Measuring Market Efficiency:

A Volatility-Adjusted Return Study of the Dow Jones Industrial Average

Race Partin

Ramapo College of New Jersey

CMPS 240: Data Analytics Using Python

Professor Lawrence D'Antonio

December 15, 2025

Abstract

This program seeks to prove the fundamental risk-return hypothesis. The premise is that higher investment risk is associated with higher investment returns. Using the UCI Dow Jones dataset, the research quantified historical volatility (risk) and average returns to calculate the Sharpe Ratio for individual stocks, aiming to verify a positive linear risk-return relationship. Contrary to expectations, linear regression analysis yielded a negative slope (-2.92), suggesting that higher risk did not correlate with higher returns over the sample period. With a R-squared value of 0.31 and a p-value of 1.00, the null hypothesis was unable to be rejected at a 5% significance level. The findings suggest that the short time horizon introduced volatility, which likely obscured the true longer-term risk-return relationship. The traditional hypothesis may not hold for shorter time horizons.

Keywords: Sharpe Ratio, volatility, linear regression

Measuring Market Efficiency

Research Question

The primary research question addressed in this project is: Does the fundamental financial theory that higher risk leads to higher returns hold for the Dow Jones Industrial Average over a specific period of time?

Specifically, this project seeks to determine if there is a statistically significant positive linear relationship between the annualized standard deviation of a stock and its annualized return. Furthermore, this project calculates the Sharpe Ratio to provide another view of the risk-adjusted returns within the Dow Jones Industrial Average.

Data Source

The dataset used for analysis was the Dow Jones Index dataset, obtained from the UCI Machine Learning Repository.

- Content: The dataset contains weekly price data in the form of open, high, low, and close for the 30 stocks that make up the DJIA. The data ranges from January 7, 2011, to March 25, 2011.

Python Libraries

The following Python libraries were utilized for data extraction, manipulation, and visualization.

- Pandas: Used for loading, manipulating, and cleaning DataFrames.
- NumPy: Used for mathematical operations such as calculating the square root. Also utilized for generating the regression line.
- SciPy: Used for performing linear regression and calculating the p-value for hypothesis testing.

- Matplotlib: Used to generate scatter plots, bubble charts, and bar charts for data visualization purposes.

Data Manipulation

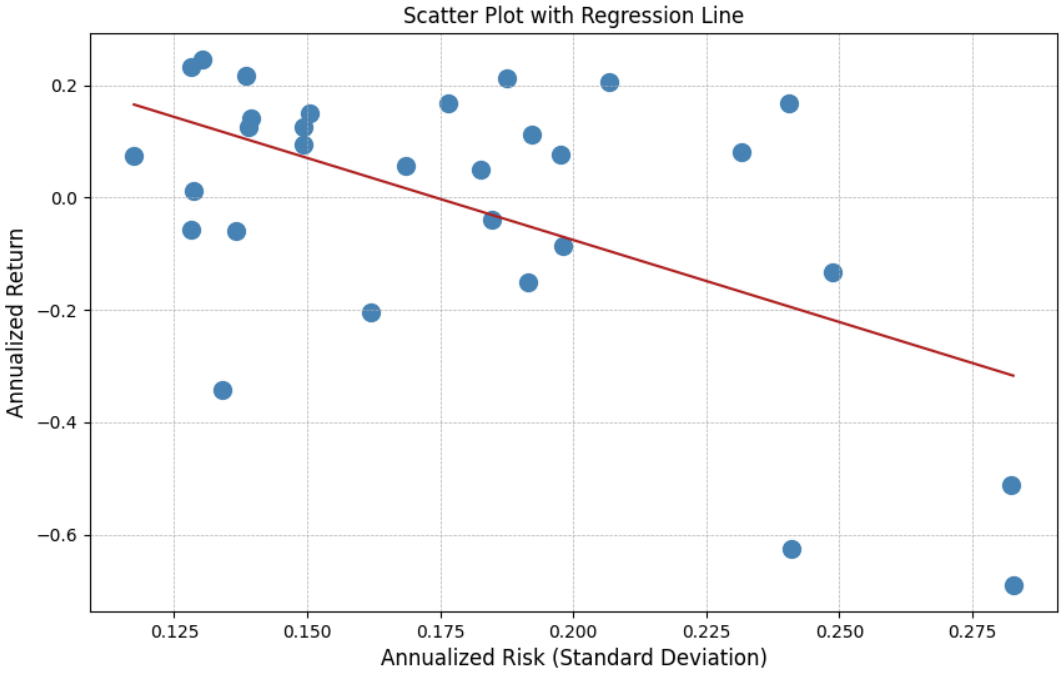
To prepare the raw data for analysis, a modular approach was used via the dedicated data.py module. The following manipulations were performed:

- Cleaning Currency Symbols: The raw dataset came with '$' symbols in the price columns. These values were stripped using the .replace in conjunction with regex.
- Date Conversion: In order to ensure proper sorting, the date column was converted to datetime objects.
- Return Calculation: The weekly return was calculated using 'pct_change' on the close of the weekly price.
- Aggregation: Data was grouped by the stock symbol, and the mean of weekly return and standard deviation was calculated.
- Annualization:
    - Returns were annualized by multiplying the mean weekly return by 52.
    - Standard deviation was annualized by multiplying the mean weekly standard deviation by the square root of 52.
- Sharpe Ratio: This metric was derived by subtracting the risk-free rate from the annualized return and dividing by the annualized standard deviation.

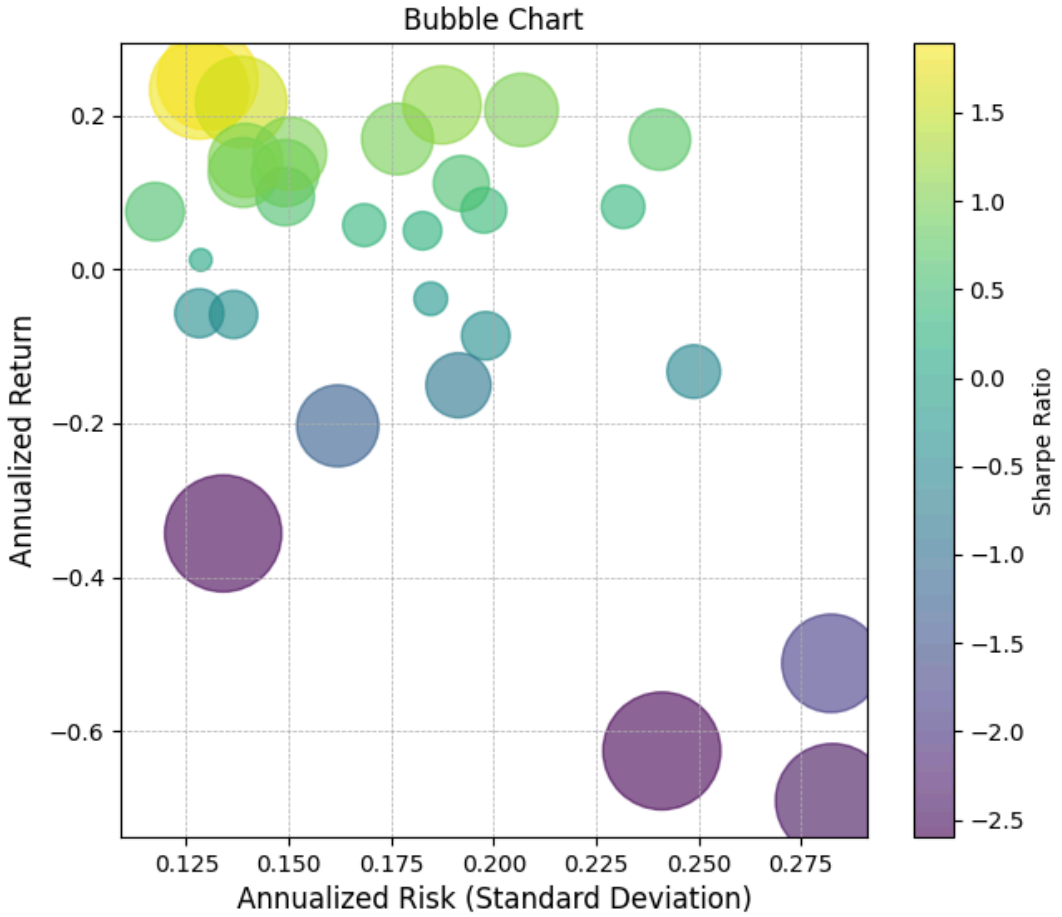
Data Visualization

Three different methods were used in order to visualize this data for interpretation.

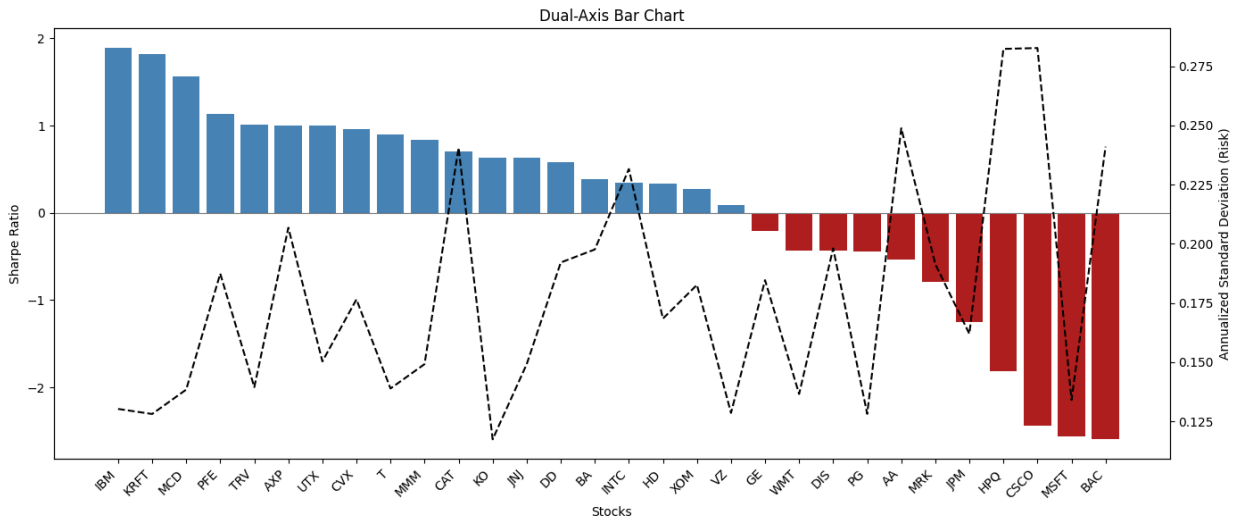
1. Scatter Plot with Regression Line: A standard scatter plot was used to visualize the direct relationship between risk (x-axis) and return (y-axis). A regression line was calculated with SciPy to find the slope of the relationship.



2. Bubble Chart: A scatter plot with a third layer of dimension. The axes remained as the risk vs. return, but the size and color of the bubbles were determined by the Sharpe Ratio. This allows quick identification of the relationship between risk and return (lighter and larger bubbles represent better risk-adjusted returns).



3. Dual-Axis Bar Chart: A sorted bar chart was created to rank the stocks by their Sharpe Ratio. The color of the bar represents whether the Sharpe was positive or negative. A secondary axis was added to plot the standard deviation as a line. This helps visualize if the higher performing stocks required a higher volatility.



Results and Conclusion

The hypothesis test aimed to reject the Null Hypothesis (H_0: Slope ≤ 0) in favor of the Alternative Hypothesis (H_a: Slope > 0).

Statistical Results:

- Regression Slope: -2.92
- R-Squared: 0.31
- One-Sided P-Value: 1.00

Conclusion: The analysis resulted in a negative slope (-2.92), indicating that for this specific dataset and time period, stocks with a higher risk provided lower returns. Due to the negative slope, the one-sided p-value was 1.00. Therefore, the project failed to reject the Null Hypothesis.

Code Structure

The code was organized into a modular Python package to ensure clean execution and separation of functions. The names of the functions allow for easy readability and the entire program can be ran with the following line (the file path needs to be correct and downloaded):

- python3 -m src.main
- To clone the full repository: https://github.com/PartinR/djia_analysis