Saudi Digital Academy

CodingDojo

Machine Learning Capstone Project

Tadawul Dataset

Members:

Ibrahim Aljadani (Leader)

Rawan Aljasser

Shahad Almalki

Ranin Alghamdi

Hamad Alodah

Supervisor:

Joseph Issa

Table OF Content :

Business Problem :

Predictions in machine learning allow businesses to make an accurate assumption as to the likely outcome of a question based on historical data. These predictions give businesses insights that result in tangible business value. And because Vision 2030 has a special program to develop the financial sector, we were interested in forecasting the data of the Tadawul market, as Tadawul is the largest capital market in the Middle East and North Africa region, and the ninth largest stock market in the world since December 2019, it has also started to attract foreign investment within the Saudi stock market as well.

Financial sector Development In Saudi's 2030 Vision :

The Financial Sector Development Program (FSDP) is one of several executive programs developed to help achieve the objectives of Vision 2030. The Financial Sector Development Program aims to develop a diversified and effective financial sector, focus on building infrastructure to support the development of the national economy, diversify its sources of income, and stimulate savings, finances, and investments. By boosting financial sector institutions, and by developing the Saudi financial market to become an advanced capital market without weakening the financial sector's stability.

The program is underpinned by the following three pillars:

- Developing an advanced capital market
- Enabling financial institutions to support private sector growth
- Promoting and enabling financial planning

There are several sub-objectives under the first pillar, which align with Tadawul's mandate. One is to facilitate the raising of capital by the government and the private sector. This includes supporting the process of privatization of state-owned entities, which will open up new opportunities for investors. Another is creating an efficient platform to encourage investment and diversify the investor base. Tadawul is contributing to this objective by drawing more foreign investors, increasing the efficiency of trading, and promoting new asset classes. A third objective is to provide a safe and transparent infrastructure, that includes an upgrade of the post-trade and risk management model, supported by the establishment of a clearinghouse based on Muqassa principles.

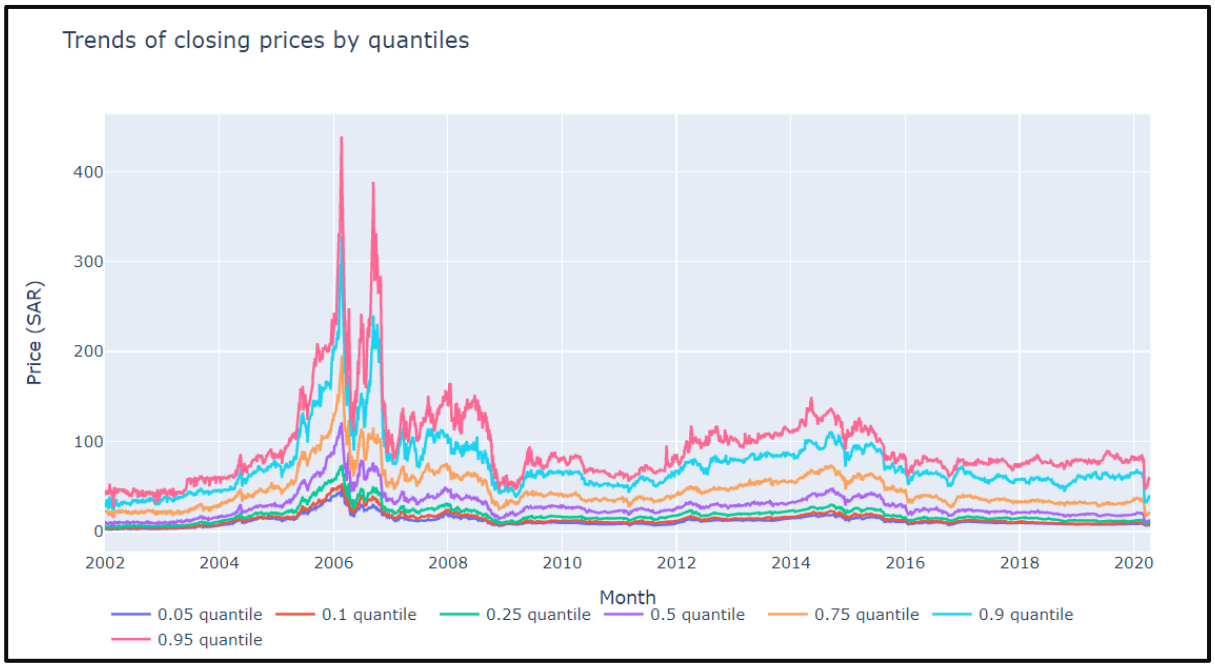
The program achieved great success after having the Saudi Stock Exchange "Tadawul" join the global indices "FTSE" and "MSCI", which contributed to the success of the initial public offering of Saudi Aramco shares in 2019.

Dataset

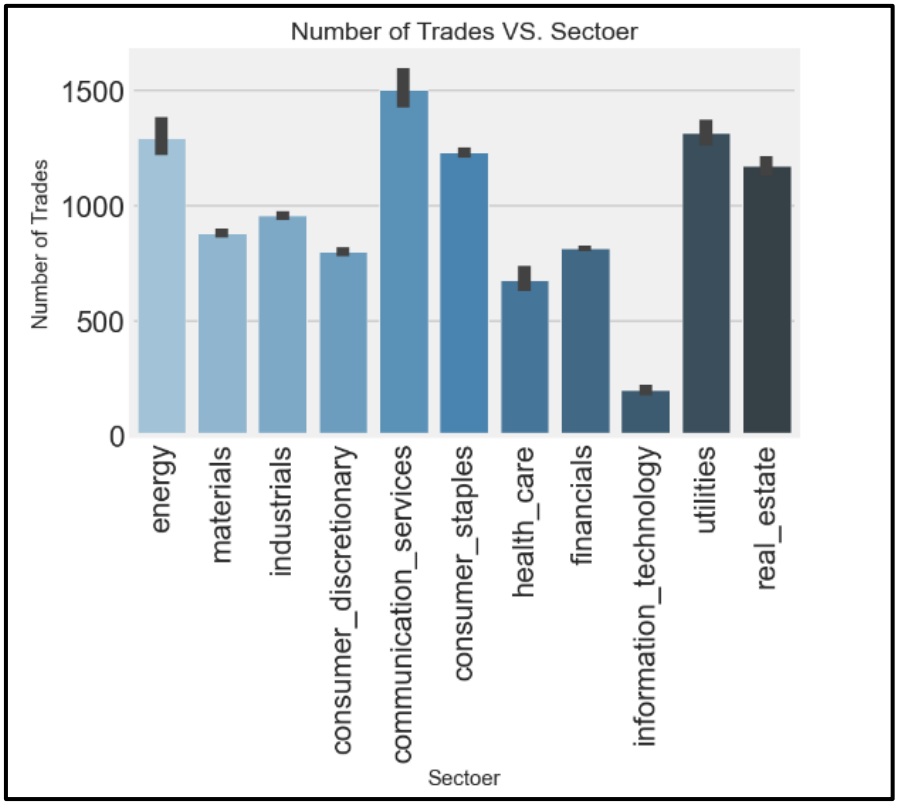
Tadawul Dataset

This is the data of Saudi stock market companies since 2000-01-01. It was collected from Saudi Stock Exchange Tadawul. Tadawul is the main stock exchange in the Kingdom of Saudi Arabia. The market value of shares of companies listed on the main market is estimated at 9.8 trillion Saudi riyals.

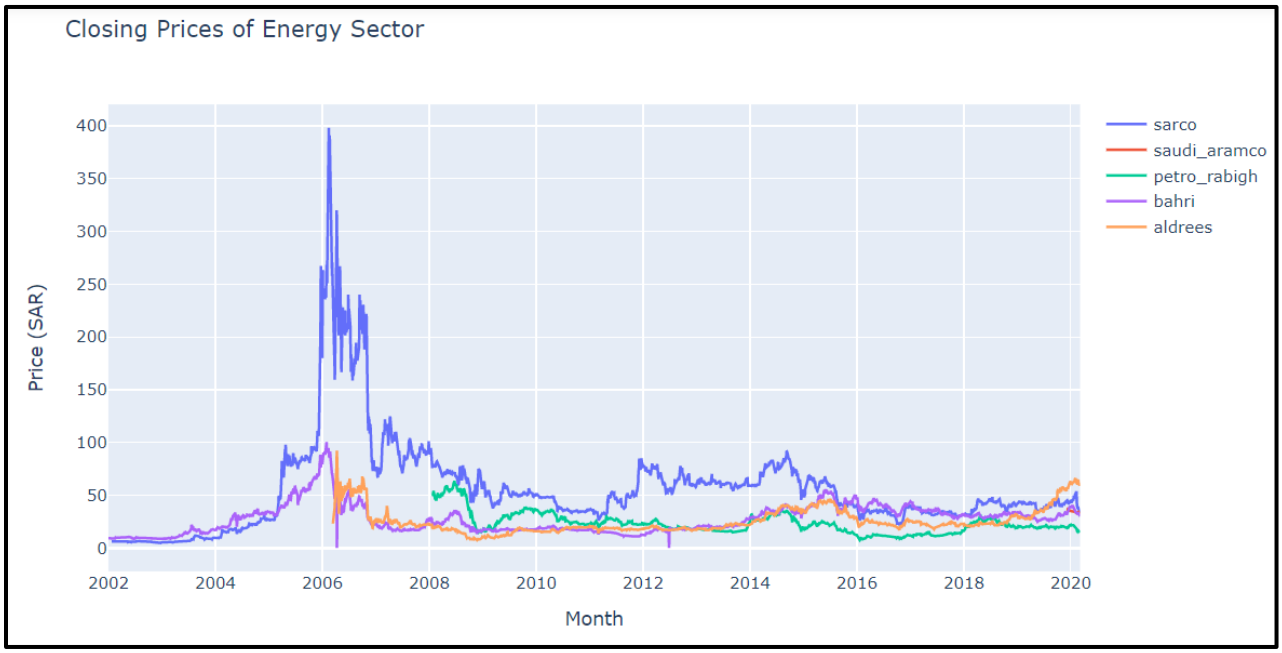
Visualize (Exploratory Data Analysis) :



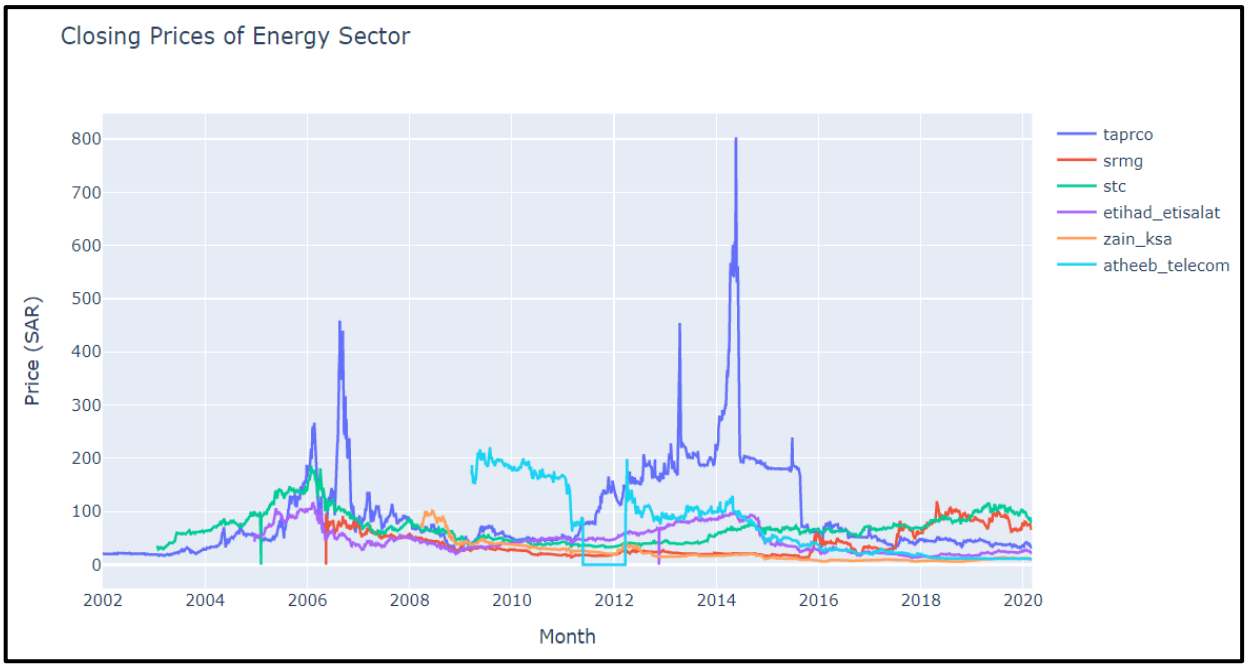
The purpose of the graph above is to quantile the close prices in a range between (0-1) to have a better vision of how the stocks that have similar price behave.



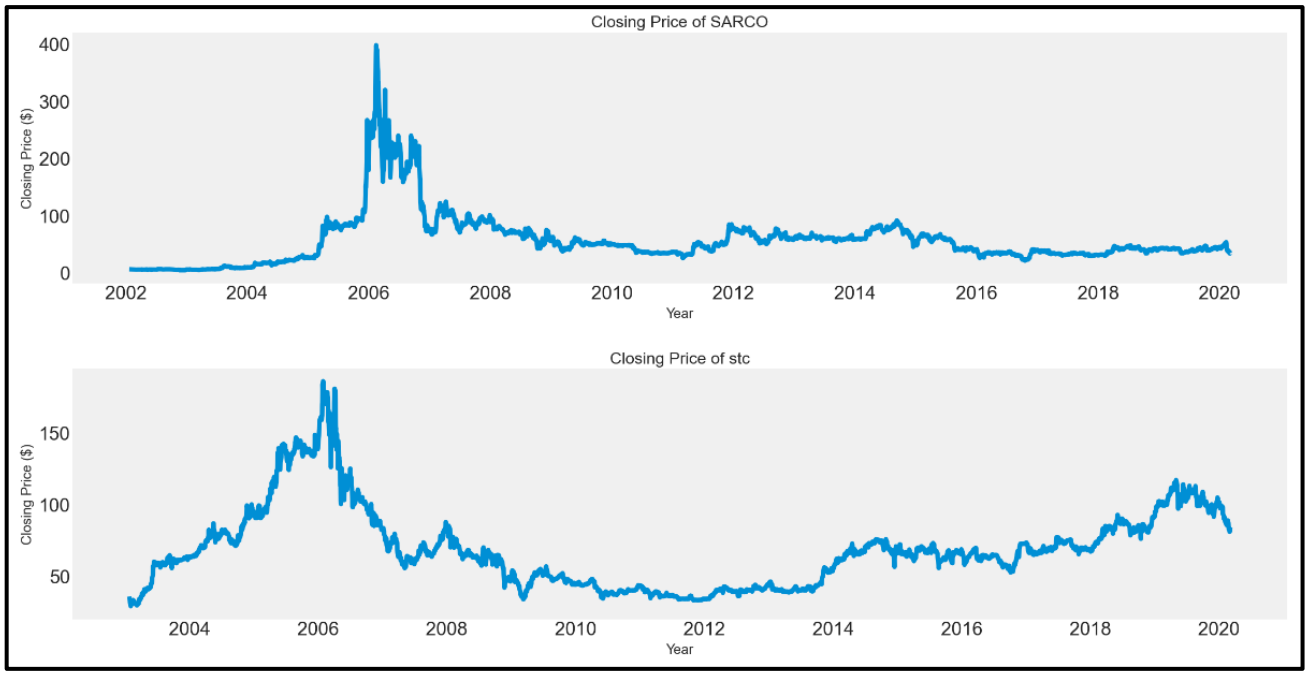
The purpose of the graph above is to compare the sectors by the No. of trades. From the graph, we choose Energy and communications sectors because they are the highest two.



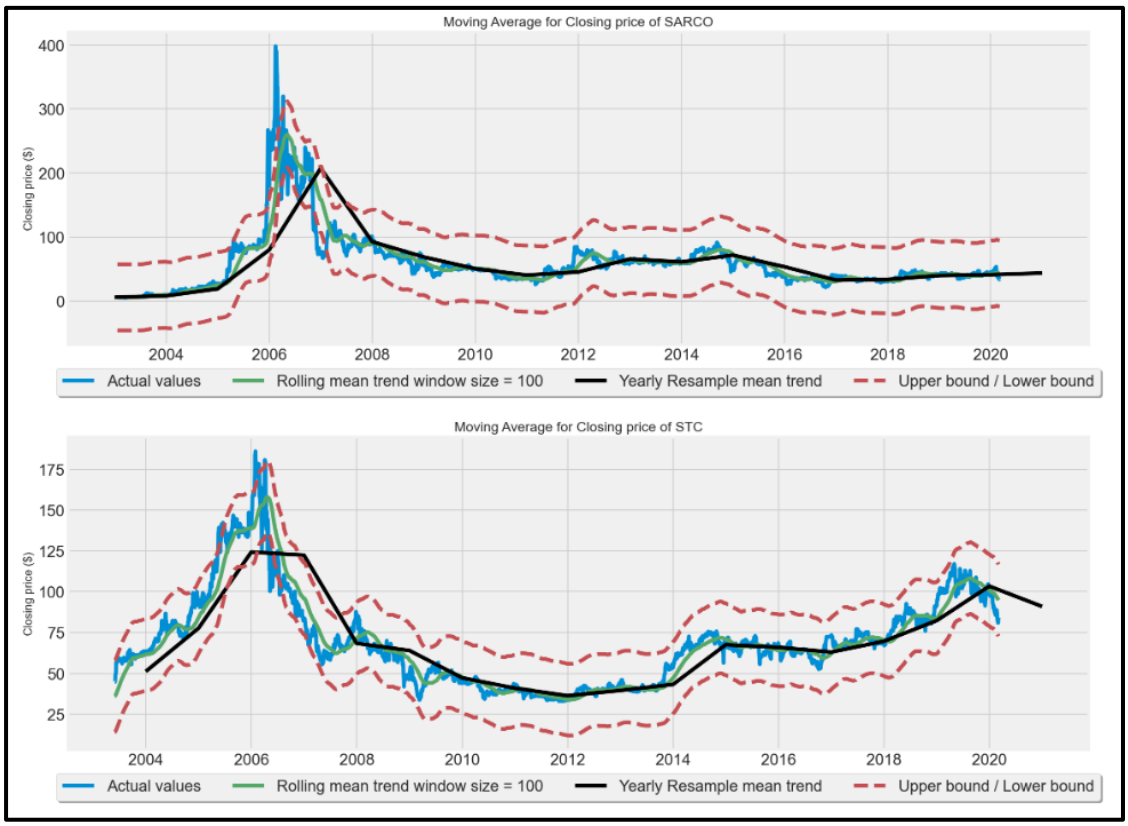
The purpose of the graph above is to show the history of the companies in the energy sector. We choose SARCO because it had better performance in 2006 and is one the oldest companies.



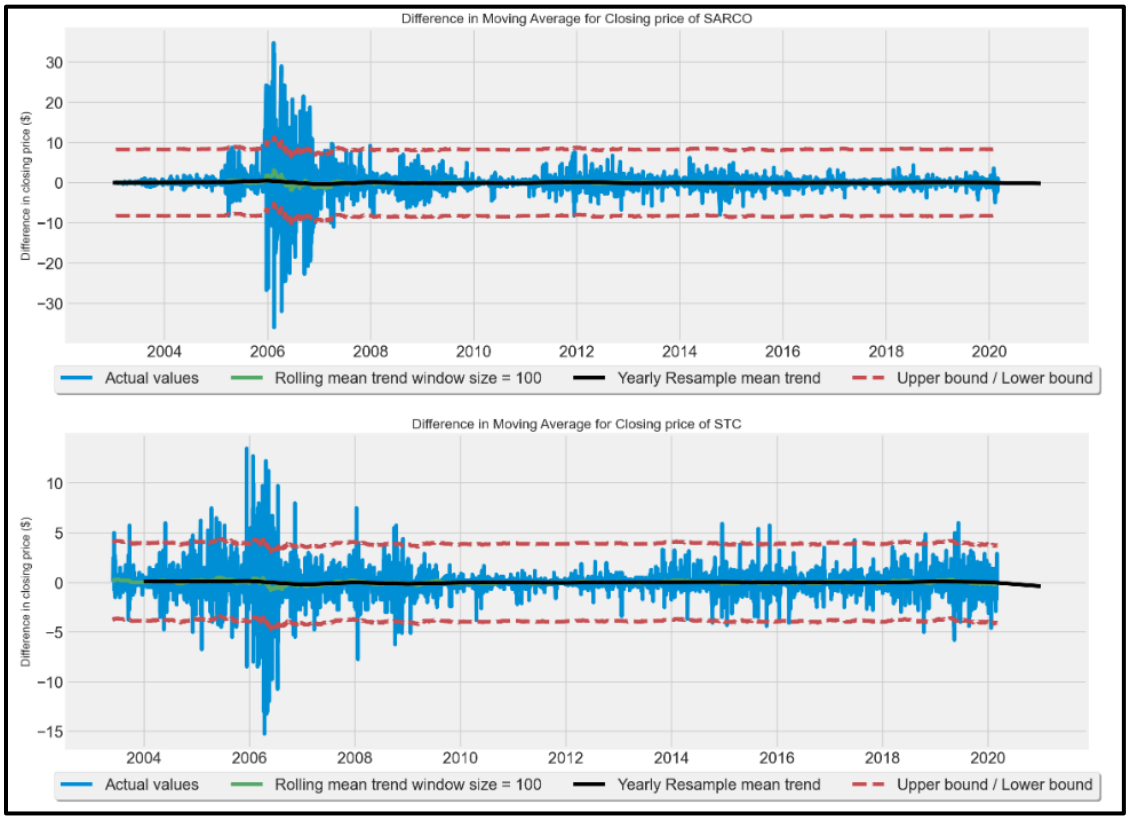
The purpose of the graph above is to show the history of the companies in the communication services sector. We choose STC because it is one of the oldest companies.



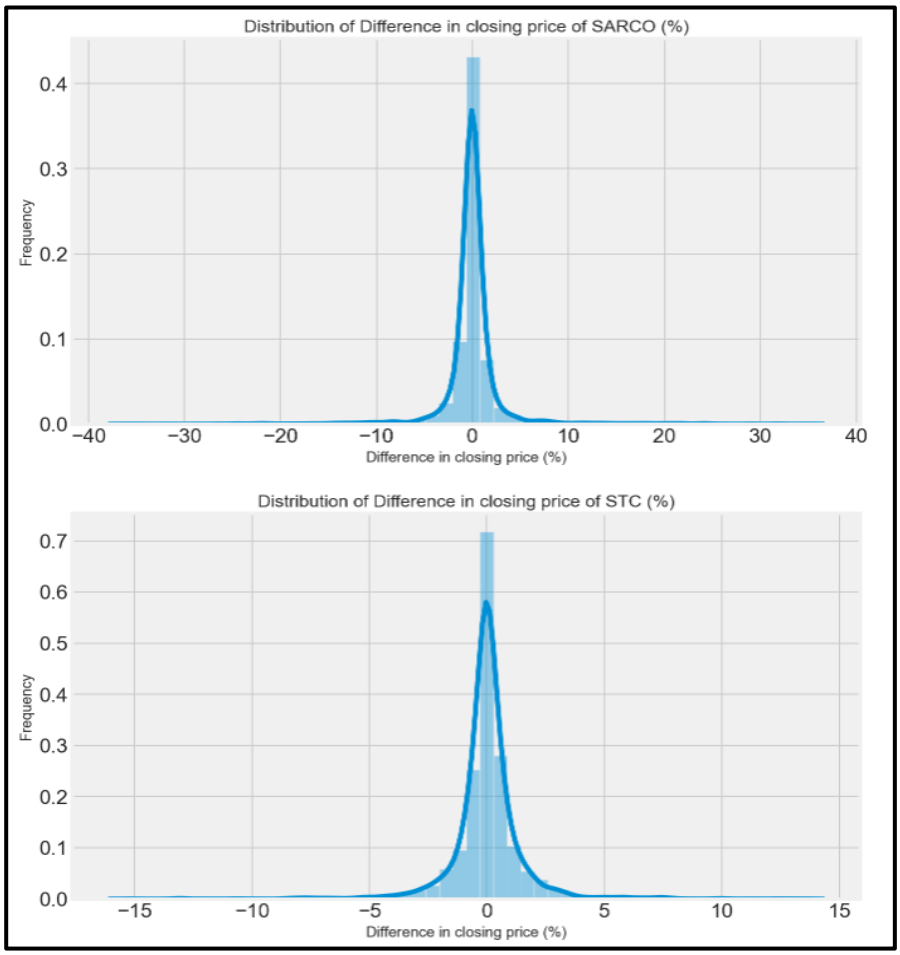
The purpose of the graphs above is to show the historical price of SARCO and STC.



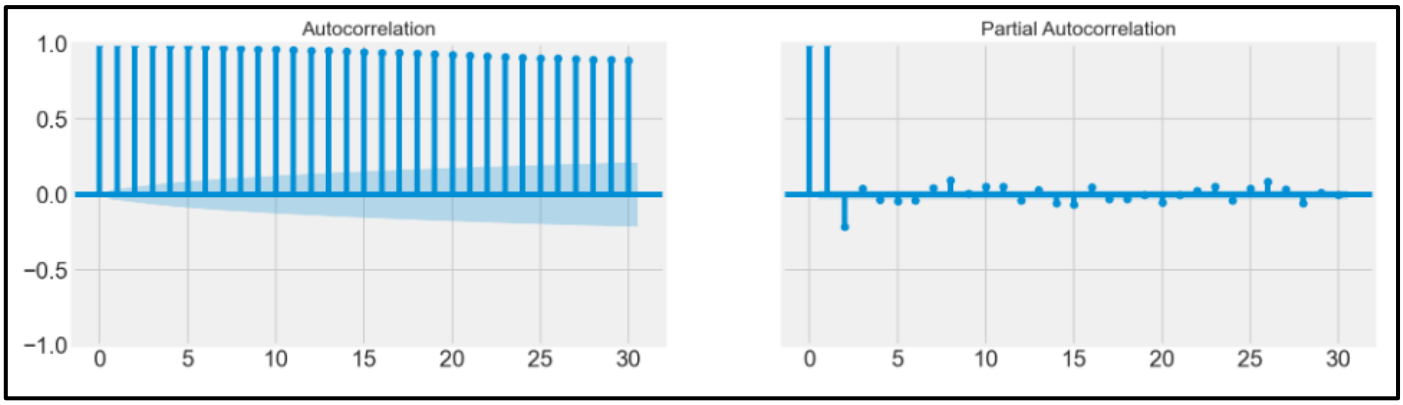
The purpose of the graphs above is to show the historical price of SARCO and STC with some Technical indicators like: (200 Envelope, Yearly Resample Mean Trend and 100 SMA)



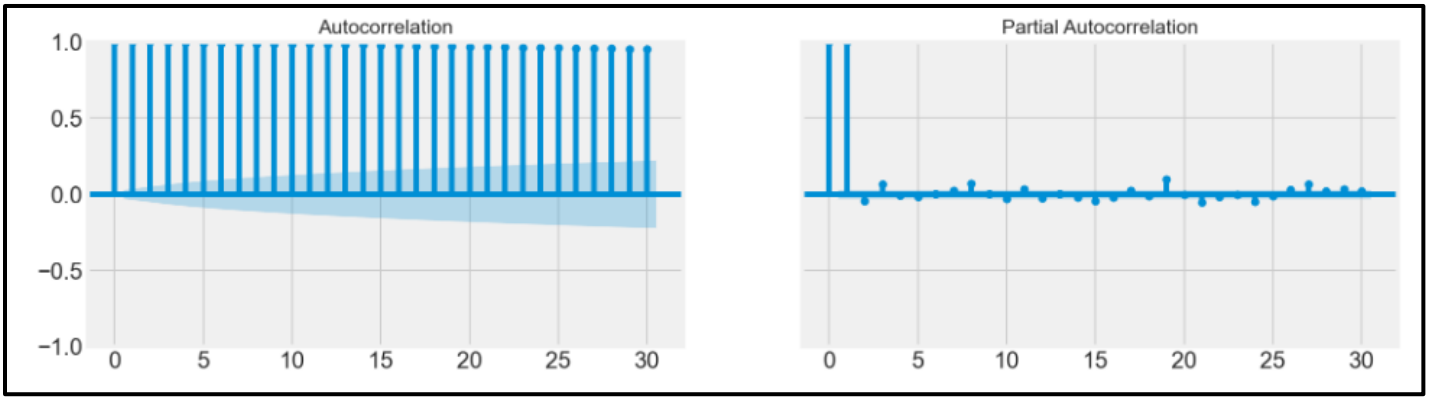
The purpose of the graphs above is to show the difference in Moving Average with the price of SARCO and STC which will show the distance between the moving average and the price.



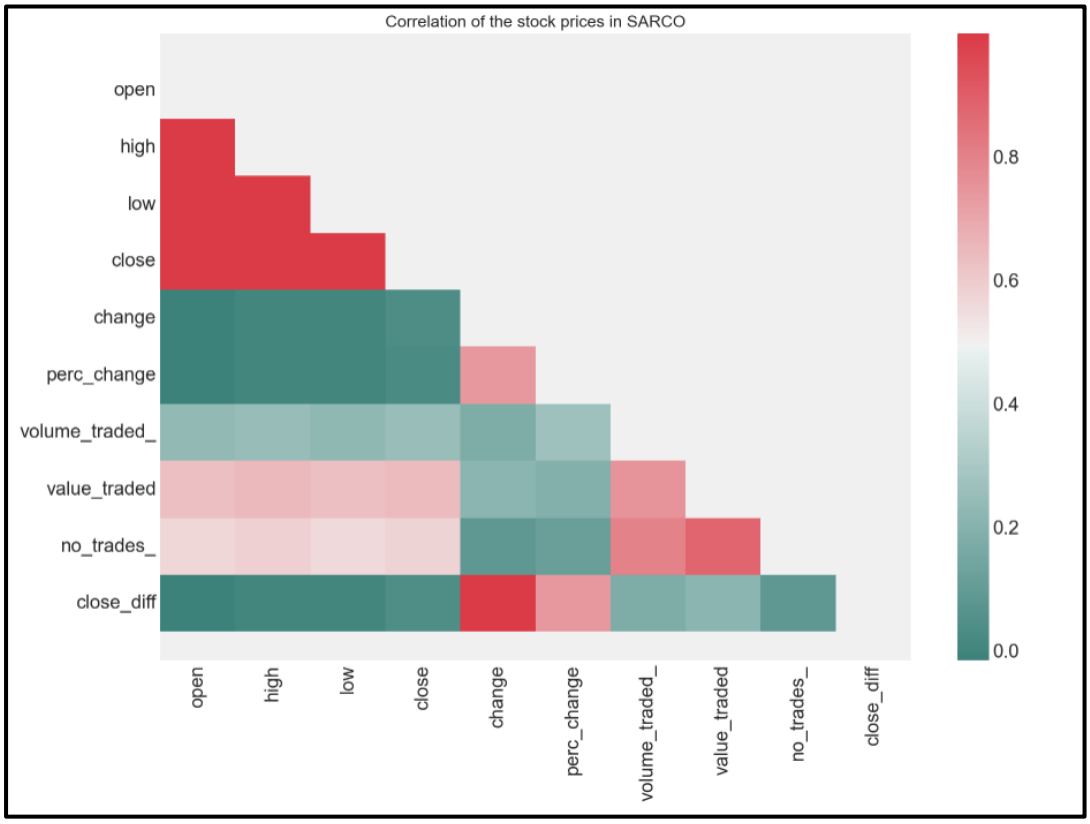
The purpose of the graphs above is to show the difference in closing price of SARCO and STC. From the graph we can include the difference in closing for STC is less than SARCO. The reason is because SARCO reach almost 400 SAR while the price of STC reaches almost 160 SAR.



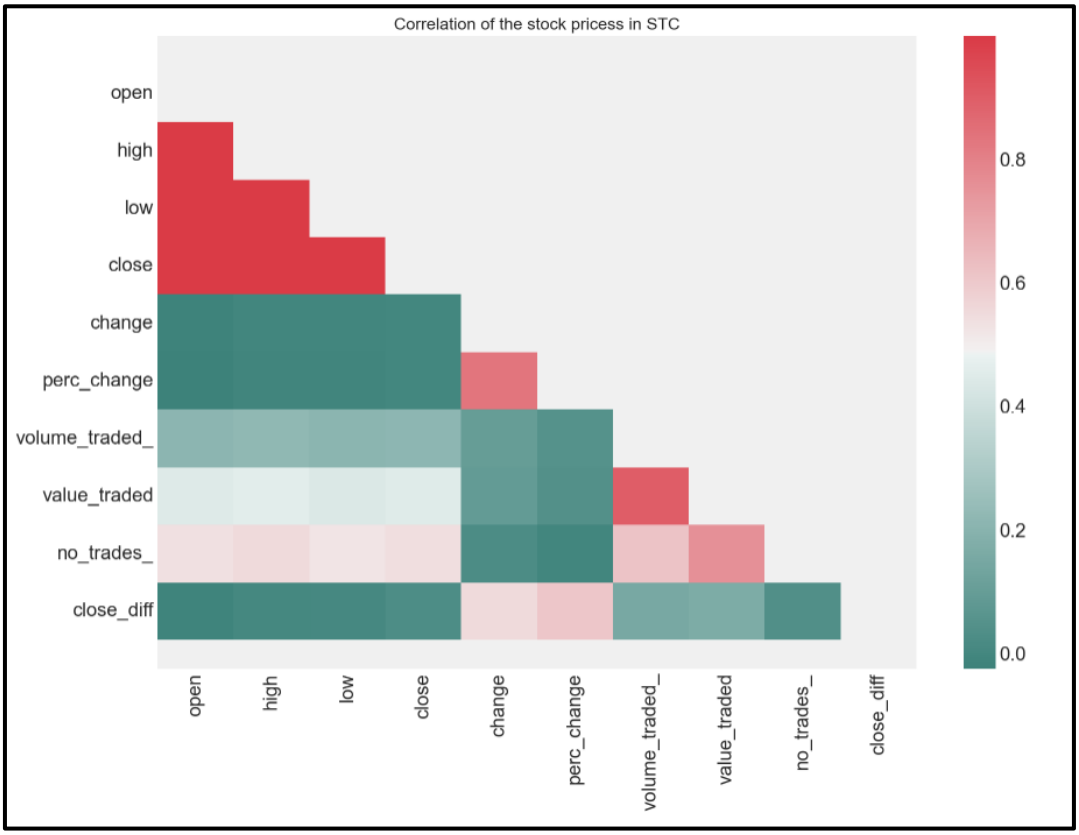
The purpose of the graph above is to show the Autocorrelation and Partial Autocorrelation of SARCO to determine which types of ARIMA model.



The purpose of the graph above is to show the Autocorrelation and Partial Autocorrelation of STC to determine which types of ARIMA model.



The purpose of the graph above is to show the relation between the columns of SARCO.



The purpose of the graph above is to show the relation between the columns of STC.

Approach :

The capstone project is managed using Agile which is an iterative approach to project management that focuses on breaking down large projects into more manageable tasks, which are completed in short iterations throughout the project life cycle. The reason for choosing is because the project is subject to change, and Agile allows to be better equipped to quickly change direction and focus. In the end, we will be able to complete work faster, adapt to changing project requirements, and optimize their workflow.

The prediction models in machine learning is the most appropriate for the selected dataset, particularly time series. The predictive models based on machine learning found wide implementation in time series projects required by various businesses for facilitating predictive distribution of time and resources. Time series forecasting is considered one of the most applied data science techniques that involve the analysis of observations collected in constant time intervals be it daily, monthly, quarterly, or yearly.

Machine learning forecasting proved to be the most effective in capturing the patterns in the sequence of both structured and unstructured data and its further time series analysis forecasting.

Results :

To solve our business problem we build three different models which are ARIMA , LSTM , Support Vector Machine .

The performance of the ARIMA model is evident in the values of each mean absolute error with score of 32.814 , mean squared error with score of 1101.554 , and R2 score of -43.485.

While the performance of the LSTM model is evident in the values of each mean absolute error with score of 0.04515 , mean squared error with score of 0.00356 , and R2 score of 0.93896 .

And the performance of the Support Vector Machine model is evident in the values of each mean absolute error with score of 0.28788 , mean squared error with score of 0.28788 , and R2 score of 0.28788 .

We have chosen the best model for the Hyperparameters application which is Long short-term memory (LSTM) . Hyperparameters are the explicitly specified parameters that control the training process. Hyperparameters are essential for optimizing the model. There are often general heuristics or rules of thumb for configuring hyperparameters. A better approach is to objectively search different values for model hyperparameters and choose a subset that results in a model that achieves the best performance on a given dataset. Grid search is the best Hyperparameter for our models. Grid-search is used to find the optimal hyperparameters of a model which results in the most accurate predictions and is useful to avoid exploring parameter combinations that make no sense or have no effect.

Future Work

There are two approaches to future expansion, one of which is data manipulation, and the other is concerned with visualizing data and making it more flexible to deal with the end user.

As for data manipulation, it is possible to add indicators in the dataset through several rules such as Moving Average, RSI, and MACD indicators. These indicators help the end user to understand the data more accurately and professionally

As for data visualization, we aspire in the future to build a site that makes it easier for the end user to see the prediction of the desired stocks quickly and simply .So that the end user can see all the stocks in the market and then choose the stock wants to predict then the forecast period is determined and finally displays a graph.