BCI Quantitative Equity Take-Home Assignment
Xiaoxin(Mark) Zhou

Introduction: In this report, we analyze ten different stock tickers from Yahoo Finance over the period from January 1, 2012, to December 31, 2021. We examine the distribution type of the stocks, assess the performance of ordinary least squares regression for predicting stock prices, compute the correlations of daily returns, perform K-Means clustering to analyze stock behaviour during the pandemic period and optimize the global minimum variance portfolio. We address each question systematically and use appropriate models to validate our assumptions.

Stocks in the energy sector:
COP (ConocoPhillips), CVX (Chevron), HES (Hess Corporation), OXY (Occidental Petroleum), and XOM (Exxon Mobil)

Stocks in the Tech sector:
AAPL (Apple), AMZN (Amazon), GOOG (Google/Alphabet), INTC (Intel), ORCL (Oracle)

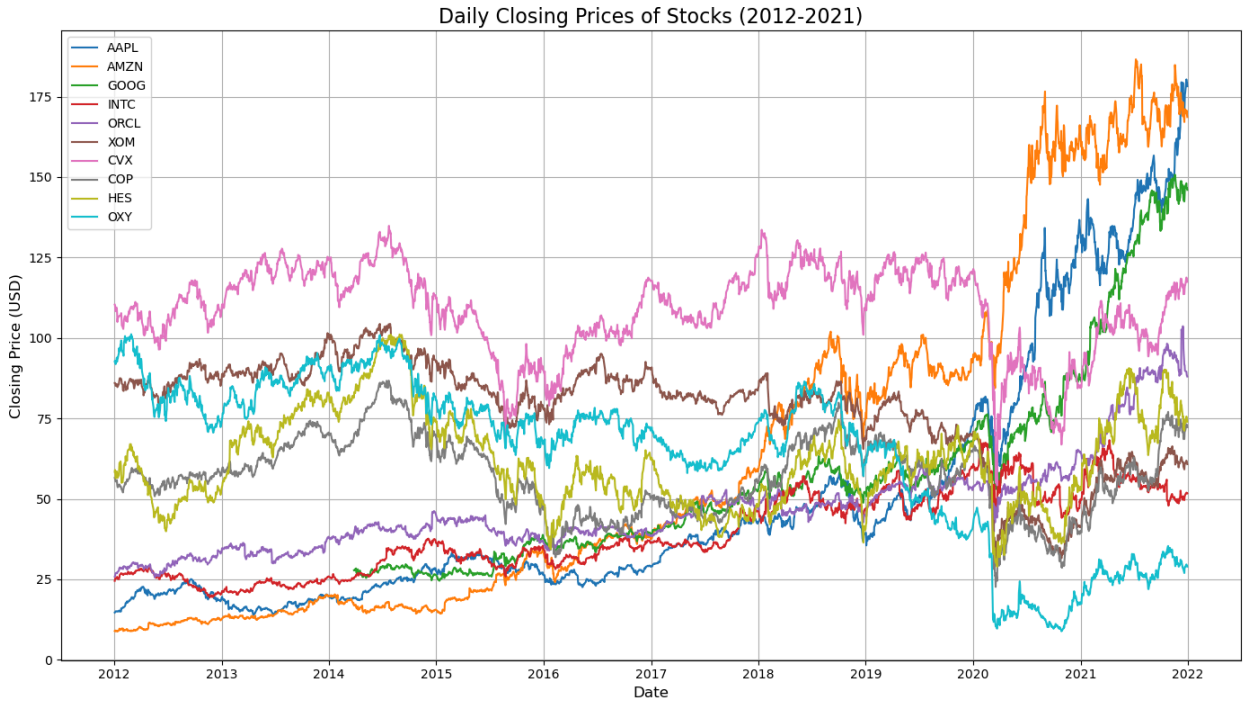

Figure 1 To visualize the daily closing prices of all the stocks over the given period from January 1, 2012, to December 31, 2021.

Question 1: Generate daily returns using the Close prices over this period for AAPL. Are the daily returns normally distributed? Provide some evidence to support your answer.

Answer: The daily returns for AAPL are not normally distributed. I will use four methods to support my solution.

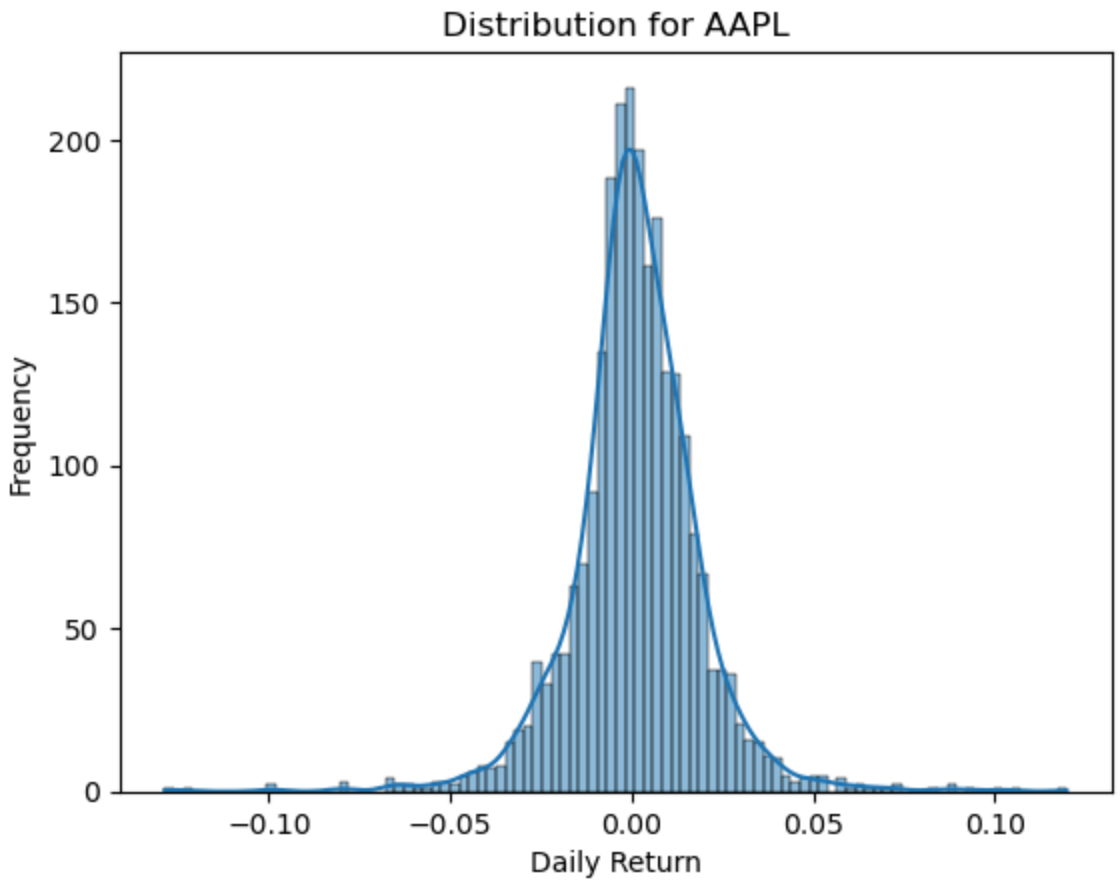

Figure 2 To visualize the daily return distribution over the given period from January 1, 2012, to December 31, 2021.

Kolmogorov-Smirnov statistic: 0.08015703010096012
p-value: 1.6717216462456285e^-14

The K-S statistic shows some deviation between the empirical distribution of AAPL's standardized daily returns and the normal distribution. The p-value is lower than 0.05, indicating that this deviation is statistically significant, meaning that the returns do not follow a normal distribution.

Figure 3 will help visualize the difference between the empirical and theoretical CDFs, providing a better understanding of the K-S statistic. We can see that most of the blue curve fits with the red curve. However, in this context, a K-S statistic of 0.0802 means that, at most, there is an 8.02% difference between the cumulative proportions of the actual returns and what would be expected under a normal distribution.

We could visually verify that the daily returns for AAPL are not in a normal distribution curve in Figure 4.

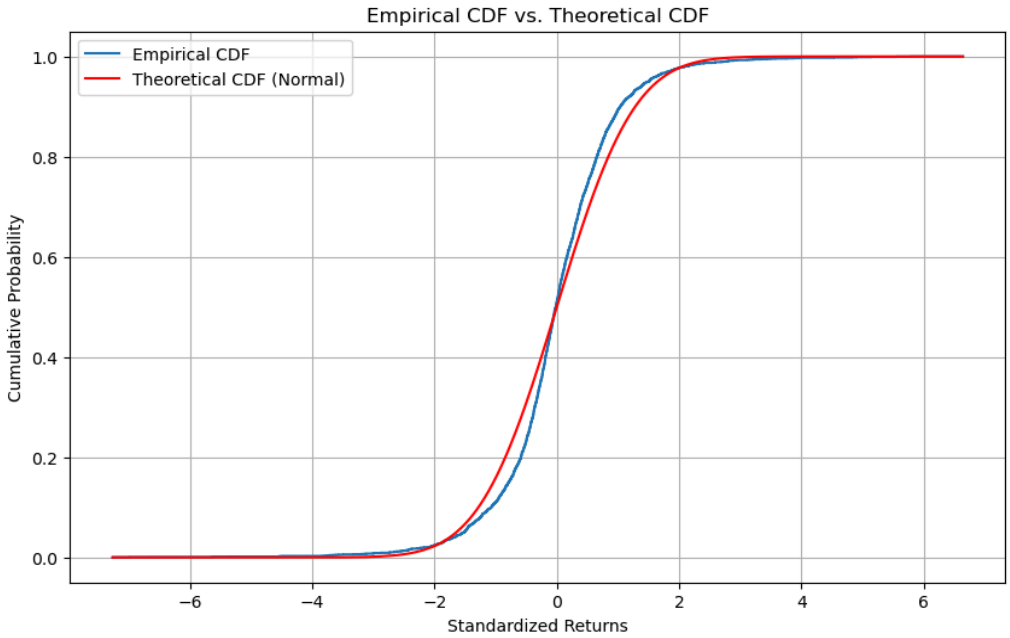

Figure 3 To visualize the empirical CDF VS theoretical CDF. Empirical CDF represent the step function that represents the cumulative probability of the sample data and theoretical CDF represents the smooth curve representing the cumulative probability of the normal distribution.

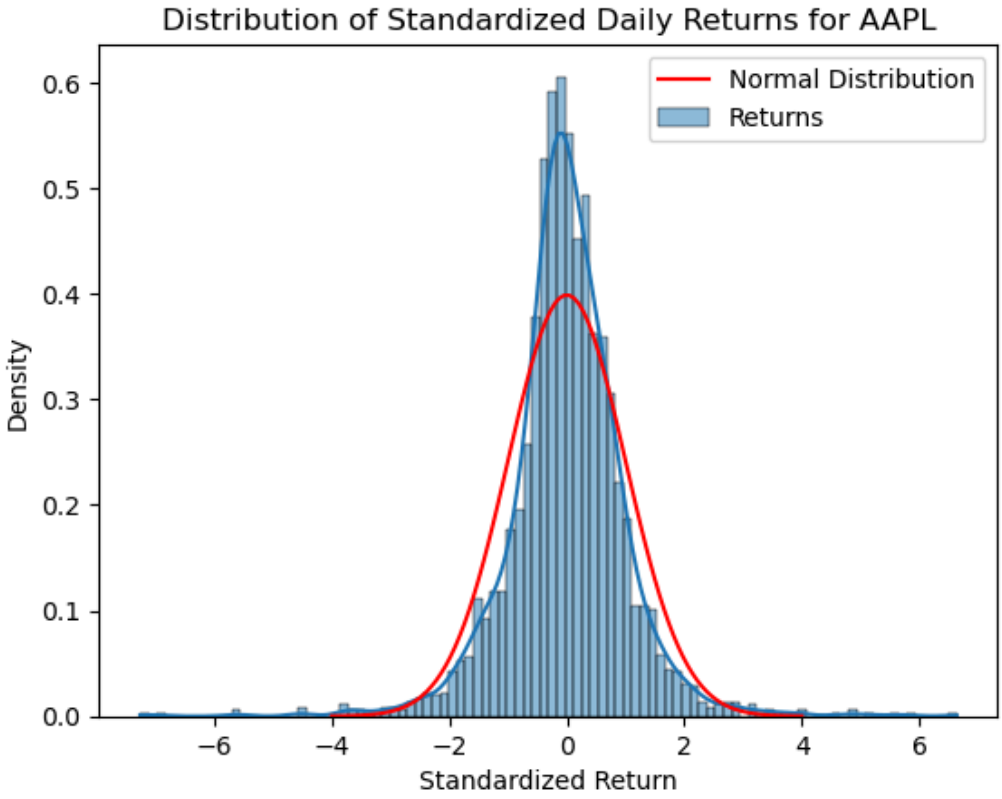

Figure 4 To visualize the daily return distribution VS normal distribution over the given period from January 1, 2012, to December 31, 2021.

We also implemented the QQ plot to support our solution. In Figure 5, the left tail shows values lower than expected, and the right tail shows values higher than expected. This suggests the presence of heavy tails, meaning there are more extreme returns than what a normal distribution would predict.

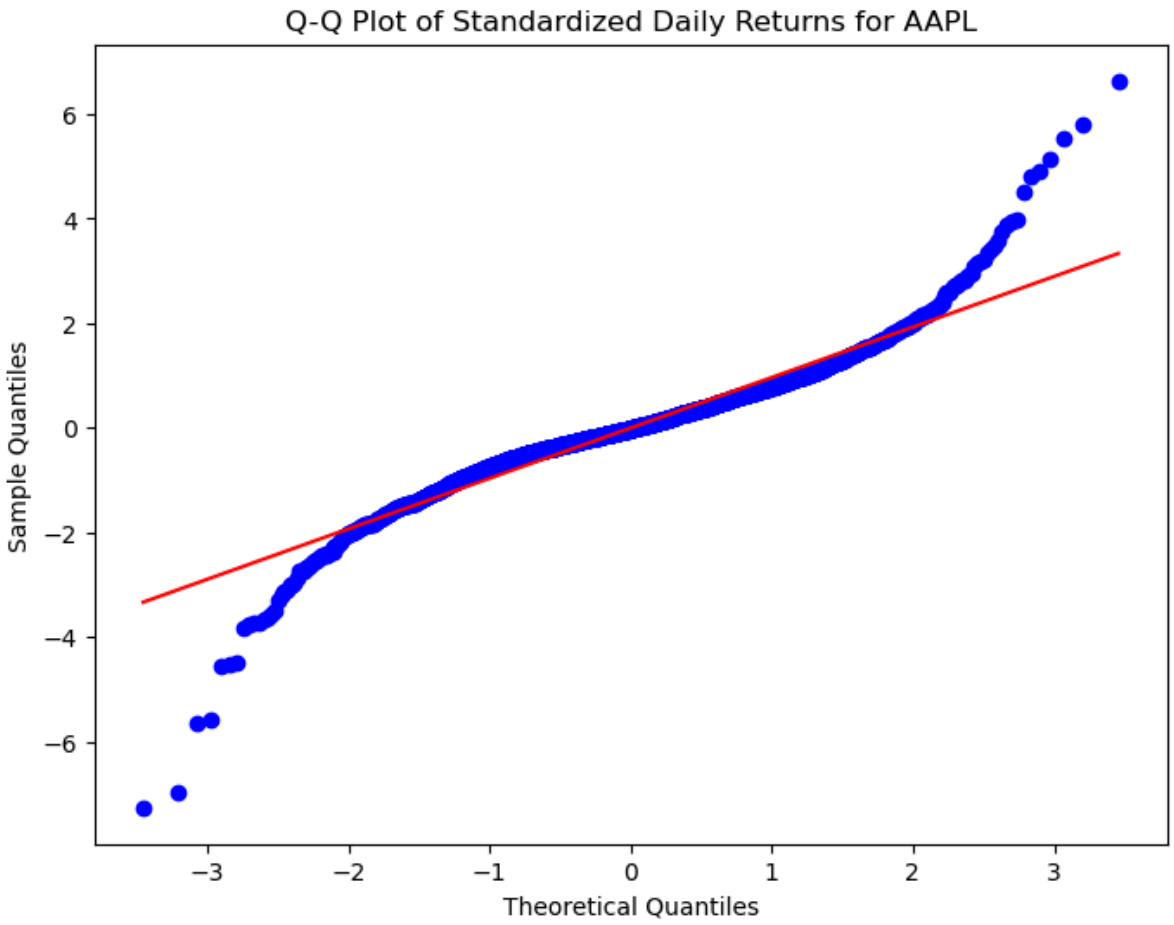

Figure 5 To visualize the daily return distribution VS normal distribution over the given period from January 1, 2012, to December 31, 2021 using Q-Q plot.

Question 2: A colleague is interested in creating a model to predict the future price of AAPL stock using ordinary least squares to regress the price on day t on the price at day t-1 and a constant:
Pt = a + b pt-1 + et

Is this a suitable model to predict the future price of AAPL? Do you think OLS is appropriate for estimating the parameters a and b? If so, estimate the parameters a and b using your data. If not, please explain why it is inappropriate.

Answer: Intercept$a$=−0.0135, Slope$b$=1.0016

OLS Regression Results

Dep. Variable:	AAPL	R-squared:	0.999
Model:	OLS	Adj. R-squared:	0.999
Method:	Least Squares	F-statistic:	2.685e+06
Date:	Wed, 17 Jul 2024	Prob (F-statistic):	0.00
Time:	14:47:18	Log-Likelihood:	-3988.9
No. Observations:	2515	AIC:	7982.
Df Residuals:	2513	BIC:	7993.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0135	0.038	-0.354	0.723	-0.088	0.061
AAPL	1.0016	0.001	1638.579	0.000	1.000	1.003

Omnibus:	579.904	Durbin-Watson:	2.186
Prob(Omnibus):	0.000	Jarque-Bera (JB):	21201.667
Skew:	-0.317	Prob(JB):	0.00
Kurtosis:	17.210	Cond. No.	101.

Table 1 The OLS regression results for predicting AAPL stock prices based on the previous day's price.

From Table 1, the OLS model is very effective at predicting AAPL prices based on the previous day's prices. The high R-squared value of 0.999 means the model explains 99.9% of the changes in AAPL prices based on the previous day's price.

The Omnibus and Jarque-Bera tests show that the residuals (errors) are not normally distributed, as we noted in Question 1. The Durbin-Watson statistic is 2.186, which suggests there is no significant autocorrelation (the residuals are not related to each other). The high kurtosis value of 17.210 indicates that the residuals have heavy tails, as we saw in Figure 5. However, since the residuals are not normally distributed ($p < 0.05$), this could affect the accuracy of the confidence intervals and other statistical inferences.

Suitability of the Model:
The high R-squared value shows that the model is very good at explaining the changes in AAPL prices based on the previous day's prices. However, the non-normality of the residuals suggests there might be patterns or outliers that the model doesn't capture. This could affect the accuracy of the predictions and the reliability of the statistical conclusions.

Appropriateness of OLS:
Even though the model explains a lot of the variability, the non-normality of the residuals and high kurtosis suggest there might be problems with the assumptions of OLS. The heavy tails

indicate that extreme values happen more often than they would in a normal distribution, which could affect the stability and reliability of the model's estimates.

In summary, while OLS is appropriate and effective in this case, we should be careful because the non-normality of residuals might affect the confidence intervals and predictions.

Question 3: Compute the correlations of daily returns for the 10 stocks over the period and comment on your findings.

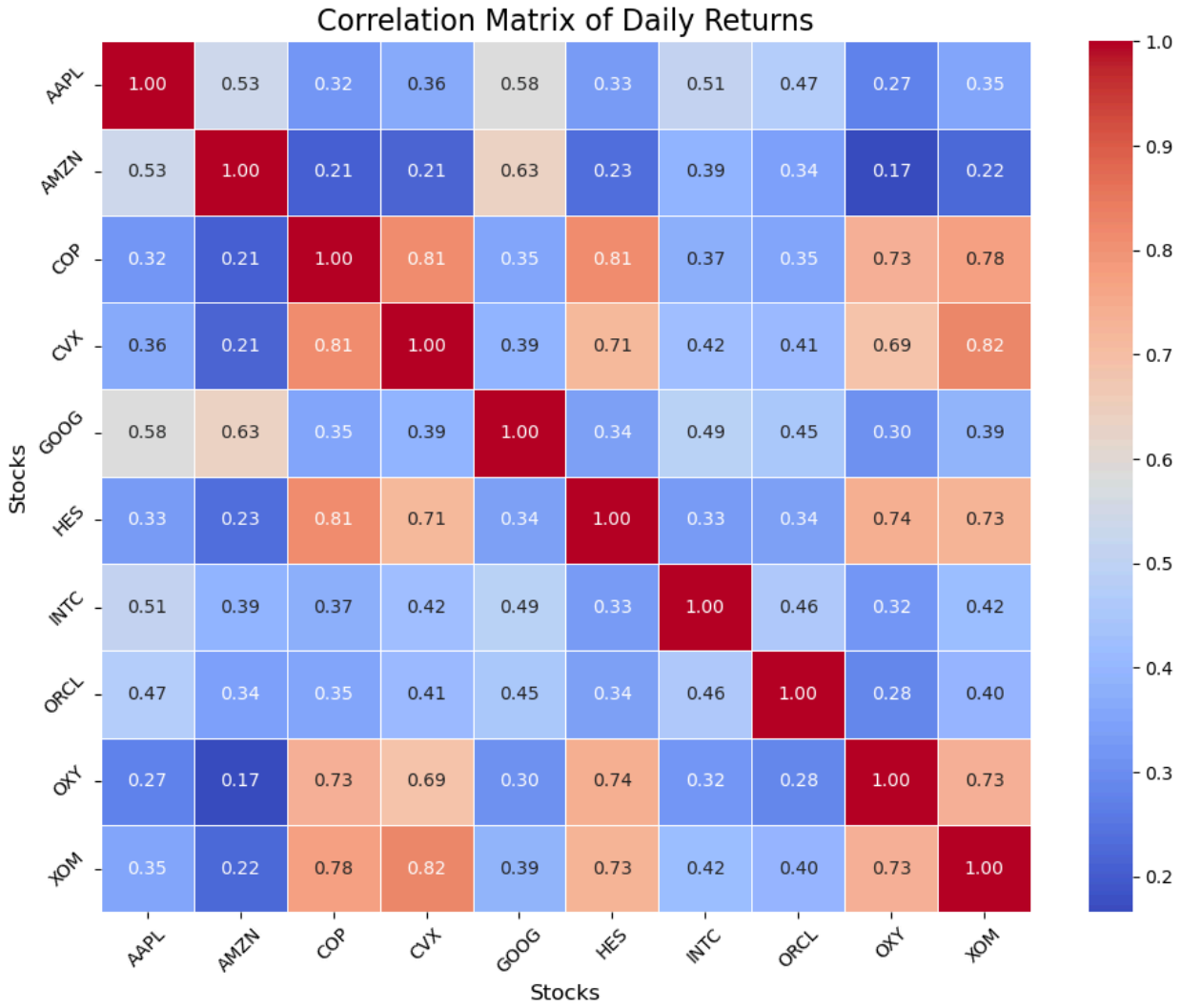

Figure 6 The correlation matrix for the daily returns of the 10 stocks over the period from January 1, 2012, to December 31, 2021, is shown in the heatmap.

Strong Positive Correlations: COP (ConocoPhillips), CVX (Chevron), HES (Hess Corporation), OXY (Occidental Petroleum), and XOM (Exxon Mobil), exhibit strong positive correlations with each other. For example, COP and CVX have a correlation of 0.81, while CVX and XOM have a correlation of 0.82. This high correlation is expected since these companies are influenced by similar market factors, such as oil prices. AAPL and GOOG have a correlation of 0.58, and

AAPL and XOM have a correlation of 0.35. This indicates that while there is some co-movement, we can find online resources that show ExxonMobil is Apple Card's exclusive fuel provider, giving Apple Card users 3% Daily Cash back on fuel (https://www.exxon.com/-/media/project/wep/exxon/exxon-retail-us/updated-pdfs/exxonmobil-apple-card-092022-qa.pdf). Perhaps it's something like a payment binding that connects the two companies.

Question 4: A colleague of yours is curious whether the 10 stocks exhibited similar behaviour during the pandemic. You think that the performance was different across two groups of stocks. Generate quarterly returns over this period for all 10 stocks. Perform a K-means clustering with two groups using quarterly returns from Q2 2020 to Q4 2021 (7 features for each stock) for each of the 10 stocks. Comment on your findings.

Answer: Cluster 0 AAPL, AMZN, GOOG, INTC, ORCL, XOM, CVX, COP, HES
Cluster 1 OXY

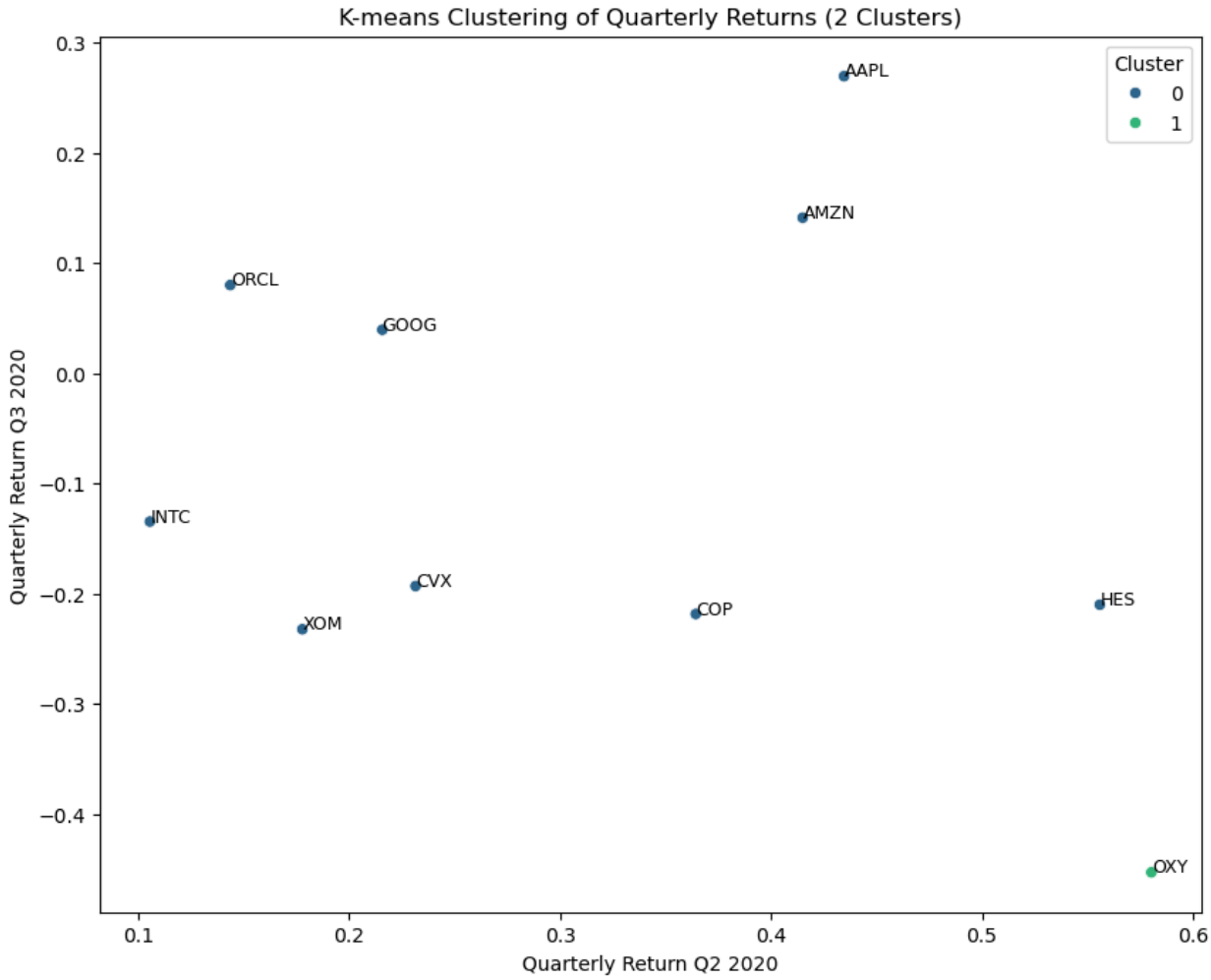

Figure 7 The scatter plot above visualizes the clustering results for the quarterly returns from Q2 2020 to Q4 2021. Each point represents a stock, and the color indicates its cluster.

The result of Figure 7 is reasonable. In Figure 1, we can see that OXY performed poorly compared to the other stocks. Before COVID-19, OXY's performance had already been declining, and the arrival of the pandemic made the stock more sensitive to external influences, leading to a greater decline than the other stocks. The reasons for the increases in AAPL and AMZN are more traceable: more people staying at home led to increased spending on games and videos, boosting Apple's revenue(https://www.investopedia.com/how-covid-affects-apple-aapl-5198334). Meanwhile, Amazon benefited from providing convenient and fast online shopping services, allowing people to purchase products without leaving their homes.

Question 5: You are deciding how to weight the 10 stocks in creating a suitable portfolio. Calculate the set of optimal portfolio weights across the 10 stocks such that you minimize the variance of portfolio returns. Comment on the optimal portfolio weights and potential issues.
Given

1. Q is the daily return covariance matrix across the 10 stocks in 2021
2. X is the vector of portfolio weights.

Global Minimum Variance Portfolio (GMVP) is the solution to the following optimization problem

$$arg \min_{x} x^T Qx$$

Subject to the following constraints:
1. All portfolio weights are positive
2. “Budget constraint” that portfolio weights sum to 1

Answer:

Optimal Portfolio Weights:
Stock AAPL: 0.0720
Stock AMZN: 0.1828
Stock COP: 0.1407
Stock CVX: 0.1151
Stock GOOG: 0.0300
Stock HES: 0.0300
Stock INTC: 0.0112
Stock ORCL: 0.1665
Stock OXY: 0.1156
Stock XOM: 0.1361

Explanation:
AMZN (18.28%) and ORCL (16.65%) have the highest weights. This suggests that these stocks have a favorable risk-return profile based on the covariance matrix. They contribute positively to minimizing the overall portfolio variance. COP (14.07%) and XOM (13.61%) also have significant weights.

INTC (1.12%) has the lowest weight, indicating that it contributes more to the portfolio's risk relative to its return or has less favorable covariance with other stocks. GOOG (3.00%) and HES (3.00%) also have relatively low weights, suggesting that they play a smaller role in minimizing portfolio variance under the given conditions.

Potential Issues:

The optimization is based on historical data from 2021. If market conditions change, these weights may no longer be optimal. It is important to regularly update the covariance matrix and re-optimize the portfolio. The financial markets are dynamic, and the relationships between stocks can change. Regular rebalancing and adjustments are necessary to maintain an optimal portfolio.