**Dataset link: https://www.kaggle.com/datasets/camnugent/sandp500**

**1. Description of the Project:**

The project is focused on volatility clustering and GARCH modeling. Volatility clustering refers to the phenomenon where periods of high volatility are followed by other periods of high volatility, and periods of low volatility are followed by other periods of low volatility. GARCH (Generalized Autoregressive Conditional Heteroskedasticity) is a popular model used in finance to capture the volatility clustering effect. The code implements various techniques to analyze and model volatility clustering using GARCH.

**2. List of Outputs with Detailed Descriptions:**

The code provides several outputs, each serving a different purpose. Here is a list of the outputs and their descriptions:

- "data": This output represents the original dataset used for the analysis. It typically includes historical financial data such as stock prices, returns, and other relevant variables.

- "returns": This output contains the calculated returns from the original dataset. Returns are typically used to measure the percentage change in the price of a financial instrument over a given period.

- "log\_returns": This output contains the calculated logarithmic returns from the original dataset. Logarithmic returns are often preferred in financial analysis as they provide more accurate estimates of the continuously compounded returns.

- "volatility": This output represents the calculated volatility estimates using the GARCH model. Volatility is a measure of the dispersion or variability of returns and is an essential component in risk management and option pricing.

- "conditional\_volatility": This output contains the conditional volatility estimates, which are the volatility estimates adjusted for the information available up to each specific time point. These estimates are obtained from the GARCH model.

- "residuals": This output represents the residuals or errors obtained from the GARCH model. Residuals are the differences between the actual returns and the predicted returns based on the GARCH model.

**3. The Main Output:**

The main output of the project is the "conditional\_volatility" series. This output provides time series data of the conditional volatility estimates obtained from the GARCH model. The conditional volatility represents the volatility estimates adjusted for the information available at each time point. It is a crucial output for analyzing and understanding the volatility clustering effect in financial time series data.

**4. Detailed Instructions for Beginners on Running the Code:**

To run the code in the provided, follow these steps:

Step 1: Open the notebook

Step 2: Upload the dataset using link above

Step 3: Run the notebook cells

- The notebook is divided into cells containing code and text explanations.

- Run each cell sequentially by clicking on it and then either pressing the "Play" button or using the "Shift + Enter" keyboard shortcut.

- Ensure that you have installed all the necessary libraries and dependencies mentioned at the beginning of the notebook.

Step 4: Observe the outputs

- As you run each code cell, the outputs will be displayed below the corresponding cell.

- Take time to understand the visualizations and analysis presented in the outputs.

- Modify the code or experiment with different parameters to explore further if desired.

By following these instructions, beginners should be able to execute the code, explore the outputs, and gain insights into clustering and dimensionality reduction techniques applied to the dataset.