S&P 500 Macro Analysis and Prediction

This project explores the relationship between the S&P 500 Index and various macroeconomic indicators using Python and advanced statistical techniques. The primary goal is to build a predictive model that can anticipate changes in the S&P 500 based on factors such as interest rates, GDP, CPI, and more.

Key Features:

- **Data Collection:** Automated collection of S&P 500 data using the `yfinance` API and macroeconomic data from the Federal Reserve Economic Data (FRED) using the `fredapi`.
- **Data Cleaning:** Comprehensive data preprocessing, including handling missing values, resampling to monthly frequency, and calculating percentage changes.
- **Exploratory Data Analysis:** Visualization of trends, monthly changes, and correlation analysis to understand relationships between macroeconomic variables and the S&P 500.
- **Lagged Variable Analysis:** Examination of how past values (lags) of macroeconomic indicators impact the S&P 500's performance.
- **Regression Modeling:** Construction of multiple linear regression models to predict the S&P 500's closing prices based on macroeconomic indicators and their lags.
- **Exhaustive Search:** An exhaustive search of all possible variable combinations to identify the best-fitting regression model with significant predictors.
- **Visualization:** Comparison of actual vs. predicted S&P 500 values to assess model performance.

Technologies Used:

- Python
- Pandas, NumPy
- Statsmodels
- Seaborn, Matplotlib
- yfinance, fredapi

Next Steps:

Future enhancements could include incorporating non-linear models, testing out-of-sample predictions, and exploring regime-specific modeling based on economic conditions.

This project serves as a comprehensive framework for analyzing and predicting market movements using macroeconomic data. It can be a valuable tool for financial analysts, data

scientists, and anyone interested in understanding the dynamic relationship between economic factors and market performance.