FINAL PROJECT REPORT

Analysis of Google's Daily Returns

Data Overview

The dataset spans from (June 28, 2022) to (November 25, 2022) covering several stock market variables for each trading day. Key columns include Open, High, Low, Close, Adjusted Close, Volume, and Daily Returns (Dis Ret).

Disclaimer:

The analysis is based on historical data and may not guarantee future market performance. The models developed are subject to limitations and should be used with caution in real-world trading decisions.

This report provides an overview of the relationships observed in the dataset, suggesting potential avenues for further investigation and refinement in predictive models.

Descriptive Statistics for Daily Returns

Mean The average daily return over the observed period.

Standard Error: The standard deviation of the sample mean, representing the accuracy of the sample mean in estimating the population mean.

Median: The middle value in the dataset, indicating the 50th percentile or the value separating the higher half from the lower half.

Mode: The value that appears most frequently in the dataset.

Standard Deviation: A measure of the amount of variation or dispersion in the dataset.

Sample Variance: A measure of the spread of data points around the mean.

Kurtosis: Describes the shape of the distribution, specifically how much data is concentrated near the mean.

Skewness: Measures the asymmetry of the distribution, determining the lack of symmetry in the dataset.

Range: The difference between the maximum and minimum values, showing the spread of values in the dataset.

Minimum and Maximum: The smallest and largest values observed in the dataset, respectively.

Sum: The total sum of all daily returns.

Count The number of data points or observations in the dataset.

- **Mean:** -0.00135915
- **Standard Error:** 0.002516534
- **Median:** -0.0026021
- **Mode:** Not Available
- **Standard Deviation:** 0.025663711
- **Sample Variance:** 0.000664959
- **Kurtosis:** 2.386700353
- **Skewness:** -0.226217408
- **Range:** 0.175919
- **Minimum:** -0.10131
- **Maximum:** 0.074606
- **Sum:** -0.142711
- **Count:** 105

Summary - Certainly! Based on the provided descriptive statistics for daily returns:

-Mean: The average daily return over the observed period is slightly negative, indicating that, on average, the returns are below zero.

- Standard Error: The standard deviation of the sample mean is relatively small (0.0025), suggesting that the calculated mean is likely a good estimate of the population mean.

- Median: The middle value in the dataset (50th percentile) is lower than the mean (-0.0026), indicating that there might be a slight left skew in the distribution of returns.

- Mode: Not Available - there isn't a specific value that appears most frequently in the dataset.

- Standard Deviation: The measure of variation or dispersion in the dataset is moderate (0.0257), signifying that there is some variability in the daily returns.

- Sample Variance: The spread of data points around the mean is relatively small (0.000664959), suggesting that data points are somewhat close to the mean on average.

- Kurtosis: The kurtosis value (2.3867) indicates a distribution that is more peaked and has heavier tails than a normal distribution (leptokurtic). This means the data has more extreme values than a normal distribution.

- Skewness: The skewness value (-0.2262) indicates a slight negative skew, suggesting that the distribution may have a slightly longer tail on the left side.

- Range: The difference between the maximum and minimum values is relatively wide (0.1759), indicating significant variability in the daily returns.

- Minimum: The smallest observed daily return is -0.10131, indicating the lowest single-day loss within the observed period.

- Maximum: The largest observed daily return is 0.074606, representing the highest single-day gain within the observed period.

- Sum: The total sum of all daily returns is negative (-0.142711), indicating that, overall, the sum of returns over the observed period is below zero.

- Count: There are 105 data points or observations in the dataset.

It suggests that the dataset might have some outliers or extreme values (as indicated by kurtosis and skewness) and demonstrates a somewhat volatile pattern in daily returns with a slightly

negative average return. The presence of a negative skew and leptokurtic distribution may indicate a less symmetric and more volatile dataset compared to a normal distribution.

Hypothesis Testing: One-Sample T-Test

The hypothesis test aims to infer whether the population mean return is less than 0 using the given sample data.

Null Hypothesis (H0): Assumes that the population mean return is greater than or equal to 0.

Alternative Hypothesis (H1): Suggests that the population mean return is less than 0.

T-Statistic: The calculated statistic that measures how far the sample mean is from the assumed population mean.

P-Value: The probability of observing a t-statistic as extreme as the one obtained, assuming the null hypothesis is true.

The hypothesis you've described is a one-tailed test comparing the population mean return against zero. Here's what the results indicate:

- **Null Hypothesis (H0):** The population mean return is greater than or equal to 0.
- **Alternative Hypothesis (H1):** The population mean return is less than 0.

Step	Description
Null Hypothesis (H0)	The population mean return is greater than or equal to 0.
Alternative Hypothesis (H1)	The population mean return is less than 0.
T-Statistic	The calculated statistic measuring sample mean's distance from the assumed population mean.
P-Value	The probability of observing a t-statistic as extreme as the obtained one, assuming the null hypothesis is true.

Given the provided statistical outputs:

- **T-Statistic:** -0.54008952 (Formula = (AVERAGE (A1:A10)-AVERAGE(B1:B10))/(Pooled Standard Deviation*SQRT(1/COUNT(A1:A10)+1/COUNT(B1:B10)))
- **P-Value:** 0.295144981 (Formula = **=**2*TDIST(ABS(T-Statistic), Degrees of Freedom, 2)

Conclusion:

- **Result:** The obtained p-value of 0.295 is greater than the typical significance level of 0.05 (assuming a common alpha), indicating insufficient evidence to reject the null hypothesis.
- **Interpretation:** With this p-value, we do not have enough statistical evidence to conclude that the population mean return is significantly less than 0 based on the provided sample data. Therefore, we fail to reject the null hypothesis

- **Model Comparison:**

- **Model 1:**
- **R-Square & Adjusted R-Square:** The R-squared and Adjusted R-squared values measure how well the model fits the data. In this model, the %Vol_t-1 explains about 5.5% of the variance in stock returns.
- **Significance (P-value):** The p-value of 0.017 for %Vol_t-1 indicates its statistical significance in predicting stock returns in this model.
- **Model 2**:
- **R-Square & Adjusted R-Square:** Utilizing %Vol^2, this model explains approximately 5.7% of the variance in stock returns. However, the increase in predictive power compared to Model 1 is minimal.
- **Interpretation:** While the addition of the squared term (%Vol^2) improves the model slightly, the improvement is not substantial compared to Model 1.
- **Model 3**:
- **R-Square & Adjusted R-Square:** Using %Vol_t-1 and its squared term (%Vol_t-1^2), this model explains around 10.8% of the variance in stock returns.
- **Significance (P-values):** Both %Vol_t-1 and its squared term (%Vol_t-1^2) are statistically significant predictors (p-values < 0.05).

- **Model 4:**
- **R-Square & Adjusted R-Square:** This model includes %Vol_t-1, its squared term, and Dis Ret_t-1. It explains 10.9% of the variance in stock returns.
- **Significance (P-values):** While %Vol_t-1 and its squared term are significant, Dis Ret_t-1 (past stock returns) does not significantly predict future returns.

Model	Predictors	R-Square & Adjusted R-Square	Interpretation
Model 1	%Vol_t-1	5.5%	%Vol_t-1 is statistically significant (p-value: 0.017) and explains 5.5% of the variance in stock returns.
Model 2	%Vol^2	5.7%	The inclusion of %Vol^2 marginally improves the model's explanatory power but not substantially better than Model 1.
Model 3	%Vol_t-1, %Vol_t-1^2	10.8%	%Vol_t-1 and its squared term are both statistically significant (p-values < 0.05) and collectively explain 10.8% of variance in stock returns.
Model 4	%Vol_t-1, %Vol_t-1^2, Dis Ret_t-1	10.9%	%Vol_t-1 and its squared term remain significant, but Dis Ret_t-1 does not significantly predict future returns. The model explains 10.9% of variance in stock returns.

Recommendations:

Further analysis and refinement of the models could enhance predictive accuracy. Exploring additional variables or refining existing ones might provide deeper insights into stock market movements.

- **Conclusion and Interpretation:**
- **Best Predictor:** Model 4 stands out as the most effective predictor, indicated by its higher Adjusted R-Square value.
- **Trading Volume's Influence:** Models 3 and 4 demonstrate the importance of trading volume data (%Vol_t-1) and its transformation (%Vol_t-1^2) in predicting stock returns.
- **Usefulness of Trading Volume Data:** The significance of trading volume-related variables in Models 3 and 4 signifies that trading volume data holds predictive power for future stock returns.
- **Limitation:** Despite trading volume's significance, the relatively low R-squared values across all models suggest the need for additional factors beyond trading volume for more accurate predictions.
- **Overall Assessment:**
- **Trading Volume's Role:** While trading volume data demonstrates predictive power, it might not solely dictate stock returns. Including other relevant variables is crucial for creating more accurate models.
- **Enhancing Predictive Accuracy:** Considering additional relevant factors alongside trading volume data could potentially enhance the accuracy of predictions for future stock returns.
- This detailed analysis emphasizes that while trading volume data is informative, a comprehensive approach incorporating various factors beyond just trading volume is crucial for creating robust models to forecast stock returns accurately.