Analyzing the relationship between stock prices and trading volume using correlation analysis

Objective: The objective of this analysis task is to determine whether there is a relationship between the trading volume and stock prices of a particular company, and to what extent the two variables are correlated.

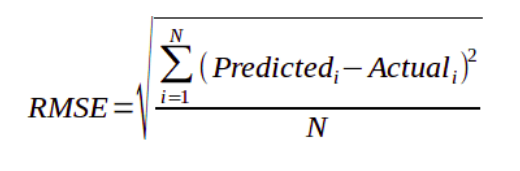
Data: The data for this analysis task will consist of historical stock prices and trading volume data for a particular company. Data is provided wherever needed.

Task Requirements:

* Use [daily](https://drive.google.com/file/d/1Moq5ki20dv_xrNyxCJeXsXEW6jm7uhcF/view?usp=share_link) data Calculate the correlation coefficient between some parameter of stock prices and some parameter of trading volume. What does the value of correlation depict? Draw necessary charts/plots for the same. You can use Excel for the same.
* Now, you can perform regression analysis in which trading volume can be used to predict the stock price. This analysis will involve fitting a linear regression model to the data, with the trading volume as the independent variable and the stock price as the dependent variable. The regression model will provide an equation that can be used to predict the stock price based on the trading volume. You can use Excel for the same. Try to learn the internal working of regression analysis.
* Identify any outliers or influential data points that may be impacting the correlation analysis, and determine whether removing these points would change the results.

You can use the Scatter plot analysis in which X-axis represents trading volume and Y-axis represents Stock price. Then you can use Cook’s distance/Leverage plots for this part. Determine whether removing these points changes the result or not. Also, comment on the purpose of this method of removing outliers.

* Use different time frames (such as [weekly](https://drive.google.com/drive/folders/1xBt3LO1Fd8mg2mboV9F_2DeQyx9Qcqeo) and [monthly](https://drive.google.com/drive/folders/1xBt3LO1Fd8mg2mboV9F_2DeQyx9Qcqeo) data) to see if the relationship between trading volume and stock prices changes over time. Perform all above sub-parts for both of these datasets. Comment on obtaining different levels of correlation between trading volume and stock prices.
* Take the [dataset](https://drive.google.com/file/d/1-1PDpEKDU4k7zXnaN8keeqAY4NXBeOtP/view?usp=share_link) (same stock). The timestamps in this dataset are a continuation of those present in the first part. Using the regression model modeled in the second part, predict the stock price for every timestamp in this dataset. You can add this predicted series as a column in the dataset. Now using two columns, (Adj Close and Predicted Price) calculate RMSE (Root Mean Squared Error) of the model. You can use Excel Root Mean Squared Error has a formula:



Suggest ways through which we can reduce the RMSE value (You may or may not implement these methods).

* Write conclusions of your study with reason for a particular correlation between trading volume and stock price.
* \*Also, add a section of your understanding with the terms- volatility and liquidity. Also, understand the relation between volatility, liquidity and traded volume with each other.

Make a report of maximum 6 pages (Max 1 page for \*marked part and Max 5 pages for rest of the parts) including your analysis, plots/charts, observations, etc. It is advised to include the references in the report.

Note: (1) Throughout the task, use Adj Close as a stock price of the stock for that particular timestamp.

(2) Feel free to use the internet to understand different terms and methodologies you encounter while solving the task.

(3) You have the option to use Excel or SPSS to complete the task, and coding language is not required but can be used if you prefer.

(4) For broader analysis and comparison of correlation between stock price movement and volume, you may do the above analysis with a few more datasets as listed below (This part is not mandatory, this is for your understanding):

(i) [data\_1](https://drive.google.com/file/d/1M2e3vdS6g80s088b1HeL_xcF2jeppBjU/view?usp=share_link)

(ii) [data\_2](https://drive.google.com/file/d/1h95KMVVsYdrL_bk-GGjxT4wL3_0rYezs/view?usp=share_link)

(iii) [data\_3](https://drive.google.com/file/d/124OQ_0FVEfv5oze66_b9lQ3ZBjzxWcPI/view?usp=share_link)

(5) Adj Close and Volume can be considered as parameters of Adj Close or Volume. By parameter we mean any mathematical function which can logically define the closest relation with the other term.