**U.S. Gasoline Prices Time Series Analysis Documentation**

**1. Project Overview**

* **Objective:** This project aims to analyze historical trends in U.S. gasoline prices and forecast future prices. The analysis will help identify key patterns, understand seasonal fluctuations, and predict potential price movements based on past data.
* **Goals:**
  + Understand the long-term trends and seasonal patterns in gasoline prices.
  + Apply statistical methods to model and forecast future prices.
  + Provide actionable insights based on historical data and model predictions.
* **Dataset Description:**
  + Source: U.S. Energy Information Administration (EIA) (.gov)
  + Columns:
    - * Date
      * U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)

**2. Data Preparation**

* **Data Loading**: The dataset is loaded into R, and necessary libraries (tidyverse, lubridate, etc.) are initialized for data manipulation and time series analysis.
* **Data Cleaning**:
  + **Whitespace Removal**: Extra spaces in the Date column are removed using trimws() to ensure consistency.
  + **Date Parsing**: The Date column is parsed using parse\_date\_time() with a month-year format (my).
* **Data Transformation**:
  + The dataset is transformed into a time series object (ts), which allows for better handling in time series modeling and forecasting. The data is sorted chronologically.

**3. Exploratory Data Analysis (EDA)**

* **Initial Visualization**:
  + A time series plot of gasoline prices shows price trends over time, including any obvious uptrends or downtrends and noticeable fluctuations.

A graph showing the price of gasoline

Description automatically generated

* **Seasonality and Trend Decomposition**:
  + The decomposition of the series into trend, seasonal, and random components highlights the underlying seasonal patterns and long-term trend in gasoline prices.

A graph of different types of waves

Description automatically generated with medium confidence

* **Stationarity Check**:
  + Using the Augmented Dickey-Fuller test, the stationarity of the series is tested. A non-stationary result indicates the need for differencing or other transformations before building a model.

**4. Time Series Modeling**

* **Differencing for Stationarity**:
  + To stabilize the mean and achieve stationarity, the series is differenced (if required), removing trends or seasonality from the series.

A graph showing a gas price

Description automatically generated

* **ACF and PACF Analysis**:
  + Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are used to identify the order of AR (AutoRegressive) and MA (Moving Average) components in the series.

A graph of a gasoline price

Description automatically generated

A graph of a gas price

Description automatically generated

* **ARIMA Model Selection**:
  + Using auto.arima(), an optimal ARIMA model is selected based on the AIC (Akaike Information Criterion). This step is followed by fine-tuning if required based on residual diagnostics.
* **Model Diagnostics**:
  + Model residuals are examined to ensure they resemble white noise (i.e., no autocorrelation or significant patterns remain in residuals), indicating a well-fitted model.

A graph of a graph

Description automatically generated with medium confidence

**5. Forecasting**

* **Future Price Prediction**:
  + The ARIMA model is used to predict gasoline prices for the next 12 months, generating point forecasts along with confidence intervals to convey forecast uncertainty.
* **Forecast Visualization**:
  + A plot of the forecast shows the predicted price trend along with confidence bands for each month. This visualization helps in understanding expected changes and their potential ranges.

A graph showing the price of gasoline

Description automatically generated

**6. Model Evaluation**

* **Accuracy Metrics**:
  + The accuracy of the model’s predictions is assessed using common metrics:
    - **RMSE (Root Mean Squared Error)**: Measures average prediction error.
    - **MAE (Mean Absolute Error)**: Average absolute differences between predicted and actual values.
    - **MAPE (Mean Absolute Percentage Error)**: Shows the error as a percentage, which helps in understanding the accuracy relative to actual values.
  + These metrics provide insight into the reliability of the model’s forecasts.

A black background with white text

Description automatically generated

**7. Results and Insights**

* **Results**

1. **Historical Trend Analysis**:
   * Analysis of U.S. gasoline prices from April 1993 to October 2024 reveals a general upward trend over the long term, interrupted by periods of volatility.
   * Significant price fluctuations occurred around 2008 (global financial crisis) and 2022 (post-pandemic recovery and global supply issues), highlighting the sensitivity of gasoline prices to economic events.
2. **Model Selection and Forecasting**:
   * The ARIMA model was selected based on statistical criteria (e.g., AIC, BIC) and autocorrelation analysis, ensuring it captures the time series characteristics effectively.
   * Using this model, a forecast for the next 12 months was generated, projecting future gasoline prices and providing confidence intervals to reflect uncertainty in the predictions.
3. **Forecasted Price Stability**:
   * The forecast indicates relatively stable gasoline prices over the next year, with no significant upward or downward trend.
   * The predicted prices suggest a range that reflects a balance between past price fluctuations and the recent declining trend after the 2022 peak.
4. **Uncertainty and Prediction Intervals**:
   * Prediction intervals (80% and 95%) were included to account for forecast uncertainty. These intervals widen as the forecast moves further into the future, highlighting increased uncertainty in long-term predictions.
   * The intervals suggest that while prices are expected to remain stable, there is a possibility of deviations, either upward or downward, due to unpredictable market conditions.

* **Insights**

1. **Short-Term Stability with Potential for Market Impact**:
   * The stable forecast suggests a possible period of calm for gasoline prices in the short term, which could be beneficial for consumers, businesses, and policymakers.
   * However, given historical volatility, any unexpected economic events or disruptions in supply chains could result in deviations from the forecast.
2. **Sensitivity to External Events**:
   * The large fluctuations observed during past crises emphasize gasoline prices’ sensitivity to economic and geopolitical events.
   * The model’s forecast assumes normal market conditions, so significant deviations could occur if there are new supply shocks, demand surges, or policy changes (e.g., environmental regulations or trade policies).
3. **Value for Decision-Making**:
   * This forecast can serve as a guide for stakeholders such as businesses (especially those in logistics or transportation), consumers, and government agencies, helping them plan for the next year.
   * However, users of this forecast should remain cautious of the inherent uncertainties and continue monitoring the market for early signs of changes that may affect prices.
4. **Future Model Improvements**:
   * To enhance forecasting accuracy, consider incorporating **external factors** (e.g., oil supply data, inflation rates, geopolitical indicators) that may provide a more comprehensive picture of price influences.
   * Seasonal or structural changes in gasoline consumption patterns might also be explored, as they could add context to the pricing trends over time.