Capstone Project (Tadawul)

Applying Machine Learning on Saudi Stock Exchange (Tadawul)

Data Divers Group

Razan Aljohani, Zahra Almahdi, Fatima Alramadan, Asma Alasmary, Lena Baeshen

September 2022

Table of Contents

Table of Figures

Introduction

Data science is the field of applying advanced analytics techniques and scientific principles to extract valuable information from data for business decision-making, strategic planning and other uses. It's increasingly critical to businesses. The insights that data science generates help organizations increase operational efficiency, identify new business opportunities, and improve marketing and sales programs, among other benefits. Ultimately, they can lead to competitive advantages over business rivals.

A stock exchange facilitates stockbrokers to trade company stocks and other securities. A stock may be bought or sold only if it is listed on an exchange. Thus, it is the meeting place of the stock buyers and sellers.

This report shows the results and finding of analyzing Saudi Stock Exchange (Tadawul) by leveraging data through exploration, cleaning, preprocessing, visualization, and application of machine learning.

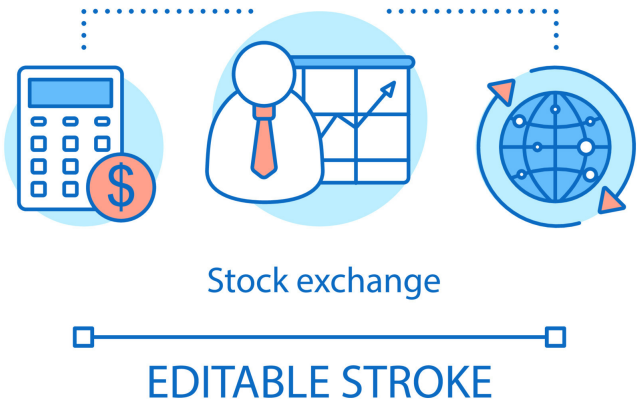

Figure 1. Stock Exchange Concept

Business problem

The Saudi Stock Exchange (Tadawul) is the sole entity authorized in the Kingdom of Saudi Arabia to act as the Securities Exchange. It is keen to provide new products and services that meet the needs of investors in addition to providing its users with the necessary resources and tools to become more capable investors.

Tadawul is providing long-term growth plans for its group, and market participants with attractive and diversified investment opportunities, so it is always required to analyze and

modeling its data for continuing its success and perfectly achieving its objectives, and to predict the change in the Saudi Stock.

Proposed solution

The project will build different effective models to predict the change in trades number per day in two sectors of Saudi Stock Exchange, Real State and Energy. It will utilize the following Machine learning Algorithms:

- **Linear Regression** It is familiar algorithm in data science field for analyzing and manipulating data. It defined a linear relationship between a dependent variable and the other given independent variables.
- **Decision Tree:** It is a supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes, and leaf nodes.
- **Random Forest:** It is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

Tadawul and Saudi's 2030 Vision

Vision 2030 is the overarching economic and social development strategy of the Kingdom. Its main themes are building a thriving economy, a vibrant society, and an ambitious nation, while one of its key pillars is for the Kingdom to become a global investment powerhouse.

Vision 2030 has many specific goals and Tadawul's activities interface with many of them including developing an advanced capital market, enabling financial institutions to support private sector growth, and promoting and enabling financial planning

The primary goal of our project is to analyze Tadawul data and use an efficient Machine Learning model to predict future progress and develop professional suggestions for Tadawul that enable them to achieve 2030's goals including:

- To raise the private sector's contribution to GDP, from 40% to 65%.
- To increase the Public Investment Fund's assets from SAR 600 Bn to over SAR 7 Tn.
- To rise from our current position of 25 to the top 10 countries on the Global Competitiveness Index.
- To move from our current position as the 19th largest economy in the world to the top 15.
- To increase the SME contribution to the GDP from 20% to 35%.

Aim and goals

The project aims to apply the learned skills from Data Science Bootcamp provided by SDA and Coding Dojo. It will build different predictive models of the Saudi Stock Exchange (Tadawul) on two sectors, Real State and Energy by achieving the following goals:

- Identifying significant attributes that increase trades number per day
- Prediction of the change of trades number per day based on its characteristics.

Scope

The Financial Sector aims to develop in a diversified and effective to support the development of the national economy, diversify its sources of income, and stimulate savings, finances, and investments. The aim is to achieve this ambition 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. Listed under the Financial Sector Development Program's umbrella are several sub-sectors, including banking, insurance, investment, and stock and debt markets.

Dataset

The data of Saudi Stock Exchange (Tadawul) was collected directly from Tadawul, and the data points in the dataset represents the price of a specific stock at a specific date.

Dataset Attributes:

- symbol: The symbol or the reference number of the company
- name: Name of the company
- trading_name: The trading name of the company

- sectoer: The sector in which the company operates
- date: The date of the stock price
- open: The opening price
- high: The highest price of the stock at that day
- low: The lowest price of the stock at that day
- close: The closing price
- change: The change in price from the last day
- perc_Change: The percentage of the change
- volume_traded: The volume of the trades for the day
- value_traded: The value of the trades for the day
- no_trades: The number of trades for the day

Reference: https://www.tadawul.com.sa/wps/portal/tadawul/home/

Analysis results

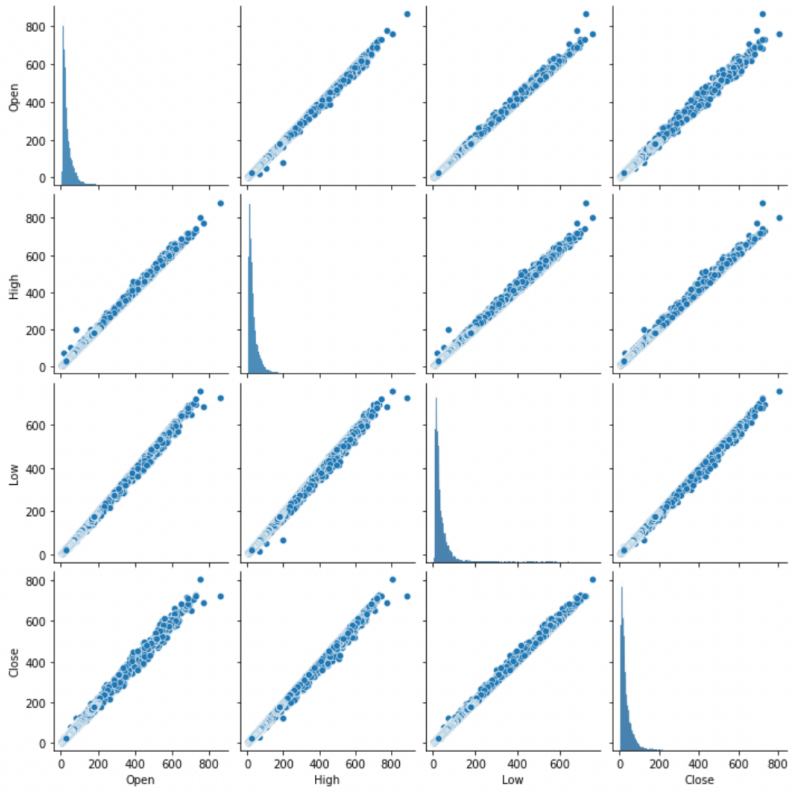

Figure 2. pair plot of Tadawul data

The above pair-plot is for the most related columns which are open, close, high, and low to see the relationship among them. We can see from the upper left the histogram that shows the distribution of the open stock prices, then we have the scatter charts that visualize the relationship between columns. Here we have positive relationships among the specified columns.

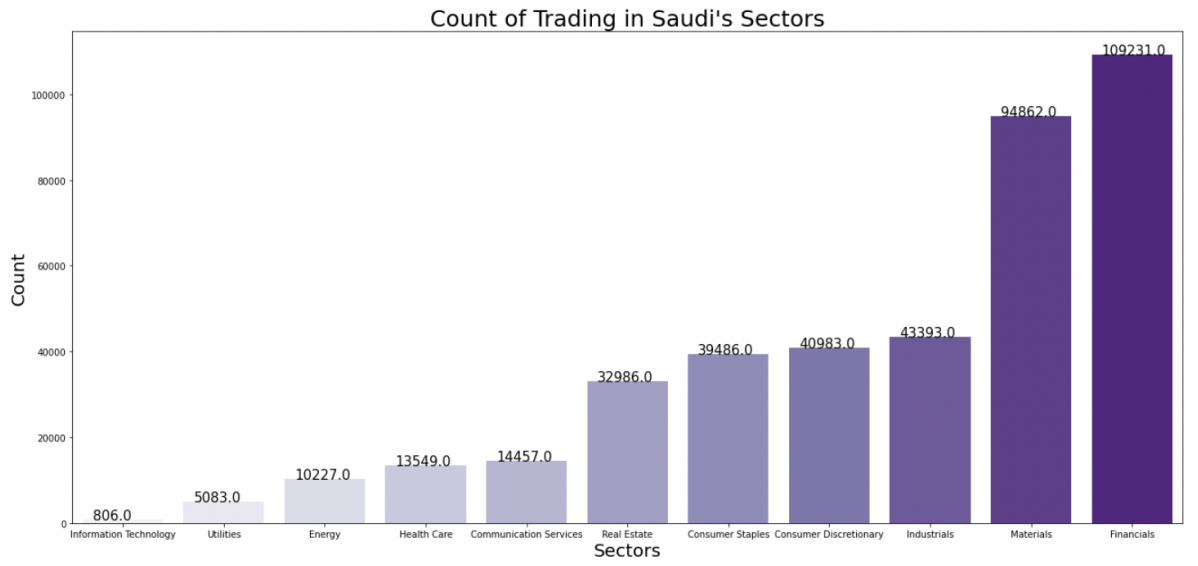

Figure 3. Bar chart for Count of Trading in Saudi's Sectors

The above bar plot aims to visualize the comparison between the different categories in Tadawl, we can see that the higher sector in the Saudi stock market is the financials while the lowest is the IT sector.

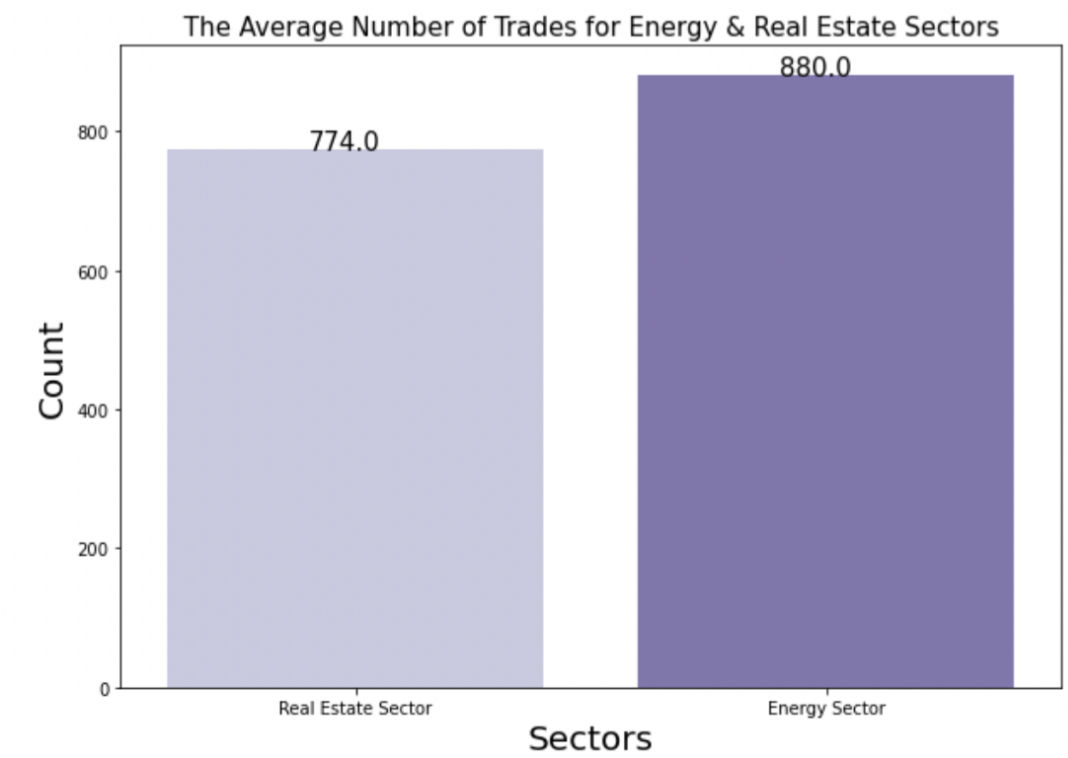

Figure 4. Bar Chart for the average number of Trades for Energy & Real Estate Sectors

The above bar chart displays the average number of trades for Real Estate & Energy. It shows that the Energy number of trades is higher than the Real Estate.

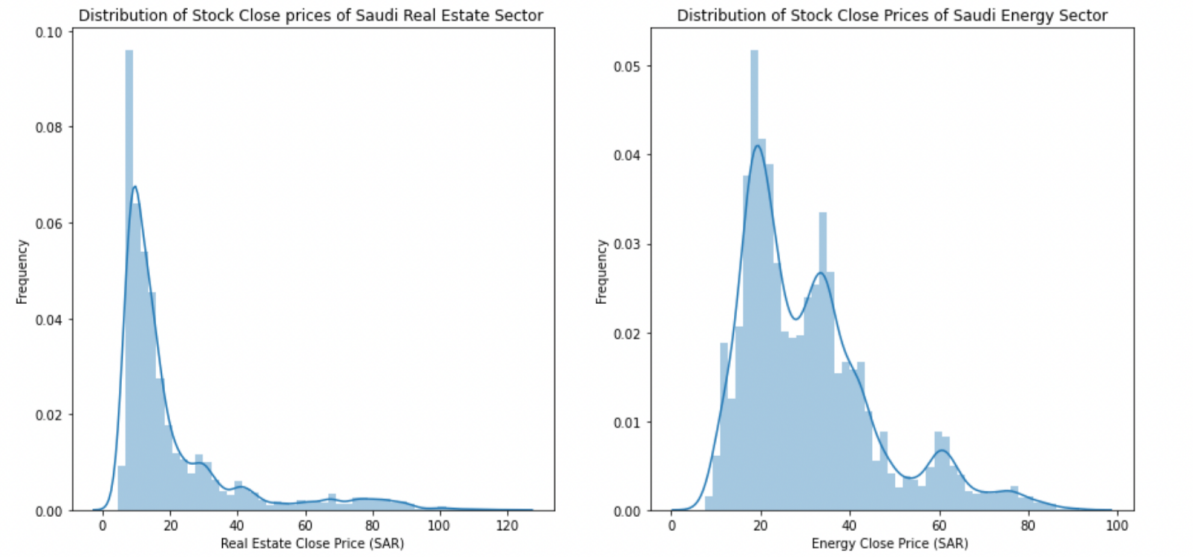

Figure 5. Distribution of Stock Close in the two sectors

The above histogram subplot is to know the difference between open/close prices in the Real Estate and Energy sectors.

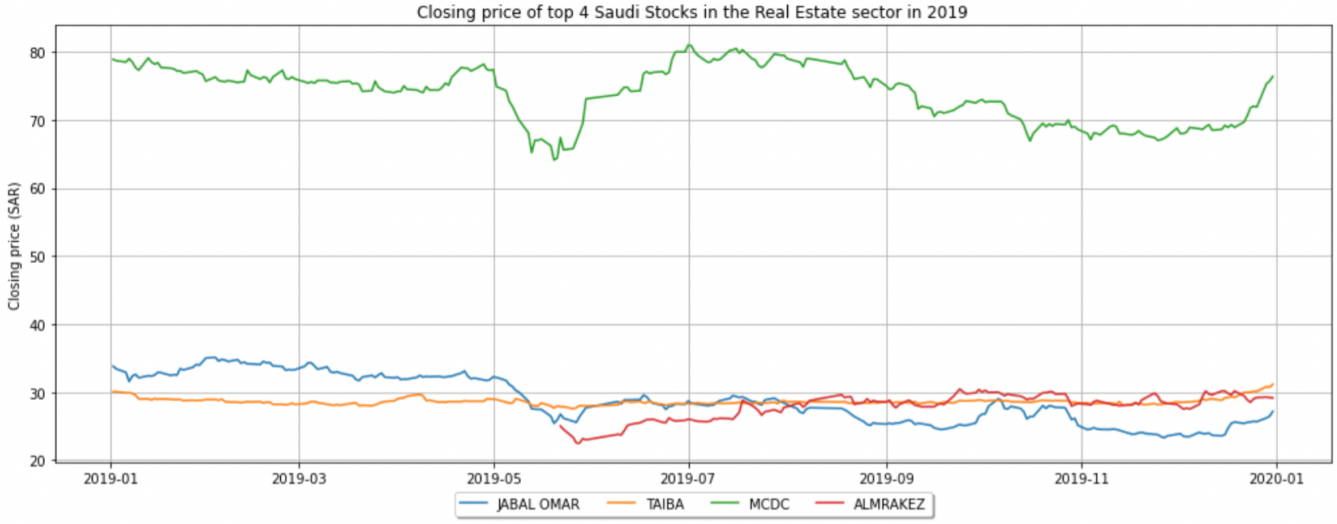

Figure 6. Line chart for Closing price of top four Saudi Stocks in the Real Estate sector in 2019

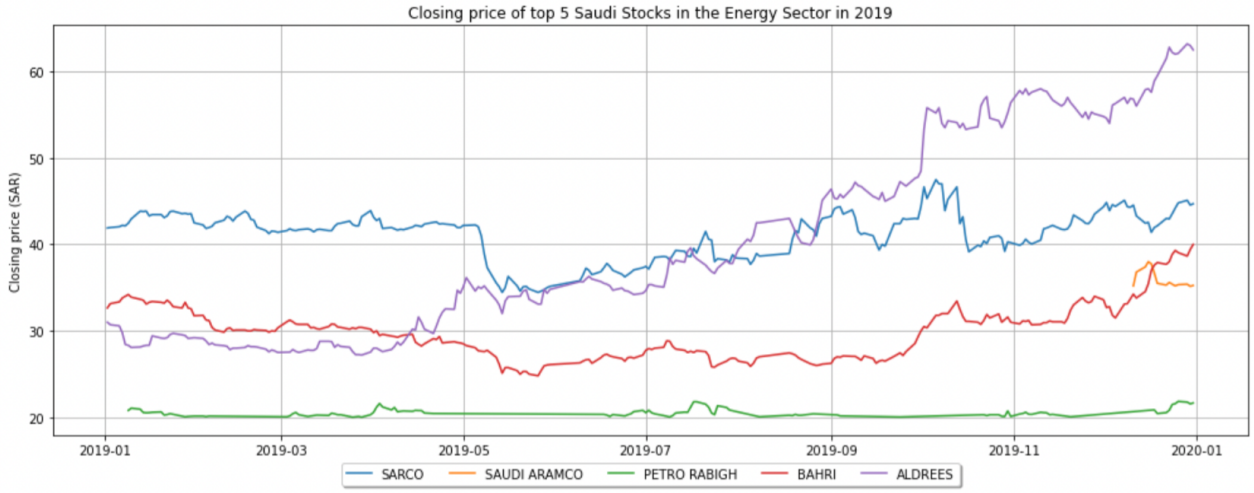

Figure 7. Line chart for Closing price of top four Saudi Stocks in the Energy sector in 2019

The above line charts are to understand and compare the stock exchange trends among the Energy and Real Estate stock sectors. We can actually see the rise of the close prices in the last quarter of 2019 with no sudden drops.

Models & Results

To complete our study, we choose only two sectors from the whole dataset which are the Real Estate and Energy sectors. There are many Machines Learning algorithm that used for modeling the data and prediction, and our project mainly used Baseline model to calculate the mean value of (the y_test), then create the calc cost function to calculate the MSE, MAE, and RMSE to make a comparison between the baseline and our prediction models i.e., Liner Regression, Decision Tree and Random Forest.

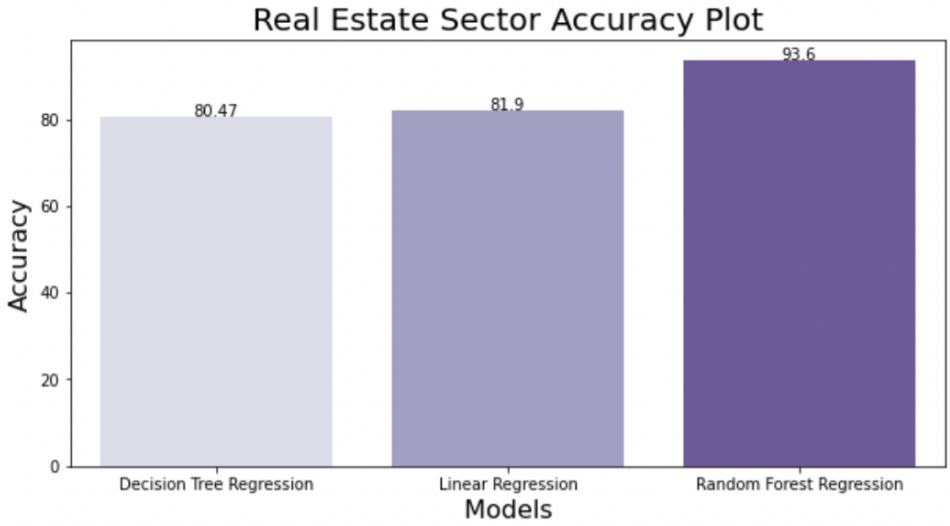

Figure 8: Real Estate Sector Accuracy Plot

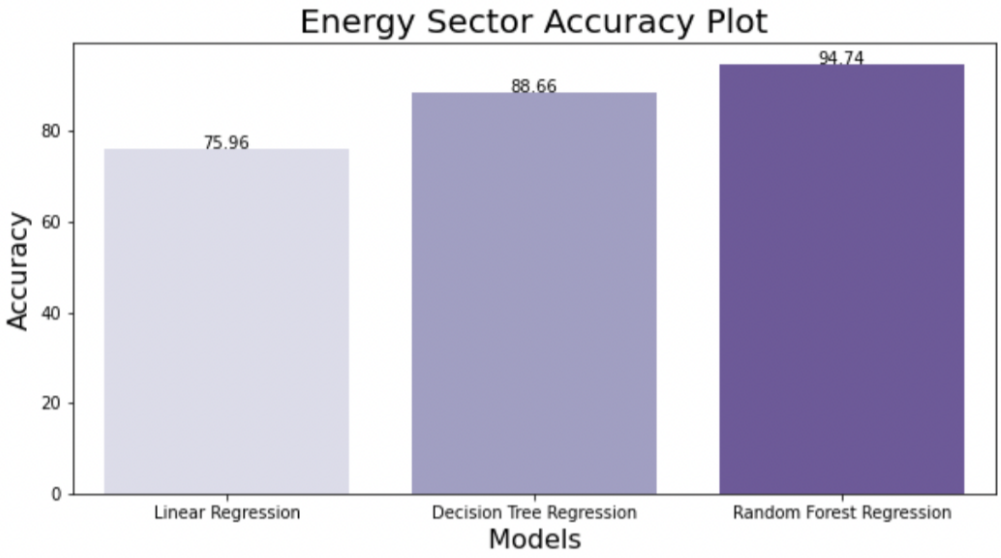

Figure 9:Energy Sector Accuracy Plot

Dashboard

A dashboard is a visual display of all data used to help track, analyze, and display data, usually to gain deeper insight into the overall wellbeing of the organization, a department, or even a specific process. It is primary intention is to provide information at-a-glance.

This project illustrates important insights through the dashboard, it uses different chart such as line chart, bar plot, pie chart, and others.

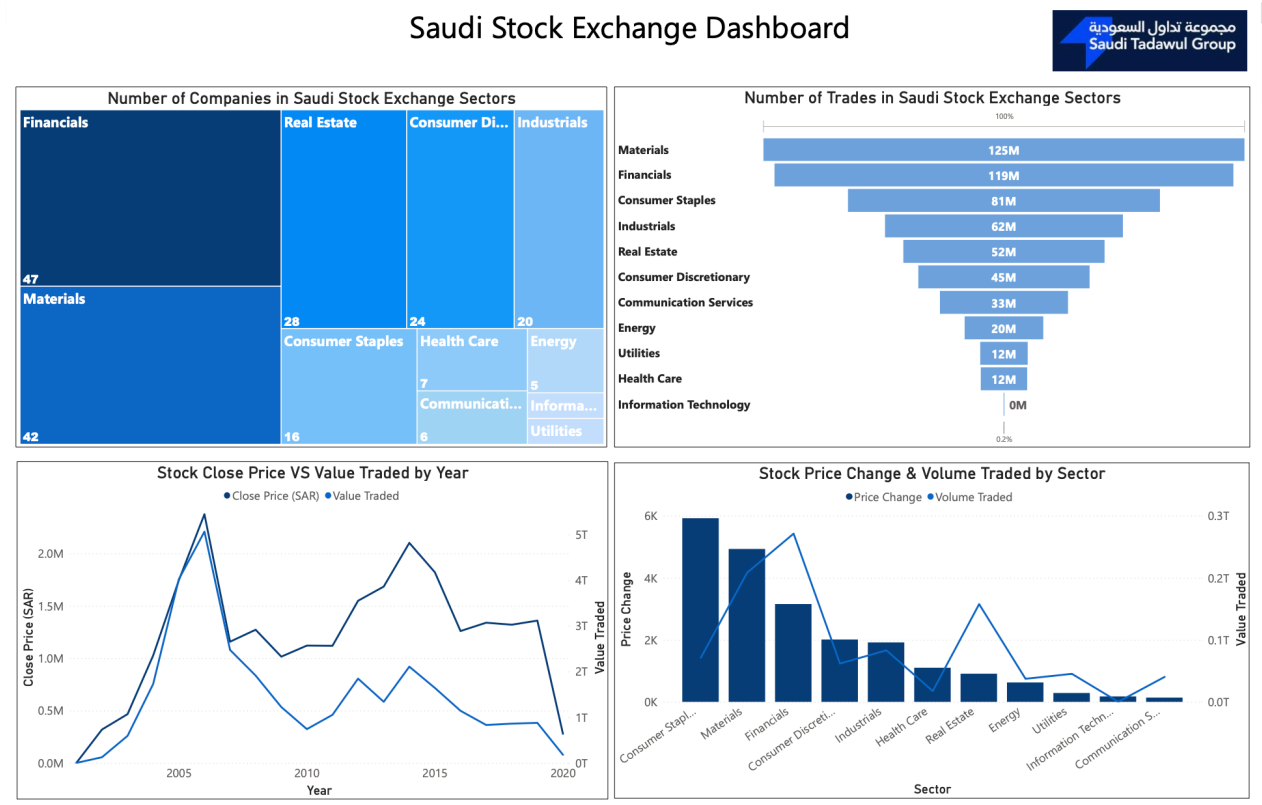

Figure 10 : Saudi Stock Exchange Dashboard

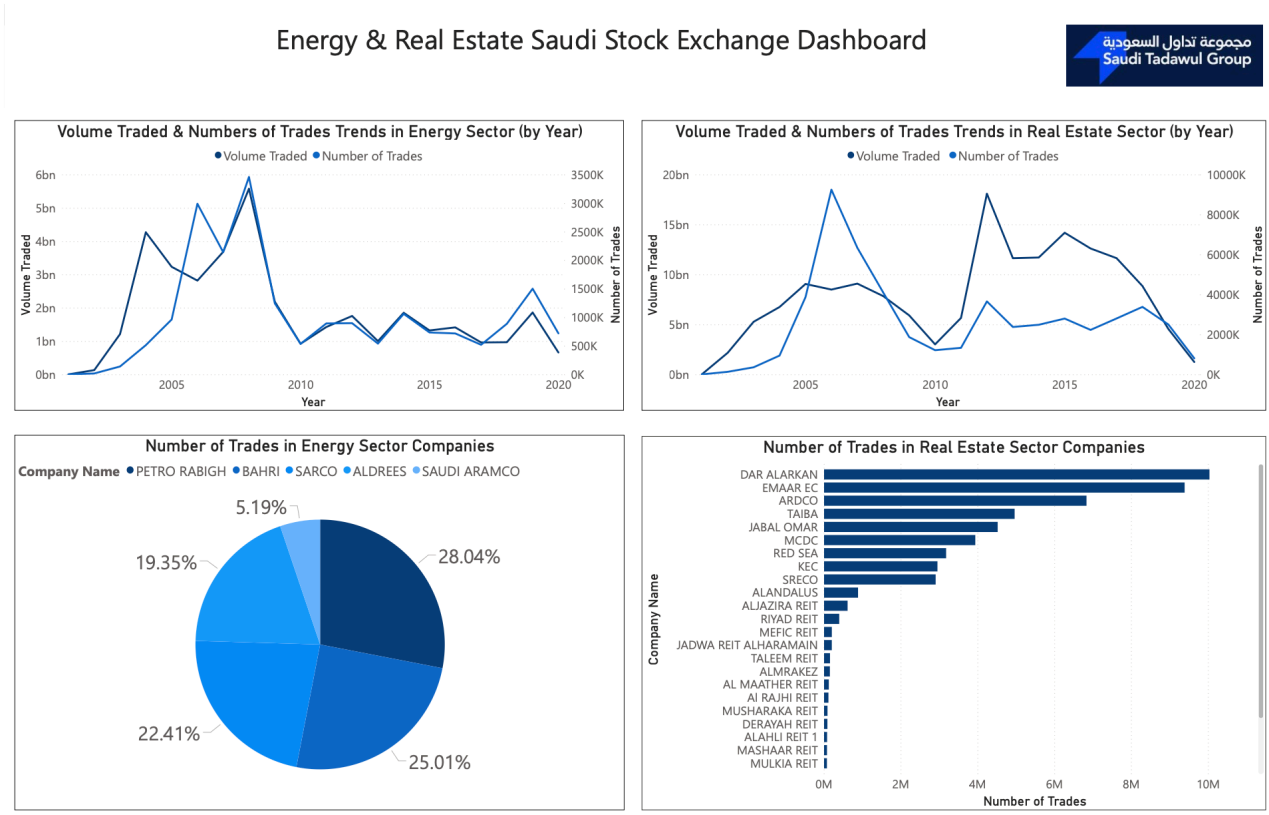

Figure 11. Energy & Real Estate Saudi Stock Exchange Dashboard

Tools

The needed tools to satisfy the project requirements:

- **Environment:** Jupyter Notebook.
- **Programming Language:** Python.
- **Data processing:** Pandas, NumPy.
- **Visualization:** Seaborn, Matplotlib, plotly.
- **Modeling:** Sickit-Learn.

Also, Microsoft Power BI is used to build the project Dashboard.

Conclusion

All in all, in this paper we applied the learned skills provided by SDA & Coding-dojo. we build different models to predict the number of trades per day of the two sectors in the Saudi Stock Exchange market and study the future progress of the stock market that will enable them to achieve the 2030 Vision.

Recommendations for future work:

1. Build different models to get higher accuracy.
2. Make a classification model to determine the best time to sell/buy stock in the market.
3. Try another target for the machine learning models.
4. Apply the study on other sectors.