UNSW
SYDNEY

FinTech Project Part B
Data Design and Analysis Report

Option #2

Nont Fakungkun

z5317972

Station #3: Model Design

The models aiming to use in the implementation

I have designed a model that uses the ARIMA model to forecast the price and contract volume of both corn and wheat contracts. I will use historical data of the corn and wheat contract retrieved from the earlier station as a source of price and volume forecast.

In addition to the ARIMA model, I will also use VADER to perform sentiment analysis on commodity news. This will allow me to return the sentiment of each contract, whether it is negative or positive.

Model assumptions

For ARIMA model, the assumptions are:

- The data is stationary. This means that the mean, variance, and autocorrelation of the data do not change over time.
- After a certain number of time steps (represented by q), the past values of the data don't help predict the future values.

For VADER sentiment analysis, the assumptions are:

- The text is written in English and has a certain length (not too short or too long).
- The text does not contain any special characters or symbols.

Restrictions, computing speed, accuracy issues

ARIMA Model

Restrictions: The ARIMA model requires the data to be stationary. This means that the mean, variance, and autocorrelation of the data must not change over time. The ARIMA model also assumes that data points beyond a certain time step have no relevance to the prediction. This means the model will not take those values into consideration when predicting future values. (Verma, 2022)

Computing speed: The ARIMA model can be computationally expensive, especially when the data set is large or when the p and q values are high. Higher p and q values can lead to a more accurate model, but also lead to a slower model. (Verma, 2022)

Accuracy: In general, the ARIMA model is more accurate for short-term forecasting than for long-term forecasting. This is because the ARIMA model assumes that the underlying trends and seasonality of the data will remain constant. Trends and seasonality that change over time can lead to inaccurate forecasts for long-term forecasting.

VADER sentiment analysis

Restrictions: VADER is a rule-based sentiment analysis tool, which means that it uses a set of rules to determine the sentiment of text. This can make it difficult for VADER to handle complex or nuanced text. (Nawaz, 2023)

Computing speed: VADER is a relatively fast sentiment analysis tool. It can analyse text in a matter of milliseconds.

Accuracy: VADER is generally accurate for short texts. However, its accuracy can decrease for longer pieces of text.

What is expected to be seen in model implementation

For ARIMA, I would expect to see a line plot showing the historical price data and the predicted price data. The predicted price line should extend into the future, showing the forecasted prices after the current data points. The volume prediction is expected to be displayed in bar charts instead of line plots.

For VADER, I would also expect to see the mean compound score, which will give a quick understanding of the overall sentiment polarity of the analysed text. Additionally, I would expect to see the sentiment prediction in a form of line plot similar to the output of price prediction from the ARIMA model.

Boundaries of the model

ARIMA Model

- ARIMA is a forecasting model, not an asset allocation model. It is designed to predict future values of a time series, not to determine how to allocate assets.
- ARIMA is a single-factor model. It only considers the past values of the data as a predictor of future values. As there are many other factors that can affect asset prices, the model should be incorporated with other prediction methods for higher prediction accuracy. (Verma, 2022)

VADER sentiment analysis

- VADER is limited in its ability to handle sarcasm or irony. These forms of expression often rely on context that VADER may not be able to understand. (Nawaz, 2023)
- VADER is not able to distinguish between different types of sentiment, such as positive, negative, and neutral. It can only determine whether the overall sentiment of a piece of text is positive or negative.

How would you define Station #3 from Data Management Perspective

Model Design is the process of defining and creating a structure that best suits the objectives of the analyst e.g., forecast price and contract volume of selected assets. The output of this process is an effective model that can generate reliable and actionable insights from the featured data streams retrieved from earlier stations.

Station #4: Model Implementation

Product Design

The product will be an application that provides the forecast of the future price and contract volume of agricultural commodities futures contracts, such as corn and wheat contracts. The application will have a user-friendly interface that is easy to navigate. The application is organised into three main sections: the homepage, the profile page, and the analysis page. Authentication process will be implemented to ensure the privacy of the users.

Features:

- Price and contract volume forecasts: The application will make forecasts for 4 ranges of time: 7 days, 14 days, 1 month, 3 months. This will allow users to make informed decisions according to their trading strategies.
- Sentiment analysis: The application will also provide sentiment analysis of news related to agricultural commodities futures contracts. There will be 4 ranges of forecast period: 3 day, 7 days, 14 days, and 1 month. This information will help users to understand the market sentiment and make informed trading decisions.
- Different visualisations: The application will display the forecasts in different visualisations, such as line charts, and bar charts. This will help users to quickly assess the market outlook. The real data and forecast data will be differentiated by the line/bar colours.

Step took when implementing this solution

ARIMA Model

1. Testing the model if it can successfully return the forecasts.
2. Try different forecast periods, assess the differences.
3. Determine the final forecast period.
4. Run the forecast for different forecast periods on each asset.
5. Combine the current and forecast data.
6. Display as line/bar plots with different colours. (See figure 1 to 4 in appendix)

VADER

1. Separate the news of different commodities so run the sentiment analysis separately.
2. Test the model if it can produce the sentiment analysis on the given dataset.
3. Put the output of the sentiment analysis in the ARIMA model to forecast the compound section of the output.
4. Try different forecast periods, assess the differences.
5. Determine the final forecast period.
6. Combine the current and forecast data.
7. Display as line plots with different colours. (See figure 5 and 6 in appendix)

Encountered difficulties

- There are numerous options for setting the input parameters of the ARIMA model, namely p, d, and q variables. This abundance of choices makes it challenging to determine the optimal combination that provides the best results.
- The sentiment analysis results are challenging to present as the sentiment scores are mostly less than 0.1. They are considered to be neutral, so it is difficult to determine whether the news articles are positive or negative. At the moment, I have decided to simply present the results as an overview of whether the sentiment is positive or negative for the given period of time.

Steps to apply when introducing this financial product to clients

1. Match client's needs and goals with the benefit of the product. Once we understand their needs, we can better explain how the product can help them achieve their goals. Make sure to highlight the key benefits that are most relevant to the clients.
2. Demonstrate how the product works. We can provide scenarios or use cases that the product will be useful rather than just telling the clients about the product.
3. Answer client's questions. Be prepared to answer any questions that the clients may have about this product.
4. Follow up with clients. After we have introduced the product to the clients, follow up with them to see how they are doing. This will help us to identify any potential problems for future improvement and to provide additional support.

Customer journey

1. The customer creates an account and enters their contact information.
2. The customer selects the commodities that they are interested in trading.
3. The customer navigate in the apps via navigation bar
4. The customer picks the forecast period and views the forecasts for the price and contract volume of the selected commodities according to their strategies.
5. The customer views the sentiment analysis and sentiment forecast of news related to the selected commodities.
6. The customer decides whether to trade the selected commodities based on the forecasts and sentiment analysis.
7. The customer trades the selected commodities and monitors their performance.
8. The customer makes adjustments to their trading strategy as needed.

FIGMA design

The interface of the application is designed in FIGMA for both website and mobile version.

The FIGMA design includes login/sign up page, homepage, profile page, and analysis page for price, contract volume, and sentiment forecast at different forecast periods.

This can be accessed via the FIGMA link
https://www.figma.com/file/j0GcdR4IDo2vSibD7M04Kd/Untitled?type=design&node-id=103%3A22529&mode=design&t=LJ0NY9PEIgsguUU8-1/

The snapshots of FIGMA design are also included in the appendix. Please see Figure 7 and 8 for desktop version and see Figure 9 and 10 for mobile version.

How would you define Station #4 from Data Management Perspective

The station is where the designed model is put into action. The output is then presented in suitable visualisations, allowing for the market to gain valuable insights. This last station ensures that the model's predictions are effectively communicated and accessible to decision-makers, empowering them to make informed choices.

Reference

Nawaz M. 2023, Unsupervised Sentiment Analysis using VADER and Flair, Soshace, viewed 5 August 2023, <https://soshace.com/unsupervised-sentiment-analysis-using-vader-and-flair/#:~:text=Limitations%20of%20VADER&text=Its%20language%20dependency%20means%20it,sentences%20or%20domain%2Dspecific%20language/ >.

Verma Y. 2022, 5 conditions when the ARIMA model should be avoided, AIM, viewed 5 August 2023, <https://analyticsindiamag.com/5-conditions-when-the-arima-model-should-be-avoided/ >.

Appendix

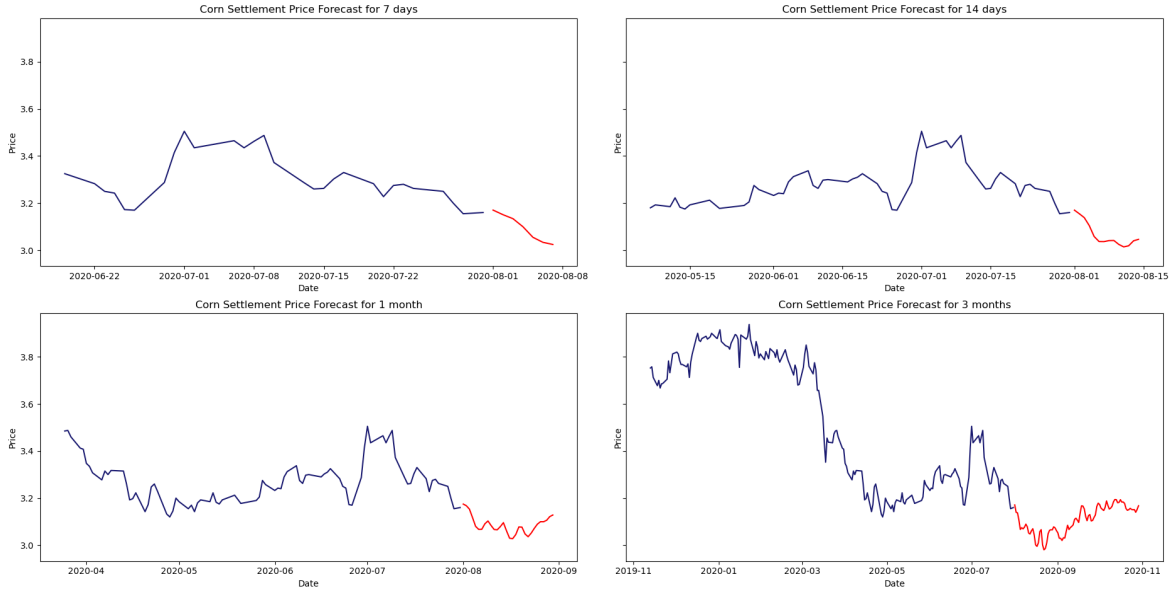

Figure 1: A snapshot of line plots of price forecasts for corn contracts with different forecast periods.

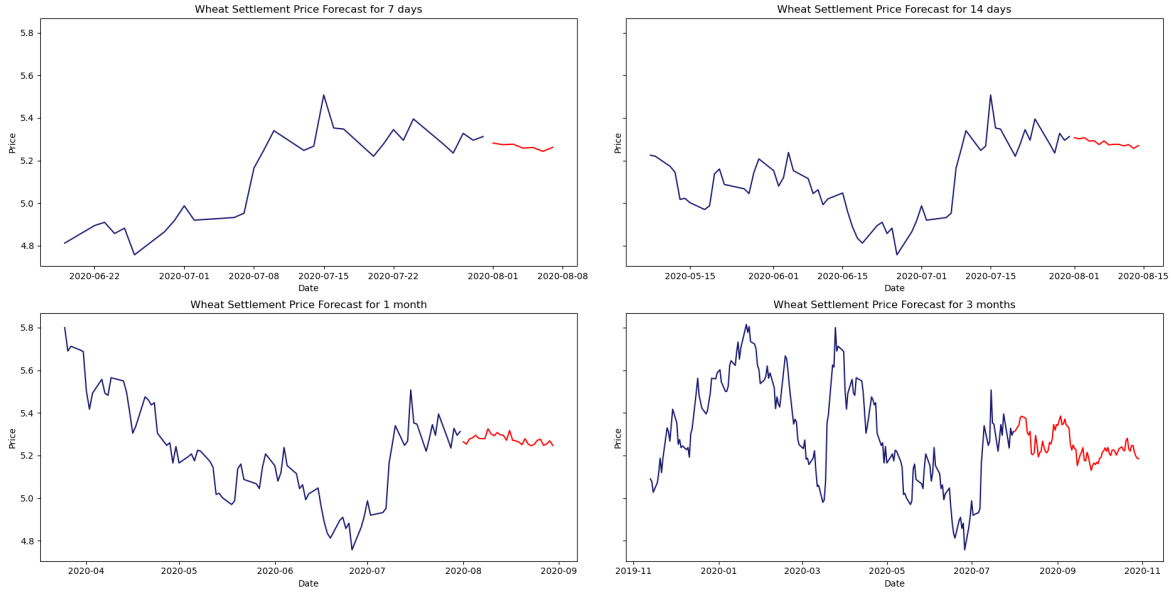

Figure 2: A snapshot of line plots of price forecasts for wheat contracts with different forecast periods.

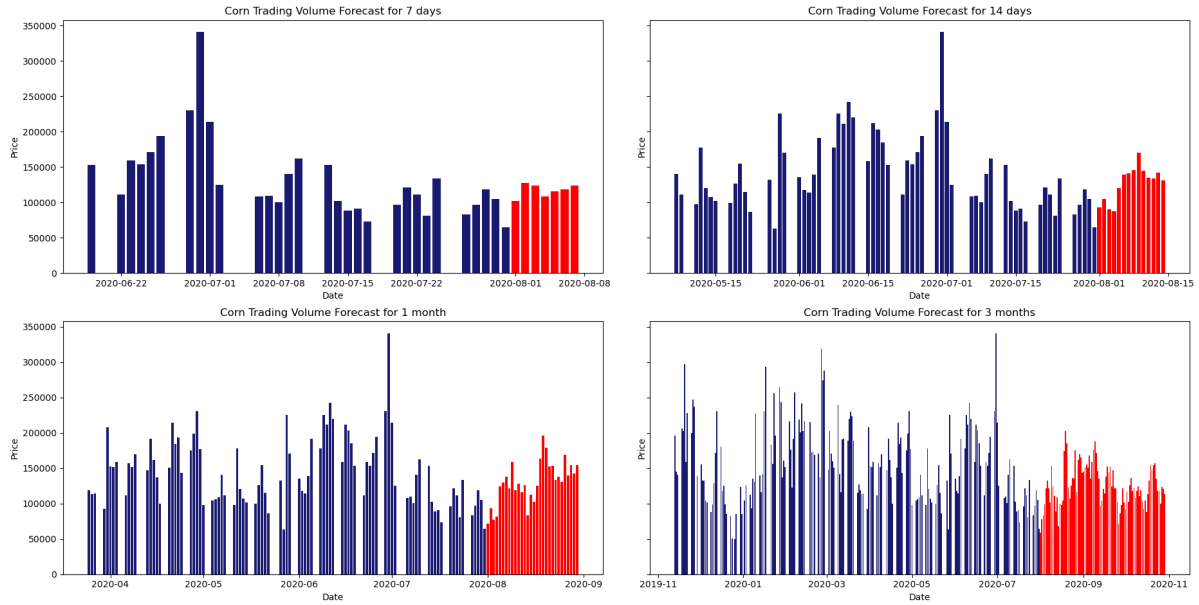

Figure 3: A snapshot of bar graphs of contract volume forecasts for corn contracts with different forecast periods.

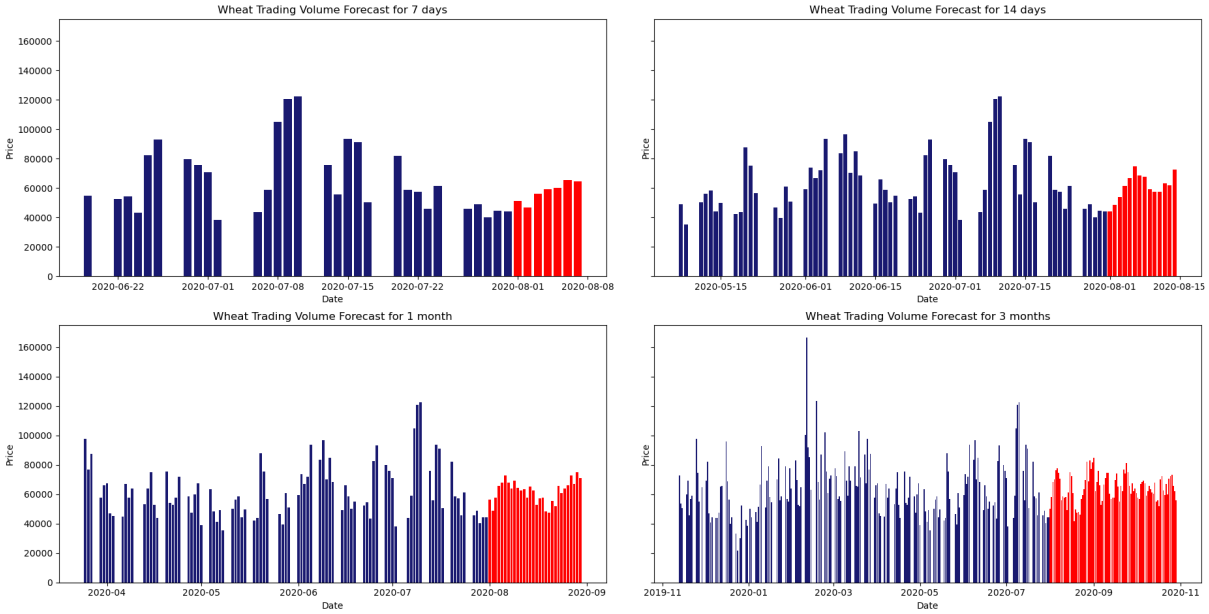

Figure 4: A snapshot of bar graphs of contract volume forecasts for wheat contracts with different forecast periods.

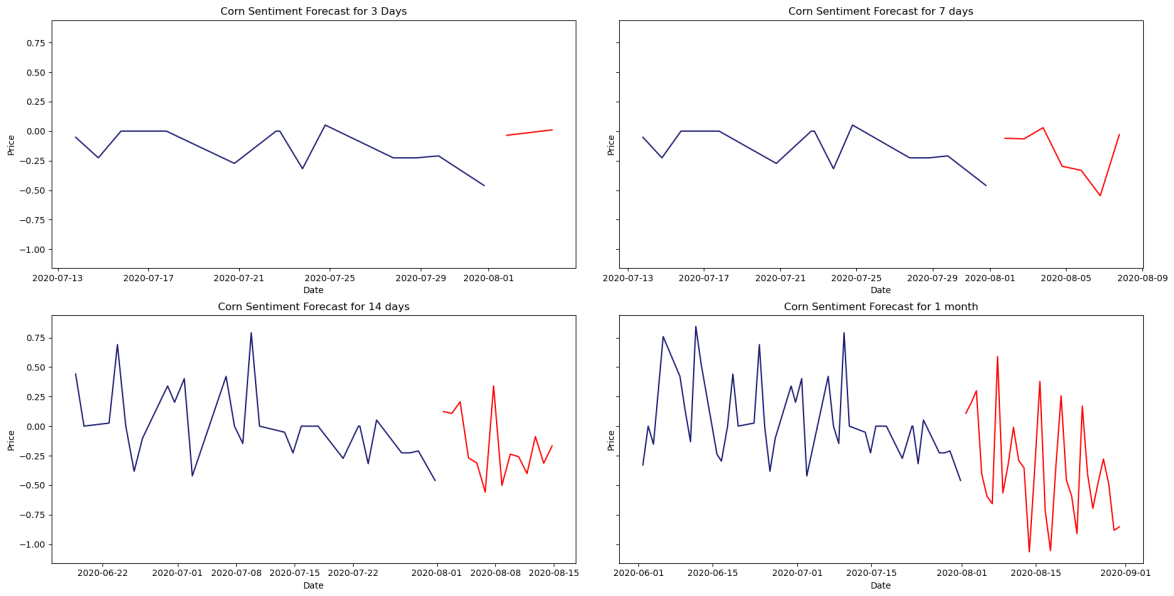

Figure 5: A snapshot of line plots of sentiment forecasts for corn contracts with different forecast periods.

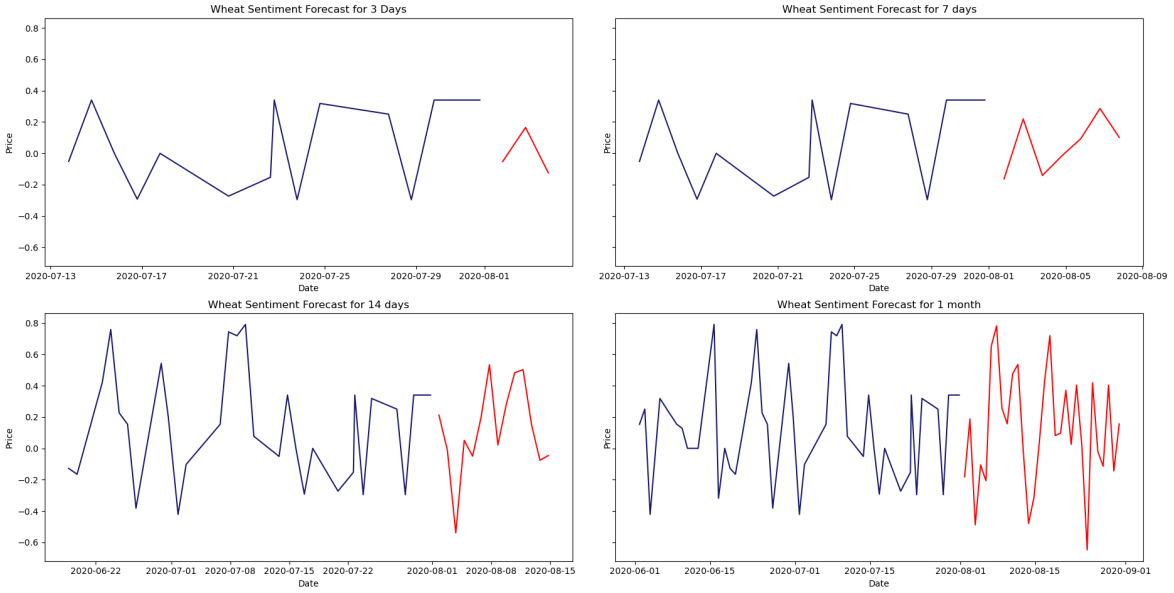

Figure 6: A snapshot of line plots of sentiment forecasts for wheat contracts with different forecast periods.

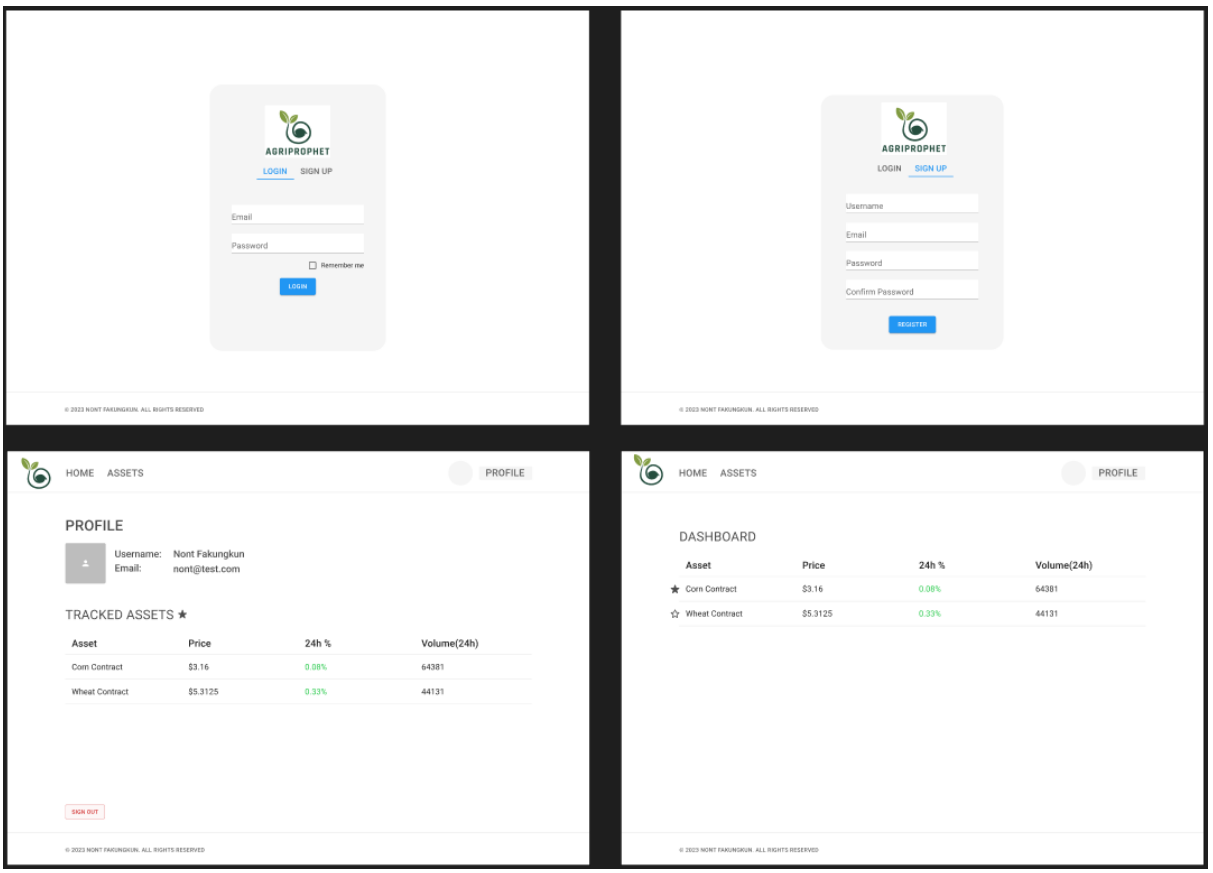

Figure 7: A snapshot of the desktop version of login page, sign up page, homepage, and profile page in FIGMA design

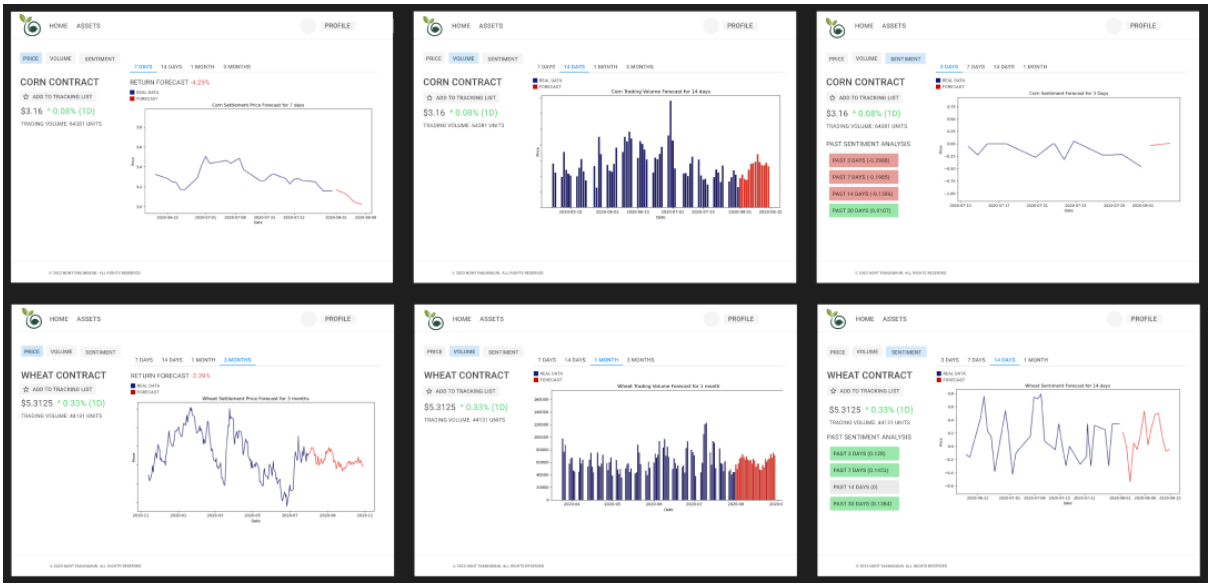

Figure 8: A snapshot of the desktop version of price, volume, and sentiment forecast/analysis pages in FIGMA design

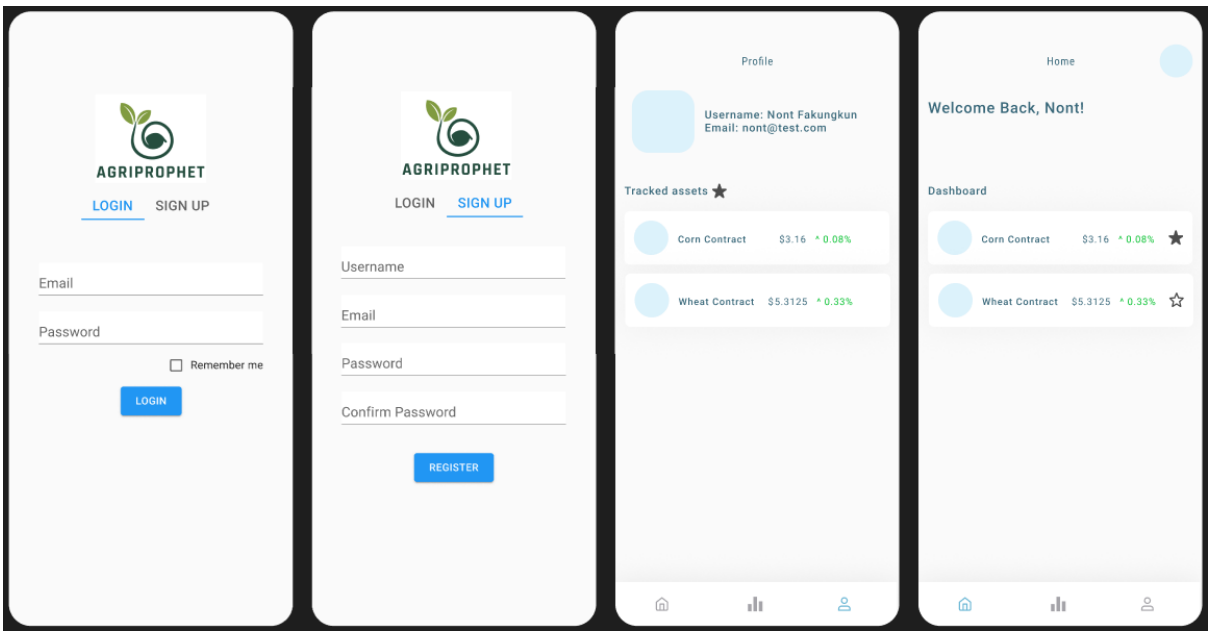

Figure 9: A snapshot of the mobile version of login page, sign up page, homepage, and profile page in FIGMA design

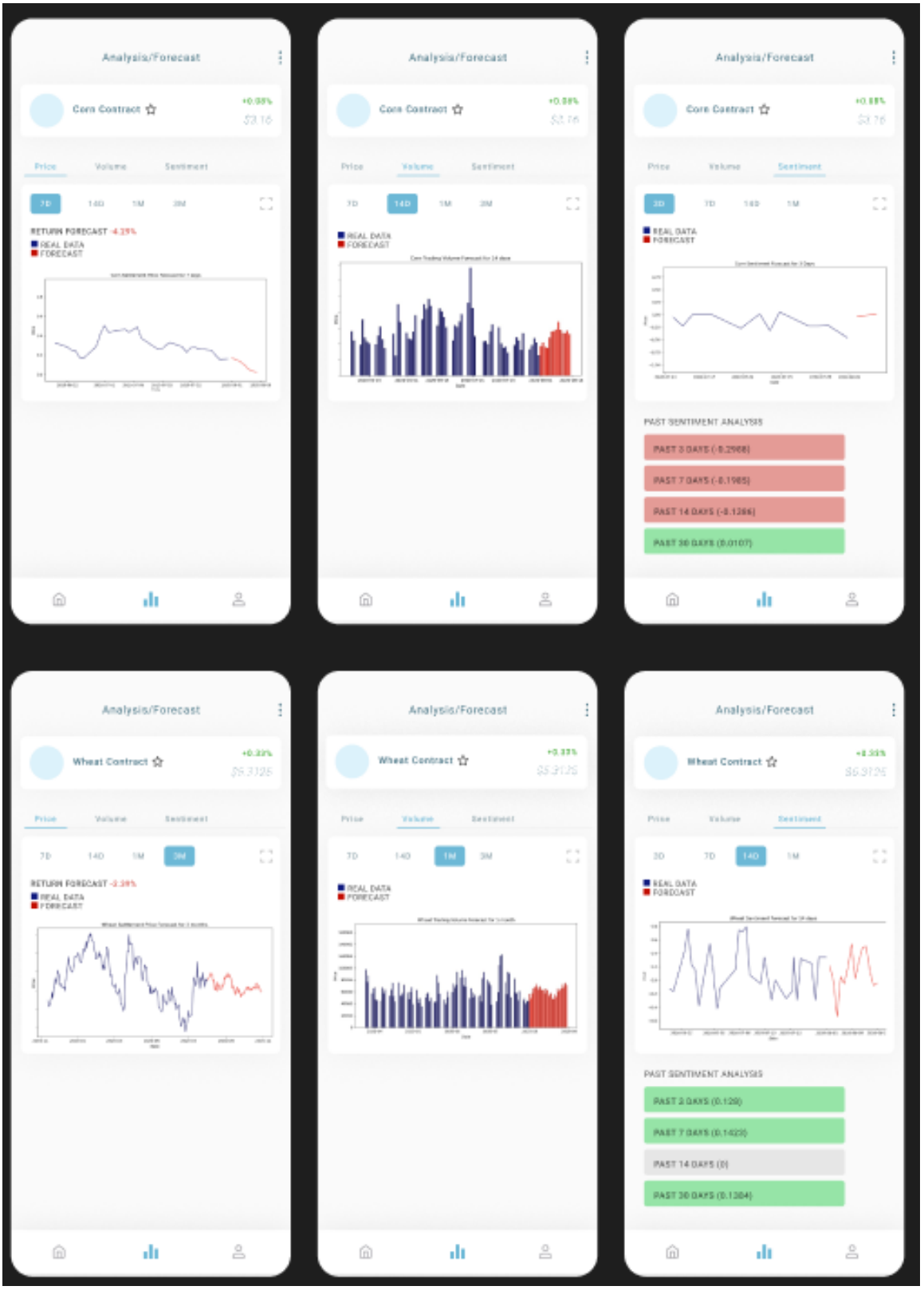

Figure 10: A snapshot of the mobile version of price, volume, and sentiment forecast/analysis pages in FIGMA design