Business Intelligence per i Servizi Finanziari

Alessandro Zanotti – 885892

A.A. 2023/2024

# Sommario

# Sommario dei dati utilizzati:

Breve descrizione di ciascun titolo e motivazione della scelta:

* Settore tecnologico:
  + Broadcom, Inc (AVGO): leader tecnologico nel settore microchip e nel settore software. Recentemente ha acquisito VMware Inc per 68 miliardi.
  + Advanced Micro Devices, Inc (AMD): AMD sviluppa processor per PC e tecnologie collegate per consumatori e imprese. Scelto perché nel 2024 aveva un YTD return del 16.85% grazie all’aumento della richiesta per soluzioni computazionali ad alte performance
* Settore sanitario:
  + Eli Lilly (LLY): azienda farmaceutica globale che sviluppa, produce e commercia nei settori endocrinologi, oncologici, immunologici. Scelta per la serie di farmaci, specialmente per diabete e pazienti oncologici, e recenti trial clinici di successo.
  + Pfizer (PFE): azienda biofarmaceutica globale che sviluppa farmaci innovativi e vaccini per diverse aree terapeutiche come immunologia, cardiologia, endocrinologia e neurologia. Scelta per il suo contributo durante la pandemia di Covid-19 del 2020 grazie al vaccino Pfizer-BioNTech.
* Settore finanziario:
  + Berkshire Hathaway Inc. Class B (BRK.B): conglomerato multinazionale amministrato da Warren Buffet. L’azienda possiede un business molto diversificato che include: assicurazioni, servizi, ferrovie, manifatturiero e possiede posizioni di equity importanti in Apple, Bank of America e Coca-Cola. Scelto per lo storico record di crescita e il portafoglio altamente diversificato le cui componenti sono scelte oculatamente.
  + Blackrock Inc. (BLK): è l’asset manager piu grande al mondo e fornisce servizi di management degli investimenti, dei rischi e servizi di advisory. Scelta per la posizione dominante nel mercato e per il suo approccio innovativo alle soluzioni di investimento.

# Funzioni utilizzate per caricare i dati finanziari

Per scaricare i dati ho utilizzato la funzione **download** della libreria **yfinance** (Yahoo! Finance).

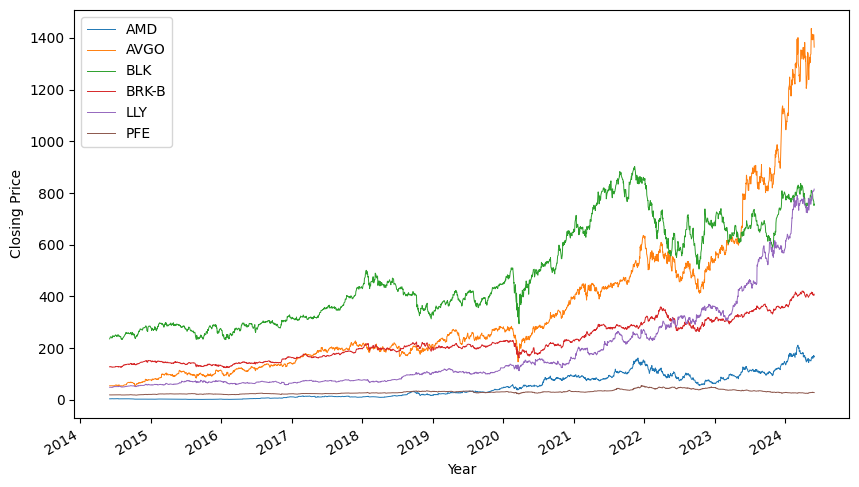
L’interfaccia della funzione è yf.download(tickers, ...) -> pd.DataFrame e possiede diversi parametri:

|  |  |  |  |
| --- | --- | --- | --- |
| Tickers | Interval | Period | Start |
| End | Prepost | Actions | Auto\_adjust |
| Back\_adjust | Repair | Keepna | Rounding |
| Group\_by | Threads | Proxy | Session |
| Timeout | Progress |  |  |

I dati Fama-Franch vengono scaricati tramite la funzione **request.urlretrieve** di **urllib** dalla Data Library di Ken French (Dartmouth University).

# Presentazione dei dati con un grafico e le prime righe del DataFrame

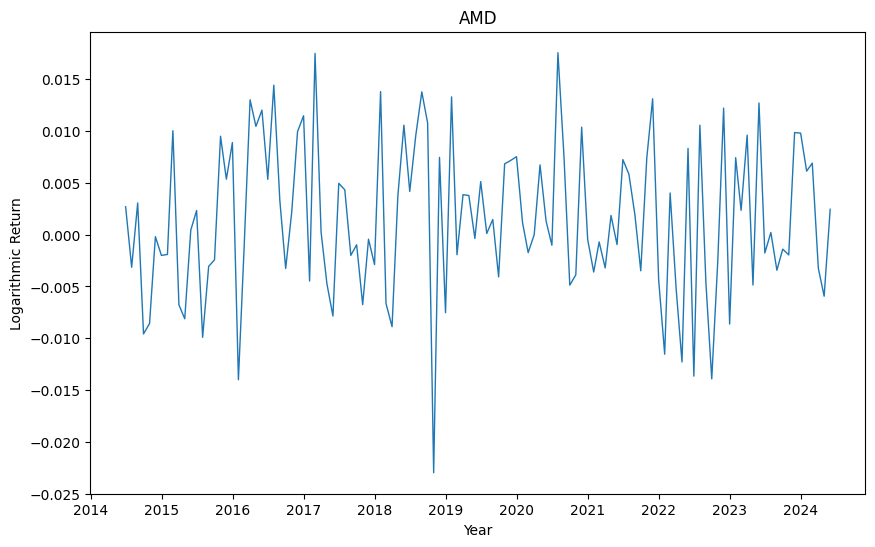


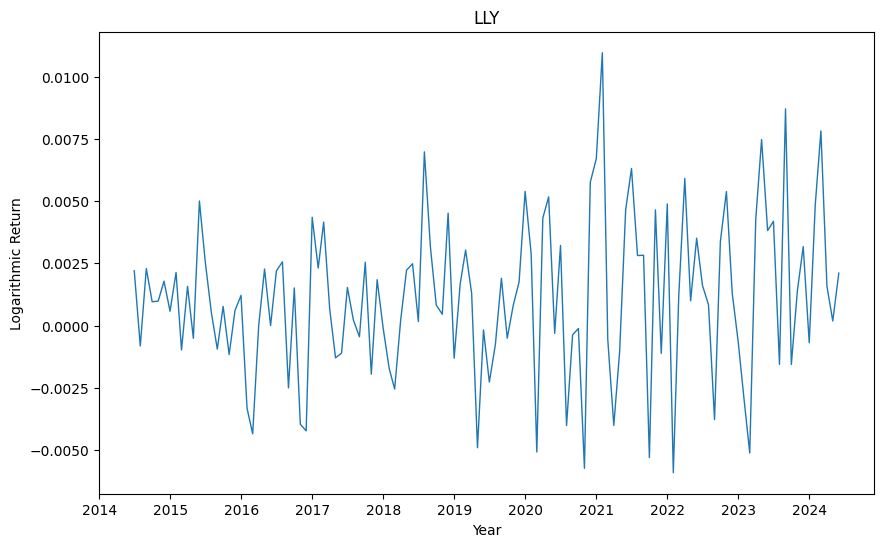


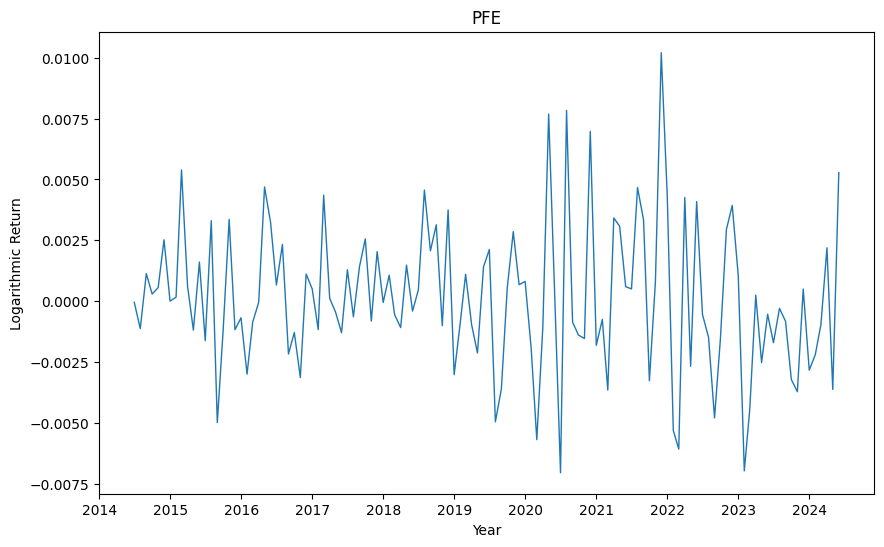
# Statistiche descrittive

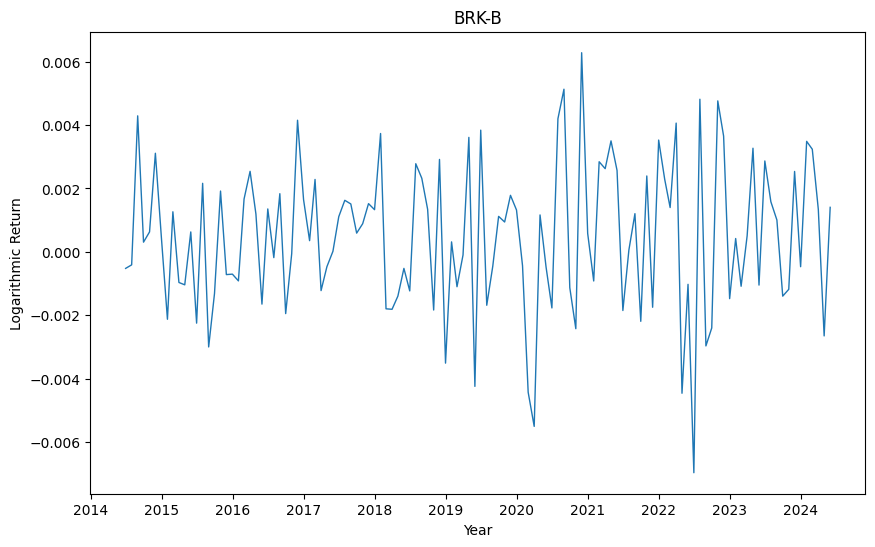
## Rendimenti semplici e logaritmici

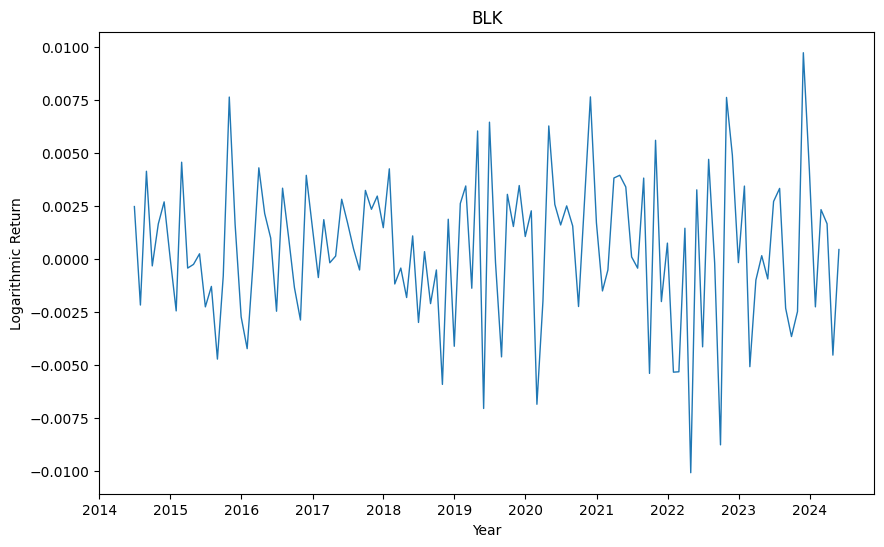












Analizzando i grafici possiamo notare alcuni andamenti e correlazioni tra i titoli scelti. Ovviamente la prima informazione che si nota è che l’andamento del mercato finanziario influenza allo stesso modo tutti i titoli, ad esempio nel 2020 a causa del Covid Crash.

Inoltre si nota anche una correlazione positiva tra le aziende che operano nello stesso settore:

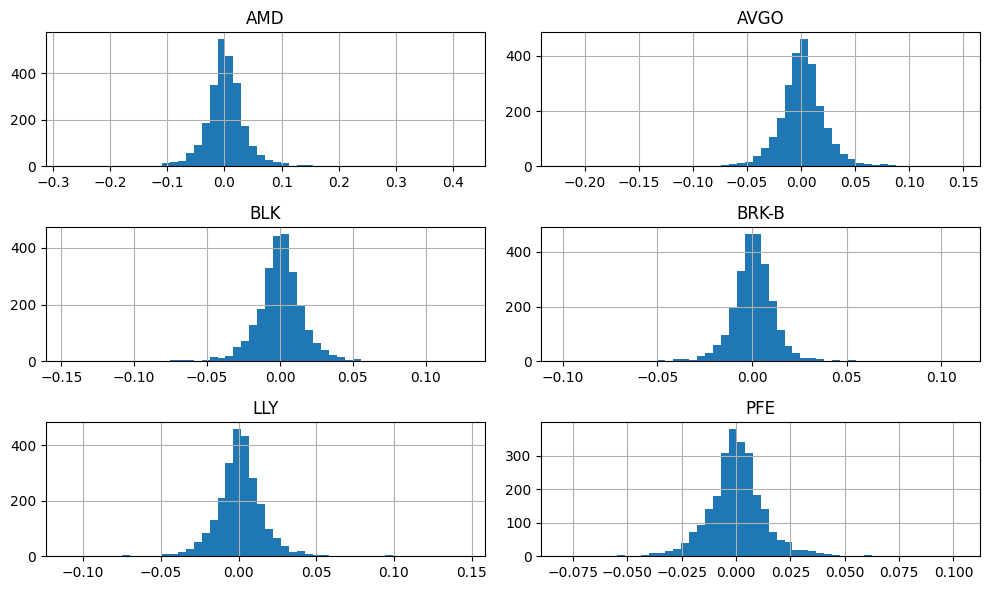
* AVGO e AMD producono entrambe chip
* LLY e PFE producono farmaci per settori sanitari che spesso sono trasversali
* BRK.B e BLK spesso investono nelle stesse aziende

### Informazioni sui rendimenti



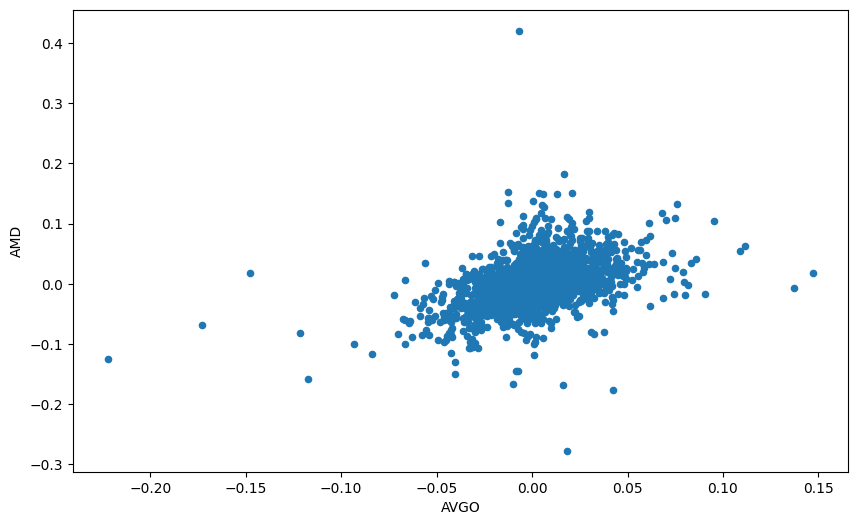
Dalla deviazione standard dei rendimenti logaritmici possiamo notare che la volatilità per i titoli dei settori Sanitari e Finanziari è molto bassa mentre per i titoli del settore tecnologico è più alta (per AMD quasi il doppio).

## Presentazione dei rendimenti con istogrammi e grafici a dispersione

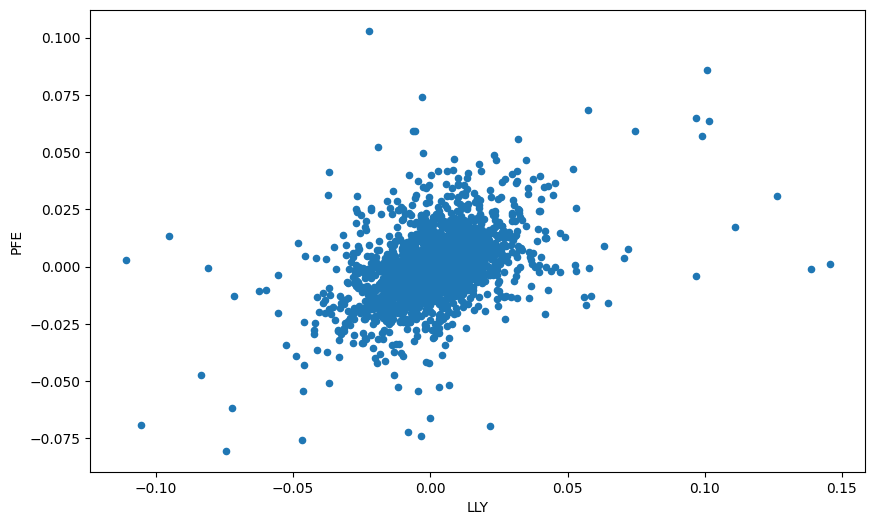


Istogrammi sulla distribuzione dei rendimenti logaritmici.

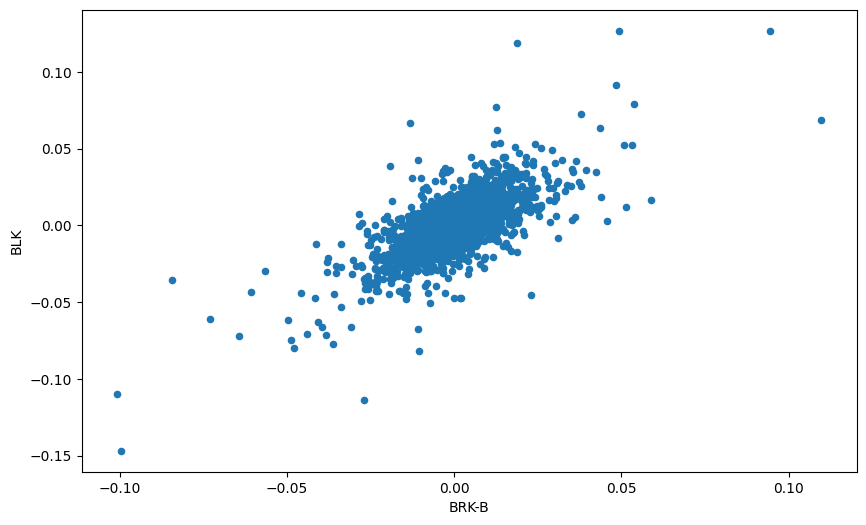
### Settore tecnologico



### Settore sanitario



### Settore finanziario



Osservando i grafici sopra riportati possiamo verificare che i dati sono altamente correlati e che esiste una forte relazione tra i titoli azionari che appartengono allo stesso settore.

## Grafici diagnostici a 3 sezioni (istogramma, boxplot e qq-plot)

### Broadcom Inc.

|  |  |
| --- | --- |
|  |  |
|  | |

Descrivimi

### Advanced Micro Devices Inc.

|  |  |
| --- | --- |
|  |  |
|  | |

Descrivimi

### Eli Lilly

|  |  |
| --- | --- |
|  |  |
|  | |

Descrivimi

### Pfizer

|  |  |
| --- | --- |
|  |  |
|  | |

Descrivimi

Berkshire Hathaway Class B

### Broadcom Inc.

|  |  |
| --- | --- |
|  |  |
|  | |

Descrivimi

### Blackrock

|  |  |
| --- | --- |
|  |  |
|  | |

Descrivimi

## Statistiche descrittive univariate

Usando le statistiche descrittive univariate possiamo ottenere informazioni su diversi aspetti chiave degli asset finanziari scelti. Le statistiche che ci interessano maggiormente sono:

* Media: rappresenta la media dei rendimenti dell’asset nel periodo di tempo considerato; ci permette di scoprire la tendenza degli investimenti nel tempo.
* Varianza e Deviazione Standard: indicano la dispersione dei rendimenti dell’asset attorno alla media; un valore maggiore indica una maggiore volatilità e un maggiore rischio dell’investimento.
* Asimmetria: indica l’asimmetria della distribuzione dei rendimenti dell’asset, un valore positivo indica un’asimmetria verso destro quindi una maggioranza di valori positivi rispetto alla media mentre un valore negativo indica un’asimmetria verso sinistra e quindi una maggioranza di valori negativi.
* Curtosi: indica la variazione della distribuzione dei rendimenti dell’asset rispetto ad una distribuzione normale, un valore elevato indica che eventi che allontanano i valori dalla media sono più probabili.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ticker:** | **AVGO** | **AMD** | **LLY** | **PFE** | **BRK.B** | **BLK** |
| **Media:** | 0.0013 | 0.0015 | 0.0011 | 0.0002 | 0.0005 | 0.0005 |
| **Varianza:** | 0.0005 | 0.0013 | 0.0003 | 0.0002 | 0.0001 | 0.0003 |
| **Deviazione Standard:** | 0.0222 | 0.0362 | 0.0169 | 0.0144 | 0.0121 | 0.0171 |
| **Asimmetria:** | -0.4277 | 0.4757 | 0.7396 | 0.1503 | -0.2482 | -0.1606 |
| **Curtosi:** | 9.3448 | 10.6656 | 10.7269 | 5.1215 | 11.2449 | 8.7427 |

Descrivimi

L’asset con il rendimento maggiore è AMD mentre quello con il rendimento minore è PFE.  
Inoltre dalla deviazione standard e dalla varianza notiamo che la volatilità dei prezzi rispetto alla media è bassa quindi gli asset scelti sono sicuri come investimenti.

## Matrici di covarianza e di correlazione dei rendimenti

La matrice di covarianza fornisce informazioni sulla relazione lineare tra i rendimenti di due asset mentre la matrice di correlazione ci fornisce informazioni riguardo alla correlazione tra i vari asset.

### Matrice di covarianza

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ticker | AMD | AVGO | BLK | BRK-B | LLY | PFE |
| AMD | 0.001312 | 0.000359 | 0.000233 | 0.000137 | 0.000104 | 0.000093 |
| AVGO | 0.000359 | 0.000492 | 0.000207 | 0.000130 | 0.000094 | 0.000081 |
| BLK | 0.000233 | 0.000207 | 0.000294 | 0.000147 | 0.000094 | 0.000099 |
| BRK-B | 0.000137 | 0.000130 | 0.000147 | 0.000147 | 0.000076 | 0.000081 |
| LLY | 0.000104 | 0.000094 | 0.000094 | 0.000076 | 0.000286 | 0.000105 |
| PFE | 0.000093 | 0.000081 | 0.000099 | 0.000081 | 0.000105 | 0.000206 |

### Matrice di correlazione

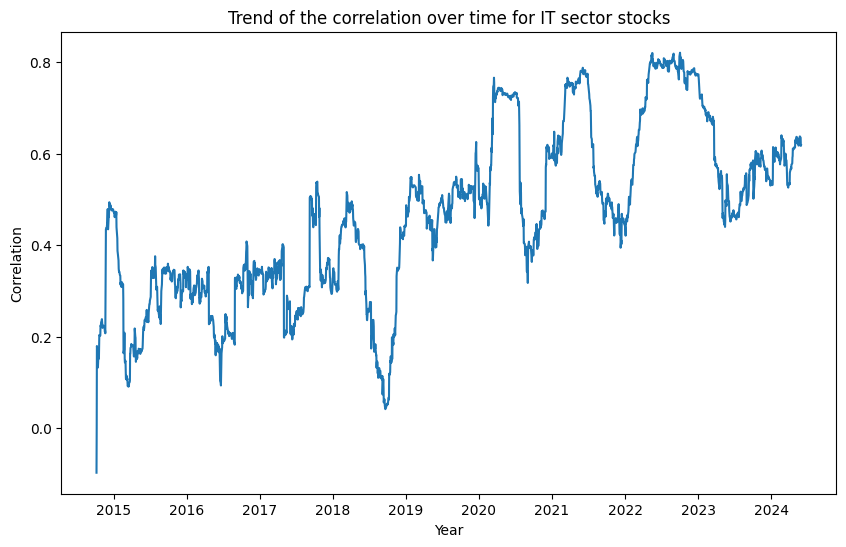
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ticker | AMD | AVGO | BLK | BRK-B | LLY | PFE |
| AMD | 1.000000 | 0.446959 | 0.374661 | 0.312397 | 0.169330 | 0.178133 |
| AVGO | 0.446959 | 1.000000 | 0.543171 | 0.485652 | 0.251777 | 0.255830 |
| BLK | 0.374661 | 0.543171 | 1.000000 | 0.706424 | 0.325222 | 0.403830 |
| BRK-B | 0.312397 | 0.485652 | 0.706424 | 1.000000 | 0.368904 | 0.463404 |
| LLY | 0.169330 | 0.251777 | 0.325222 | 0.368904 | 1.000000 | 0.434079 |
| PFE | 0.178133 | 0.255830 | 0.403830 | 0.463404 | 0.434079 | 1.000000 |

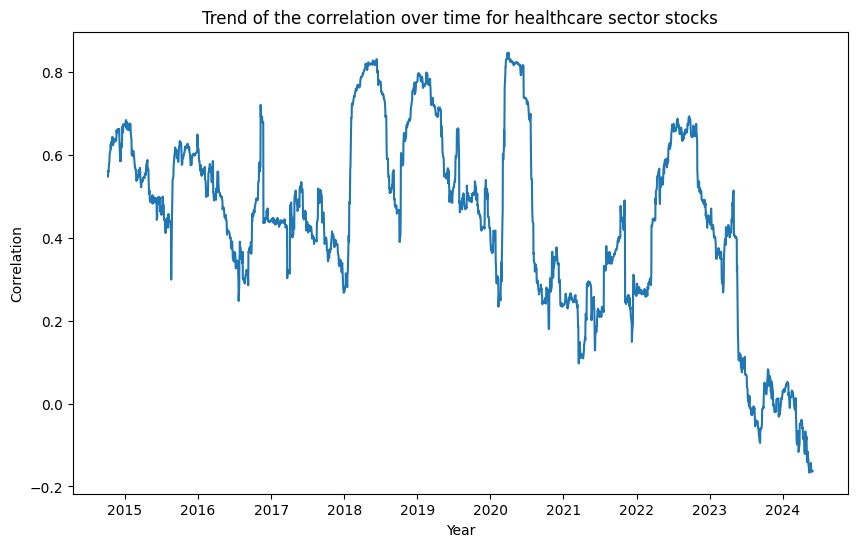
Descrivimi

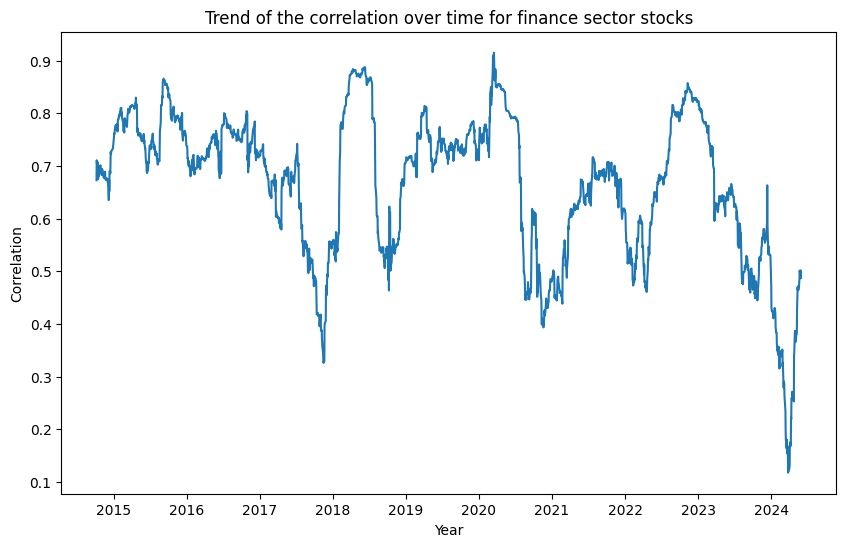
Bisogna tenere in considerazione che diversi fattori possono influenzare la correlazione tra i titoli:

* Settore di appartenenza
* Mercato globale
* Azioni intraprese dall’azienda
* Variazioni nel tempo
* Interazioni tra asset

## Andamento nel tempo delle correlazioni tra gli asset e le correlazioni medie

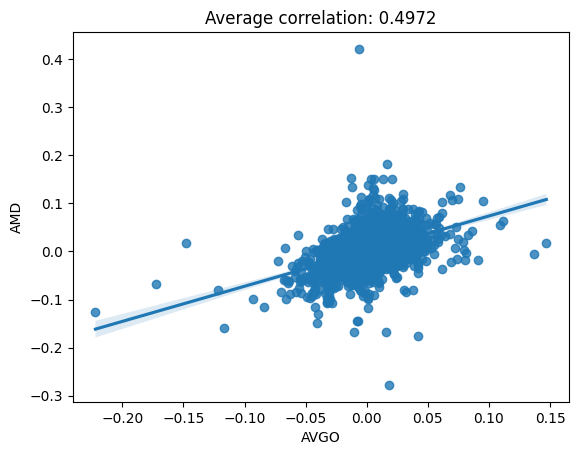






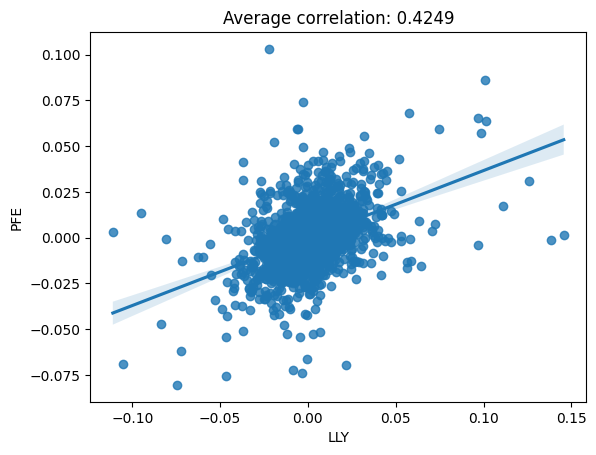
Descrivimi

### Correlazione media dei titoli del settore tecnologico



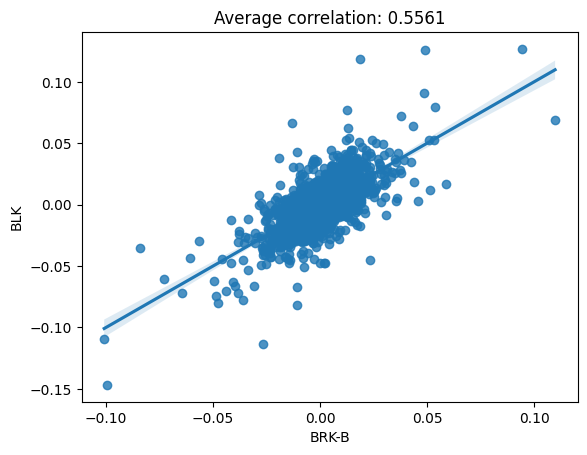
Descrivimi

### Correlazione media dei titoli del settore sanitario



Descrivimi

### Correlazione media dei titoli del settore finanziario



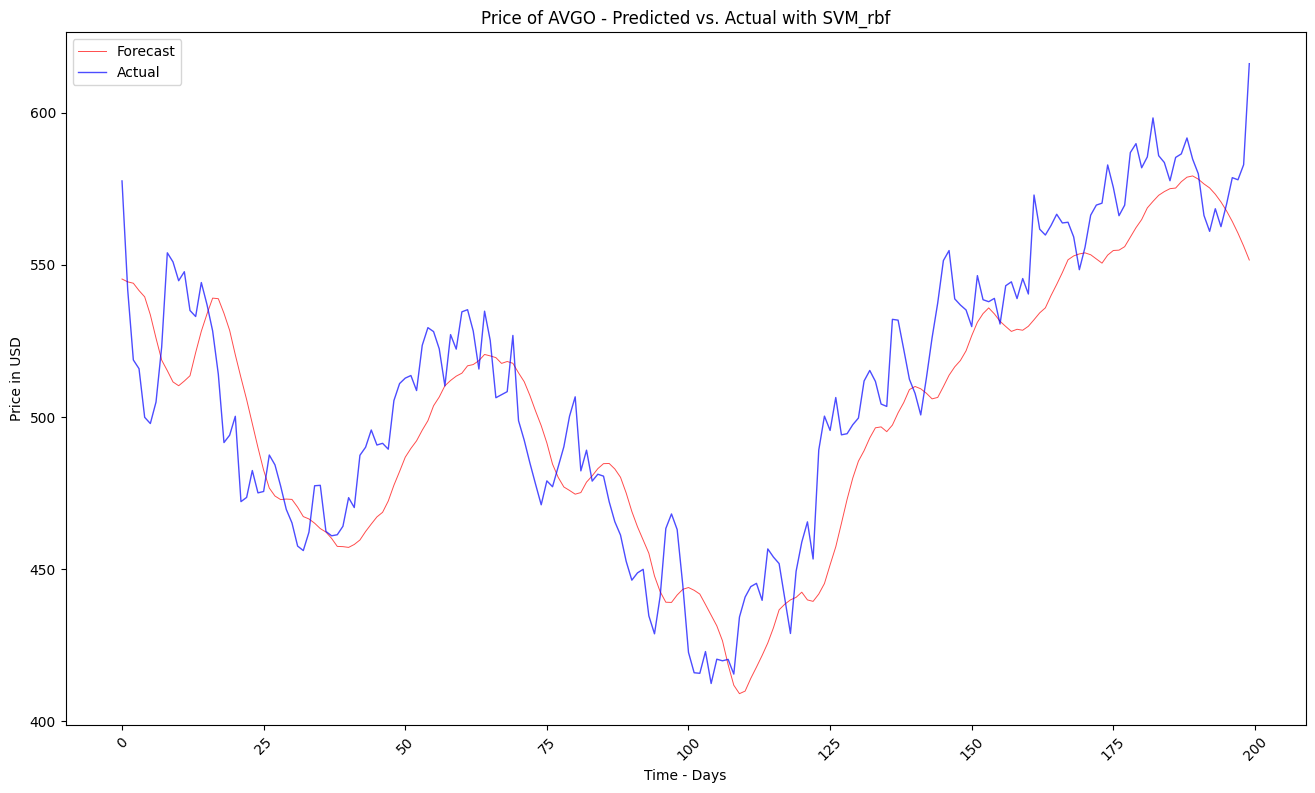
Descrivimi

# Analisi di previsione

Per l’analisi di previsione ho scelto di usare il modello Support Vector Machine (SVM) che è un algoritmo di apprendimento supervisionato.

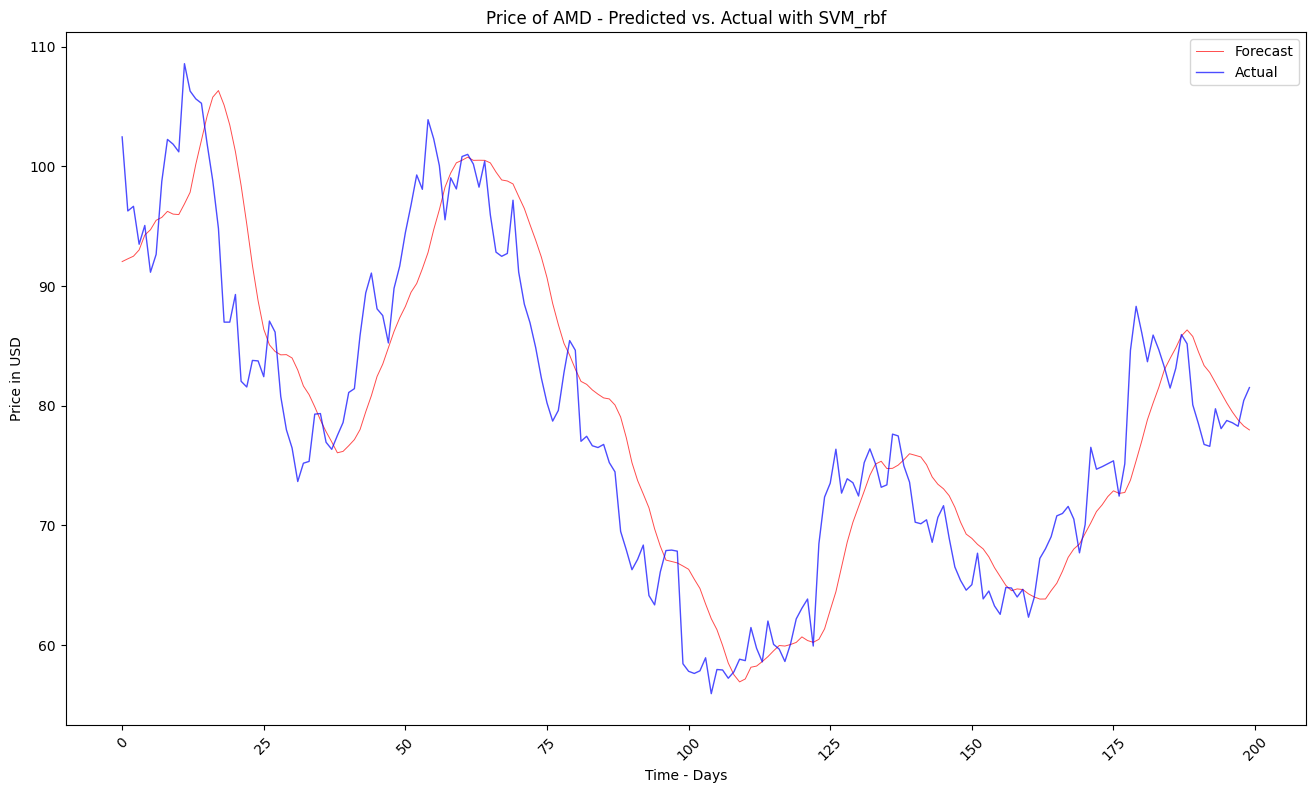
## Analisi AVGO

|  |  |
| --- | --- |
| SVM Model Confidence Score | 0.9309357373334644 |
| Mean Absolute Error | 17.098402559903455 |
| Mean Squared Error | 468.49933752317145 |
| R2 Score | 0.9020057684693716 |



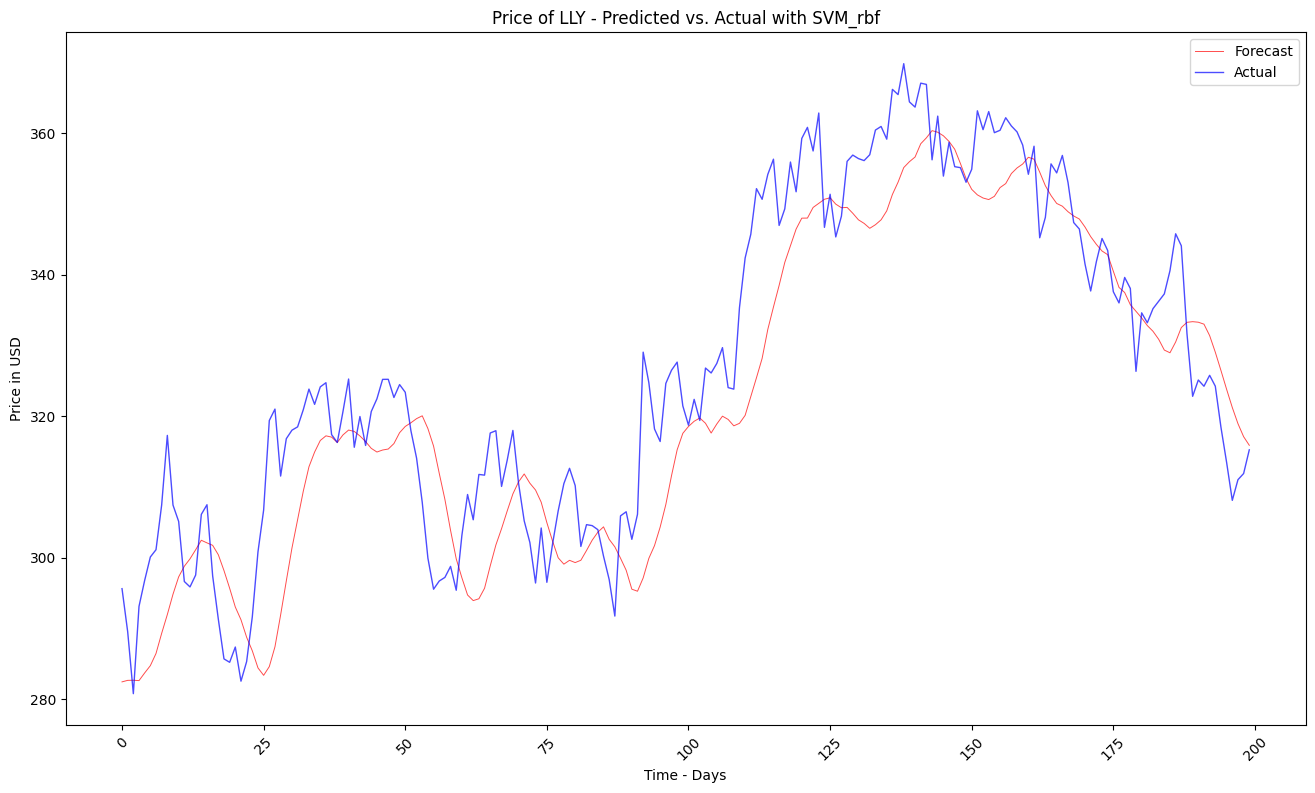
## Analisi AMD

|  |  |
| --- | --- |
| SVM Model Confidence Score | 0.9295317141583804 |
| Mean Absolute Error | 5.252784451862971 |
| Mean Squared Error | 47.571228788361566 |
| R2 Score | 0.9002271303214141 |



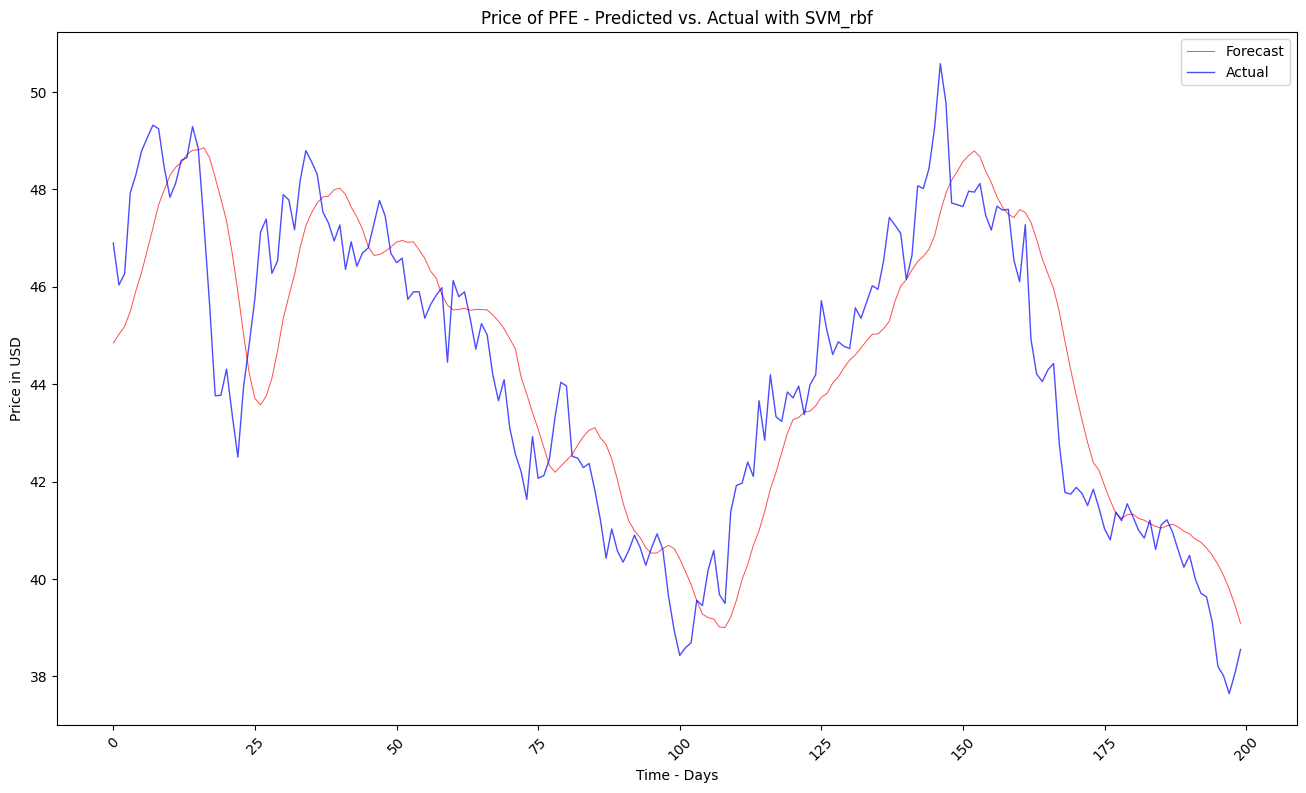
## Analisi LLY

|  |  |
| --- | --- |
| SVM Model Confidence Score | 0.9203654052400172 |
| Mean Absolute Error | 8.293077427277545 |
| Mean Squared Error | 115.44470118646156 |
| R2 Score | 0.9720466435497603 |



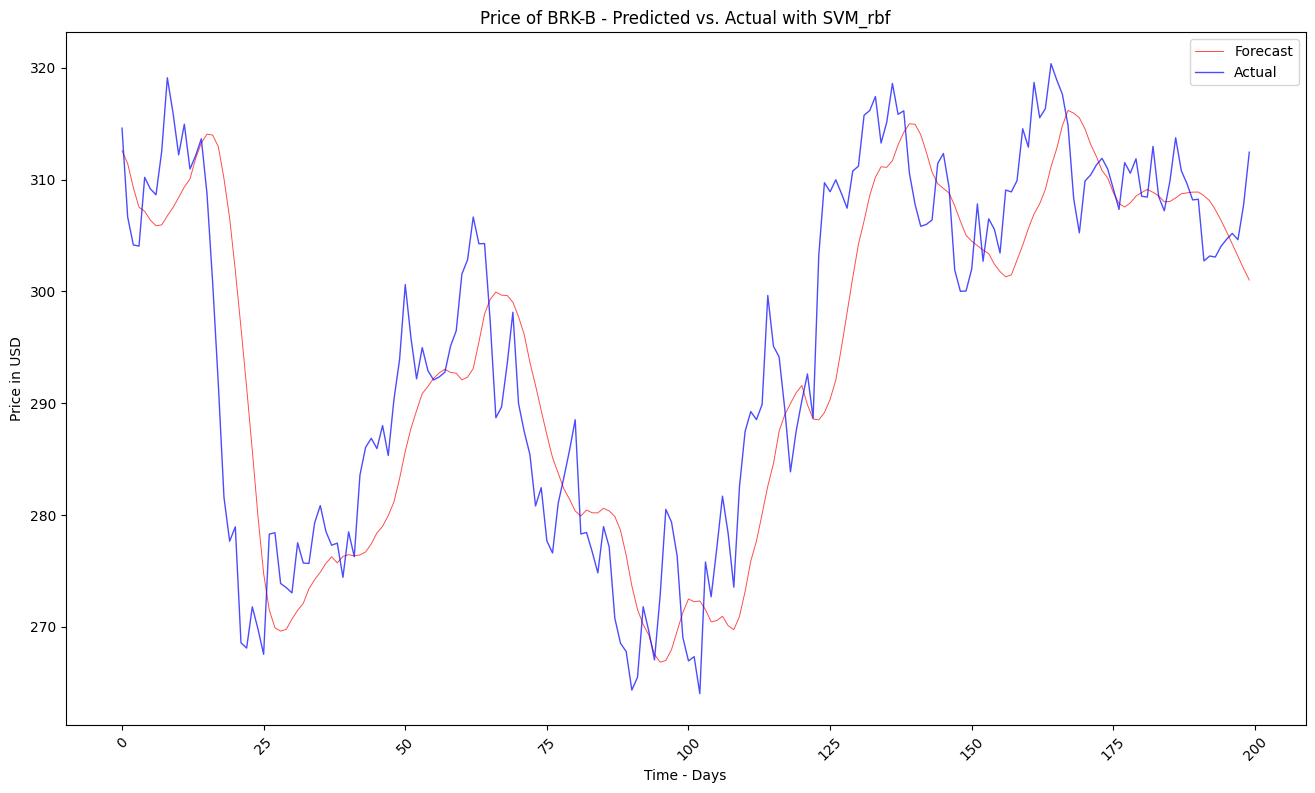
## Analisi PFE

|  |  |
| --- | --- |
| SVM Model Confidence Score | 0.945351981328591 |
| Mean Absolute Error | 1.2343845410129441 |
| Mean Squared Error | 2.7518007466754315 |
| R2 Score | 0.9334179018585117 |



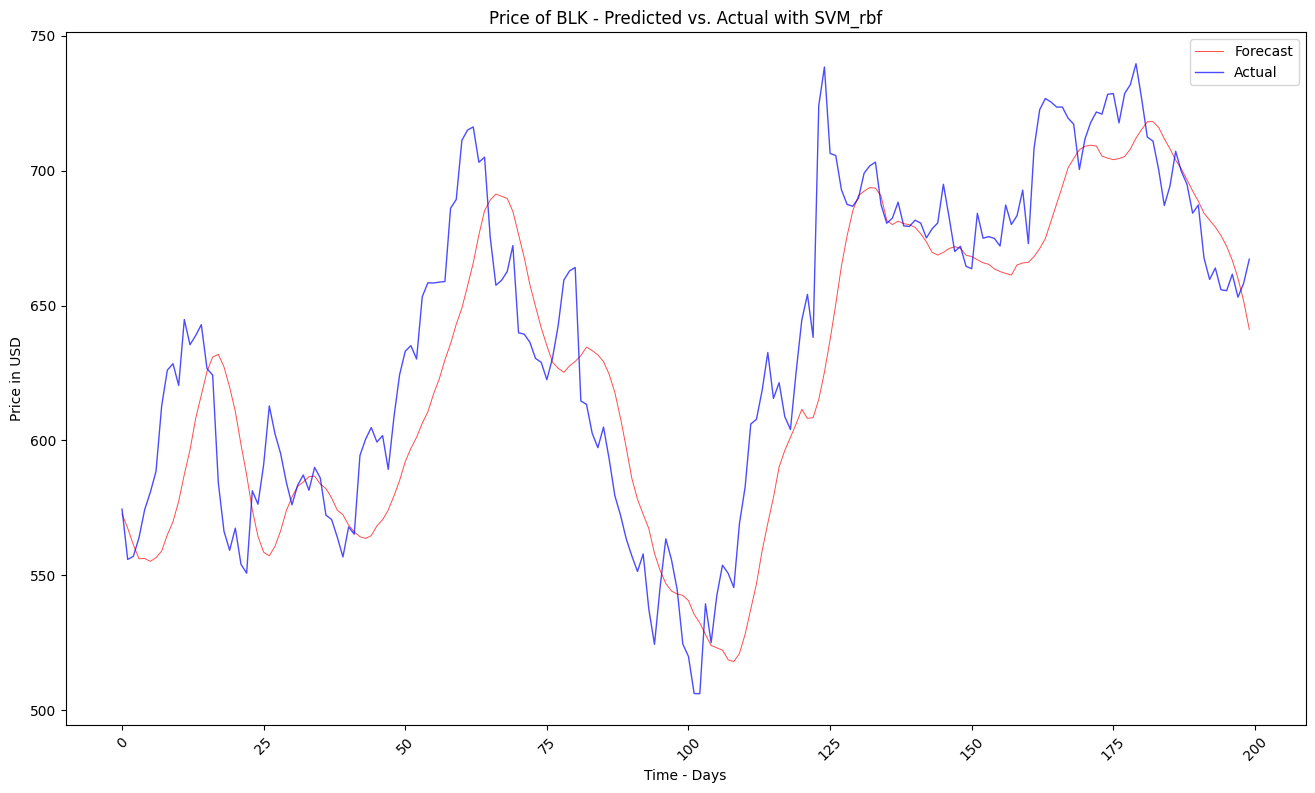
## Analisi BRK.B

|  |  |
| --- | --- |
| SVM Model Confidence Score | 0.9418194634492808 |
| Mean Absolute Error | 6.057625538634839 |
| Mean Squared Error | 63.76416645169634 |
| R2 Score | 0.9372949005767744 |



## Analisi BLK

|  |  |
| --- | --- |
| SVM Model Confidence Score | 0.8633001374790357 |
| Mean Absolute Error | 24.34896287579768 |
| Mean Squared Error | 923.5820371272531 |
| R2 Score | 0.8993176069907192 |

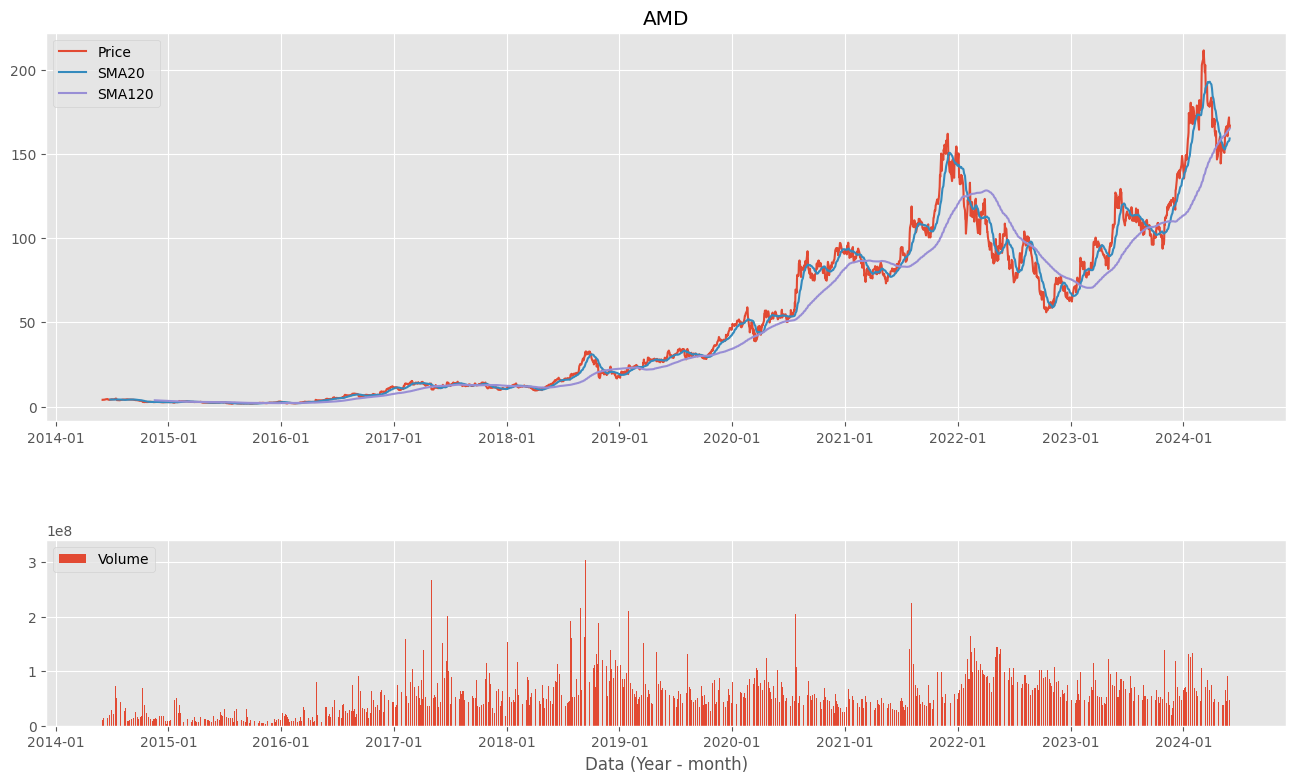


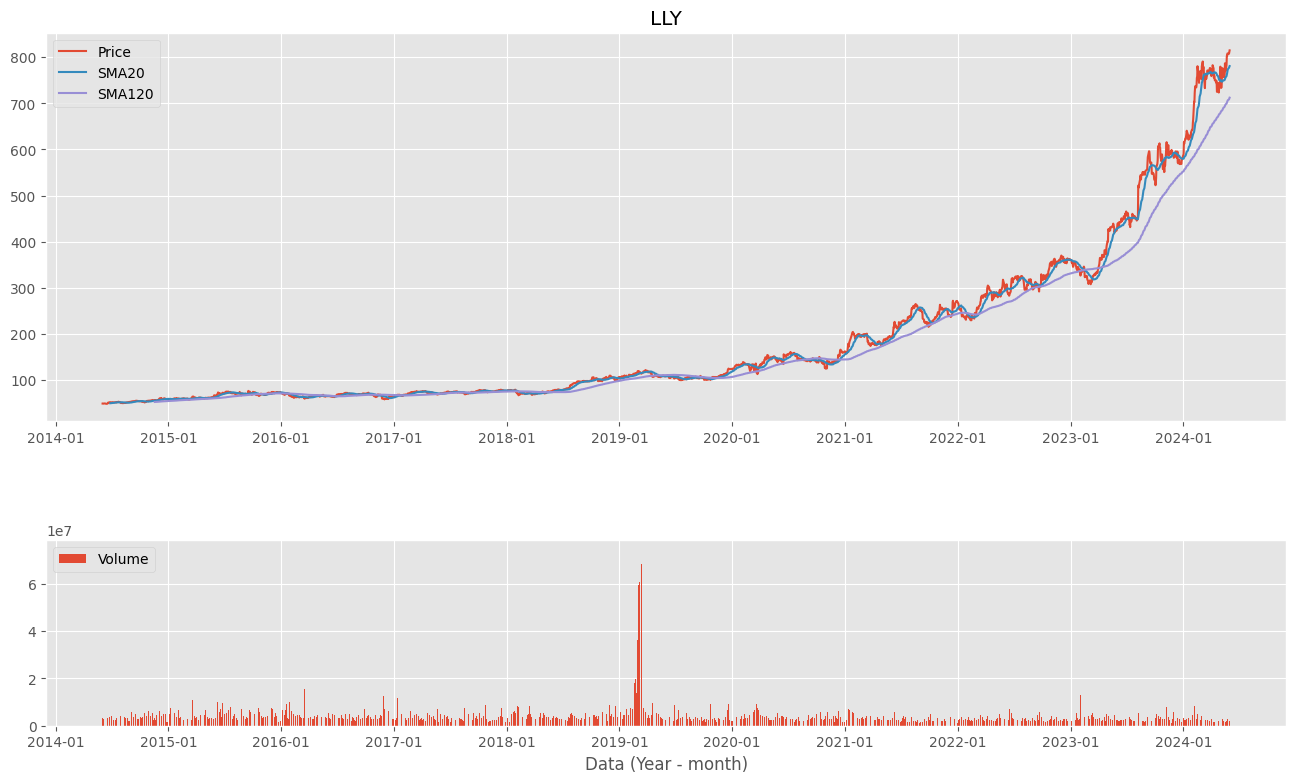
Descrivimi

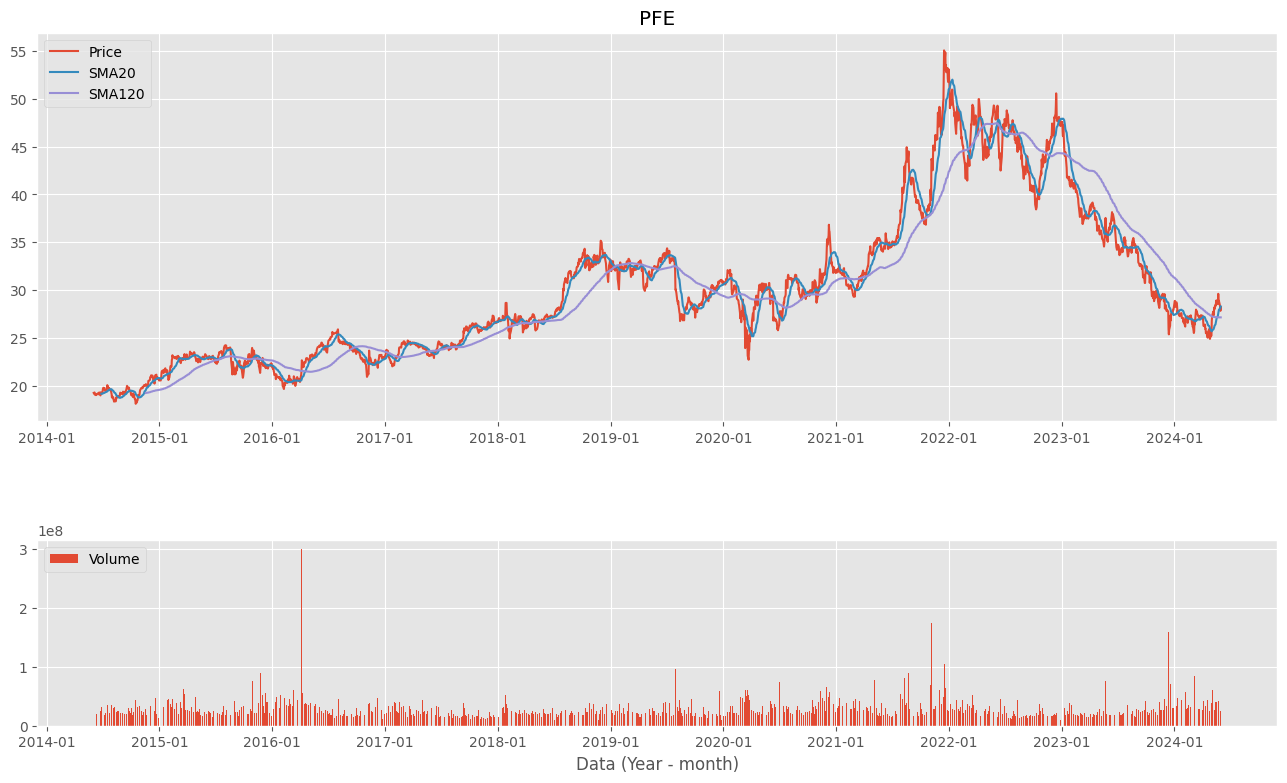
# Strategie di trading e backtesting

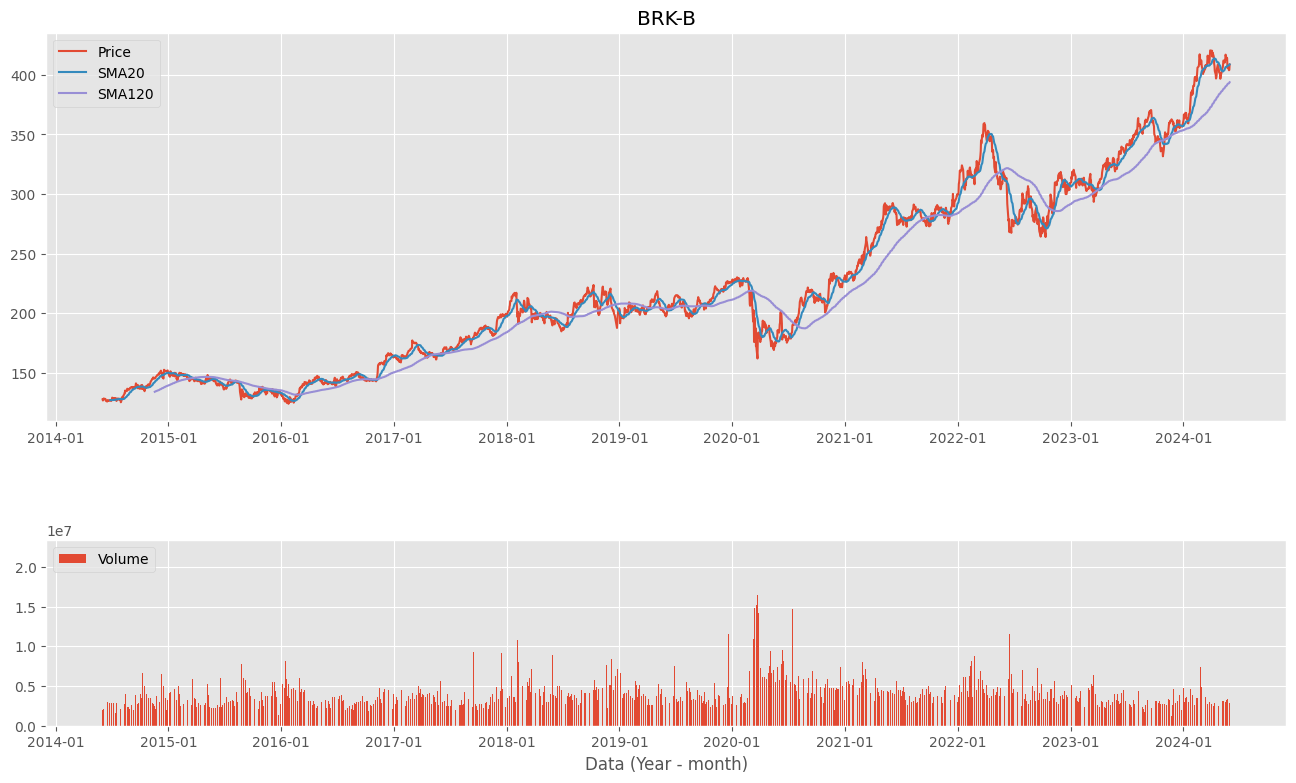
Ho utilizzato Moving Average Convergence Divergence (MACD) come indicatore per identificare i segnali di acquisto/vendita di titoli sui mercati finanziari per confrontarlo con la strategia base Buy and Hold.

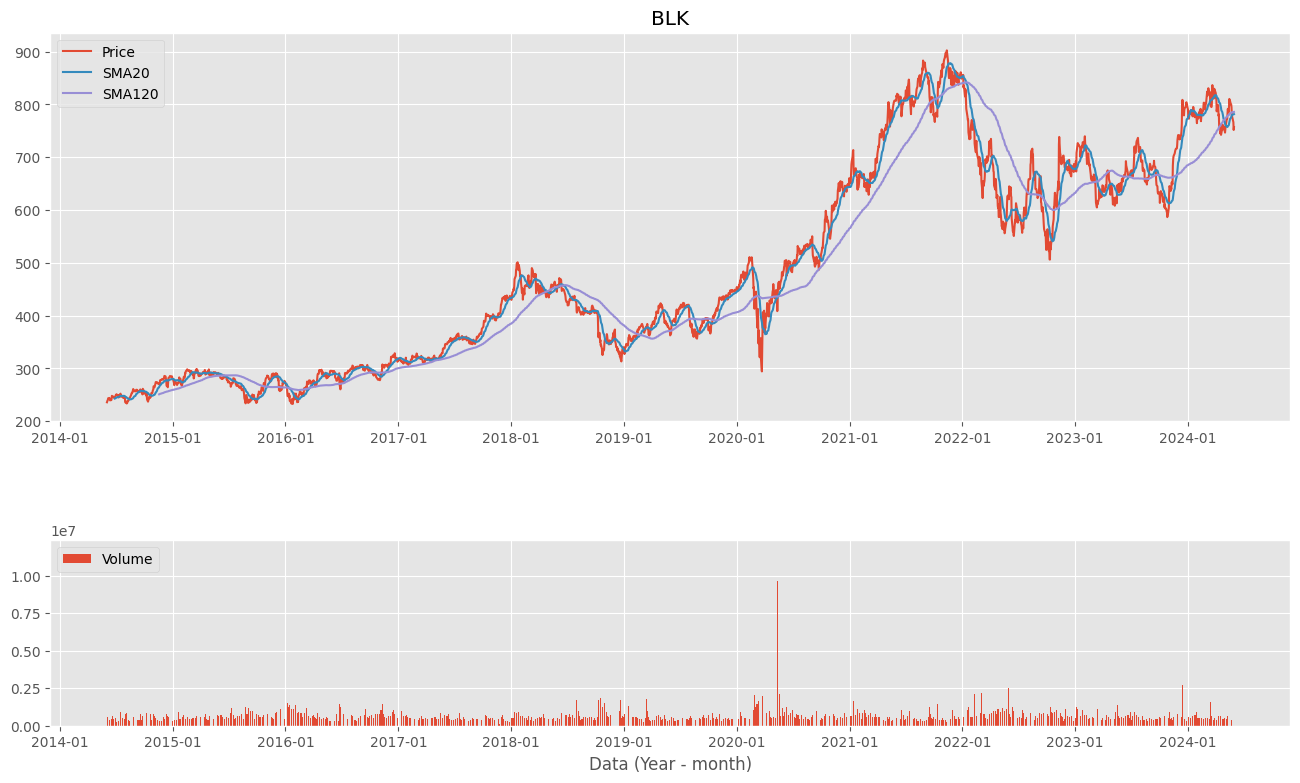








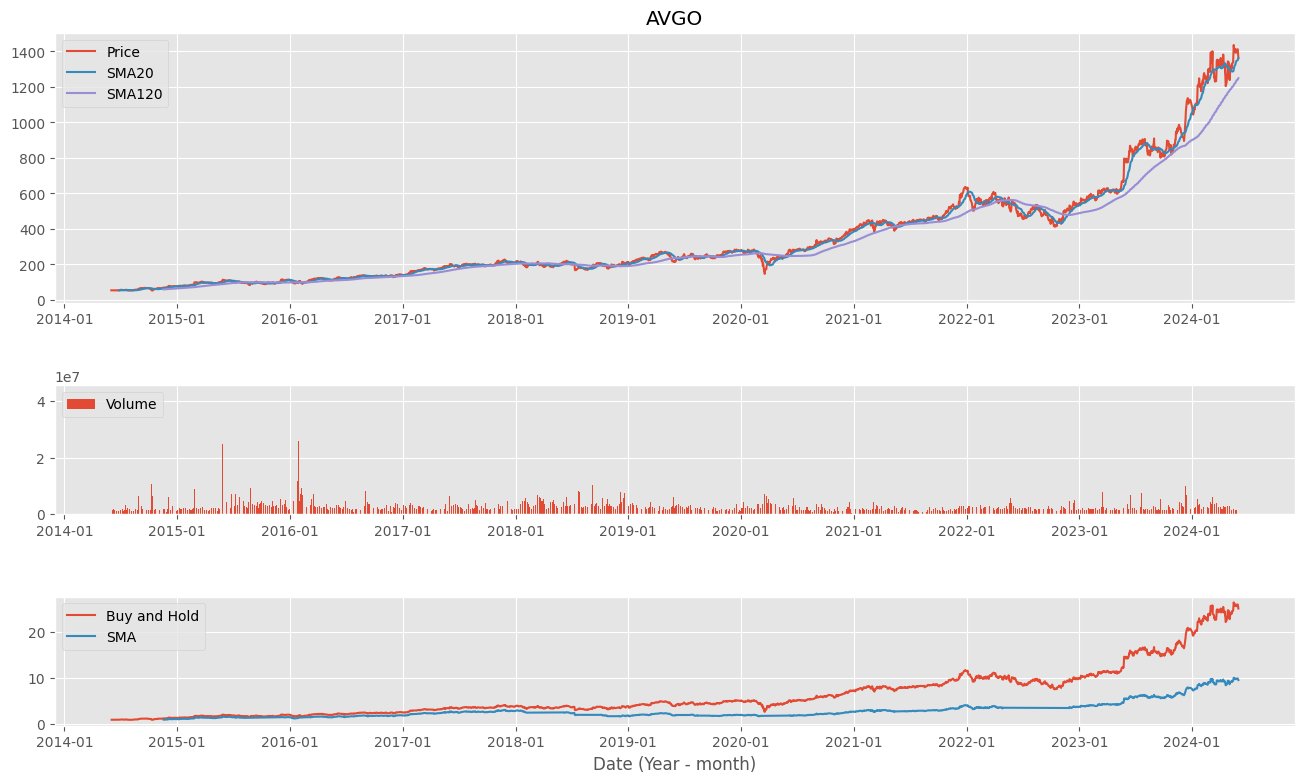


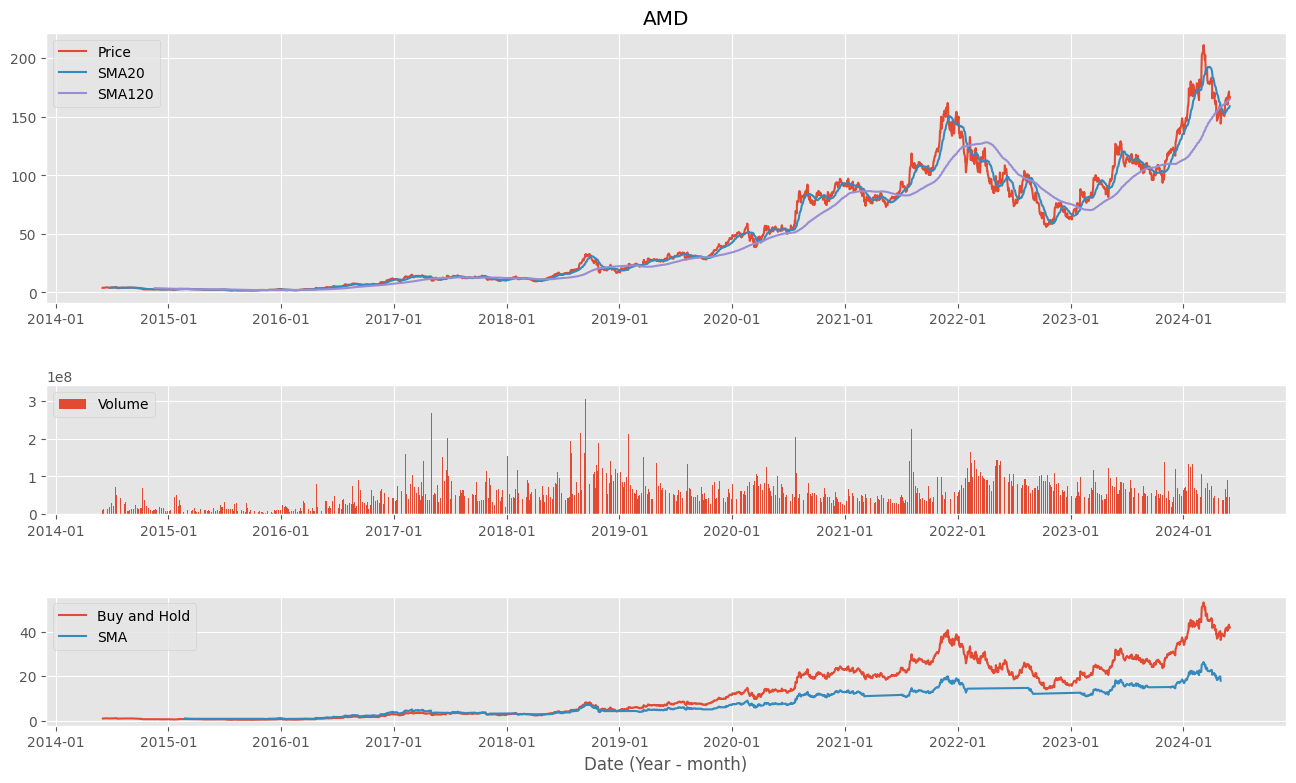


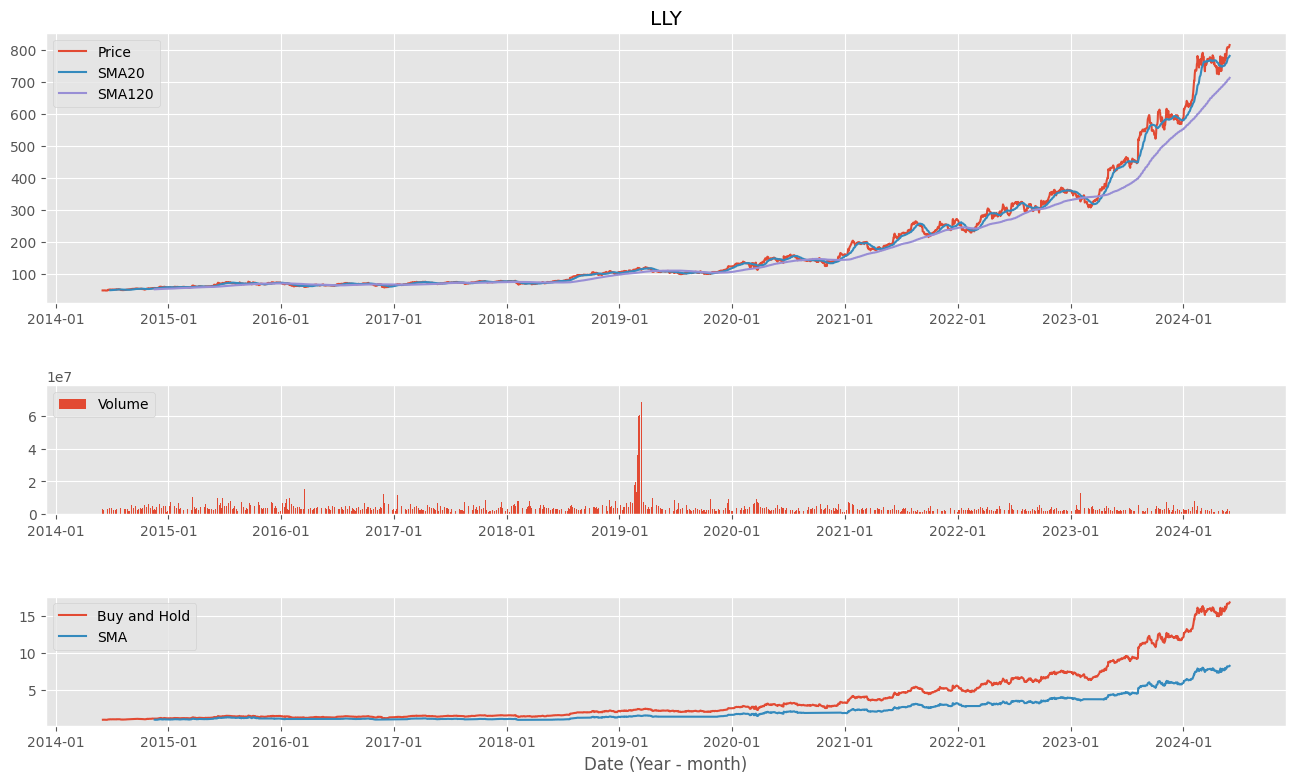
La strategia è stata implementata su due periodi:

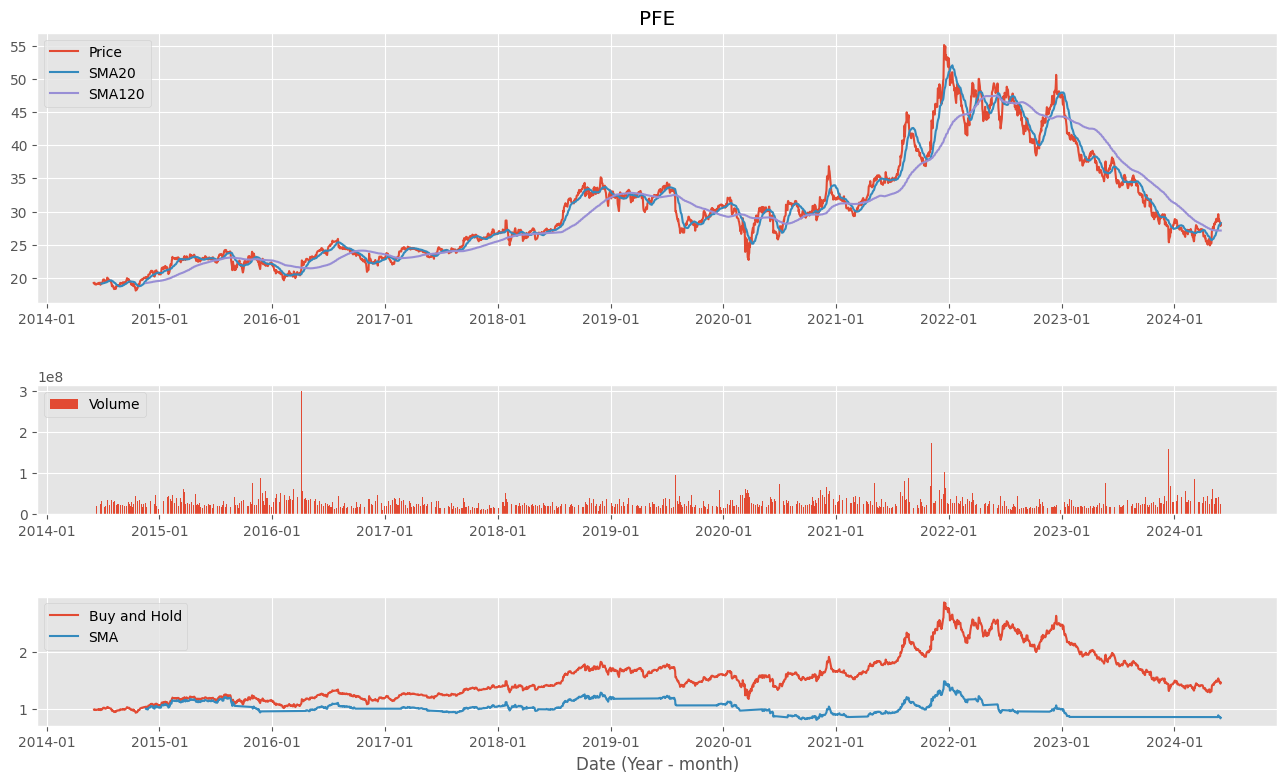
* SMA20: calcolata sulla base degli ultimi 20 periodi prezzo
* SMA120: calcolata sulla base degli ultimi 120 periodi di prezzo

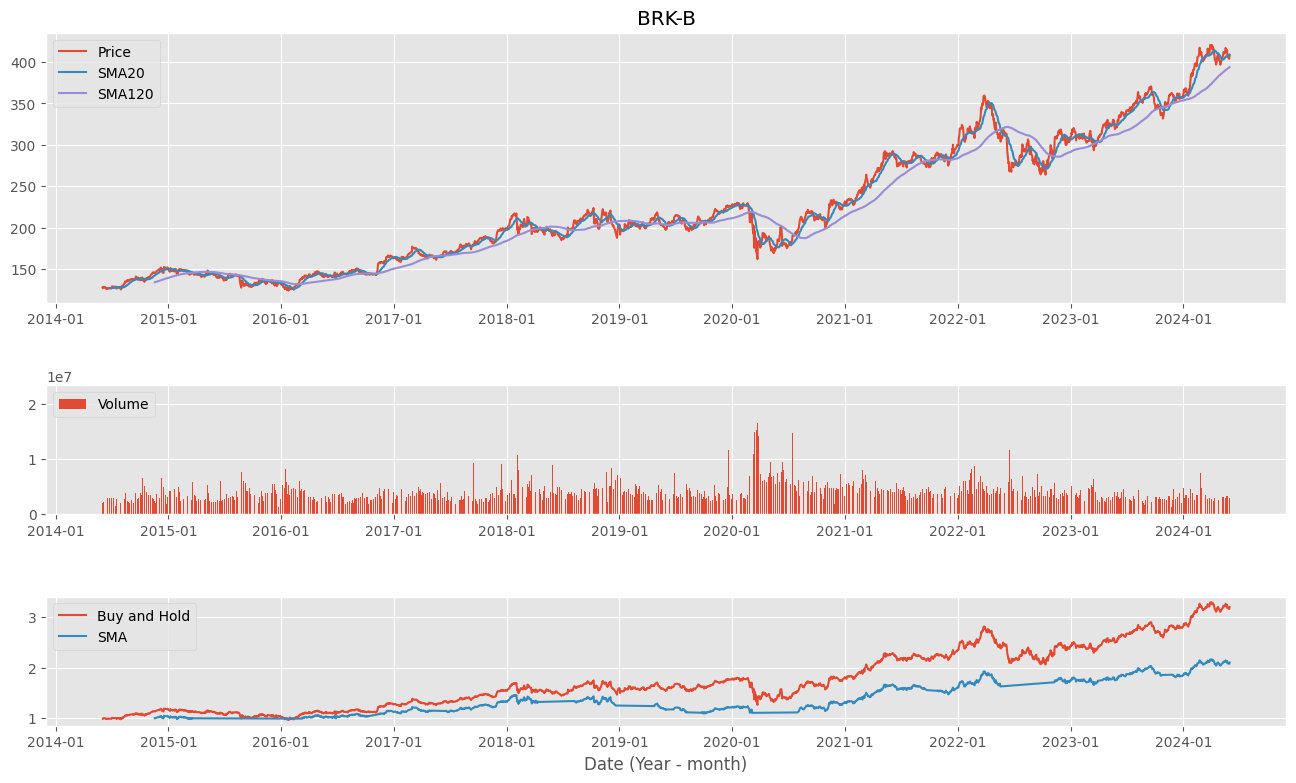
## Confronto con la strategia Buy and Hold

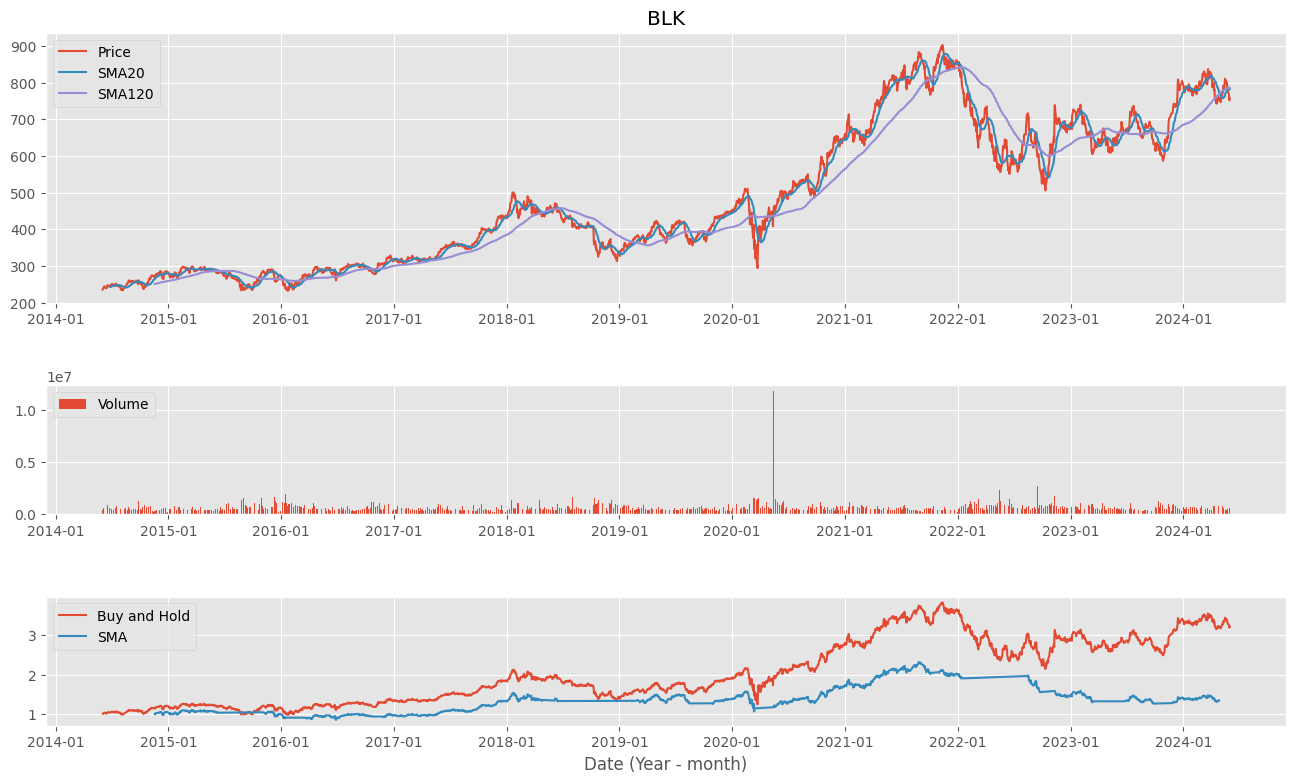










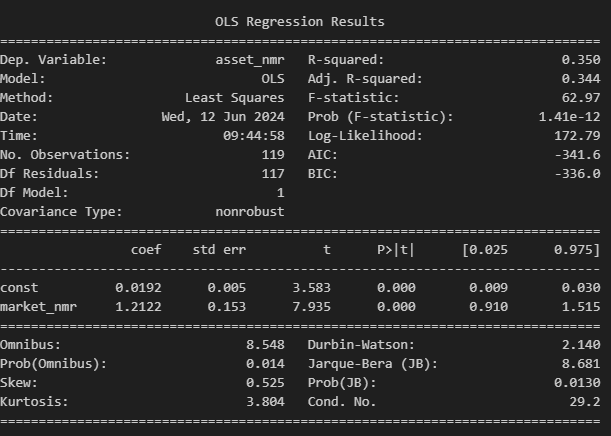


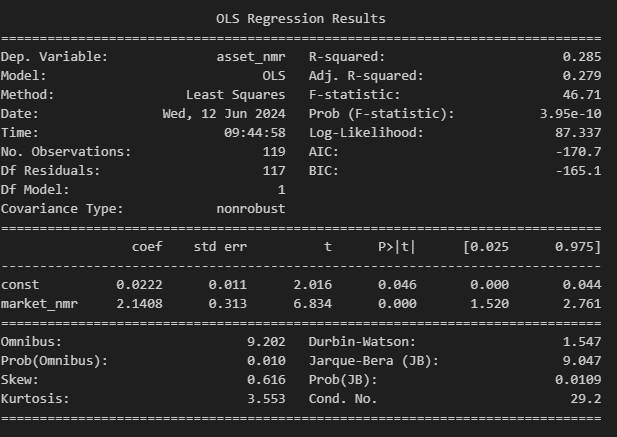
È evidente che la strategia Buy and Hold è una strategia migliore e piu sicura rispetto alla strategia basata su media mobile.

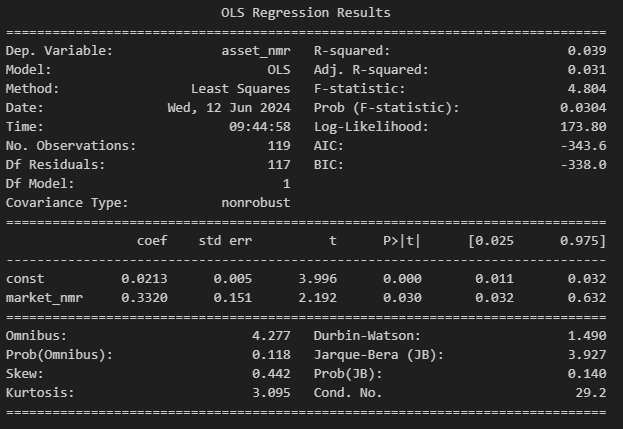
# Capital Asset Pricing Model

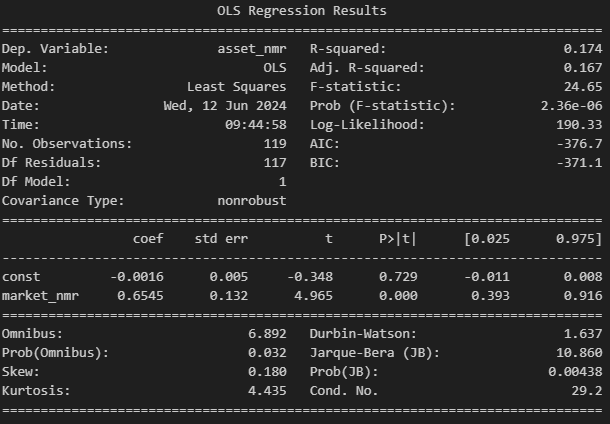
Procederemo con l’utilizzo del modello Capital Asset Pricing Model (CAPM) per analizzare l’utile atteso dei nostri asset. I dati che verrano utilizzati sono:

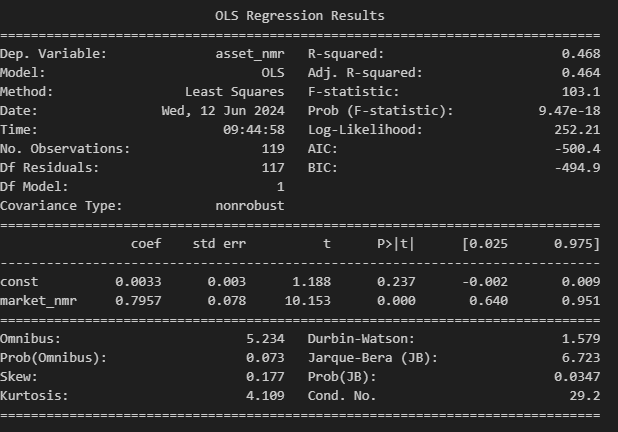
* Df Residuals: I gradi di libertà residui, che rappresentano il numero di osservazioni meno il numero di variabili nel modello.
* Df Model: Il numero di variabili indipendenti nel modello, che in questo caso e 1.
* R-squared: Il coefficiente di determinazione, che misura la proporzione di varianza della variabile dipendente.
* Adj. R-squared: E una versione aggiustata del coefficiente di determinazione che tiene conto del numero di variabili indipendenti e del numero di osservazioni.
* F-statistic: E una statistica che valuta l'importanza complessiva del modello. Un valore piu elevato indica una maggiore significativita del modello.
* Prob (F-statistic): E il valore p associato alla F-statistic, che indica la probabilita che il modello non sia significativo.
* Coef: I coefficienti stimati per le variabili indipendenti nel modello.
* P>|t|: Il valore p associato al test di t per la significativita dei coefficienti.
* [0.025 0.975]: Questo intervallo di confidenza indica l'intervallo entro cui e probabile che cada il vero valore del coefficiente con un determinato livello di confidenza.
* Omnibus, Prob(Omnibus), Jarque-Bera (JB), Skew, Kurtosis: Queste statistiche valutano la normalita dei residui del modello. In generale, si desidera che i residui siano distribuiti normalmente per garantire l'affidabilita delle stime

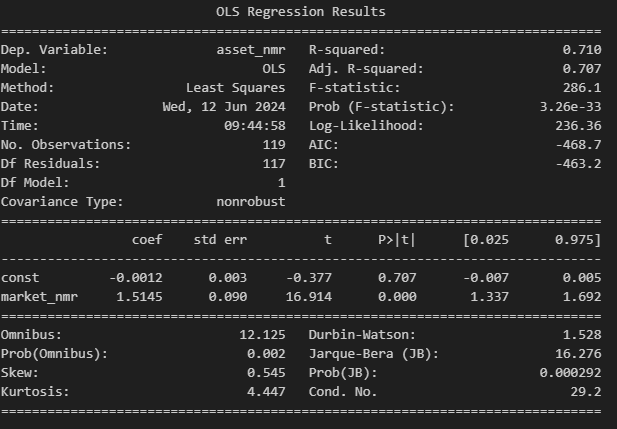












Grazie al modello CAPM possiamo calcolare i Beta dei titoli rispetto al mercato:

|  |  |  |
| --- | --- | --- |
| **Ticker** | **Beta** | **Volatilità** |
| AVGO | 1.212226745218554 | Maggiore rispetto al mercato |
| AMD | 2.140762152610683 | Maggiore rispetto al mercato |
| LLY | 0.33198758301982945 | Minore rispetto al mercato |
| PFE | 0.6544596583916281 | Minore rispetto al mercato |
| BRK.B | 0.7956732278617329 | Minore rispetto al mercato |
| BLK | 1.5144788798458553 | Maggiore rispetto al mercato |

Grazie ai Beta possiamo stimare i rendimenti attesi per ogni asset

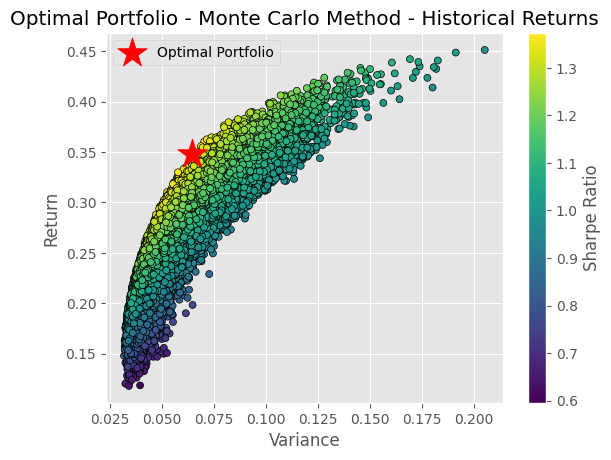
|  |  |  |
| --- | --- | --- |
| **Ticker** | **Rendimento Atteso** | **Rendimento Atteso Percentuale** |
| AVGO | 0.14871092438713726 | 14.9% |
| AMD | 0.25222311526498586 | 25.2% |
| LLY | 0.050582750221353424 | 5.1% |
| PFE | 0.08653161200187678 | 8.7% |
| BRK.B | 0.10227395832616595 | 10.2% |
| BLK | 0.1824056875633203 | 18.2% |

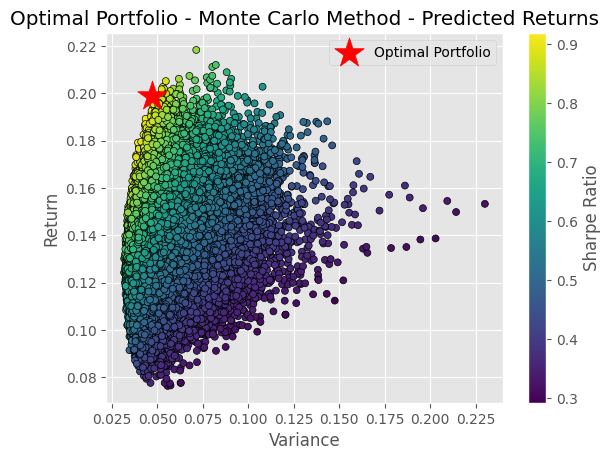
Vediamo che il range dei rendimenti è compreso tra il 5.1% e il 25.2%, ovviamente si tratta di una stima e quindi per piccole variazioni mercato queste stime potrebbero venire invalidate.

# Costruzione di un portafoglio

In questa sezione viene costruito un portafoglio in termini di media-varianza sugli ultimi 108 mesi di dati.

Per il metodo analitico usiamo la frontiera efficiente per visualizzare il portafoglio ottimale basato sui rendimenti passati e su quelli attesi.





# Conclusioni

Possiamo concluedere dicendo che i dati esaminati sono distribuiti normalmente, come visto nella parte sulle statistiche descrittive, inoltre come possiamo vedere dalla deviazione standard dei rendimenti logaritmici in generale notiamo che hanno una bassa variabilita rispetto alla media.

Gli asset facenti parte dello stesso settore sono fortemente correlati tra loro.

Grazie al modello costruito per fare l’analisi di previsione siamo in grado di prevedere in modo approssimativo l’andamento del mercato, non tenendo in considerazione grossi sconvolgimenti dei mercati finanziari, e basandoci esclusivamente sull’andamento dei dati usati per fare il training.

Riguardo al Beta di ciascun titolo abbiamo scoperto che la maggior parte degli asset hanno una volatilita maggiore rispetto al mercato.

Mentre per la strategia presa in considerazione, abbiamo visto come le previsioni ci danno dei risultati peggiori rispetto alla classica strategia di “Buy & Hold”.

Per quanto riguarda la creazione del portafoglio abbiamo trovato i migliori portafogli in termini di rapporto rendimenti/varianza fatto sui rendimenti passati e sui rendimenti previsti calcolati precedentemente.