A Project Final Report on

OPTIMAL ASSETS ALLOCATION

Submitted in Fulfillment of the Requirements for
the Apprenticeship of **Financial Data Analyst Apprentice Program**
under CodeRush

Submitted by:-
Kashindra Mahato
(Financial data analyst apprentice)

Date:
9th October 2022

CodeRush
Bhanimandal, Lalitpur, Nepal

INTRODUCTION

Asset allocation means spreading investments across various assets. The goal of allocating assets is to minimize risk while meeting the level of return expected.

The asset allocation task involves predicting the weights to be assigned to each of the assets based on the risk of that asset. For asset allocation, the concept of Modern Portfolio Theory by Harry Markowitz in his paper “Portfolio Selection” is used in this project. According to Markowitz, any investment’s risk and return characteristics should not be viewed alone but be evaluated by how it affects the overall portfolio’s risk and return. Therefore a plot of 50,000 different combinations of weights for portfolios is plotted in an efficient frontier plot and hence determined the set of optimal portfolios that offer the highest expected return for the defined level of risk and the lowest risk for a given level of expected return. To predict the returns and volatility using machine learning models, namely LSTM and GARCH models are used respectively.

What would be the average returns of each stock in a portfolio, How volatile are the stocks, what is the risk associated with each stock, and which combination of weights for stocks in the portfolio provides the expected return are the important questions answered by this work. The further part of the document will elaborate upon these questions and the methods used to get the result and also the significance of the results. To summarize the expected results, an optimal set of portfolios would provide a risk-free return or return with the expected level of risk.

The securities that are used for analysis are listed as follows:

AAPL : Apple Inc

ANF : Abercrombie & Fitch Co

GIS : General Mills, Inc

HRL : Hormel Foods Corp

K : Kellogg Company

PIM : Putnam Master Intermediate Income Trust

SBUX : Starbucks Corporation

VZ : Verizon Communications Inc.

ANALYSIS

A. What are the returns(daily, weekly, monthly, and annual) over a specified period of time?

For the analysis below a period of 5 years is taken i.e we have datasets containing the 'Close' price of 8 securities from '2016-01-01' to '2021-12-30'.

Before getting the returns we first get the data on a weekly, monthly, and annually basis. We do so by resampling the data accordingly and taking mean. Then we calculate the average returns:

Table 1: Table representing the average daily, weekly, monthly, and yearly percentage returns of each security listed above.

	Symbol	Avg_Daily_Return	Avg_Weekly_Return	Avg_Monthly_Return	Avg_Yearly_Return
0	AAPL	0.585	1.495	2.970	36.677
1	ANF	1.253	2.044	1.521	29.552
2	GIS	0.193	0.315	0.478	3.047
3	HRL	0.137	0.286	0.525	5.503
4	K	0.097	0.117	0.101	-0.230
5	PIM	0.063	0.148	0.478	5.496
6	SBUX	0.243	0.533	1.159	15.205
7	VZ	0.098	0.184	0.512	5.253

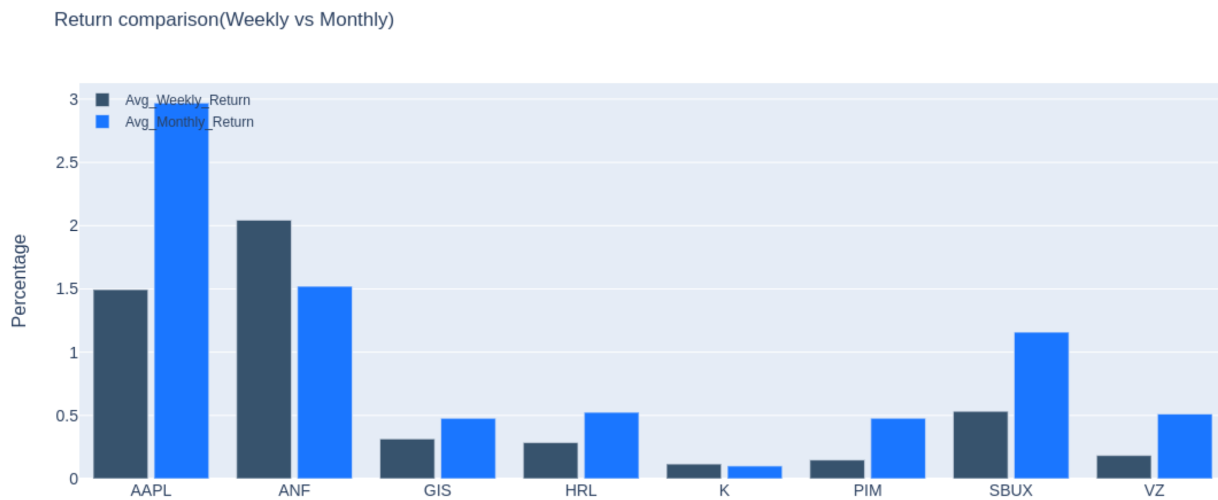

Figure 1: Bar chart of the average weekly and monthly percentage returns of each security.

The average weekly and average monthly values from the above table and comparison bar chart above show how AAPL, ANF, and SBUX provide the most returns among all other securities in the portfolio. Also, it can be seen that the monthly return for AAPL, SBUX, PIM, and VZ are almost double the weekly returns.

The stocks that performed poorly and have even fewer returns on a monthly basis are ANF and K.

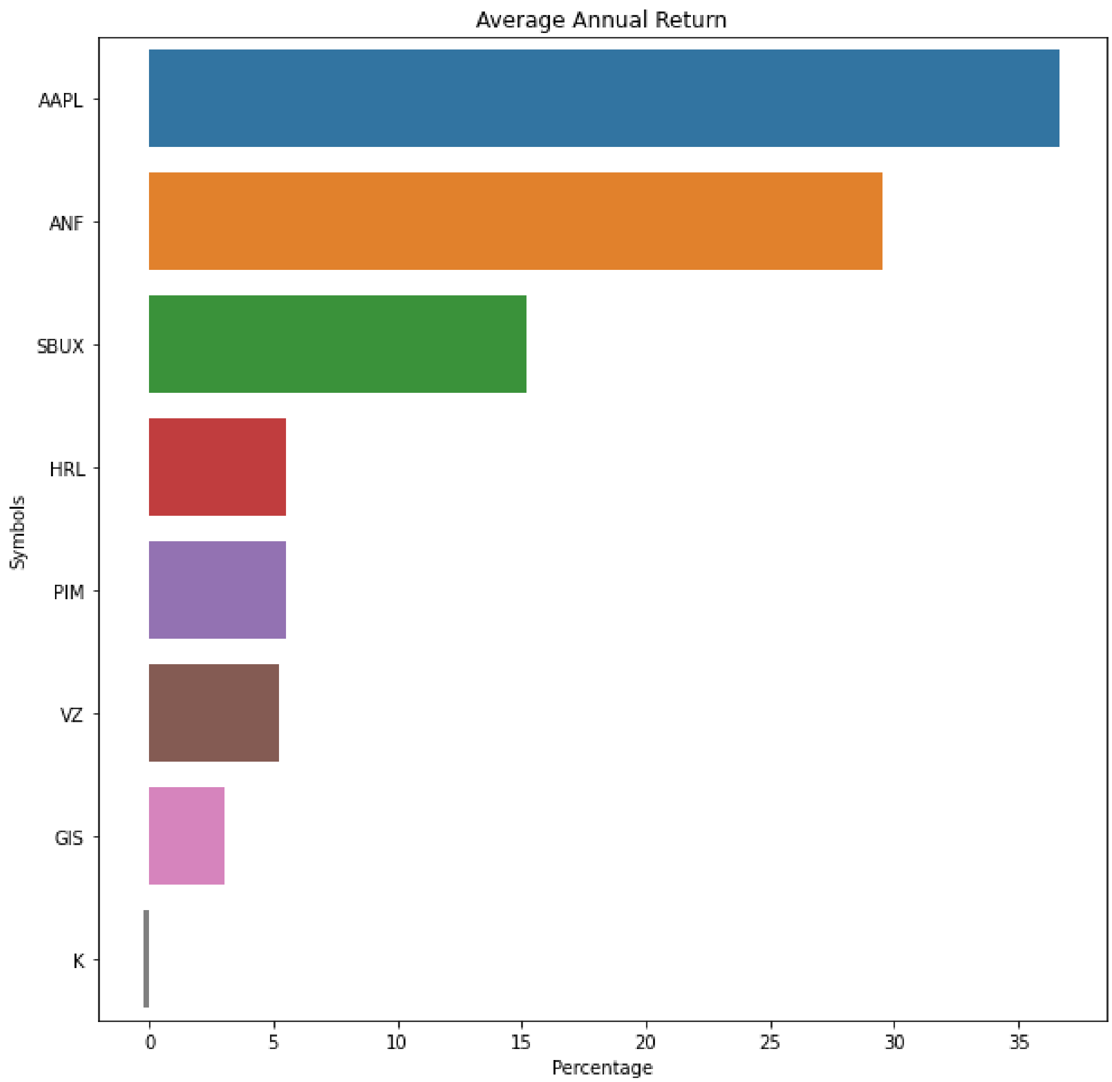

Figure 2: Horizontal bar plot showing the annual average percentage return.

AAPL, ANF, and SBUX have the most return over 5 years on yearly basis. The return of K has gone negative as we can see from the table above and also from figure 2.

To expect the most return, one must invest in the stocks that have provided maximum return over a recently given period of time. If one expects more returns on a short-term basis, one should invest in stocks that provide better returns on a weekly and monthly basis. From our analysis, we have AAPL, ANF, and SBUX as the top 3 stocks that would provide a good return.

Similarly for long-term investors, an yearly analysis would provide a base for investment decisions. In our case coincidently the same top 3 stocks provide the best returns on yearly basis as well.

B. What is the volatility(variance, standard deviation, Sharpe ratio) of each ticker specified in the portfolio?

Volatility is a statistical measure of the dispersion of returns for given security. It is a measure of how riskier a security is. The volatility of given stocks is calculated using **monthly data.**

Table 2: A table representing the volatility of securities.

	Variance	Std_Deviation	Sharpe_Ratio
AAPL	1684.826	41.047	0.252
ANF	73.230	8.557	0.620
GIS	43.798	6.618	0.252
HRL	37.189	6.098	0.300
K	13.077	3.616	0.097
PIM	0.132	0.363	4.601
SBUX	494.673	22.241	0.182
VZ	40.990	6.402	0.279

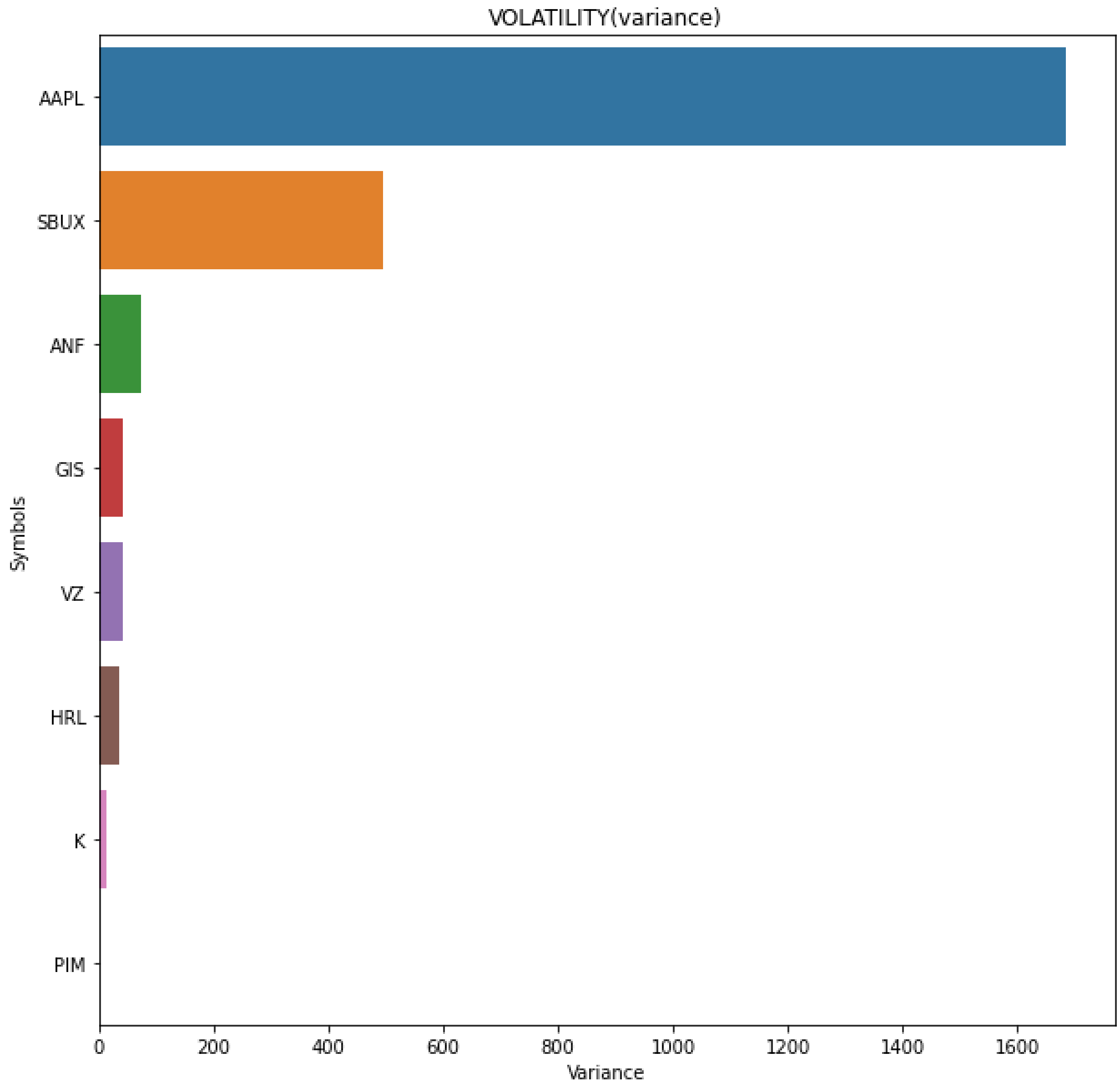

Figure 3: Horizontal bar plot showing the variance of given securities.

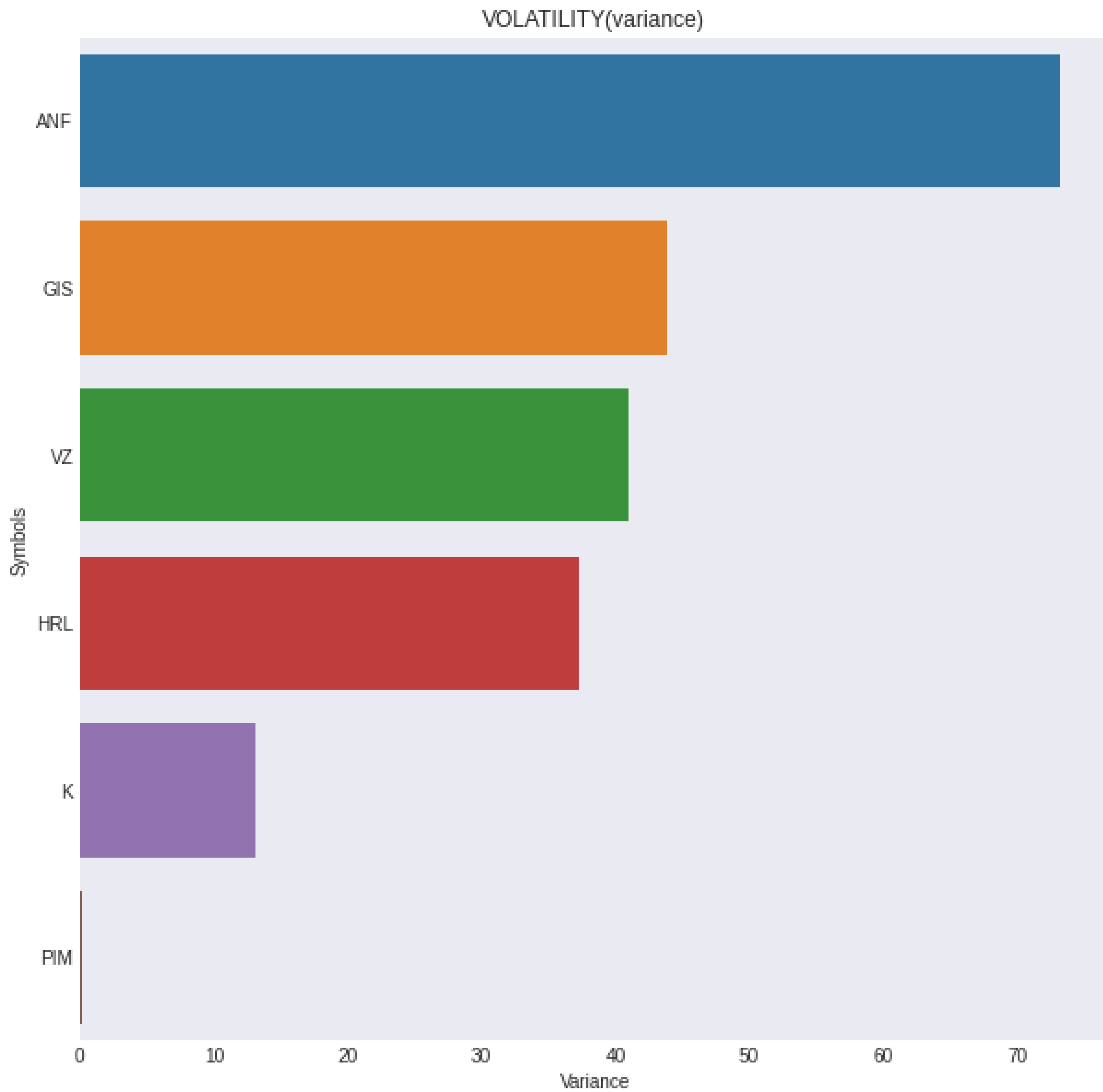

Figure 4: Horizontal bar chart showing the variance of stocks(except AAPL and SBUX).

Variance measures how far each number in the set is from the mean, and thus from every other number in the set. If a security has a greater variance, it may be interpreted as more risky or volatile.

Here, AAPL and SBUX have the greatest variance and hence the most volatile of them all. Therefore all other stocks seem to have a lesser variance from figure 3 but after plotting the same graph after removing AAPL and SBUX, we can see that ANF also has a reasonably larger variance.

From the above returns graphs and variance graph, we can see there seems to be a direct relationship between return and variance.

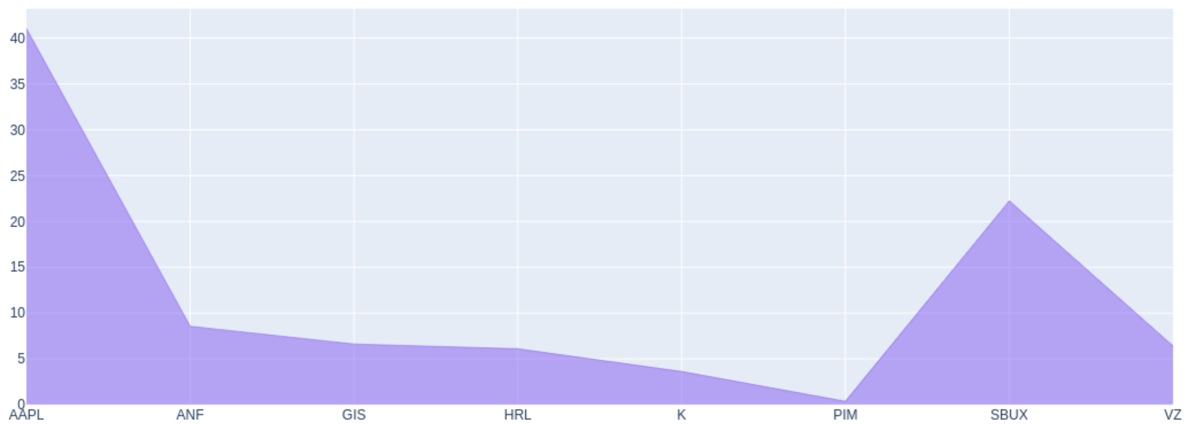

Figure 5: Area plot showing the standard deviation of given stocks.

The standard deviation is calculated as the square root of the variance. The standard deviation is more reliable than the variance calculation since variance calculation is prone to error because of outliers. Although the results and effect of standard deviation on the volatility are the same in our case.

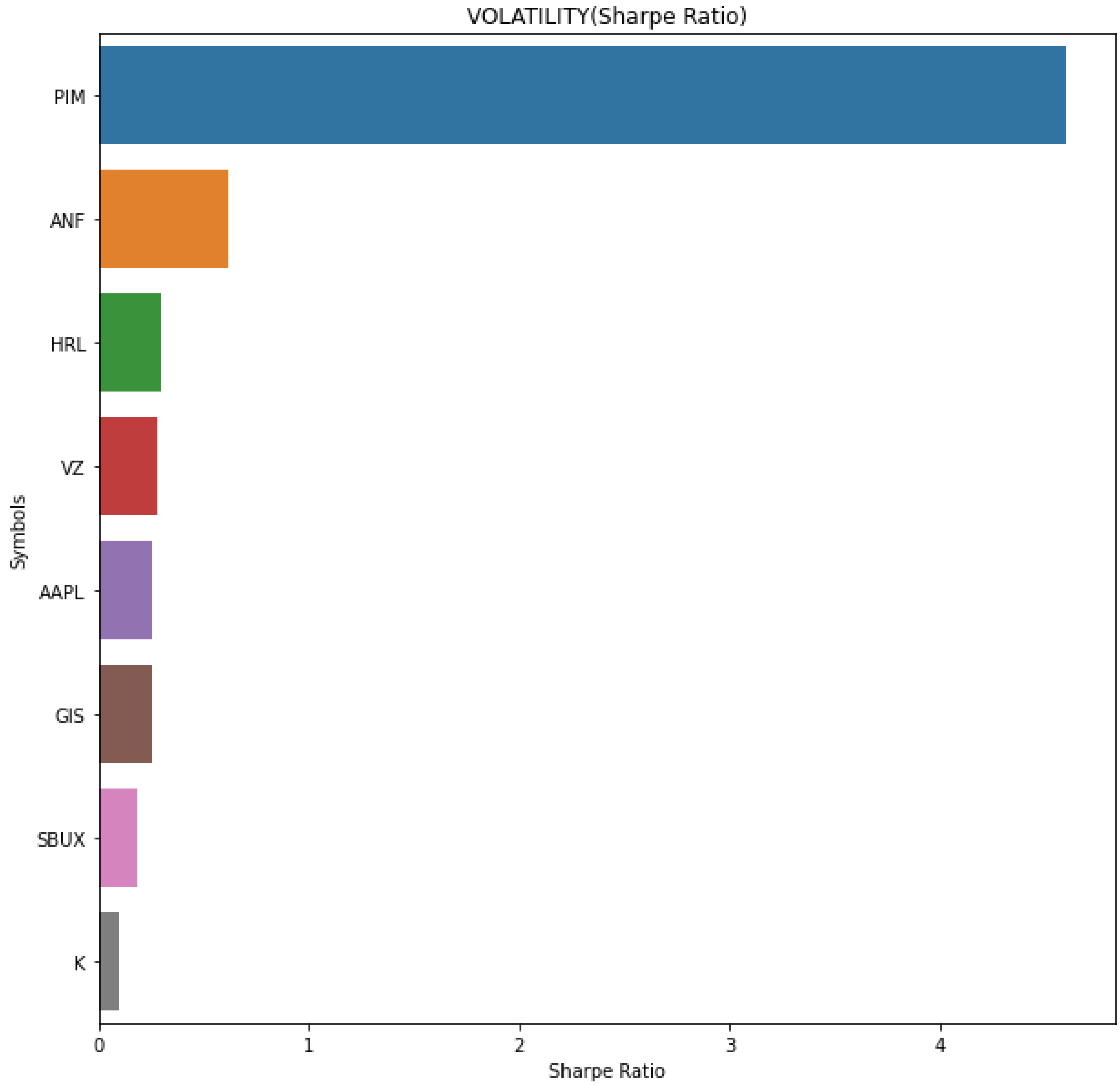

Figure 6: Horizontal bar plot showing the Sharpe ratio of given stocks.

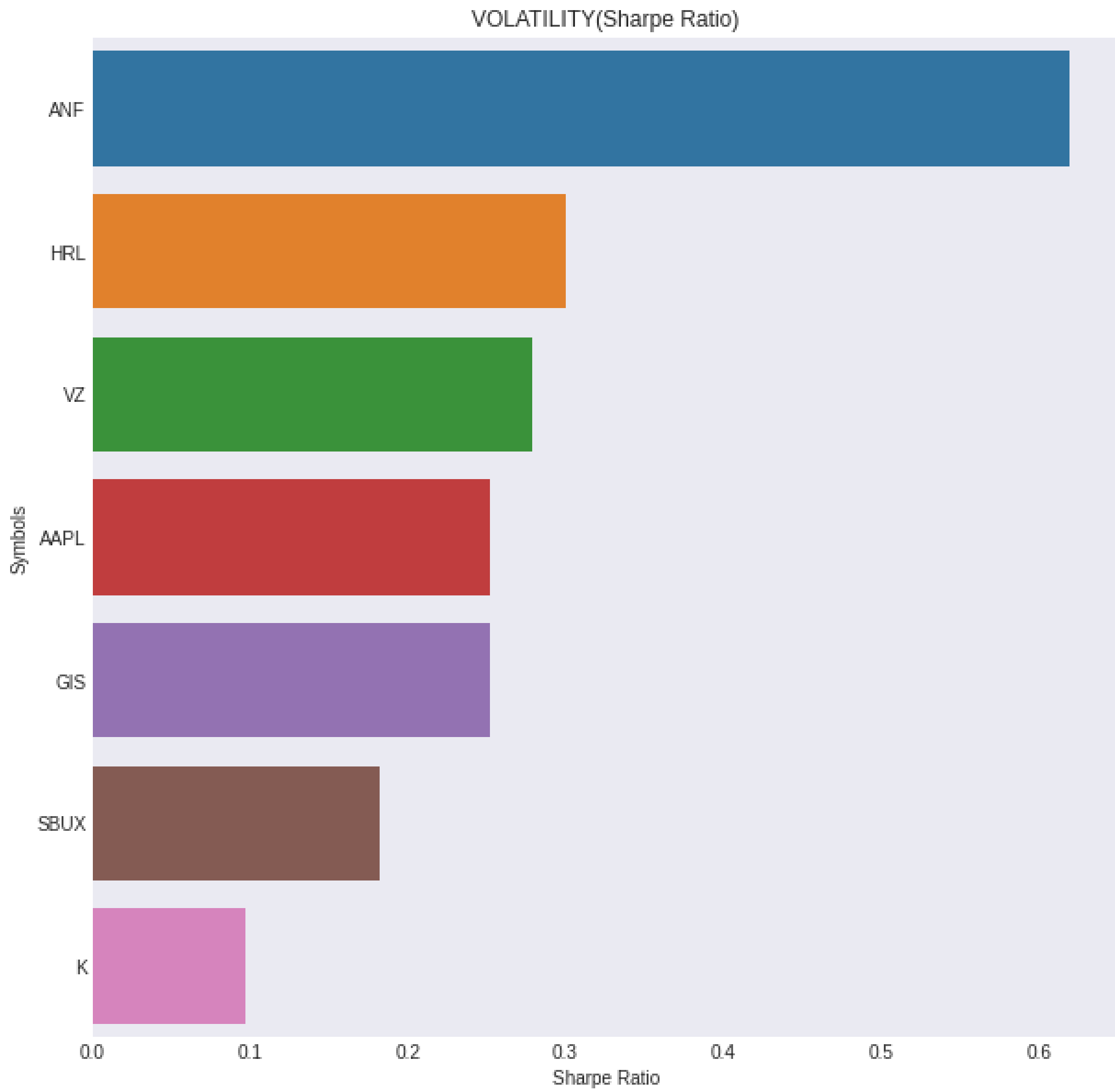

Figure 7: Bar plot of the Sharpe ratio(except PIM).

The Sharpe ratio compares the return on investment with its risk. A higher Sharpe ratio is better when comparing similar portfolios.

Sharpe Ratio = (return of portfolio - risk-free rate) / standard deviation of the portfolio's return

From table 2 and figure 6 we can see PIM has the largest Sharpe ratio compared to others. Therefore a 2nd graph is plotted excluding PIM, which makes the graph more readable and comparable. ANF provides the 2nd largest Sharpe ratio. Given the amount of returns, AAPL and GIS provide but they have a lesser Sharpe ratio because of their higher variance/standard deviation.

A stock that is more volatile may provide more returns but also is riskier, here although AAPL and SBUX provided more returns they are riskier. ANF and PIM have better Sharpe ratios and also provide reasonable returns.

C. What are the values at risk at 95% and 99% confidence intervals for each ticker and also their covariance?

Table 3: Risk management table showing value at risk.

	VAR_at_99p	VAR_at_95p
AAPL	0.119	0.068
ANF	0.216	0.182
GIS	0.088	0.055
HRL	0.066	0.043
K	0.066	0.046
PIM	0.049	0.018
SBUX	0.091	0.047
VZ	0.056	0.038

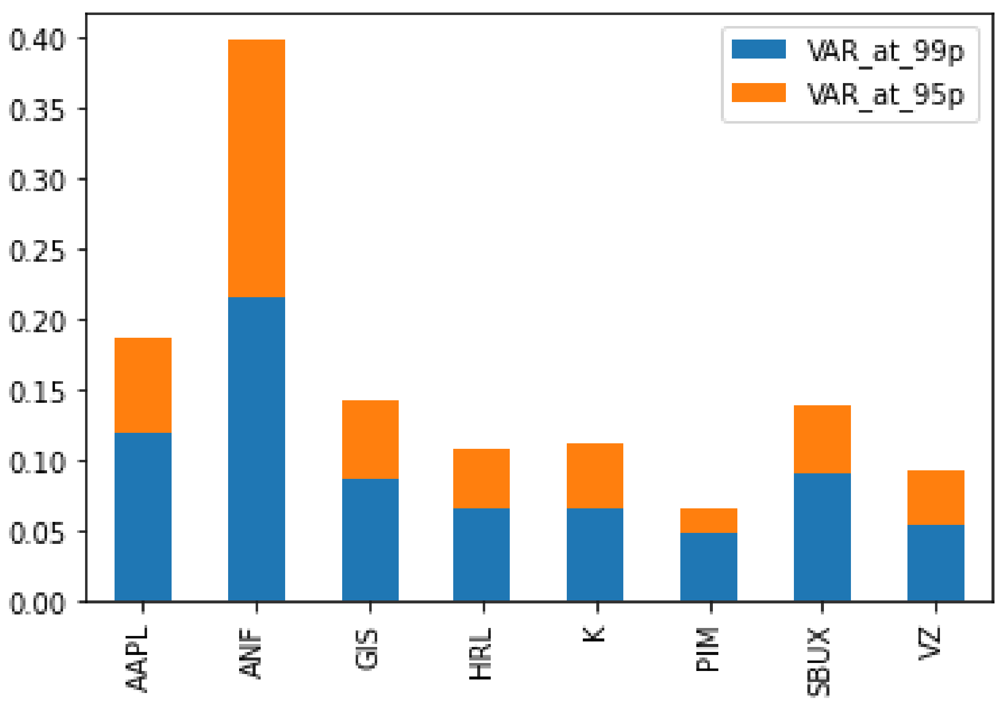

Figure 8: Stacked bar chart showing the comparison of value at risk at 95% and 99% confidence intervals.

Value at risk(VAR) is a way to quantify the risk of potential losses for a firm on an investment. From the above bar chart, insight can be gained that ANF has the most value at risk at both confidence intervals.

Table 4: Table representing covariance matrix of monthly returns of stocks.

	AAPL	ANF	GIS	HRL	K	PIM	SBUX	VZ
AAPL	4.723e-03	3.635e-04	6.988e-04	6.608e-04	7.216e-04	7.660e-04	6.764e-04	1.383e-04
ANF	3.635e-04	1.601e-02	-2.972e-04	4.578e-05	-3.198e-05	5.250e-04	1.589e-03	4.105e-04
GIS	6.988e-04	-2.972e-04	1.670e-03	4.958e-04	6.640e-04	2.090e-04	5.347e-04	5.855e-04
HRL	6.608e-04	4.578e-05	4.958e-04	1.207e-03	3.353e-04	7.476e-05	4.386e-04	3.847e-04
K	7.216e-04	-3.198e-05	6.640e-04	3.353e-04	1.021e-03	9.162e-05	4.250e-05	2.301e-04
PIM	7.660e-04	5.250e-04	2.090e-04	7.476e-05	9.162e-05	4.908e-04	5.522e-04	2.020e-04
SBUX	6.764e-04	1.589e-03	5.347e-04	4.386e-04	4.250e-05	5.522e-04	1.965e-03	4.556e-04
VZ	1.383e-04	4.105e-04	5.855e-04	3.847e-04	2.301e-04	2.020e-04	4.556e-04	9.712e-04

Covariance is a measure of the relationship between two or more variables. In finance, it is used to measure the relationship between two asset returns. Covariance applied to a portfolio can help determine what assets to include in the portfolio. It measures whether stocks move in the same direction(a positive covariance) or in opposite directions(a negative covariance).

From the above table, we can see that ANF, GIS, and K have a negative covariance which means they will not move in the same direction. And hence allowing us to minimize the risk.

D. Show the correlation between securities mentioned in the portfolio.

Correlation is a statistic that measures the degree to which two securities move in relation to each other. Correlation and covariance differ in a way that covariance tells you that two variables change the same way while correlation reveals how a change in one variable affects a change in the other.

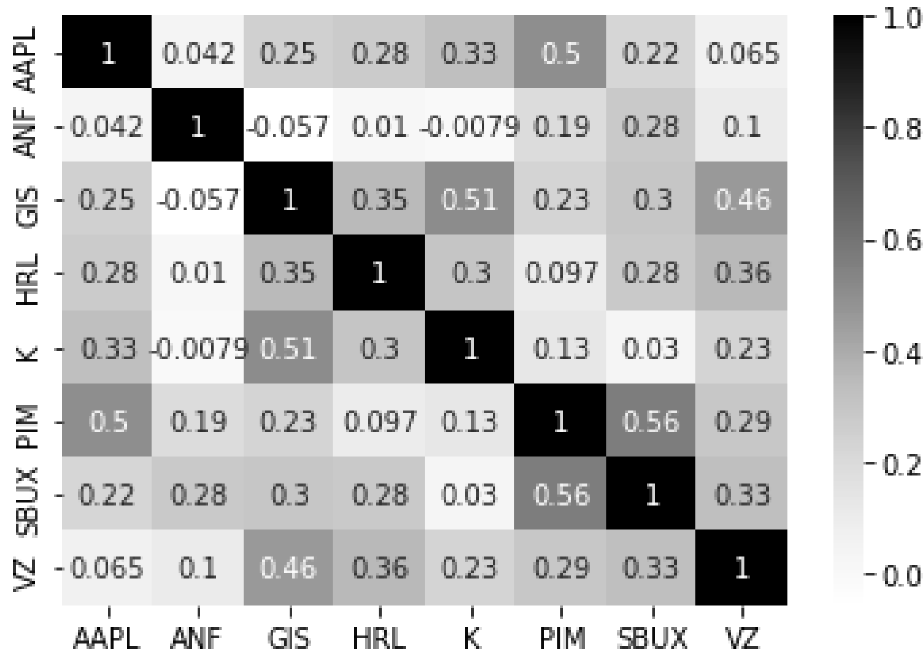

Figure 9: A heatmap showing the correlation between the security's monthly return.

The heatmap above the dark grey colored box represents a high positive correlation while a brighter box means a low correlation or negative correlation.

ANF is negatively correlated with APPL, GIS, and K. There also seems to have a positive correlation between PIM and SBUX, GIS and VZ, AAPL and PIM, and so on with a value greater than 0.5.

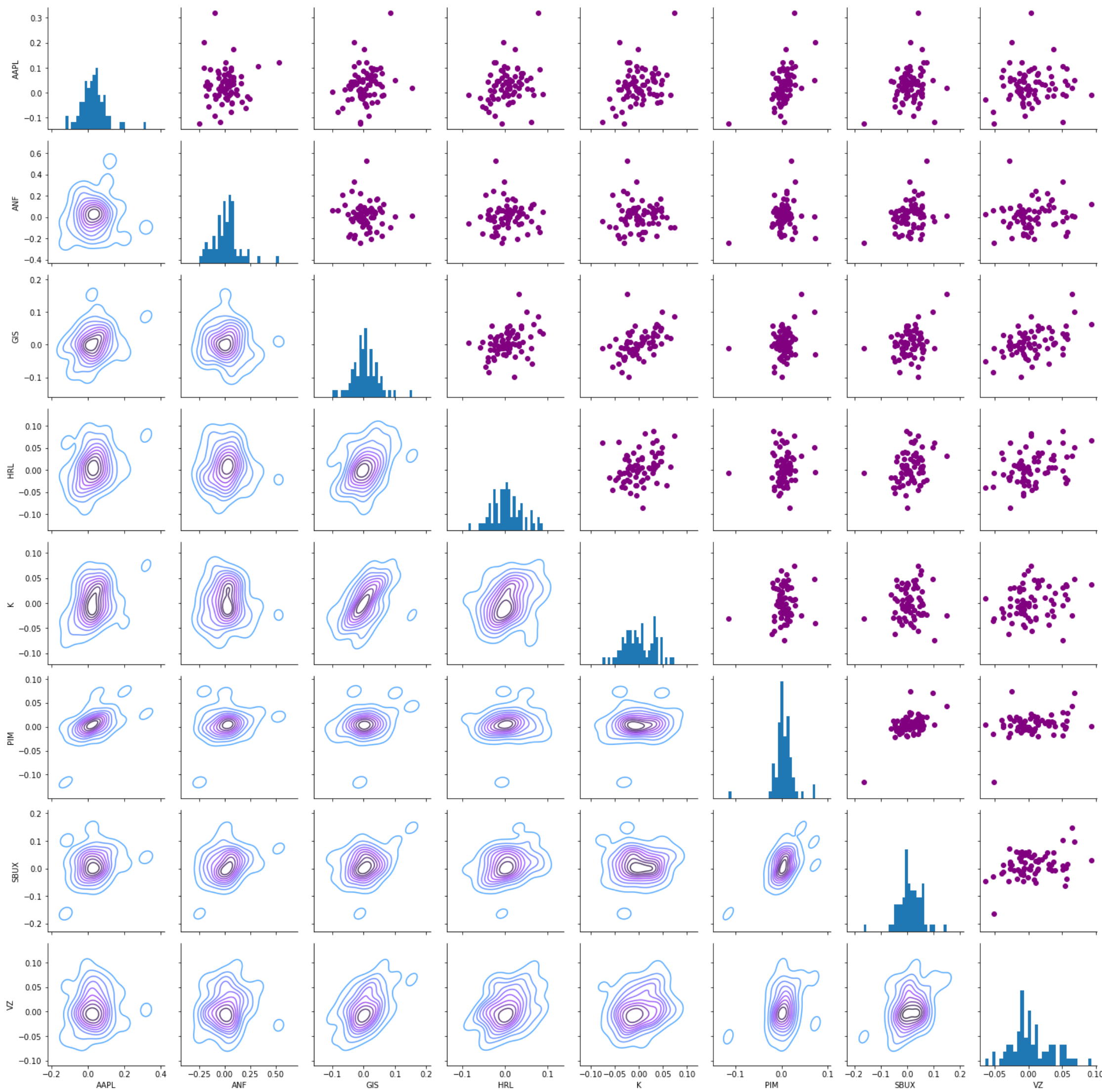

Figure 10: Pair plot of monthly return data of stocks with scatter plot in the upper triangle, KDE plot in a lower triangle, and histogram in the diagonal.

The scatter plot shows how two variables are related ie if there exists a linear relation or other relation or no relation at all. Similarly, a bivariate kernel density estimation plot shows the probability density function. From the pair plot, we can see the scatter plot forming an almost linear pattern for those variables which are correlated which we saw from the previous correlation table. From the contour plot, we can also see the density of each pair, which is almost centered.

Highly correlated stocks are riskier if considering a portfolio return. Therefore in our case, we must select the ones that have less correlation or the ones that are negatively correlated to minimize the risk.

E. Which portfolio provides the minimum variance and which portfolio provides the maximum sharpe ratio?

An efficient frontier is used for this task. The efficient frontier is the set of optimal portfolios that offer the highest expected return of a defined level of risk or the lowest risk for a given level of expected return.

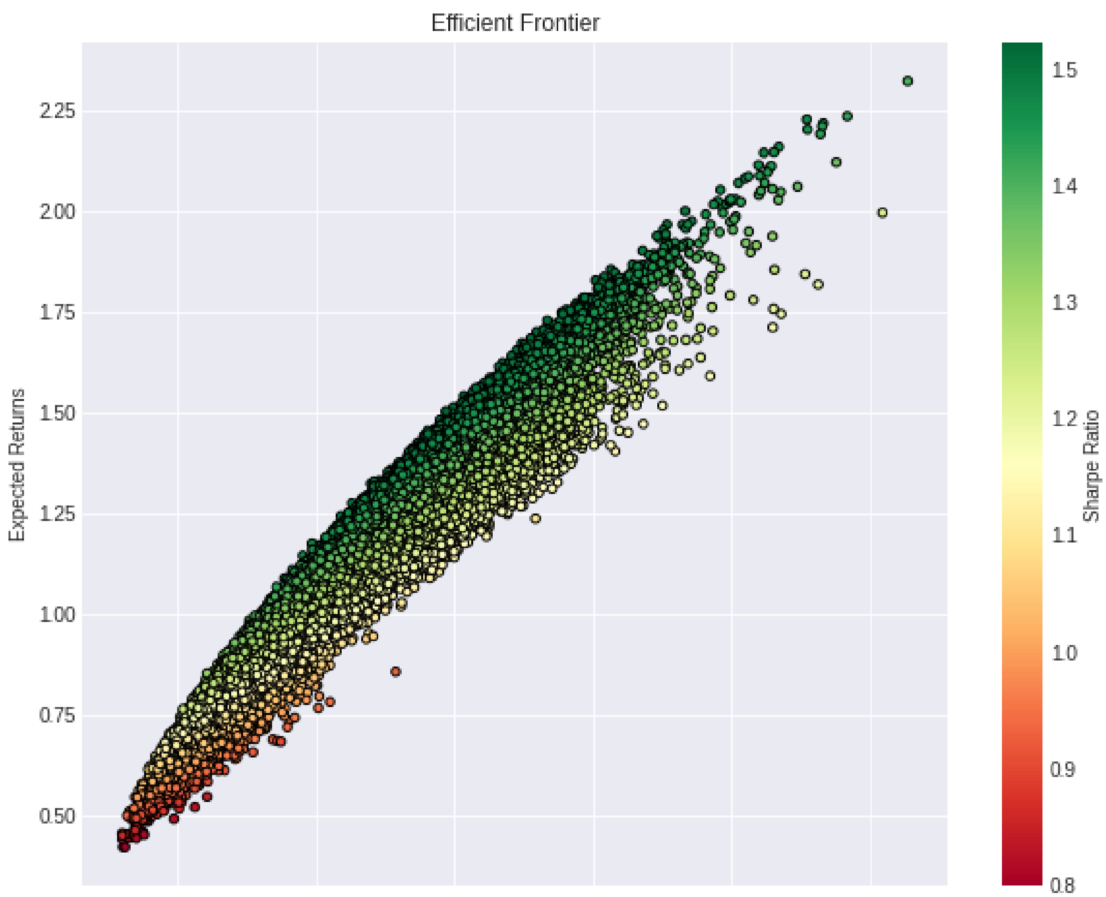

Figure 11: An efficient frontier plot between expected return and volatility along with the sharpe ratio.

To plot this graph, daily and annual returns and covariances are calculated. Then we randomly assign weight to the 50,000 different combinations of securities in the portfolio to get their combined return, volatility, and sharpe ratio. The weights along with return and volatility are then combined to form a dataframe, a sample of which is shown below in table.

Table 5: Components of an efficient frontier.

	Returns	Volatility	Sharpe Ratio	AAPL Weight	ANF Weight	GIS Weight	\
0	1.336	0.967	1.382	0.096	0.244	0.188	
1	1.377	0.997	1.381	0.162	0.191	0.006	
2	1.215	0.867	1.401	0.103	0.177	0.146	
3	1.526	1.092	1.398	0.135	0.233	0.059	
4	1.197	0.943	1.269	0.015	0.220	0.224	

	HRL Weight	K Weight	PIM Weight	SBUX Weight	VZ Weight
0	0.154	0.040	0.040	0.015	0.223
1	0.261	0.224	0.057	0.049	0.049
2	0.098	0.207	0.047	0.099	0.124
3	0.152	0.176	0.014	0.180	0.051
4	0.187	0.071	0.023	0.159	0.102

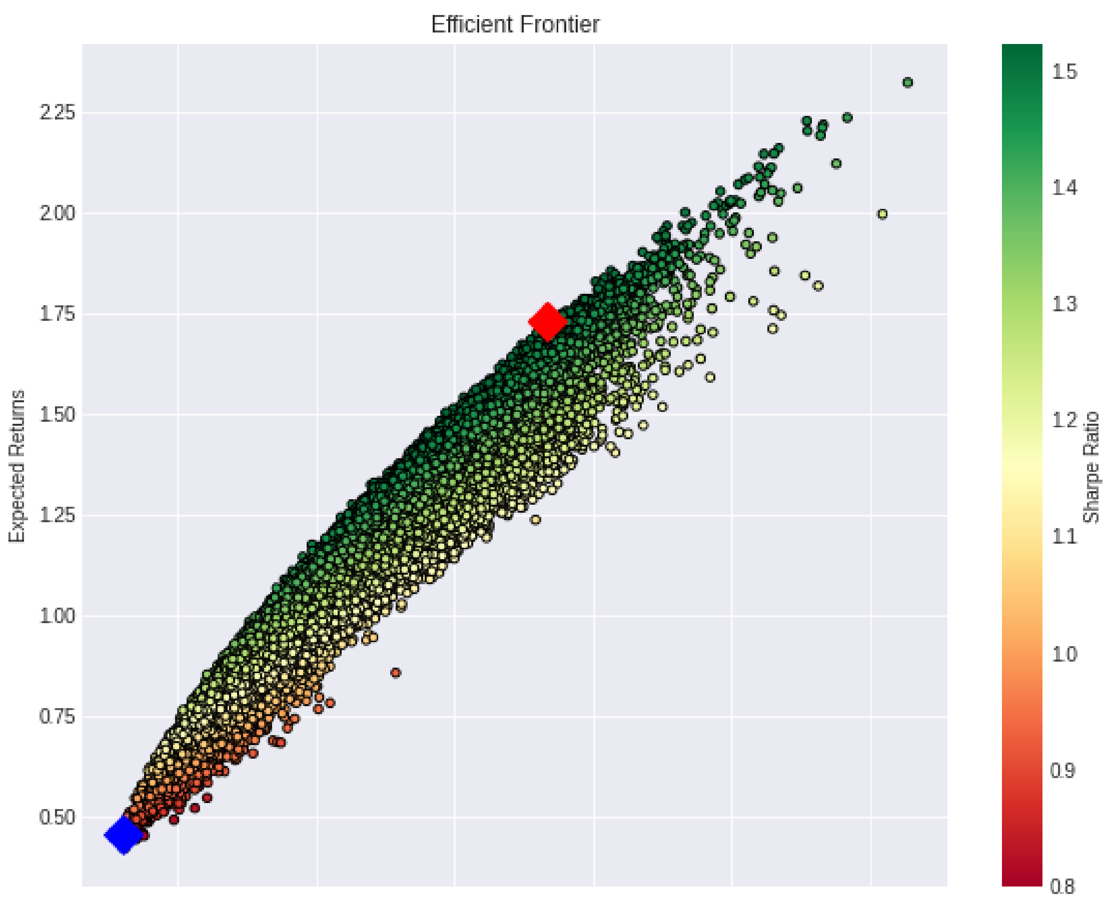

Figure 12: An efficient frontier showing the minimum variance(blue) portfolio and maximum sharpe ratio(red) portfolio.

The portfolios that lie below the efficient frontier are sub-optimal because they do no provide enough return for the level of risk.

```
Minimum variance portfolio:

Returns        0.453
Volatility     0.520
Sharpe Ratio   0.871
AAPL Weight    0.025
ANF Weight     0.004
GIS Weight     0.150
HRL Weight     0.094
K Weight       0.183
PIM Weight     0.267
SBUX Weight    0.009
VZ Weight      0.268


Maximum Sharpe ratio portfolio:

Returns        1.727
Volatility     1.135
Sharpe Ratio   1.523
AAPL Weight    0.314
ANF Weight     0.180
GIS Weight     0.299
HRL Weight     0.041
K Weight       0.022
PIM Weight     0.008
SBUX Weight    0.099
VZ Weight      0.037
```

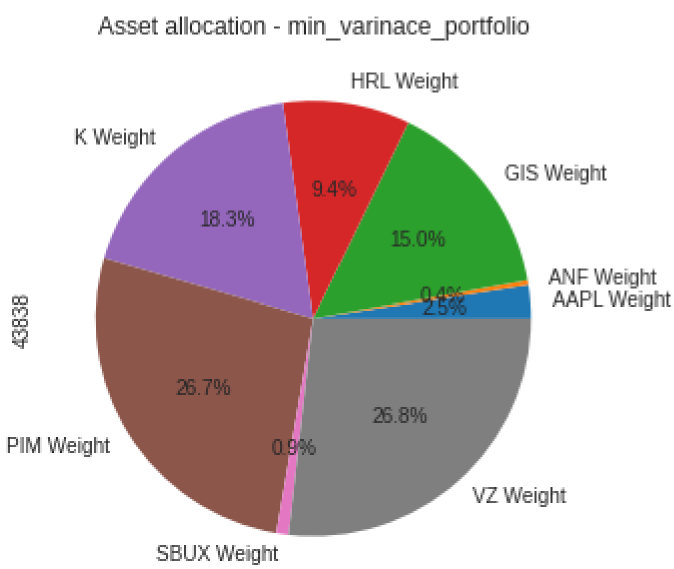

Figure 13: Pie chart showing the percentage weight allocation of minimum variance portfolio.

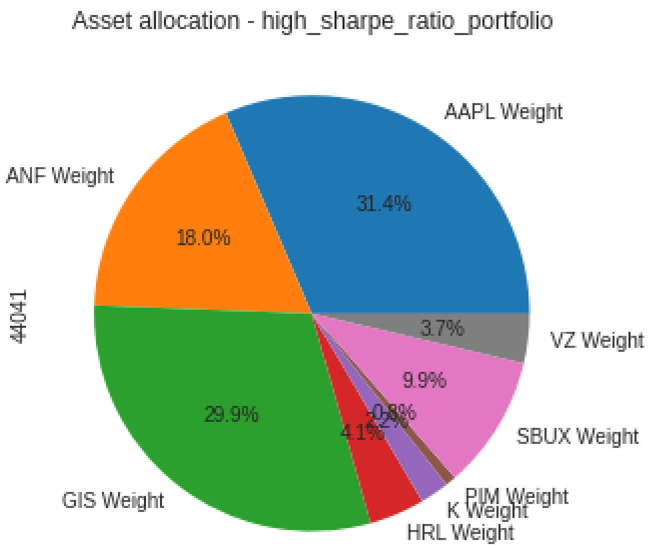

Figure 14: Pie chart showing the percentage weight allocation of max sharpe ratio portfolio.

F. What are the predicted daily returns for a week into the future?

For a forecasting model, an LSTM network is trained on a dataset containing the closing price of 8 different securities from the date '2000-03-01' to '2020-12-30'.

Table 6: Table representing a forecasted daily return to 7 consecutive days in the future from the training date.

Date	AAPL	ANF	GIS	HRL	K	PIM	SBUX	VZ
2020-12-30	-1.374e-04	1.161e-03	-1.708e-03	1.074e-03	4.378e-04	3.513e-03	-1.390e-03	-5.853e-04
2020-12-31	2.920e-03	3.492e-03	7.974e-04	2.120e-03	1.737e-03	-7.378e-04	-1.113e-03	8.299e-04
2021-01-01	7.775e-03	3.865e-03	1.485e-05	3.154e-03	3.056e-03	-3.659e-03	5.810e-04	2.044e-03
2021-01-02	1.135e-02	-6.962e-04	-2.267e-03	3.518e-04	1.802e-04	-2.796e-03	2.164e-03	-1.974e-03
2021-01-03	1.169e-02	2.579e-03	-7.421e-04	2.545e-03	2.605e-03	3.985e-03	2.263e-03	4.161e-04
2021-01-04	1.041e-02	2.074e-03	-2.565e-03	-2.926e-04	2.246e-04	3.286e-03	1.598e-03	-1.415e-04
2021-01-05	4.897e-03	2.718e-03	-3.492e-04	1.072e-03	3.732e-04	2.290e-02	4.537e-03	1.636e-03

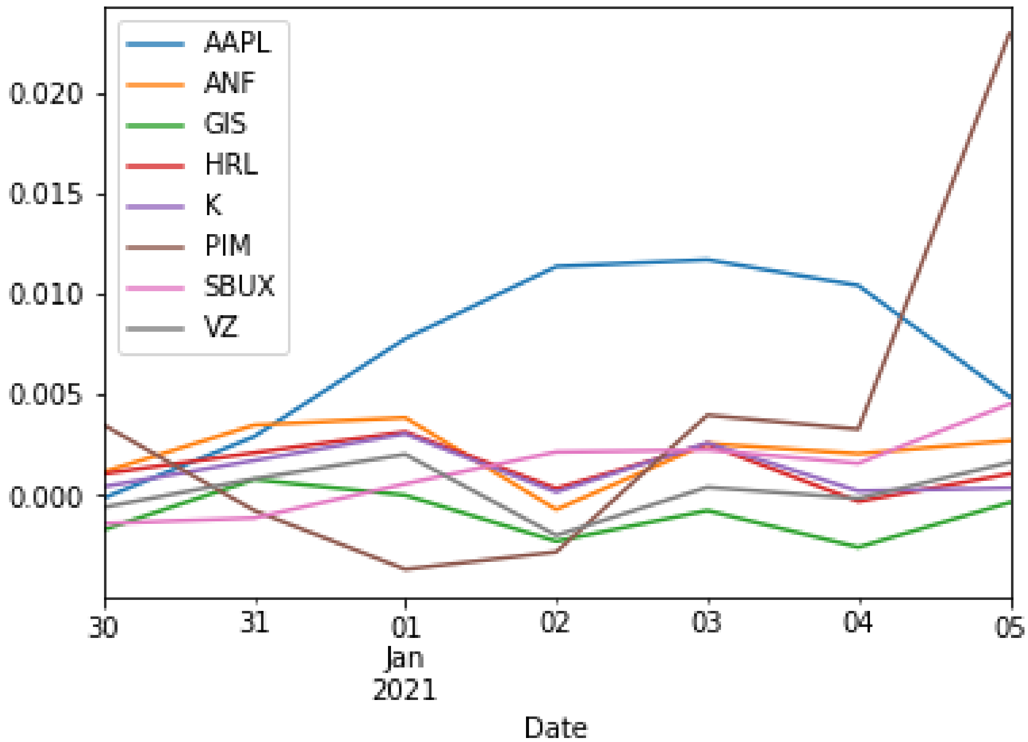

Figure 15: Line plot of the forecasted returns.

G. What is the predicted covariance for a week?

To forecast covariance, we need forecasted volatility and correlation of given securities. The correlation matrix can be calculated from the forecasted return but volatility needs to be predicted using past data. Therefore a GARCH model is used to forecast the volatility.

Table 7: A table of daily forecasted volatility using the GARCH model.

Date	AAPL	ANF	GIS	HRL	K	PIM	SBUX	VZ
2020-12-30	4.123e-04	0.001	1.145e-04	1.279e-04	1.190e-04	3.901e-05	2.683e-04	8.083e-05
2020-12-31	3.661e-04	0.001	1.156e-04	1.341e-04	1.192e-04	3.858e-05	2.600e-04	8.370e-05
2021-01-01	3.727e-04	0.001	1.172e-04	1.342e-04	1.221e-04	3.974e-05	2.625e-04	8.690e-05
2021-01-02	3.913e-04	0.001	1.167e-04	1.388e-04	1.210e-04	4.145e-05	2.706e-04	8.485e-05
2021-01-03	3.914e-04	0.001	1.165e-04	1.413e-04	1.212e-04	4.162e-05	2.703e-04	8.673e-05
2021-01-04	3.920e-04	0.001	1.169e-04	1.436e-04	1.221e-04	4.150e-05	2.705e-04	8.885e-05
2021-01-05	3.901e-04	0.001	1.173e-04	1.464e-04	1.226e-04	4.201e-05	2.711e-04	9.023e-05

Table 8: Table representing the average daily covariance of the securities for next seven days.

	AAPL	ANF	GIS	HRL	K	PIM \
AAPL	1.000e+00	-1.178e-07	-1.655e-08	-5.609e-09	7.364e-09	-3.512e-09
ANF	-1.178e-07	1.000e+00	1.316e-07	1.346e-07	1.196e-07	7.819e-09
GIS	-1.655e-08	1.316e-07	1.000e+00	1.286e-08	9.297e-09	4.298e-10
HRL	-5.609e-09	1.346e-07	1.286e-08	1.000e+00	1.572e-08	-1.227e-09
K	7.364e-09	1.196e-07	9.297e-09	1.572e-08	1.000e+00	-1.815e-09
PIM	-3.512e-09	7.819e-09	4.298e-10	-1.227e-09	-1.815e-09	1.000e+00
SBUX	5.659e-08	-4.082e-08	-4.740e-09	-8.398e-09	-6.522e-09	7.107e-09
VZ	-7.252e-09	1.161e-07	7.641e-09	7.691e-09	6.218e-09	1.209e-09

	SBUX	VZ
AAPL	5.659e-08	-7.252e-09
ANF	-4.082e-08	1.161e-07
GIS	-4.740e-09	7.641e-09
HRL	-8.398e-09	7.691e-09
K	-6.522e-09	6.218e-09
PIM	7.107e-09	1.209e-09
SBUX	1.000e+00	2.966e-09
VZ	2.966e-09	1.000e+00

The code snippet to generate above matrix is mentioned below in appendix.

H. Predict the portfolio for a week that provides the minimum variance and which portfolio provides the maximum sharpe ratio.

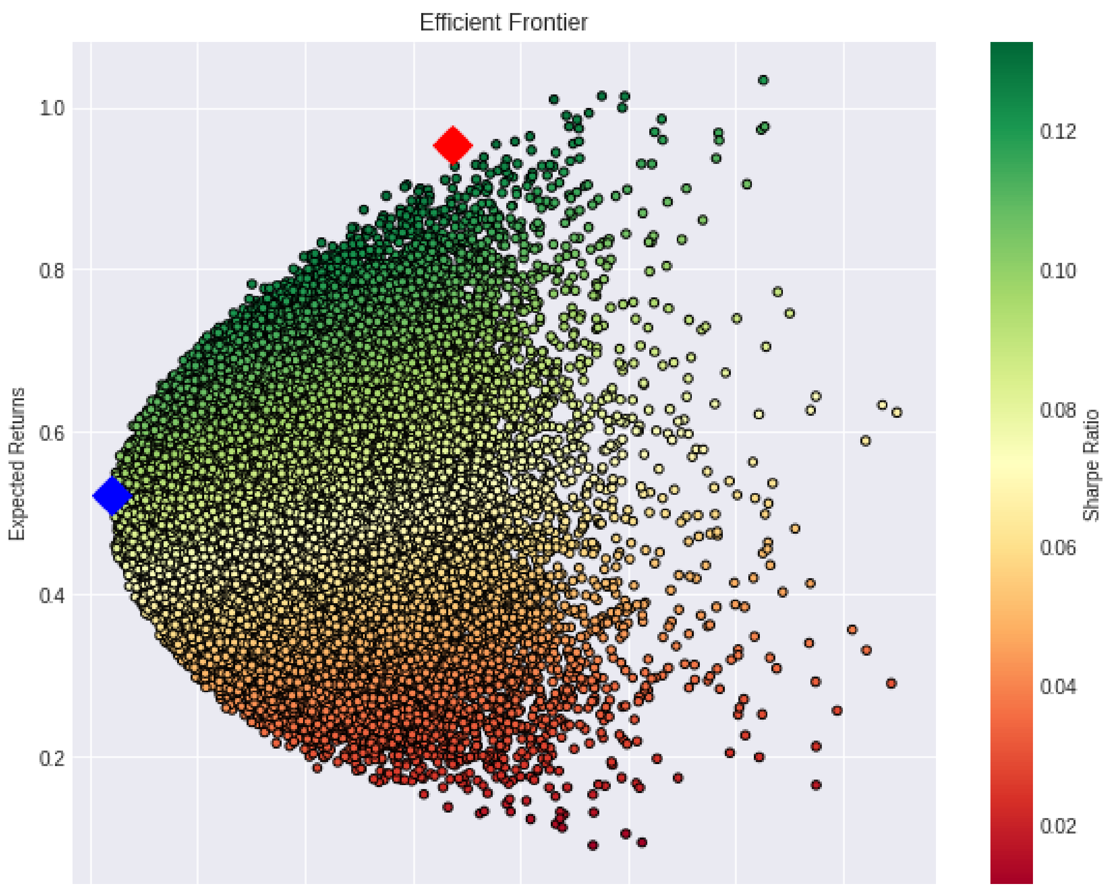

Figure 16: An efficient frontier showing the minimum variance and maximum sharpe ratio portfolios.

```
Minimum variance portfolio:

Returns         0.524
Volatility      5.601
Sharpe Ratio    0.093
AAPL Weight     0.136
ANF Weight      0.121
GIS Weight      0.114
HRL Weight      0.115
K Weight        0.126
PIM Weight      0.123
SBUX Weight     0.132
VZ Weight       0.133
```

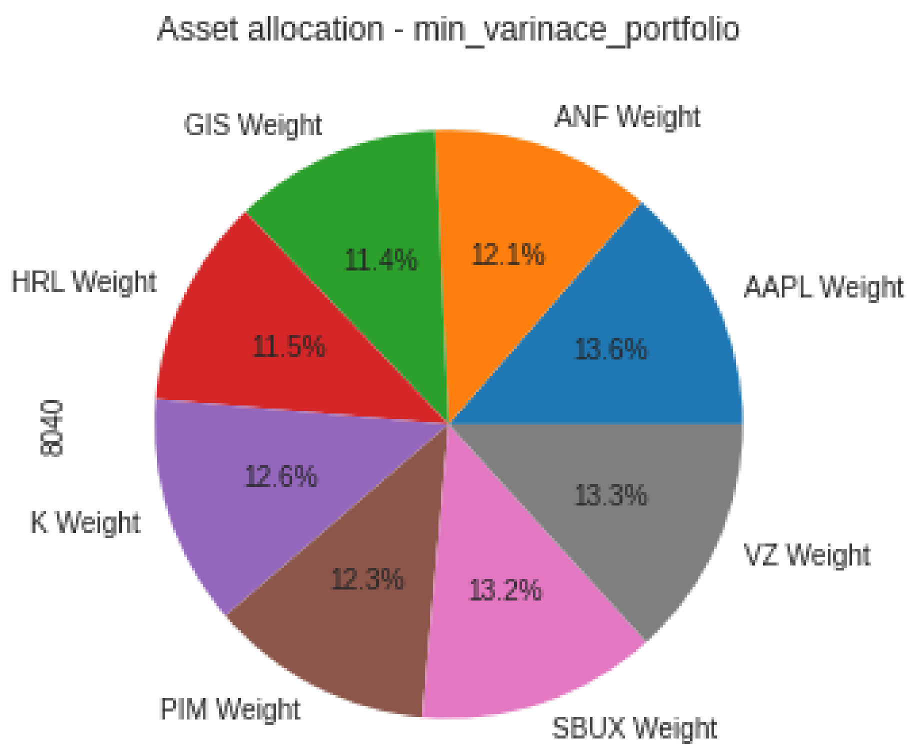

Figure 17: Pie chart showing the weight allocation for forecasted portfolio having minimum variance.

```
Maximum Sharpe ratio portfolio:

Returns         9.537e-01
Volatility      7.183e+00
Sharpe Ratio    1.328e-01
AAPL Weight     3.230e-01
ANF Weight      1.263e-01
GIS Weight      8.261e-04
HRL Weight      1.152e-01
K Weight        7.777e-02
PIM Weight      2.337e-01
SBUX Weight     1.096e-01
VZ Weight       1.362e-02
```

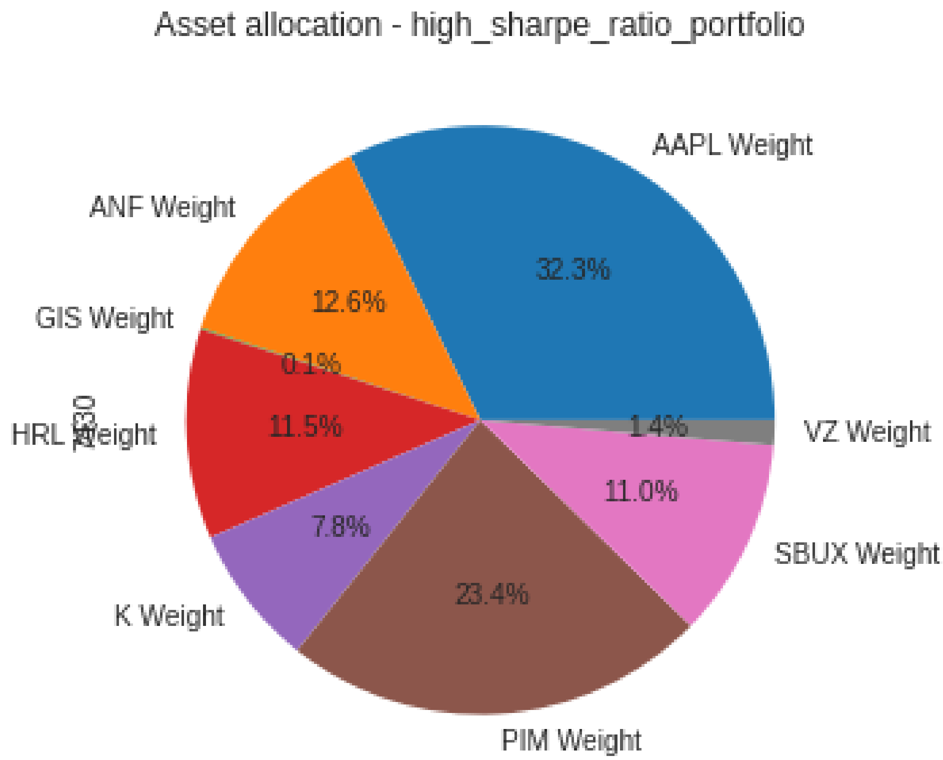

Figure 18: Pie chart showing the weight allocation of the maximum sharpe ratio forecasted portfolio.

CONCLUSIONS AND DISCUSSIONS

The descriptive and predictive analysis of the assets allocation task gives us the base to invest properly and gain risk-free returns. From the above analysis, we have weights percentages based on historical returns and covariance as well as predictive weight assignment for risk-free returns and minimum variance portfolios.

Future work might include proper testing of the LSTM model and tuning hyperparameters to generate better predictions, backtracking in order to verify the results, although LSTM handles the seasonality pretty well one can remove the seasonality, transform the data, or take the log return instead of simply calculating the percentage change in order to improve the prediction result.

APPENDIX

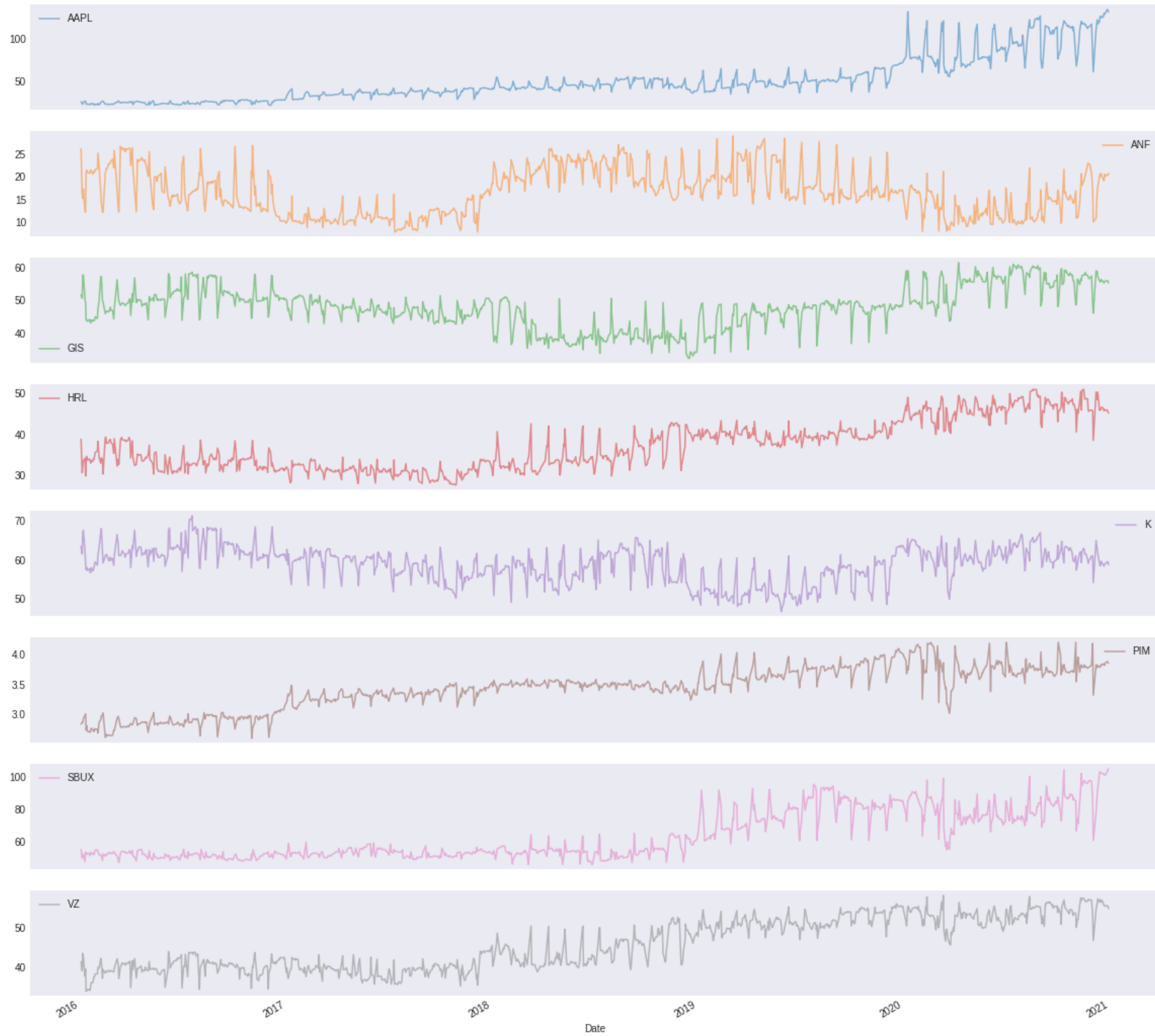

Figure 19: Line plot showing the daily close price.

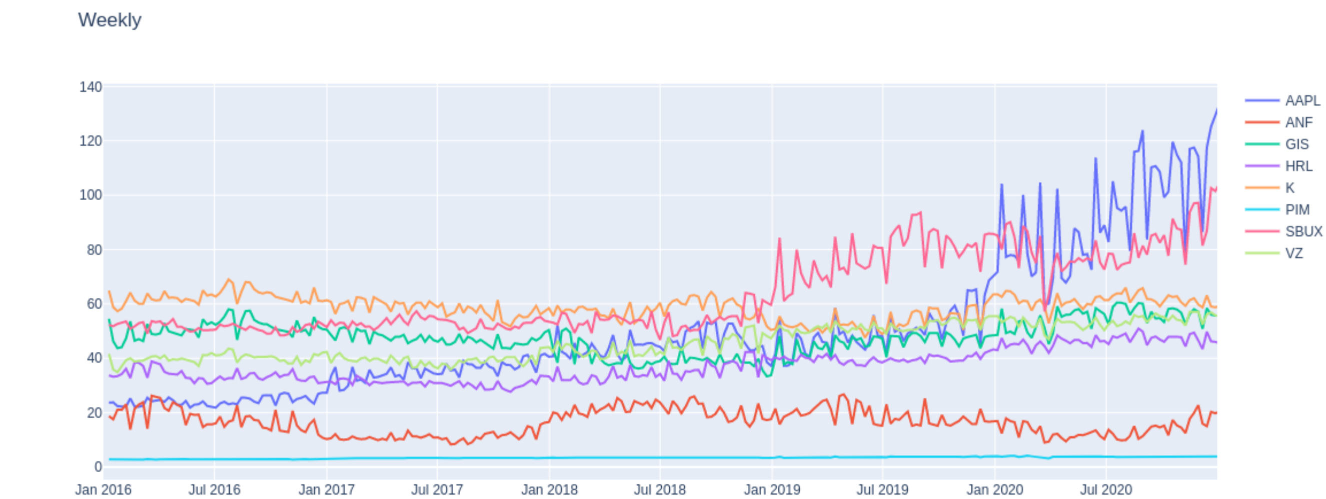

Figure 20: Line plot showing the weekly close price.

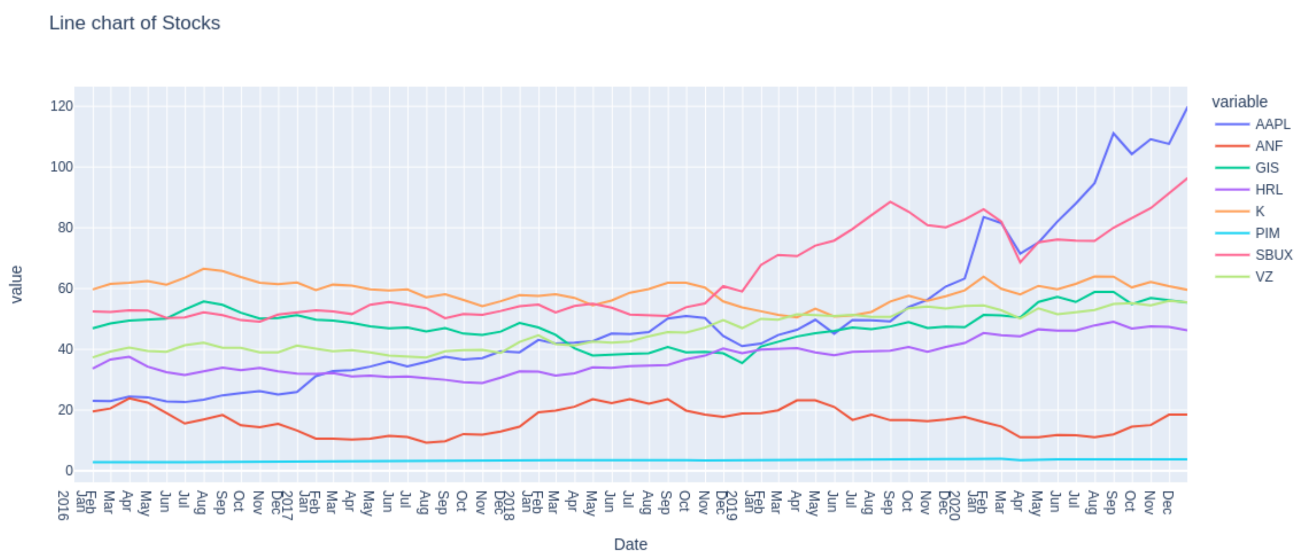

Figure 21: Line plot showing the monthly close price.

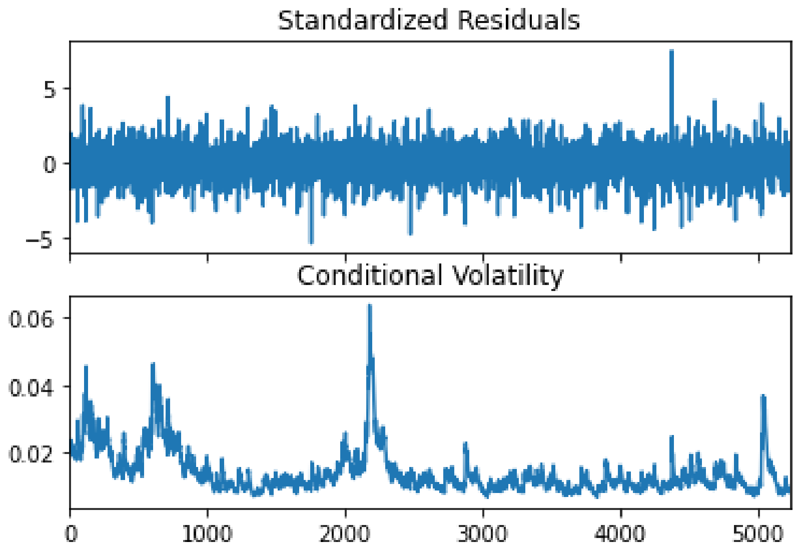

Figure 22: GARCH model result's residuals and volatility.

Code to find covariance

```
# find correlation
corr = forecast_df.corr().to_numpy()

# Get the upper triangle elements
corr1 = corr[np.triu_indices(len(tickers), k=1)]

# Find the recursive product of returns in order to calculate covariance i.e eg:
ret of AAPL * ret of ANF and so on
mul = []
temp = gm_forecast_df.reset_index().drop('Date', axis=1)
for i in range(len(gm_forecast_df)):
  for j in range(len(gm_forecast_df)):
    if (j+i+1) <= len(gm_forecast_df):
      mul.append(temp.iloc[:,i].mean()*temp.iloc[:,j+i+1].mean())

# Find the upper triangle part of covariance matrix
covarr = [corr1[i] * mul[i] for i in range(len(mul))]

# Replicate the upper triangle to lower as well
cov_mat = forecast_df.corr()

cov_mat1 = cov_mat.to_numpy()
# i_lower = np.tril_indices(n, -1)
cov_mat1[np.triu_indices(len(tickers), k=1)] = covarr
cov_mat1[np.tril_indices(len(tickers), -1)] =
cov_mat1.T[np.tril_indices(len(tickers), -1)]

covariance_df = pd.DataFrame(cov_mat1)
```

REFERENCES

[1]. Investopedia

[2]. Kaggle datasets

[3]. Minimum Variance portfolio

[4]. Derek Snow (2020).Machine Learning in Asset Management- Part1: Portfolio Construction-Trading Strategies.The Journal of Financial Data Science Winter 2020, 2 (1) 10-23;

[5]. Thomas Raffinot (2018).Hierarchial Clustering-Based Asset Allocation.The Journal of Portfolio Management Multi-Asset Special Issue 2018, 44 (2) 89-99;

[6]. Portfolio Optimization with Python using Efficient Frontier with Practical Examples

[7]. Introduction to ARCH models