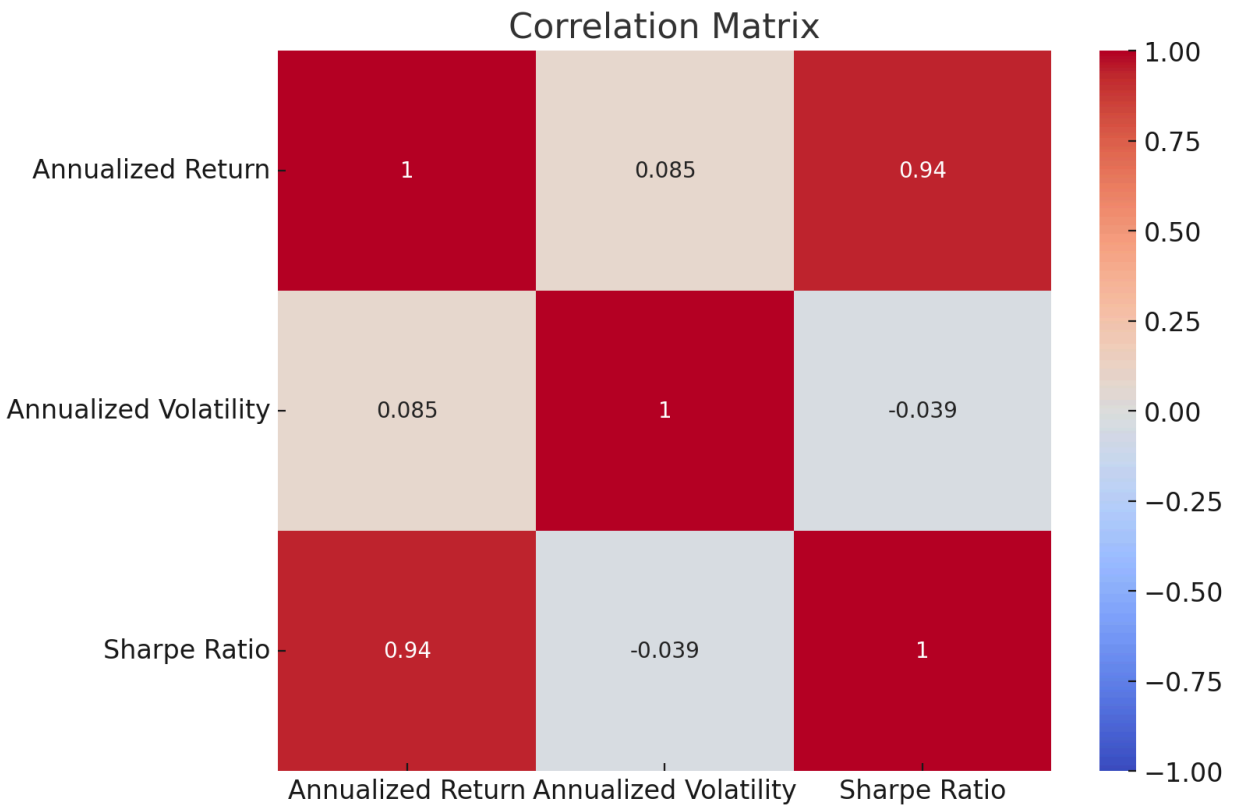

1. Data function

Variable Additions:

a. Based on the dataset, other variables like trading volume, beta, or sector classification could provide more insights.
b. Including external factors like economic indicators or sentiment scores from news could enhance the analysis.

Correlation Analysis:

1. Annualized Return and Sharpe Ratio: High positive correlation (0.94), shows that stocks with higher returns usually have higher risk-adjusted returns.
2. Annualized Volatility and Other Metrics: Low correlation with both annualized return and Sharpe ratio, implies that volatility is not a strong predictor of return or risk-adjusted return.

2. Model Enchancement

Outlier Removal:

- Ensures that the data fed into the clustering model is free from anomalies, leading to better clustering results.

ARIMA vs. LSTM:

- **ARIMA**: Good for linear and seasonal patterns. Easier to interpret but may not capture complex patterns.
- **LSTM**: Capable of capturing non-linear relationships and long-term dependencies. More complex and computationally intensive.

-