**Dow Jones Industrial Average**

Table of Contents

[Introduction: 3](#_Toc89519239)

[Literature Reviews: 3](#_Toc89519240)

[Data Collection: 4](#_Toc89519241)

[Methodology and Findings Analysis: 6](#_Toc89519242)

[Conclusion: 14](#_Toc89519243)

[References 15](#_Toc89519244)

[Appendix: 16](#_Toc89519245)

# Introduction:

The Dow Jones Industrial Average (DJIA) is a stock market index that monitors 30 big, publicly traded blue-chip businesses that trade on the New York Stock Exchange and the Nasdaq Stock Exchange. The Dow Jones is named after Charles Dow, a business colleague of Edward Jones who founded the index in 1896.

The Dow Jones Industrial Average is the country's second-oldest stock market index, after the Dow Jones Transportation Average (DJTA). The DJIA was created as a proxy for the US economy's overall strength. The DJIA, also known as "the Dow," is one of the most widely followed stock market indices in the world. The Dow Jones Industrial Average is made up of a diverse group of companies, all of which are blue-chip companies that routinely generate substantial profits.

# Literature Reviews:

The **GARCH** process, also known as generalized autoregressive conditional heteroskedasticity, is an econometric term developed by economist Robert F. Engle, who earned the Nobel Memorial Prize in Economics in 2003. The GARCH approach is a method for assessing the volatility of financial markets.

GARCH modelling can be done in a number of methods. Financial specialists frequently utilize the GARCH process to estimate the prices and rates of financial instruments since it provides a more realistic backdrop than other models(Franses, et al., 1996).

Many scholars have published their findings and proposed other models based on the GARCH model.

The ARCH (Autoregressive Conditional Heteroskedasticity) and Generalized ARCH (GARCH) models have emerged as the most popular techniques for assessing volatility due to their ability to describe the unpredictable movement of financial data series. Many researchers have looked at how well GARCH models explain the volatility in mature stock markets over time, but only a few have used daily data from stock markets in Central and Eastern Europe to test them (see, **for example,** Kash-Haroutounian and Price, 2001; Poshakwale and Murinde, 2001; Murinde and Poshakwale, 2002; Patev and Kanaryan, 2006; Lupu and Lupu, 2007; Miron and Tudor, 2010; Harisson and Moore, 2011). Our research examines how to forecast stock market volatility in Romania, a market that has gotten little attention. According to the findings of various studies, asymmetric GARCH models fit better stock market returns volatility for emerging CEE countries. According to Lupu and Lupu, an EGARCH (Exponential GARCH) model is adequate for the logarithmic returns of the Romanian composite index BET-C spanning the period 03/01/2002-17/11/2005. (2007). Miron and Tudor (2010) tested multiple asymmetric GARCH-family models using daily stock return data from the United States and Romania from 2002 to 2010. (EGARCH, PGARCH, and TGARCH). They discovered that the EGARCH model's volatility estimates have a lower forecast error and are hence more accurate than the estimates of the PGARCH and TGARCH models. Harisson and Moore (2011) examined stock market volatility in 10 CEE stock exchanges from 1991 to 2008 (Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Poland, Romania, Slovenia, and the Slovak Republic). According to their findings, models that allow for asymmetric volatility outperform all other models studied.

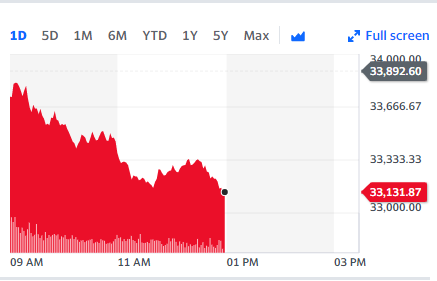
In 2017 Sulaeman Rahman Nidar et. al. presented a research article regarding the dow jones effect on global stock index.The results obtained from this study indicate that the stock market index index and therefore the STI index have a major positive result on the movement of foreign investments within the Stock Exchange. In contrast, the movement of world oil costs and charge per unit of the IDR/USD have a significant negative effect on the movement of foreign investments in the BEI.

In 2017 Yrd.Doç.Dr. Hakan YILDIRIM presented a research article which showed the effects of Dow Jones on BIST100 index, thus showed the volatility of the data.

# Data Collection:

The given assignment was to analyze the data on the basis of market risk and do the exploratory data analysis. The data which was given to us was Dow Jones latest data which was based on the current market rates and risks. It contains the DJIA calculations and the years. The dataset also contains some null values which were denoted by NA in the dataset for that we have to run the command na.rm(dataset) which removes the NA values in the datasets and gives us the values like mean, median, variance, correlation without those of the NA values thus this how calculated the results after removing the NULL values and to get NULL values we ran the command na.omit(dataset) and it gave us all the attributes with the NULL values thus this is how the data collection process worked.

The DJIA dataset shows us the current market trends and the volatility of the market time by time, it even tells how much share market goes up and down. The plot below is given by the yahoo research of share markets with the help of the DJIA formula, the plot for the industrial volatility by the time looks like below.



The above plot shows the Dow Jones index values the index values of the above plot can be calculated. S&P Dow Jones Indices calculates multiple return types which vary based on the treatment of regular cash dividends. The classification of regular cash dividends is determined by S&P Dow Jones Indices.

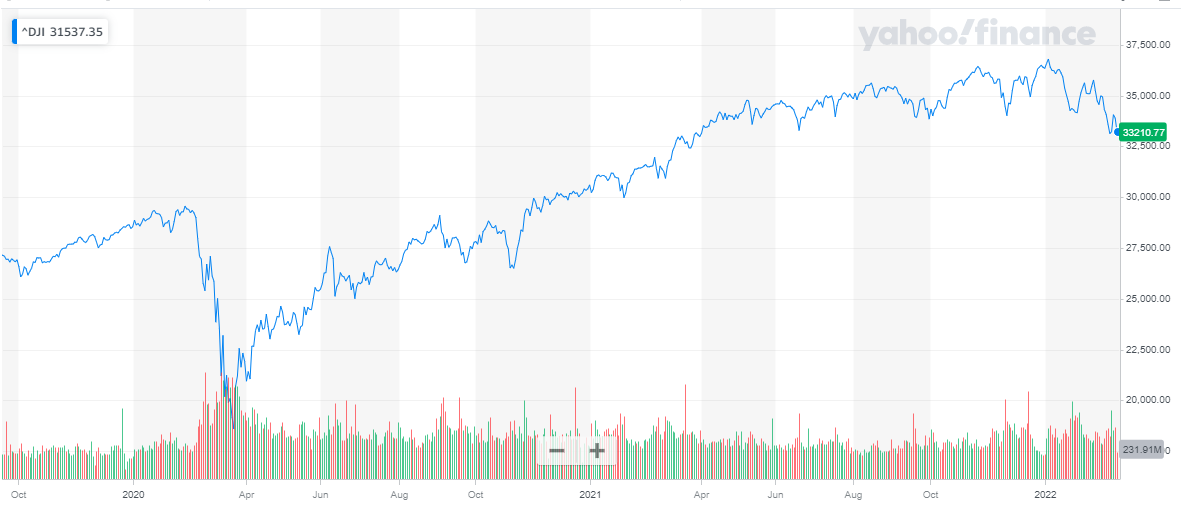
• Price Return (PR) versions are calculated without adjustments for regular cash dividends.

• Gross Total Return (TR) versions reinvest regular cash dividends at the close on the ex-date without consideration for withholding taxes.

• Net Total Return (NTR) versions, if available, reinvest regular cash dividends at the close on the ex-date after the deduction of applicable withholding taxes.

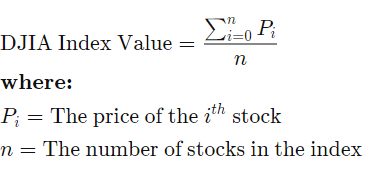
In the event there are no regular cash dividends on the ex-date, the daily performance of all indices will be identical.

The Dow Jones Industrial average plot contains the data of the market and this year as well, it shows us how much volatile the market can be in few hours of the day. The plot looks like below.



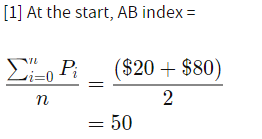
The above plot shows the market average and industrial trend of the last year and this year as well, by seeing the above graph we can clearly identify the current market trends and industrial ups and downs as well.

The formula to calculate Dow Jones Industrial average index values is dependent on several parameters. These parameters are the price weighted index, market cap weighted index for example S&P 500. The index is calculated by adding the stock prices and then we divide it by the divisor. The formula for the index value calculation is below.

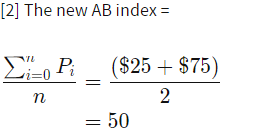


Assume that the exchange constructs a mathematical number represented by "AB Index," which is being measured on the performance of the two stocks (A and B). Assume that stock A is trading at $20 per share and stock B is trading at $80 per share on day 1.

Applying the initial concept of Dow to our hypothetical example of AB index.



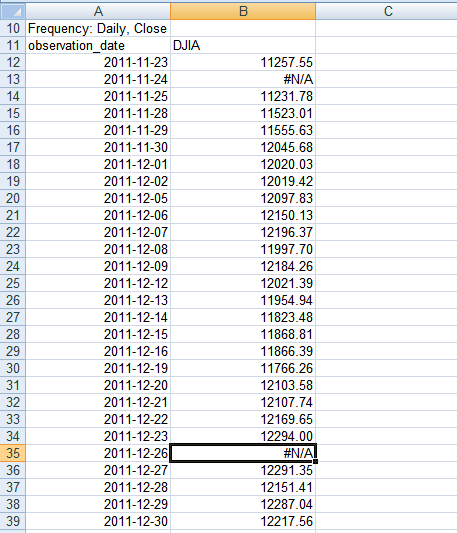
Similarly for the Day 2 calculation goes like below,



Here we can clearly see that the outcomes are similar but the ranges change time by time. Thus this is how we calculate the dow jones index values.After the collection, we did the conversion of the dataset to XTS data which is a time series notion of the dataset and gives the results with respective times and dates. After this we again did the same work which was finding the NULL values in the current time series dataset, removing those NULL values from the time series dataset, and preparing the data for the further analysis part which the calculation of risk, confidence intervals, and other processes like calculating the skewness and kurtosis, returning the weekly, monthly and yearly data from the whole dataset. Thus, this is how our data collection and data preprocessing work is completed.

The dataset was based on the market financial conditions and its ups and downs on the basis of the DJIA coefficients and tells whether the market will go up or down according to the stipulated time thus this is the way we can say that the dataset tells the market risk analysis and tells which market is safe and which is not.

This dataset looks like below we did the preprocessing of the data using some basic R commands which are to make the dataset suitable for the analysis process. The image below tells us how the dataset looks like.



The above image shows the dataset which is known as the FRED graph, FRED stands for the federal reserve financial statistics graph. FRED was founded in the early 1990s as an extension of the Federal Reserve Bank of St. Louis' long-standing tradition of giving financial information to help understand better the Fed's policy actions. The data was delivered in list form on a dial-in, electronic bulletin board system prior to the widespread use of the World Wide Web. The information was divided into categories with about 300 data series each, and then extended from that. Interestingly, FRED did not start out as part of a great plan or strategic goal. Rather, it grew in a really natural way over time. Staff from the St. Louis Federal Reserve, whether directly involved in the FRED project or operating on the outskirts, developed ad hoc tools for the database.

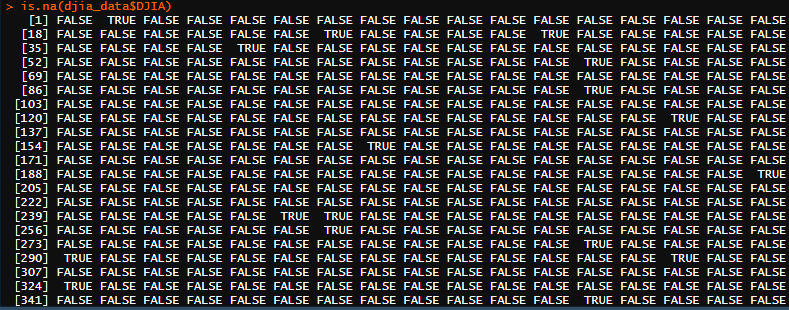
The Scientific Department is committed to the database's expansion. Some of the most well numbers released by the Governing board, since its inception, FRED has featured data from the Bureau of Economic Research, the Bureau of Labor Statistics, and the Census. Over time, FRED's collection has expanded to include a wide range of global, national, and local data. More recently, it's become clear that data relevant to additional themes and geographies must be included if FRED is to continue serving its users, guaranteeing that it develops and evolves. Considerable care will be used to enter data logically and methodically.

# Methodology and Findings Analysis:

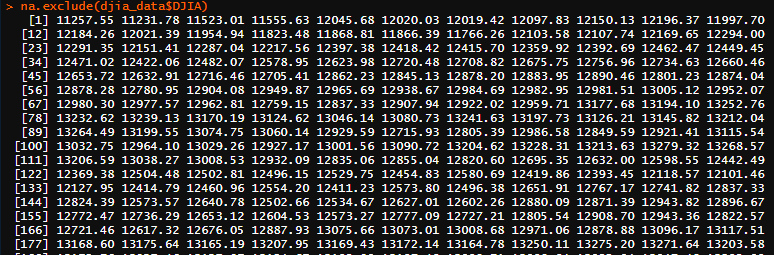
The analysis method takes place with the basic statistical operations but before performing the operations we have to prepare the dataset for the operations like calculating mean, median, modes, variance, deviations, etc. For these operations, we first converted our dataset to a simple time series dataset which can tell the difference in the output by the difference of the particular time differences thus we can analyze the dataset on the basis of those time series operations.

At first, we converted the dataset using the code which is for the conversion of a simple dataset to a proper time series dataset, for this process we ran the code.

The code gives the result below which tells the null values first then it tells us the dataset information which gives us the data for the given particular date and time. The output looks like below.



The result after removing the NULL values looks like below, we ran the remove NA command which na.excludes (dataset), the below output shows the results of the command we ran to remove NULL values.



After the above processes, we ran an analysis of the dataset and plotted the chart series data for the Dow Jones Industrial Average dataset & the command which we ran for the plotting of the dataset is **chart series (DJIA, type = "bars", theme="white")** this command gives us the following plot which tells us the volatility of the market and how much ups and downs are the part of the market and risk analysis in the market.



The expansion and volatility of the market looks below which we calculated from the dataset; the plot looks like below.

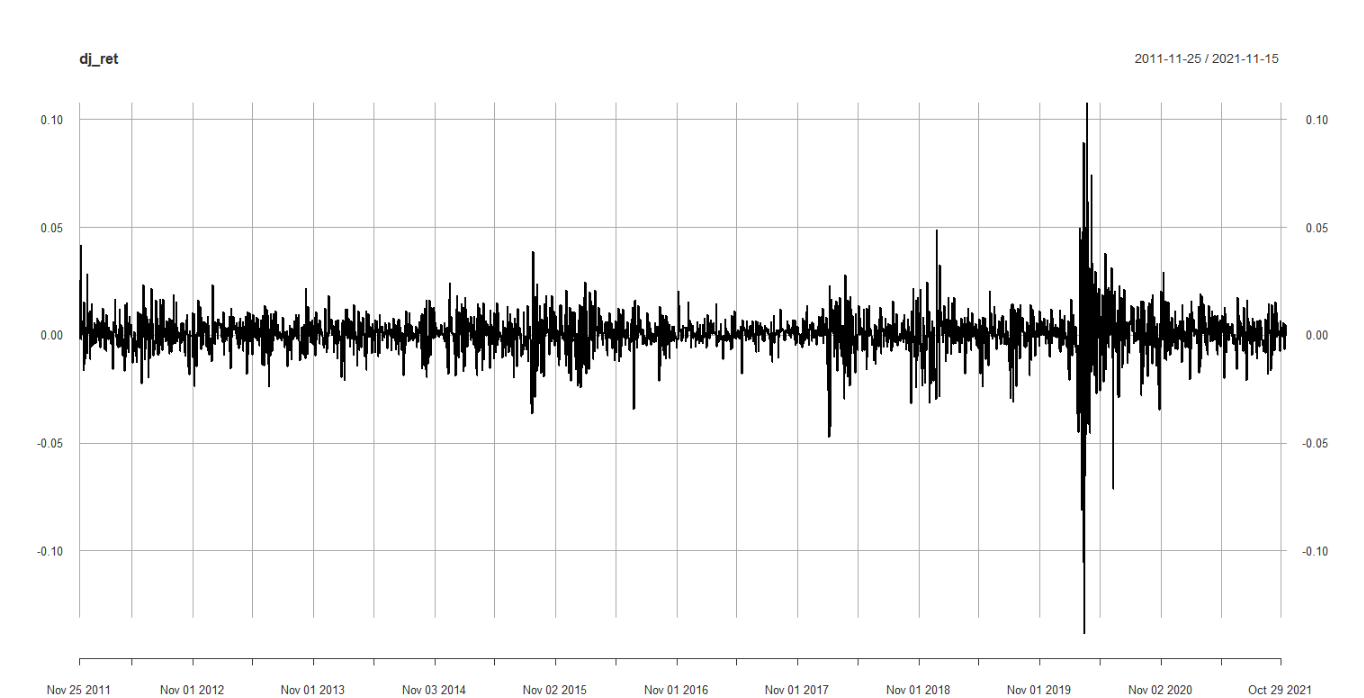
The formula for the log return, which shows the volatility of the market goes like,



**The above plot gives us a clear picture of the market and how much volatile it is with respect to risks and demands, thus we can clearly analyze the volatility of the market. In the above graph we can see that market is volatile and the increasing demands are increasing the level of risk in the maket. Thus on the basis of above we can clearly say that volatility and risk are directly proportional to each other as per the market scnerio,**

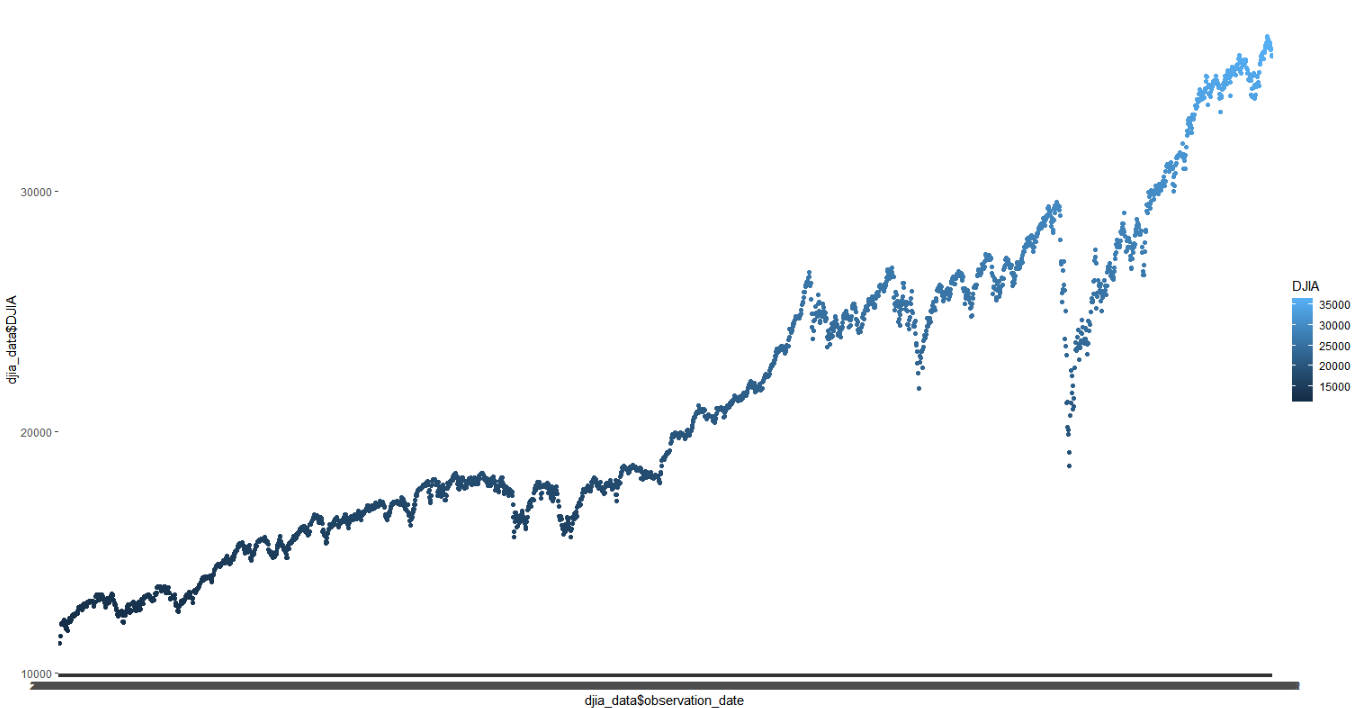
After the above process, we calculated the returns given by the current volatile market, for this process we again created another data frame from the processed dataset and ran the below code for the process.

After this process, we also plotted the returns in the volatile market which looks like below.



**We can see that market gives us the results as per the volatility of demand for any commodity if the demand is constant the plot gives constant going but as soon as the demand grows and shows the spike in the volatility of the demand and thus tells us that the risk level is somewhat higher but return is good and hefty. Thus this how we can see the volatility of the demand in the market.**

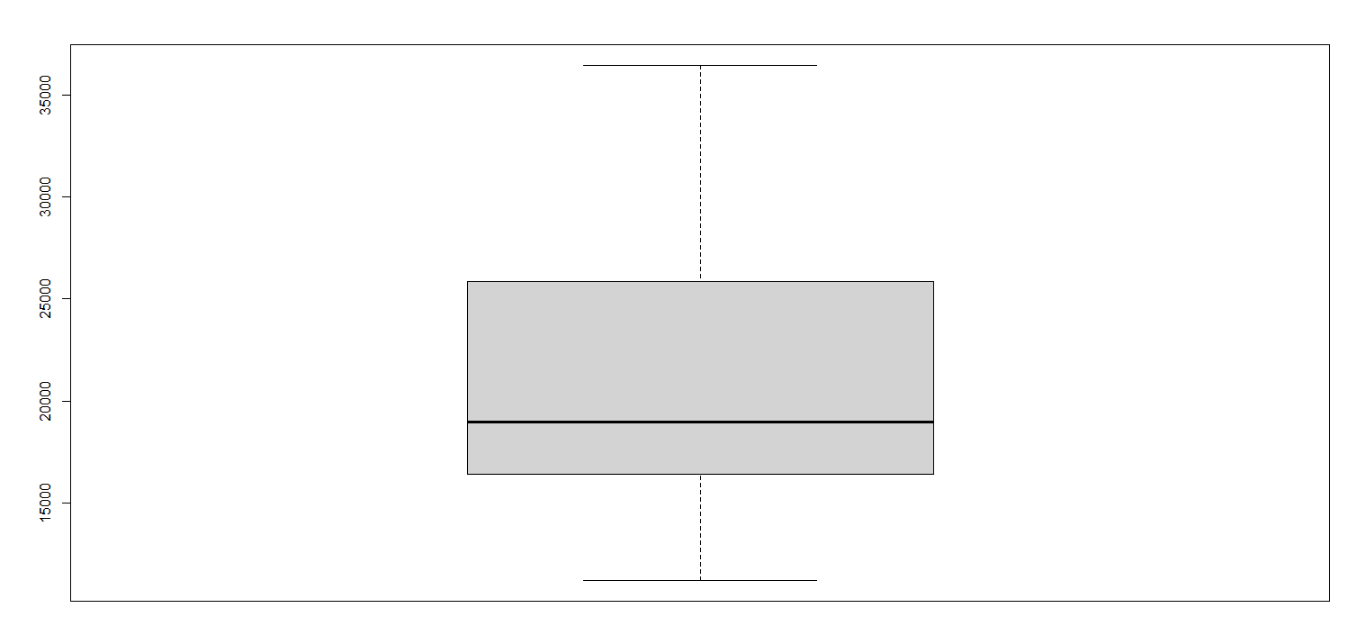
We have also plotted our results using the date time data which gives us the below results, this shows how the market changes by the time and so the index values go up and down accordingly.



After the above analysis, we ran the basic statistical operations and ran the commands to calculate the mean, median, and other statistics related data from our newly created data frame for that we ran the following commands.

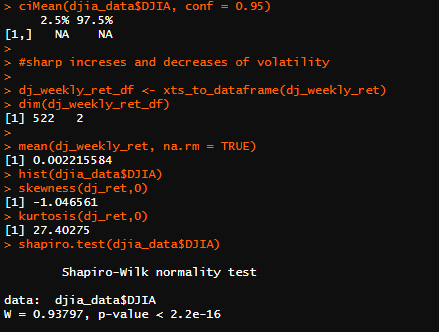
The results for the commands are below which shows the results of all the statistical calculations and operations.

After the processes we plotted the boxplot for the statistical operations the boxplot looks like below.



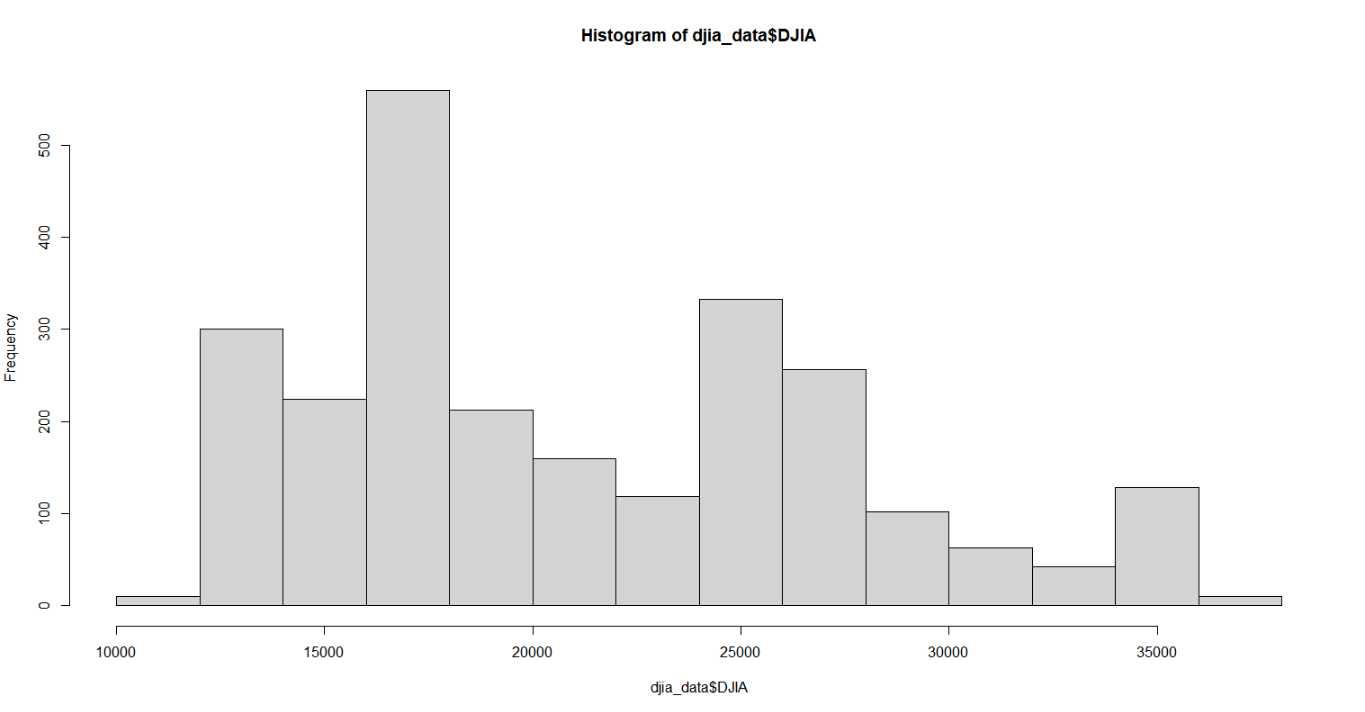
**The above boxplot basically shows the results of the various statistical operations which tells that the rate of statistical operations is constant and thus these operation will give the data regarding the partular market.**

After the above processes, we ran the command to generate the confidence intervals and did the Shapiro Wilk test to test the confidence at 95 percent thus this is how we have done our final process for our DJI dataset, the operations of the confidence intervals look like below.  
The operation gives the following result of the Shapiro Wilk test and confidence intervals of the given dataset and our processed data frame.



The above output shows the result of the Shapiro wilk test and it shows that there is 97.5 percent confidence interval, which tells us that the frequency of demands and return rates are 97.5 percent accurate and there is no overfitting at all.

The plot of our final processes looks like below.



Thus after the Shapiro Wilk Test plotted the final graph, in the above graph we can see the Dow Jones Industrial average and the change of the values as per the frequency of demand which clearly shows us the volatility of market and thus we can say that market of today’s era is always in flux because the volatility in demand can make large changes.

# Conclusion:

The process of data analysis is something which requires a lot of prerequisites to be fulfilled like data pre-processing, removal of NULL values, removing the useless attributes, dimension reductions and most importantly converting the dataset to the suitable time series data frame, hence it is the process which needs patience and knowledge of statistical operations apart from the testing of our results and confidence all these processes are important and they make the output more robust for the random inputs. As the finding we can see the last plot which is directly connected to the frequency of demand and change in the dow jones industrial average, the increase in demand is directly proportional to the increase in industrial average values. Thus this is how we completed the market risk analysis of the dataset of the Dow Jones index and its industrial market risk analysis.

# References

* Orlowski, L.T., 2012. Financial crisis and extreme market risks: Evidence from Europe. Review of Financial Economics, 21(3), pp.120-130.
* Karagiannis, N. and Tolikas, K., 2019. Tail risk and the cross-section of mutual fund expected returns. Journal of Financial and Quantitative Analysis, 54(1), pp.425-447.
* Ni, Z., Wang, L. and Li, W., 2021. Do fund managers time implied tail risk?—Evidence from Chinese mutual funds. *Pacific-Basin Finance Journal*, p.101590.
* Cui, W. and Yao, J., 2020. Funds of hedge funds: Are they really the high society for little guys?. *International Review of Economics & Finance*, *67*, pp.346-361.
* Team, R.C., 2013. R: A language and environment for statistical computing.
* Barker, C., 2009. The mean, median, and confidence intervals of the Kaplan-Meier survival estimate—computations and applications. *The American Statistician*, *63*(1), pp.78-80.
* Reimann, C., Filzmoser, P., Garrett, R. and Dutter, R., 2011. *Statistical data analysis explained: applied environmental statistics with R*. John Wiley & Sons.
* Demirtaş, H., VARDAR ACAR, C., Gao, R., Aydemir, İ. and Yabaci, A., 2020. An Improved Textbook Rule on the Mean-Median Inequality for Discrete Data. *Turkiye Klinikleri Journal of Biostatistics*, *12*(2).
* Franses, P.H. and Van Dijk, D., 1996. Forecasting stock market volatility using (non‐linear) Garch models. *Journal of Forecasting*, *15*(3), pp.229-235.

# Appendix:

