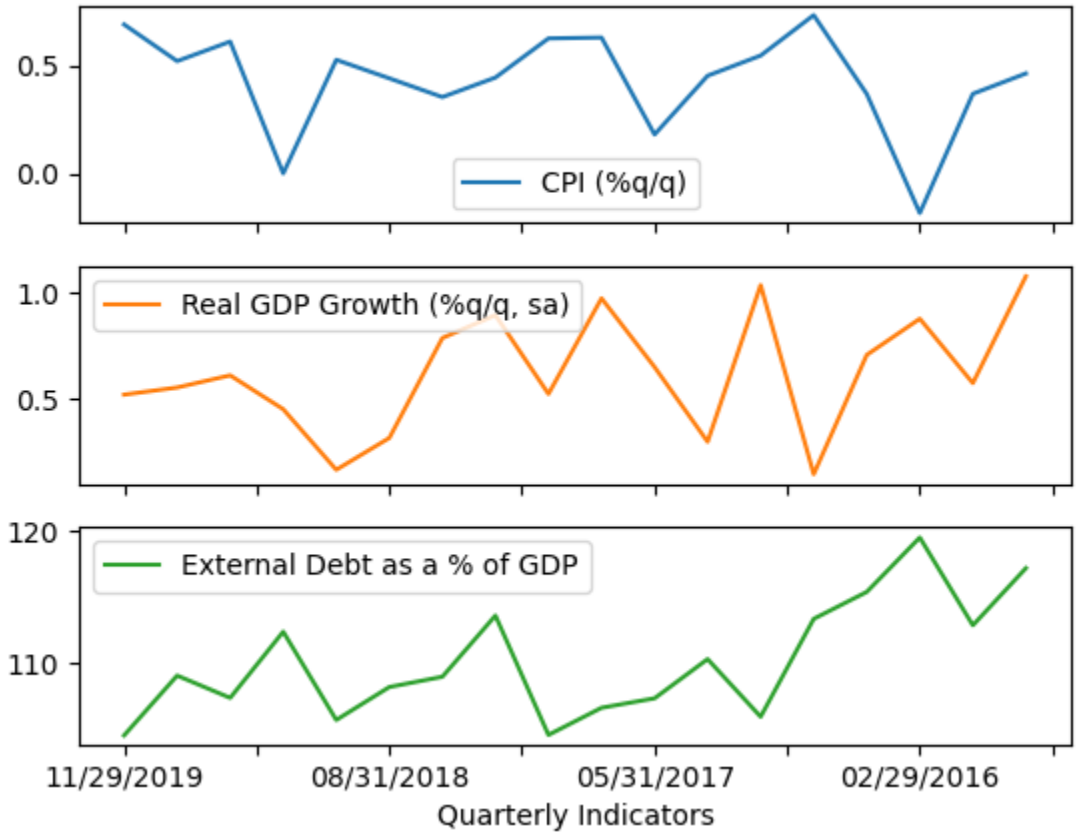
Data collection and data format requirement

Data collected will be for the three excels with be in a data frame format following station 1. Data format requirement is data frame input as well as output. Having a data frame format will better facilitate graph creations as shown to the right.

Core Features

The overall features that this software will have are; portfolio generation as well as uses listed in station 1. It will be based on a wide variety of applications such as browning model and self-learning algorithms. Sentiment valuation will also be done through looking at the news dump Json as well as any extra sources such as economic indicator.

Within this station certain features have been selected and implemented in preparation of station 3. The first feature is that for the ASX200Top10 excel file, all data excluding the prices has been omitted, making it ready for portfolio analysis. Another Feature is that for the economics indicators file certain features have been selected to be used as input in station 3 as shown below. These data have been selected based on relevancy as a way to anticipate future consumer spending and market performance.



Preparing for model creation and implementation

Station 2 is all about preparing for model creation and implementation by identifying and isolating key information that can be used to construct financial models. This is done through features selection which involves predicting models that are going to be used in station 3 and then isolating and capturing all relevant data into one data frame so that it is ready to be featured in models and predictive analytics. Station 2 is also used as a final way to check the validity and relevance of data by constructing various models and performing tests such that we begin to see some patterns begin to emerge. Examples of feature engineering can be found in the appendix as well as the code file.

Model design and implementation

The model design that has been chosen is Markowitz Modern Portfolio theory as base model and then additional predictive analytics to supplement and reinforce the model. The model will have the option to be based on solely historical analysis or it will have the option to have predicted future stock prices included in the model. This customizability is a key feature in our model implementation as we live in a society which is unpredictable and so it may be better to omit predictions during unpredictable periods. For example, including predictive analytics during the mass Covid outbreak would be less beneficial and even detrimental as the training models will not have any idea or scope of when restrictions would end and so would not take it into account. This would leave either a model which regards the sudden loss in stock prices as an indicator of a long-term recession or it could regard it as part of the business cycle and increase the price in the next business cycle.

Modern portfolio Theory

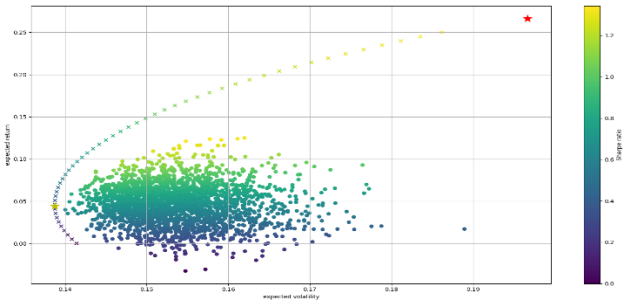
Modern Portfolio Theory or Markowitz portfolio theory is a relevant financial model that has widespread applicability and strong feature usage determining whether stocks are profitable as well as the optimal composition to ensure it minimizes additional market risks by lowering beta. Modern portfolio theory is an optimal model to be used as a base model for our design as it incorporates Sharpe ratio which allows clients to customize the level of risk, they are willing to take. Markovitz proposed the concept of an “Optimal Risky Portfolio” or “Optimal Tangent Portfolio,” which is a combination of assets that maximizes the Sharpe Ratio. The Sharpe Ratio measures the performance of an investment adjusted for risk and is calculated using the formula:

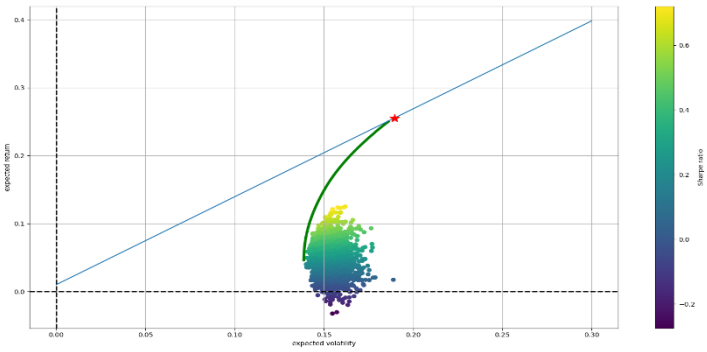
$$SR = \frac{\mu_p - r_f}{\sigma_p}$$

where E(rp) is the expected portfolio return, rf is the risk-free rate (assumed to be 1% per annum in the code), and σp is the standard deviation of the portfolio return.

This portfolio, denoted as P*, lies on the “Efficient Frontier,” a line connecting the portfolio with the highest return for each level of risk. In theory, all investors should invest in P* regardless of their risk tolerance. An individual’s risk profile determines what proportion of their capital should be invested in the Optimal Risky Portfolio. The more risk-averse an individual is, the greater the proportion of their funds that should be invested in a risk-free asset such as a government bond.

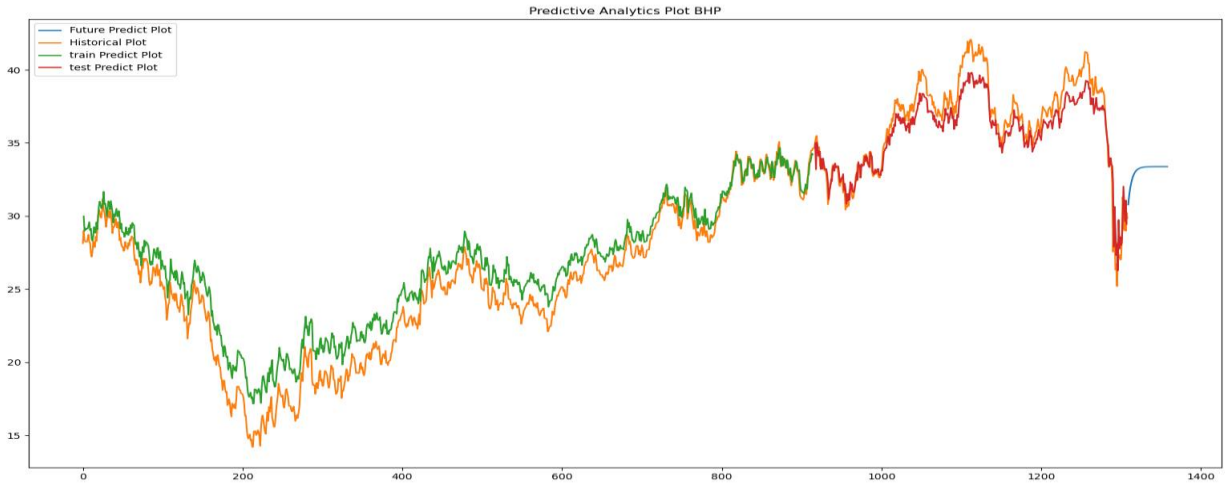
In our model design we incorporated this by creating random portfolio weights equal to the number of assets we are analyzing and then simulated 2500 combinations of returns and deviations in order to create an efficient frontier and optimal risky portfolio.



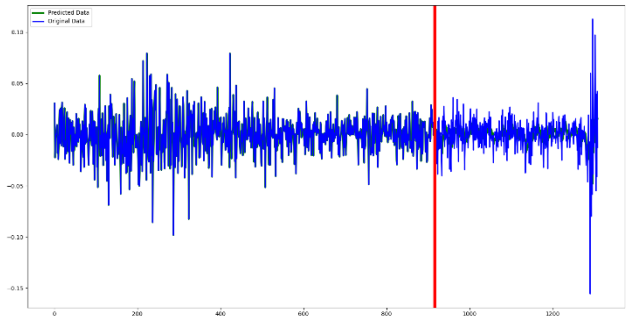


Predictive Modelling (TensorFlow, ARIMA)

TensorFlow is used in predictive modelling to create a model that trains and predicts future stock prices based on the current stock price. The training data and testing data is personalized for that particular stock so environmental conditions are controlled. For example, a stock which would be affected by Covid would not be used to train a stock which would not be affected by Covid, such as Coles and Quantas as they have very different conditions. All stocks have been trained and tested such that they all predict the next 50 future stock price changes and are subject to change in the future. All models have been trained 100 times.

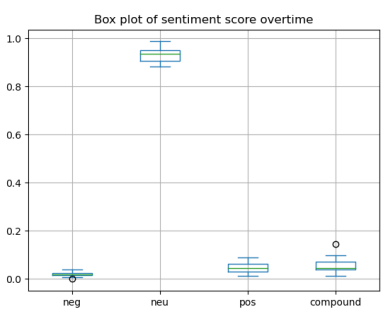


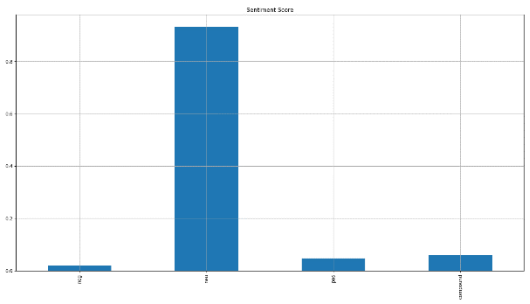
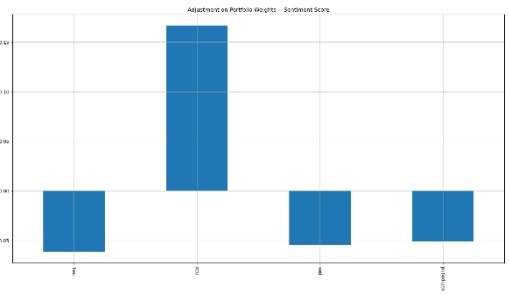
Arima has also been used as a predictive model to give an indicator of returns and whether or not certain stocks are predicted to yield high returns in the future. Currently, Tensorflow is used as the main predictive model and the final model will input data from TensorFlow but Arima will be used to supplement the data as a secondary opinion and provide insight on stock profitiability.



Sentiment analytics Modelling (NLTK VADER, NLP TensorFlow)

Two different models of Sentimental analytics were created using different predictive neural network frameworks, VADER and NLP. Vader was trained using its pre-trained NLTK training model, which is regarded as gold standard training. Although both Vader and NLP models have been created for sentimental analytics only Vader will be applied to the final model implementation. This is because the NLP model has been trained using movie review data and so will be less applicable to financial application, unlike Vader which has wider applicability. Instead, NLP, usage in the overall product design will be as quality assurance and a checkpoint providing secondary opinion on whether there is generally a positive or negative sentiment. In cases when NLP finds a negative sentiment. It will raise a flag and check whether that particular stock is within the portfolio and if so, then makes a warning that further research on the stock should be performed.



Model assumptions

- There are no brokerage fees which although minor may cause discrepancy in data.
- Short selling is sometimes the optimal portfolio but, in our model, we assume the client does not want to short sell and the model cannot short sell.
- Risk free rate is 1%, however can be changed in future iterations and upon client request.
- Models do not prioritize and give extra weight to more recent headlines and news.
- In the model there is currently no government bond as part of the portfolio and instead we assume that clients will always want some degree of risk as opposed to storing in a bank

Restrictions; computing speed, capacity, accuracy issues

For running the models and implementing the solutions there is a lot of initial cost in initial training of the models, however the model training is saved, and so future runs of the model will have drastically lessened computing time. The longest model to computer is the TensorFlow NLP in computing IMDB movie reviews to train the model, taking around 40 minutes. However, once it has been run once the runtime decreases drastically to less than 2 minutes, making the overall runtime of all models around 5 minutes to completely run all models.

What is expected to be seen in model implementation

Model implementation will seek to bring all these models together, whether it be through one model or providing many models for clients such that they are able to use their own judgement with the models to see whether the recommended portfolio is worth it. Our product is also flexible in that while it does provide optimal portfolios for clients, it also is used to aid customers' pre-existing portfolios by supplementing stock information with data and self-learning predictive analytics.

Model Boundaries

Currently boundaries in the model exist because of the assumptions that we make when considering the use case of the product and the clients needs. Boundaries that are hence created is that the model does not incorporate a risk-free asset in its portfolio management which very risk-averse customers might want. Another boundary is that the model cannot short sell for customers who are extremely risk-seeking as short-selling is a way to propel expected returns at the cost of higher Sharpe ratios and risks.

Product Design

The final product design will be an automated advanced portfolio management product for clients covering equity asset class, based on a broad spectrum of market and sentiment indicators, news and heterogenous data streams. It will be implemented by taking into consideration all the different models such that they all play a part in weighting and final modifications after the base modern portfolio has created an optimal portfolio.

Predictive analytics using data stations has been implemented in the final product design by having an additional option where customers can choose to use a customized model that includes predicted stock prices at the end of the dataset such that modern portfolio theory is created with future stocks in consideration.

Sentimental analytics can then be taken into consideration by looking at sentiment statistics and then having the weightings of stocks multiplied by one plus the compound result of predictive analytics. This way stocks that perform well using sentimental analytics will have a higher compound and thus a higher multiplier increasing the weighting of the stock. For example, if a portfolio had 20% AMC, it would then be multiplied by 1.042304 (1 + compound on right) Once all sentimental analytics have been done then the total weight will be greater than 100% and so we will rebalance it by multiplying all values by a factor such that all values go back to cumulatively summing up to 100%.

Equity	neg	neu	pos	compound
AMC	0.040077	0.909192	0.050731	0.042304
BHP	0.015137	0.945112	0.039750	0.066181
BXB	0.018308	0.907769	0.073923	0.098092
CBA	0.035871	0.897271	0.066814	0.072631
CSL	0.027538	0.883462	0.089000	0.145949
FPH	0.000000	0.988231	0.011769	0.033877
RIO	0.019765	0.936137	0.044098	0.047176
TLS	0.007182	0.966091	0.026697	0.043233
WES	0.020690	0.935286	0.044024	0.039519
WOW	0.017316	0.954395	0.028289	0.012092

Once both predictive analytics of future stock prices and sentimental analytics have been taken into consideration two final portfolios will be generated for the client allowing them to choose which one, they prefer in terms of their risk level as shown below. With each index corresponding to different stocks and the below array showing additional data such as beta and alpha and Sharpe ratio.

```
***Maximization of Sharpe Ratio***
[0.    0.376 0.027 0.    0.    0.    0.    0.    0.    0.597]
[0.266 0.197 1.354]
****Minimizing Variance***
[0.    0.003 0.098 0.03  0.14  0.064 0.167 0.23  0.08  0.187]
[0.044 0.139 0.321]
```

Steps to introduce this financial product to clients

When introducing this financial product to clients, we will convey our strengths and relevancy in our modern-day financial markets through stressing three key points.

- **The importance of artificial intelligence**

Artificial intelligence and neural networks are the future of financial analytics and portfolio management as they incorporate future data and so are key in creating portfolio's that wish to stand the test of time and generate short term returns as well as long term returns. Without predictive analytics as generated in this financial product, clients experience a time lag in information over other firms and so will be behind other investors in investing in future lucrative assets, thus buying at a higher cost premium than others. With our help through predictive analytics clients will be able to buy cheaper and earlier.

- **The importance and power of sentimental analytics**

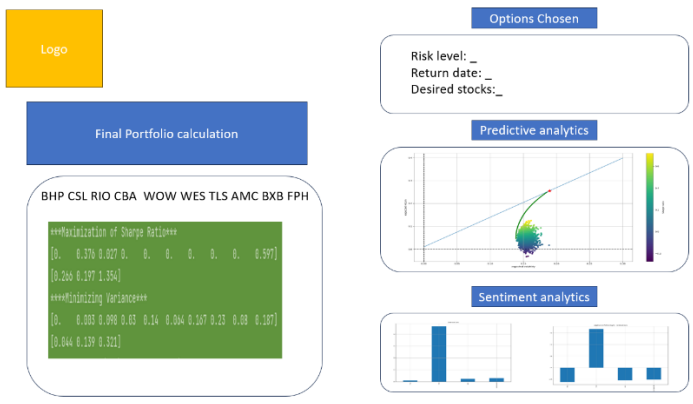
Financial models have traditionally relied on historical returns and data generated from numerical structural data. However, in modern day finance we can delve into unstructured data to make greater, more informed decisions. Finance companies and investors who don't incorporate sentimental analytics have only tapped a small portion of available data and with the help of our product, the accuracy and relevancy of future expected stock prices can be vastly improved and reinforced, such that as soon as major events occur, our models can quickly adapt and inform clients.

Customer journey:

Background: Client has no current assets and has no preference for what stocks must be included. However, he wants a low-risk portfolio with a level of 5.

1. Login
2. Select new portfolio.
3. Webpage will show up with input boxes:
4. (optional) Choose stocks the client wants included and stocks the client doesn't want included.
5. Enter the amount that the client is willing to invest.
6. Selection scale for the amount of risk the client wants from 1-10
7. Portfolio will then be generated and will be displayed on a new screen, with the option of allowing us to create an actual portfolio for them at an added cost.

Hence, the customer journey will be seeing three different models and then choosing and customization what models they want to include in their final portfolio as clients need to consider real world applicability and nuances that may exist causing our financial model to be inaccurate at times.



Appendix:

Station 1:

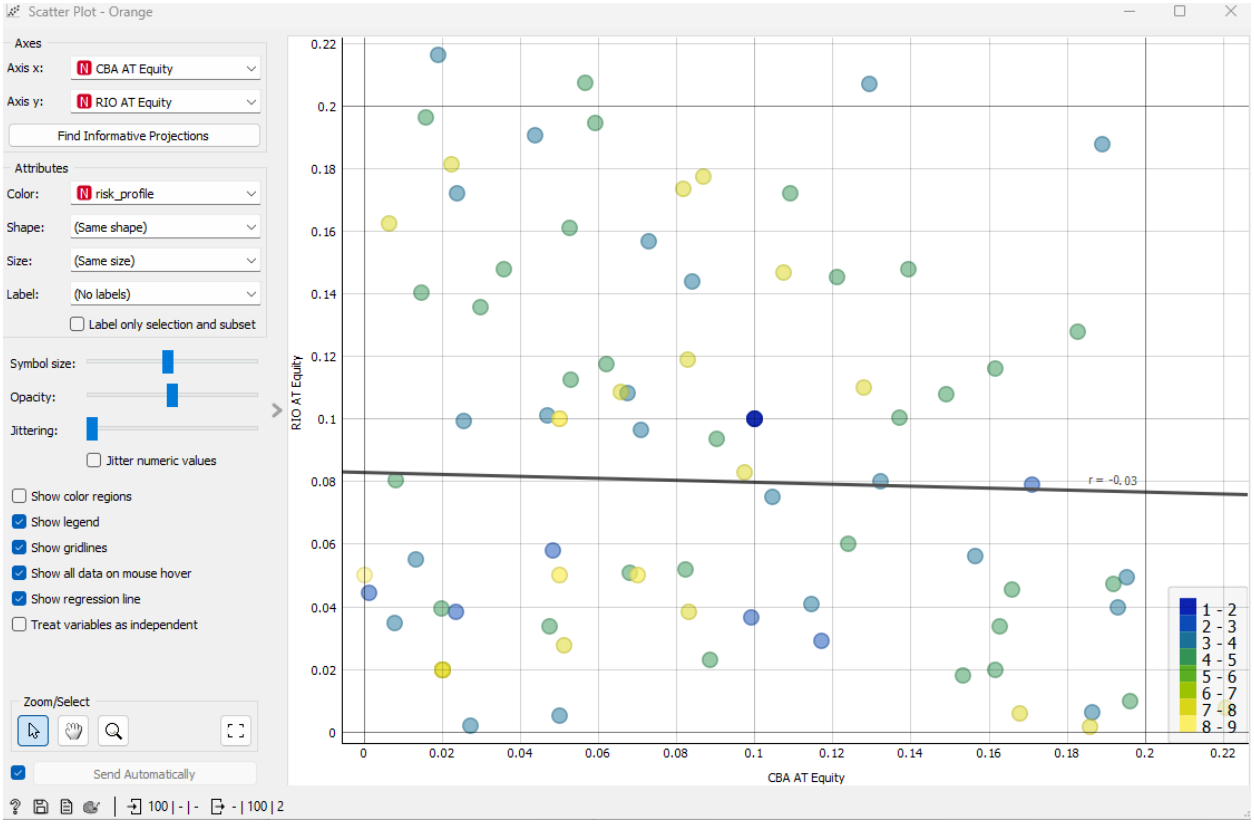


Figure 1 ETL Data Exploration

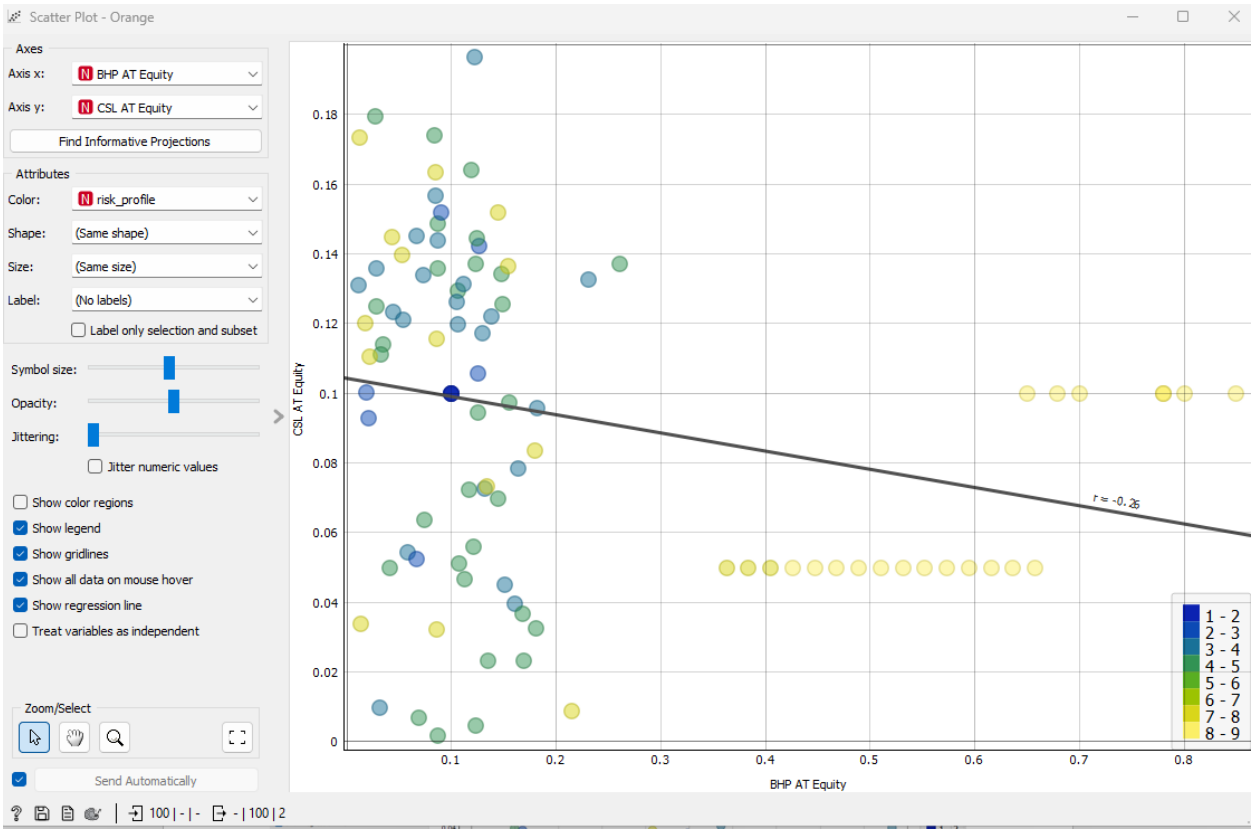
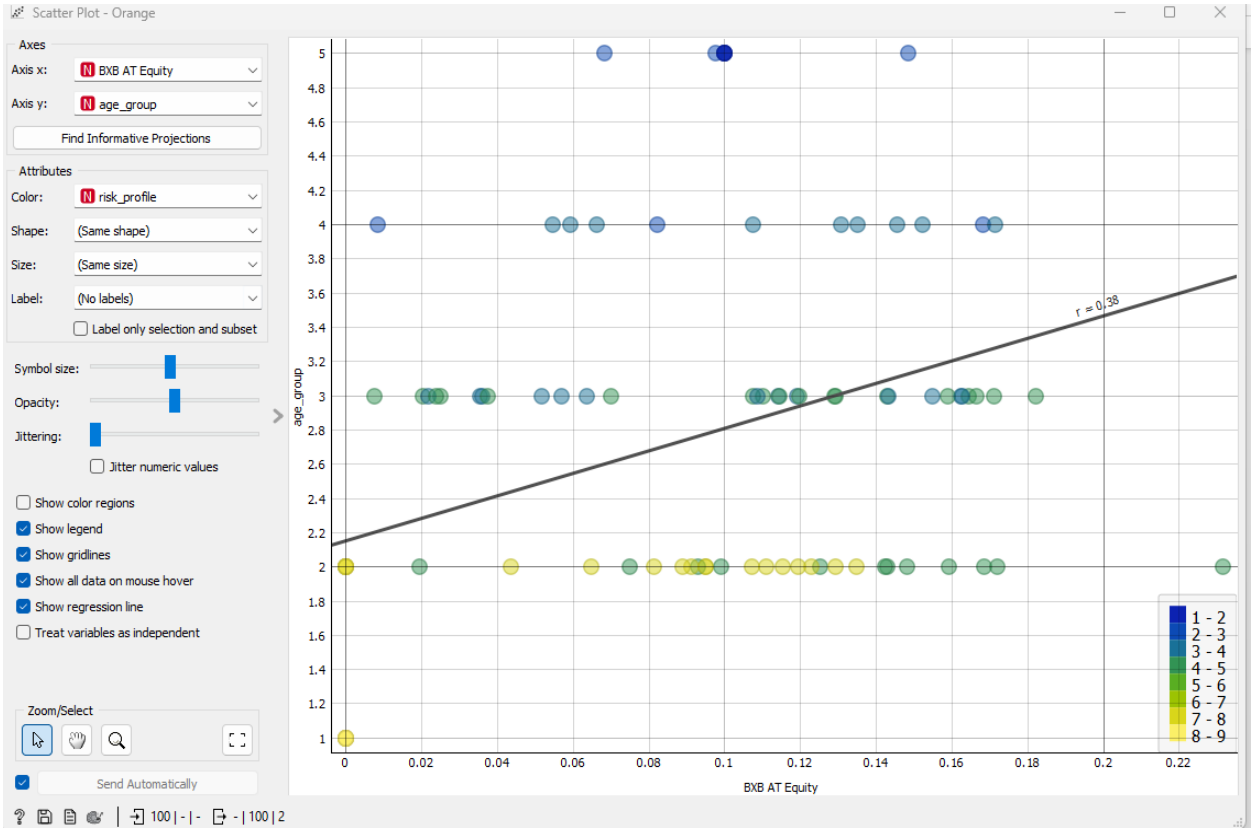


Figure 2ETL Data Exploration correlation



Station 2:

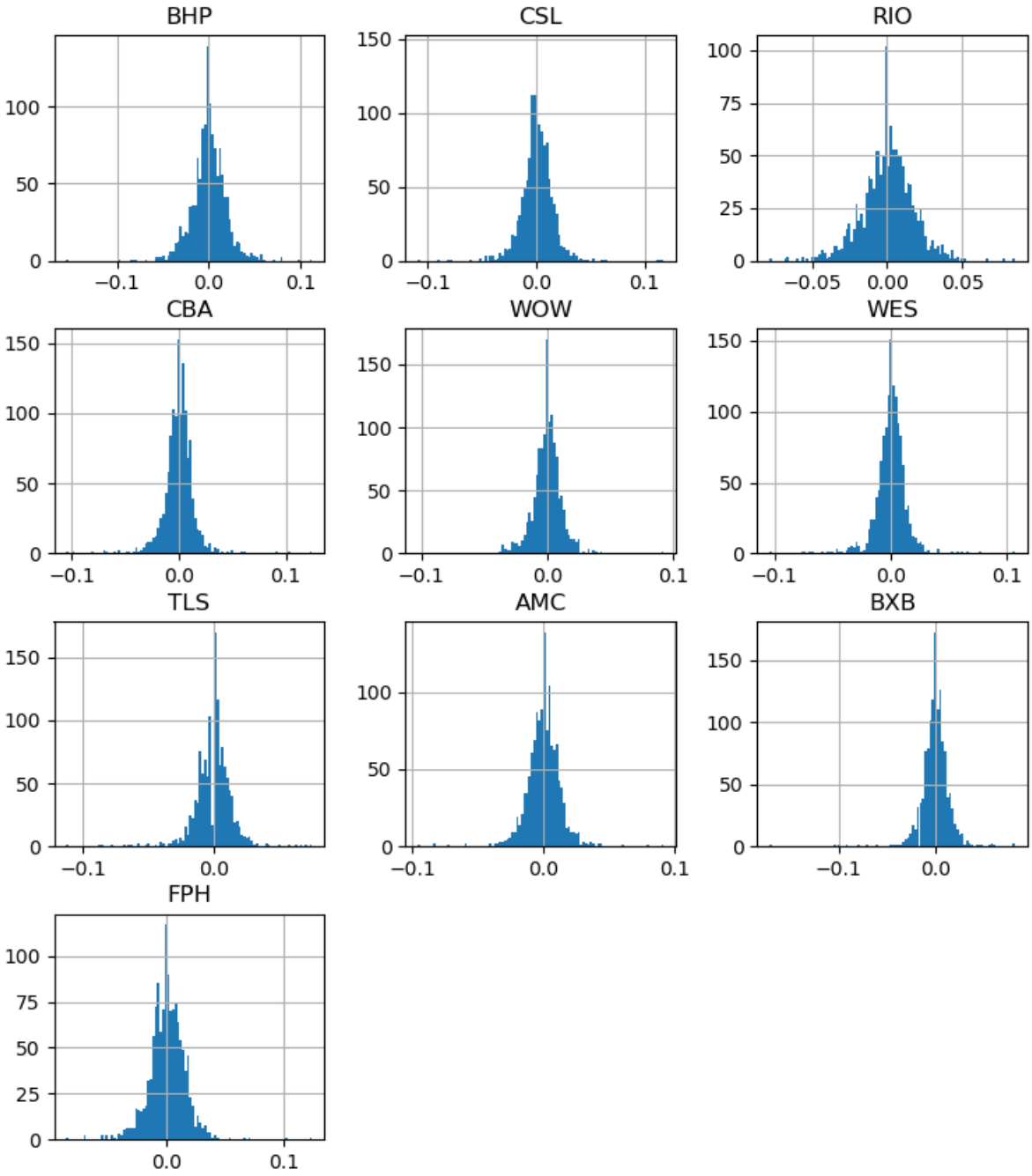


Figure 3Feature Engineering normal distribution of stock price

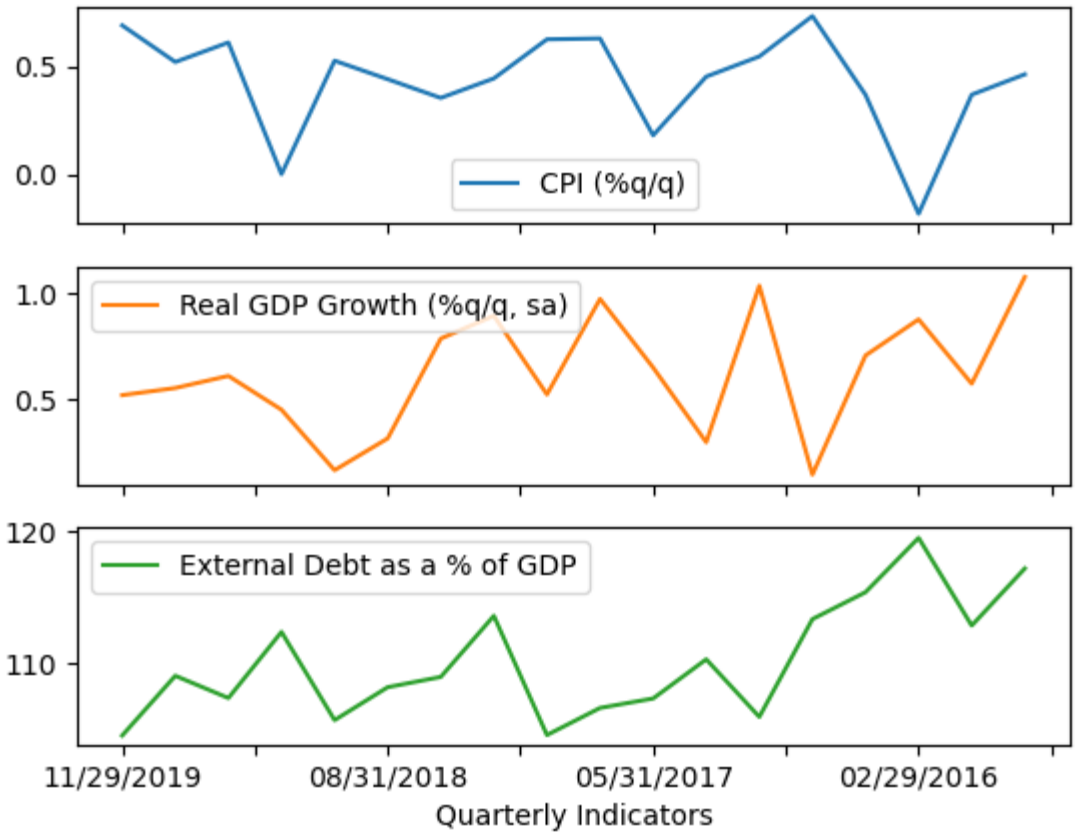


Figure 4 Feature Engineering, Core economic indicators

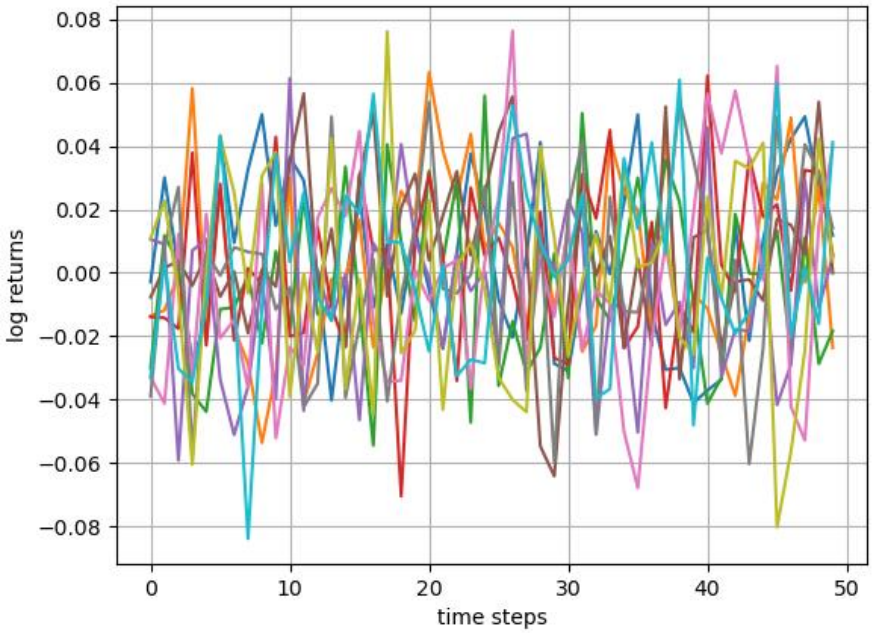


Figure 5 Feature Engineering Brownian motions log

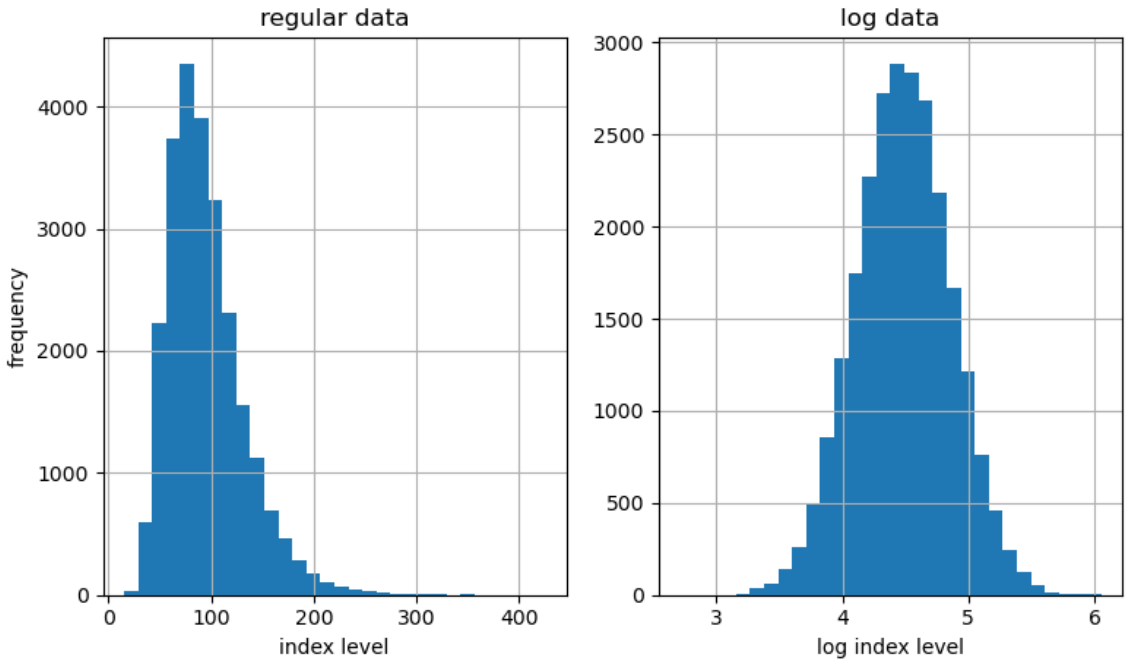
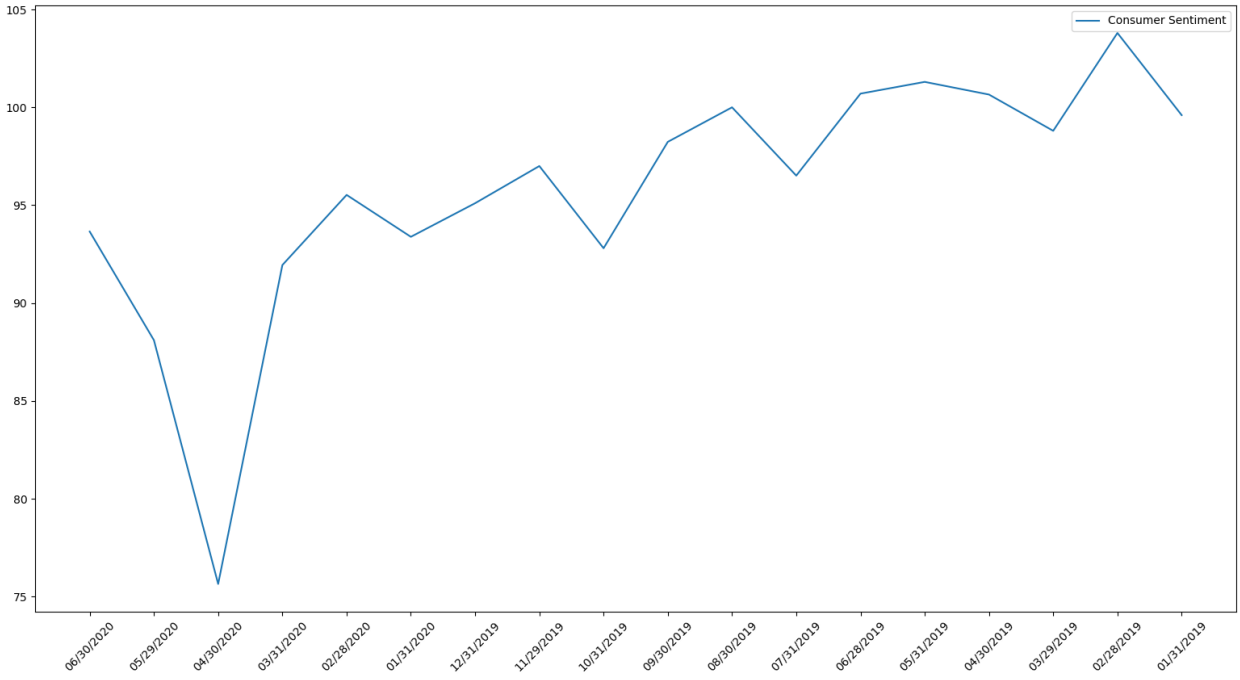
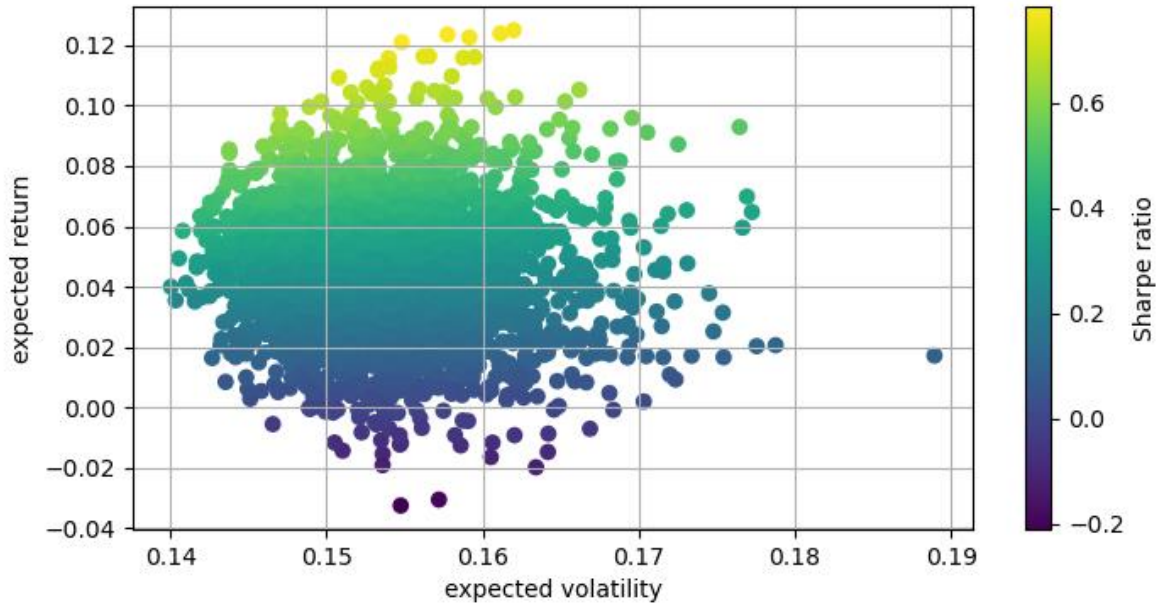
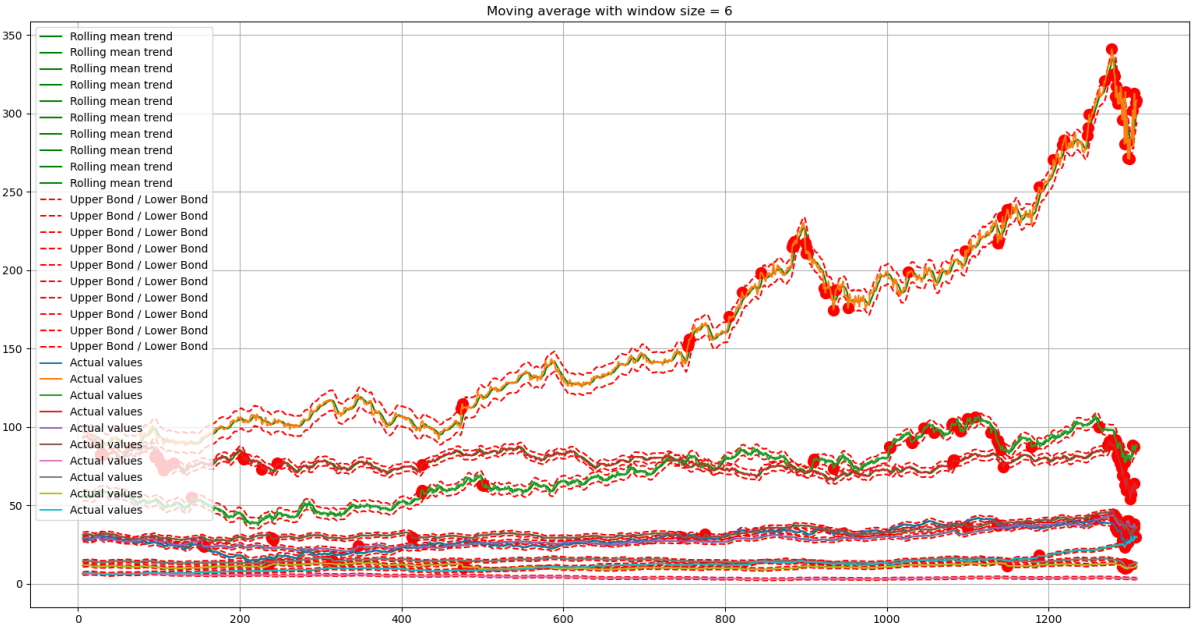


Figure 6 Feature Engineering Normally Distributed Returns sample



Scatter plot modern portfolio





BHP price

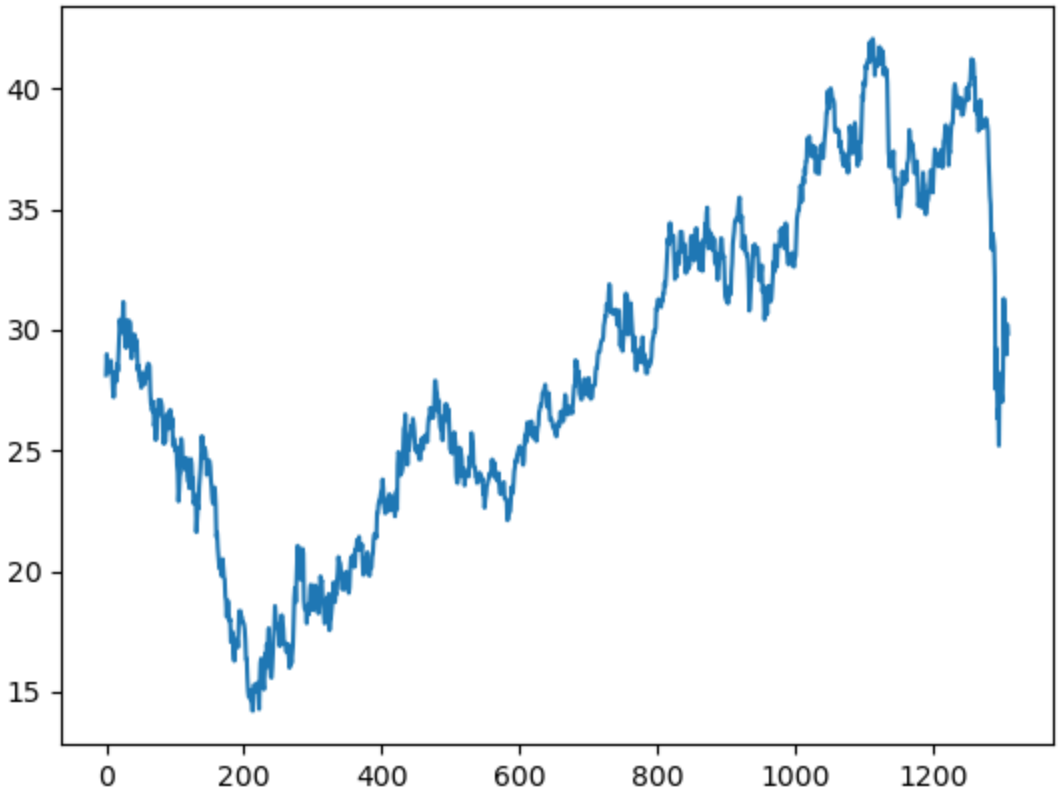


Figure 7 Feature Engineering Consumer Sentiment