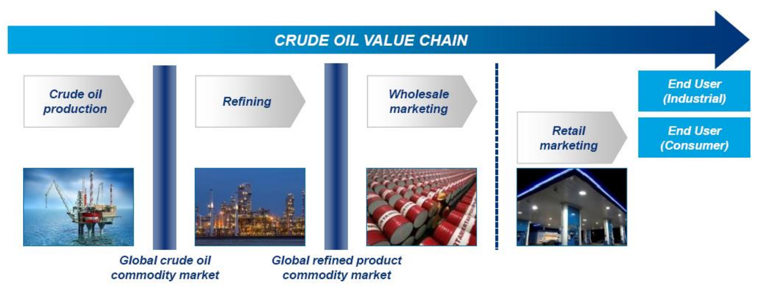
SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Computing and Information Systems

Shiny-RFSA: A Web-enabled Visual Analytics Application For Retail Fuel Market Analysis In APAC

Ranice TAN Hui Qi
Supervisor: Associate Professor KAM Tin Seong

1. Introduction



The commodity market has a profound impact on the global economy, including the retail fuel sector (gasoline, diesel). Recent disruptions, such as the post-pandemic supply chain crisis in 2021 and geopolitical conflicts in 2022, have significantly affected retail fuel sales volumes and led to record-high refining margins. In response, governments in Asia have implemented temporary subsidies and price freezes to reduce burden on consumers. Looking ahead, the long-term outlook indicates a peak in retail fuel demand in the coming decade, driven by electrification of vehicles. Amidst the global energy crisis and imminent energy transition, operators must reconsider strategies and transform their businesses swiftly.

2. Objective

The Retail Fuel Service Application (RFSA) product offered by information provider Woodmac provides valuable insights into the dynamics of retail fuel markets. However, the current analysis of RFS data is conducted using conventional spreadsheets, which, although easy to use, can be manual and time-consuming. Furthermore, analysts face challenges in handling complicated calculations that involve multiple data sources and links, leading to longer processing times and reduced efficiency. To navigate these challenges, the proposed interactive Shiny-RFSA will consolidate data and streamline the analysis process. The tool aims to provide a seamless experience for market analysts by offering features that:

- Benchmark retail fuel markets in selected Asia Pacific,
- Forecast retail fuel prices and margins in various markets, and
- Analyse the impact of regulatory frameworks on prices or margins

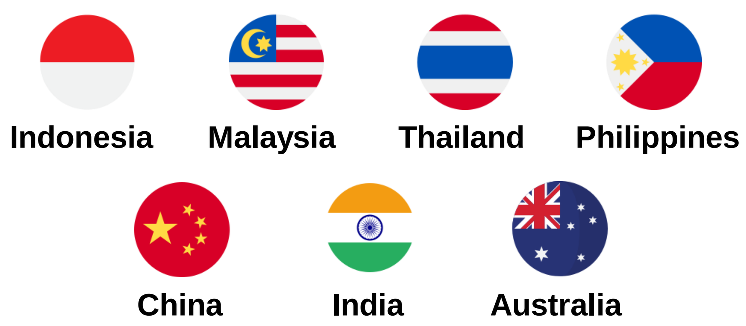
3. Research Methods

1. Data Preparation

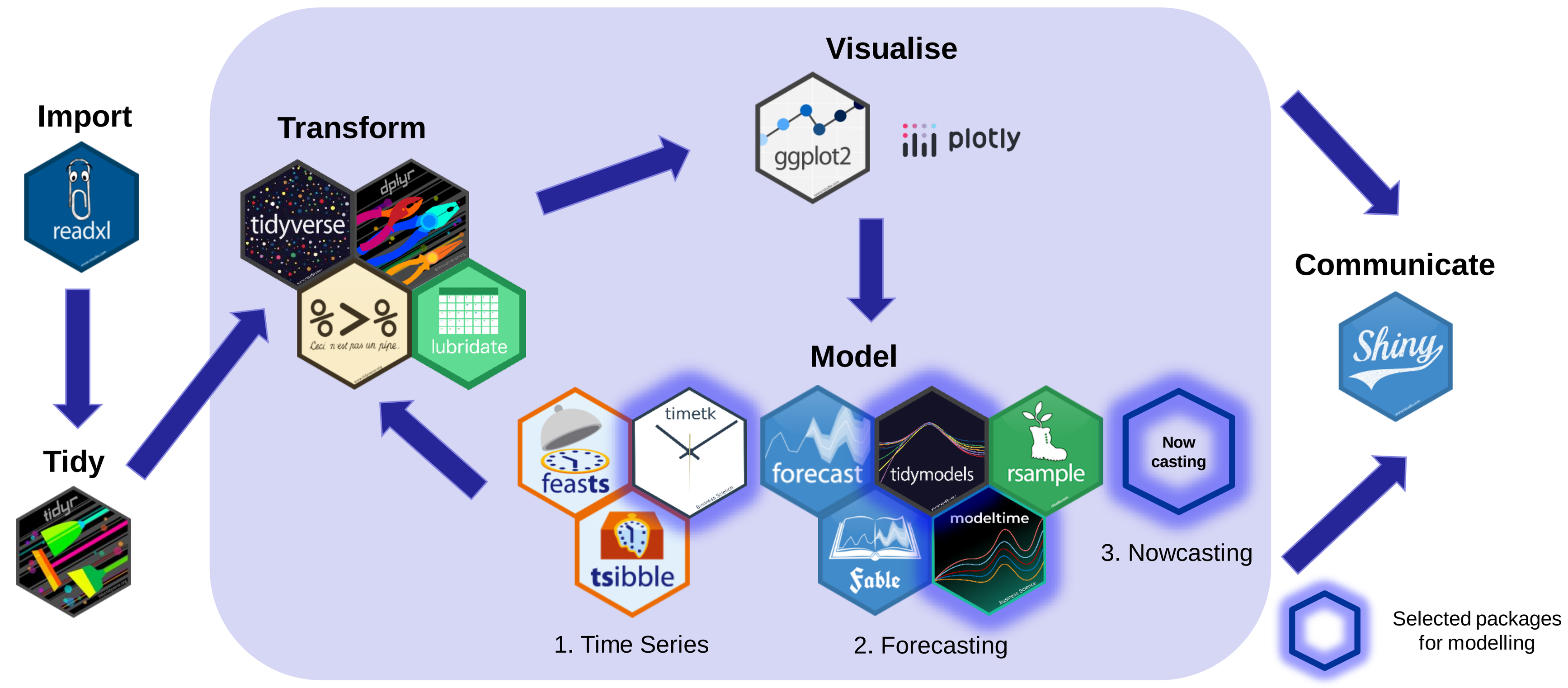
The following datasets were scrapped:

Data	Source
Retail pump pricing at the panel	National Statistics
Pricing policies (tax)	National Statistics
Price of product exiting refinery	Internal
Exchange rates (vs. USD)	Internal
Freight rates	Internal

BeautifulSoup PyPDF2

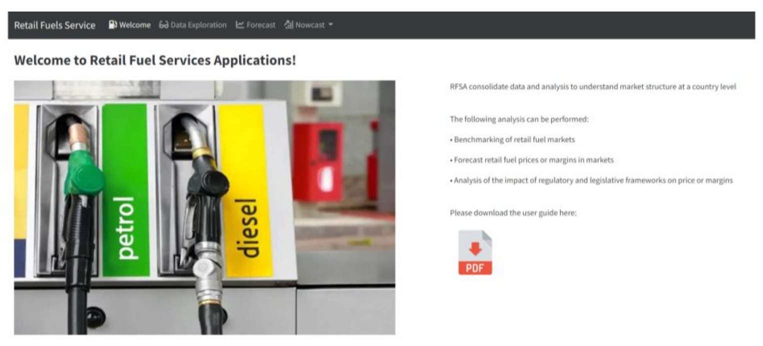


2. Data Transformation and Modelling Process



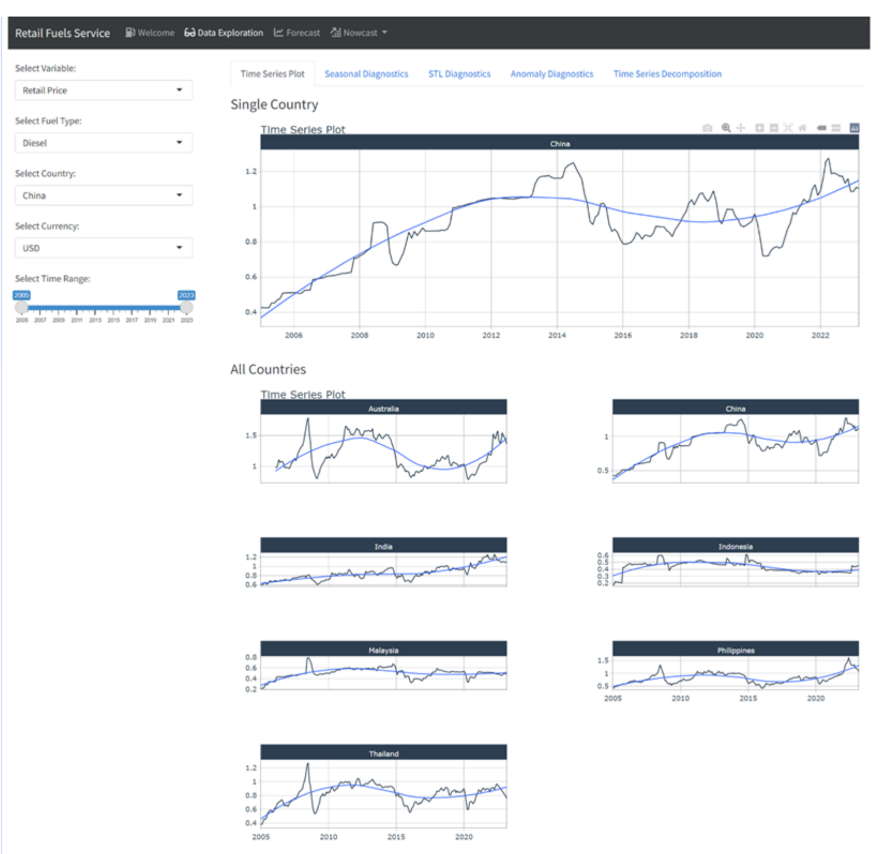
3. Application Architecture

To achieve a user friendly and informative dashboard that provides interactive features based on user fuel and market preferences, the application follows a minimalist design for user interface, user control and is configured to match data between the digital and real-life applications closely. (Molich & Nielsen, 1990)



3. Results

3.1 Data Exploration



Exploratory Data Analysis (EDA) plays a crucial role in understanding and extracting meaningful insights into pricing trends, patterns, and factors influencing the fluctuations in prices.

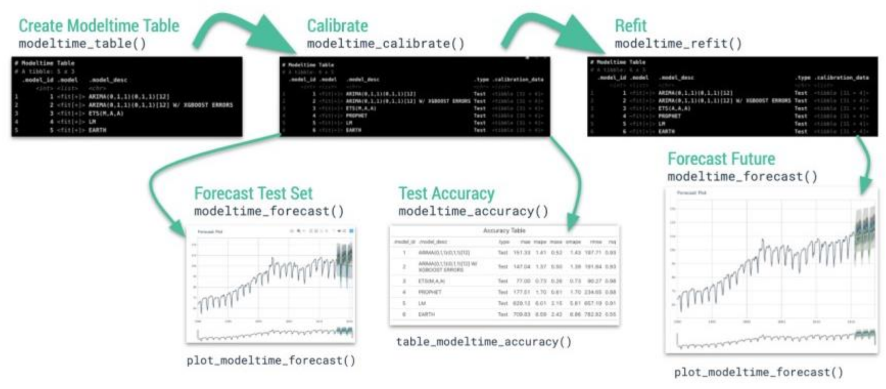
EDA is split into 5 comprehensive analyses:

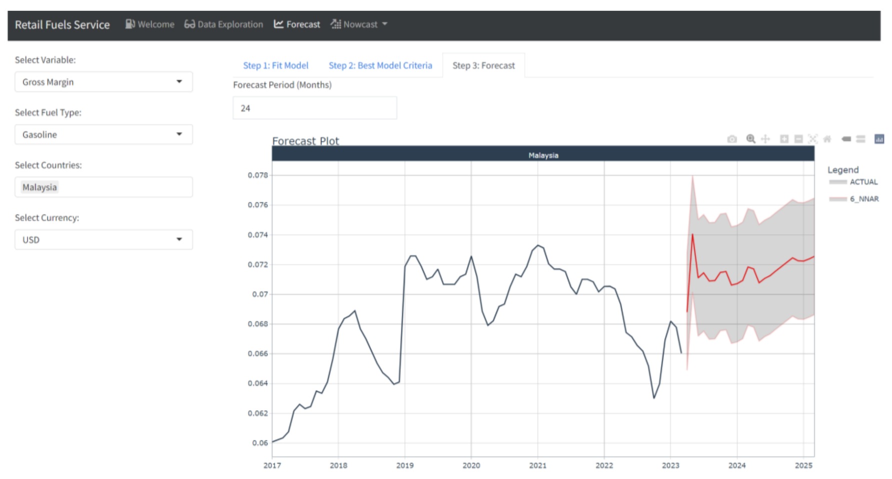
1) Time Series Plot: For benchmarking
2) Seasonal Diagnostics: Understand and predict patterns that occur at regular intervals
3) STL Diagnostics: Study and analyze trend, seasonal and remainder component separately
4) Anomaly Diagnostics: Detecting outliers that signify special events
5) Time Series Decomposition: Analyze the autocorrelation structure, identify significant lag values for forecasting model selection

Users can select the data to visualize on the left pane, and navigate through the tabs. The size of the visualisations are designed to fit within the user's screen without having to scroll.

For example, we select the retail price of diesel for China from 2019 to 2023. From the red dots in the anomaly plot, we can verify 0.1 level of significance, that 2020 and 2022 exhibit unusual patterns that deviate from the expected behavior of the data. This confirms that significant events like the global lockdown in 2020 and energy crisis in 2022 are indeed atypical occurrences that resulted in disruptions to the global retail fuel market.

3.2 Time Series Forecasting





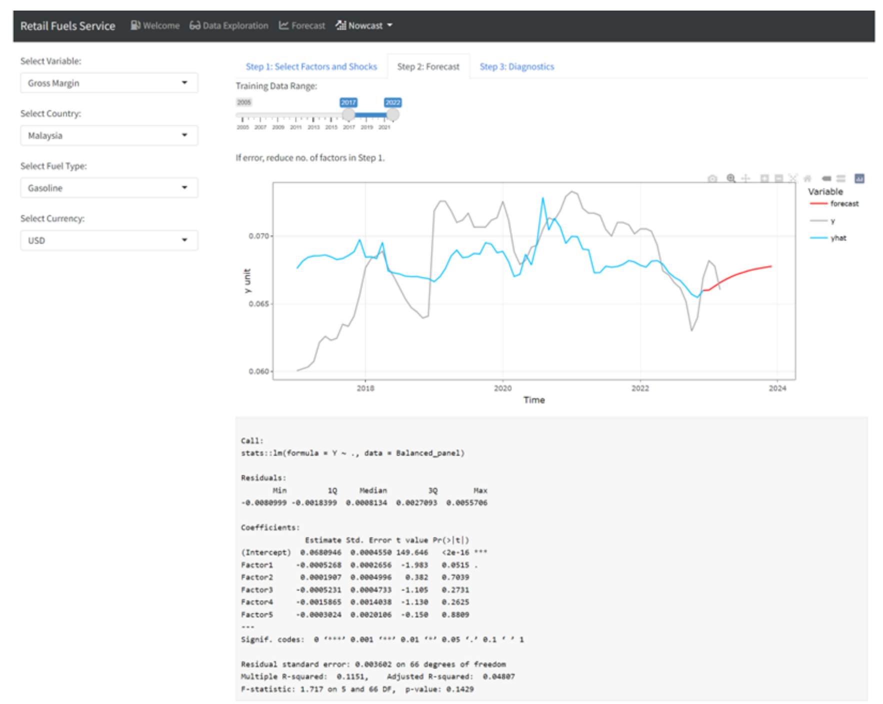
The *modeltime* workflow on the left, is designed to streamline the process of time series forecasting, making it accessible to users to someone who do not have deep knowledge into complicated models, like market researchers. Hence, the tabs on the forecast page is designed to follow the *modeltime* workflow:

1. Pre-selected models are fit and trained on the historical time series data
2. Evaluate the performance of the models based on the appropriate evaluation metric
3. Refit for forecasting

A forecast of Malaysia's gasoline gross margin is shown on the left, based on training data from 2017 to 2020, and RMSE criteria. It shows a general upward trend in gross margin with spikes in the middle of 2023. This suggests positive expectations for gasoline producers in Malaysia and could be a result of stabilisation of diesel post the energy crisis. However, the magnitude of the spike may be overstated in the 2nd quarter of 2023 due to the lag in data updated (Latest data is Mar'2023). As forecasting models rely on historical data and assumptions, and they may not always accurately capture the exact magnitude of future fluctuations.

3.3 Nowcasting

To address the challenges of data lags in forecasting process, nowcasting focuses on generating real-time estimates and adjusts its forecasts, can be used (Giannone, 2008). Nowcasting also leverage real-time data sources e.g. web scrapping, for more timely and accurate forecasts. (Bańbura, Modugno, 2010)



The sequence of the tab panels in the nowcasting tab is designed based on its workflow:

1. Select the numbers of factors and shocks
2. Fit and forecast the model. Recalibrate to select the best fit model.
3. View the model diagnostics

A nowcast of Malaysia's gasoline gross margin is shown on the right, also based on training data from 2017 to 2020. In this case, the estimated values in blue is close to the actual. Similar to forecast, the nowcast indicates a general upward trend in gross margin, but a much gentler increase. Comparing the performance summary to forecast, nowcast has a similar residual standard error but a higher r-squared. This could be attributed to the its ability to address the data gap in the 2nd quarter of 2023.

4. Discussion

The Shiny-RFSA application offers a comprehensive set of analysis tools that streamline the market research process for analysts. The design of the application is simplified to cater to users who do not have vast data science and programming knowledge. Data exploration enables analysts to benchmark markets, identify seasonal changes, detect unusual events and identify autocorrelations. The forecast tab offers a variety of models that the analyst can use for forecasting. However, as forecasting models typically rely on historical patterns and trends to make predictions, they do have limitations when it comes to capturing and predicting shocks in the market. Hence, nowcast value-adds in providing real-time estimates by capturing current trends and real-time information instead of using historical data. The model performance of nowcast is generally better than the forecast models, but it is only good for short-term use. For longer-term forecast, forecast will still have to be used. Thus, a combination of data exploration, forecast and nowcast can help market analysts gain a better understanding of market dynamics and identifying growth opportunities.

5. Conclusion and Future Work

Overall, the development of Shiny-RFSA is a substantial advancement in enhancing the efficiency and productivity of market analysts in delivering valuable insights into the retail fuel markets across Asia.

As this is the initial prototype, there are still several potential avenues for enhancements:

- Integrating additional tuning parameters and advanced models to enhance the accuracy of forecasting, once analysts have gained proficiency with data science models
- Implementation of real-time data integration from multiple sources to provide analysts with up-to-date information for more timely recommendations
- Additional development of capabilities can be included to segment the market, to enable more targeted insights and recommendations and expand the reach of Shiny-RFSA beyond Asia.

References

Molich, R., & Nielsen, J. (1990). Heuristic evaluation of user interfaces. *CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249 - 256. doi:https://doi.org/10.1145/97243.97281

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics, 665-676.

Bańbura, M., & Modugno, M. (2010). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. Frankfurt: European Central Bank.