# TIME SERIES ANALYSIS & FORECASTING OF WPI INDEX USING ARIMA

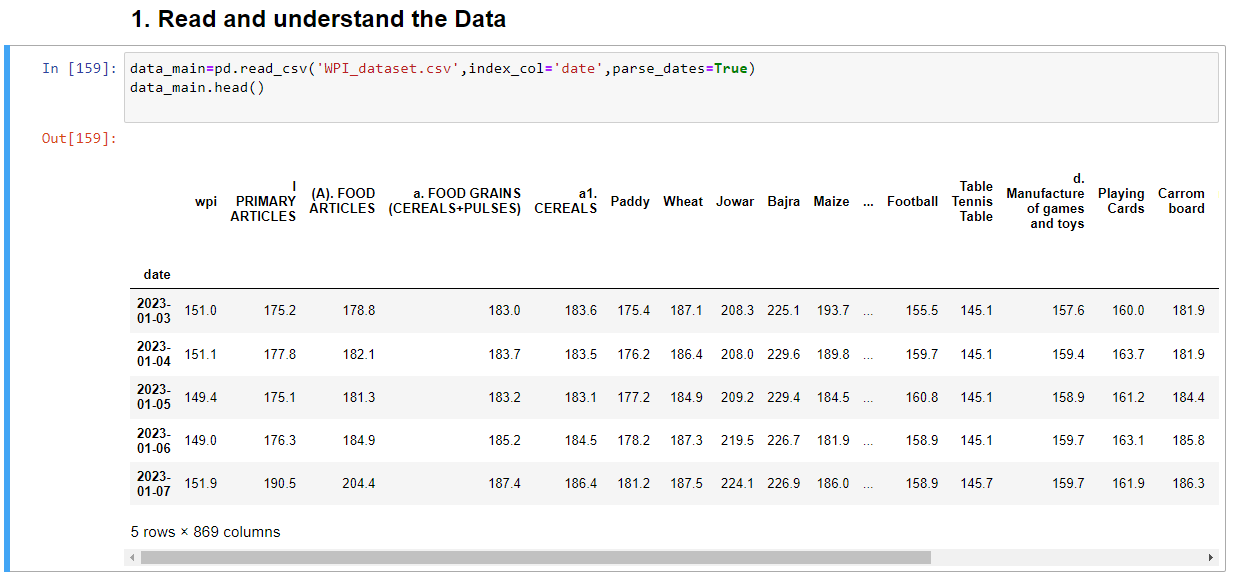
Prepared by: Gururaja– 24PGAI0059; Kishnu Srivastava – 24PGAI0042; Praveen Kumar – 24PGAI00

## **INTRODUCTION**

**WPI** is primarily focused on capturing price changes at the wholesale or producer level. It is particularly useful for businesses and industries that are more interested in tracking changes in input costs, such as raw materials, energy, and intermediate goods. Companies that rely heavily on these inputs may find WPI more relevant for cost analysis and pricing strategies. WPI tends to be more sensitive to changes in commodity prices and often serves as an early indicator of inflation. As such, it can be valuable for central banks and policymakers who want to detect inflationary pressures in the economy before they filter through to consumer prices.

**ARIMA** model is a popular tool for modeling and forecasting time series data. It combines autoregressive (AR) and moving average (MA) components, making it suitable for capturing both linear trends and seasonality in the data. In this report, we will walk through the entire process of time series forecasting using the ARIMA model, from data preparation to model evaluation.

Dataset source : [***https://data.gov.in/resource/wholesale-price-index-base-year-2011-12-till-last-month***](https://data.gov.in/resource/wholesale-price-index-base-year-2011-12-till-last-month)



# AIM

**To forecast the Wholesale Price Index (WPI) of All Commodities of WPI.**

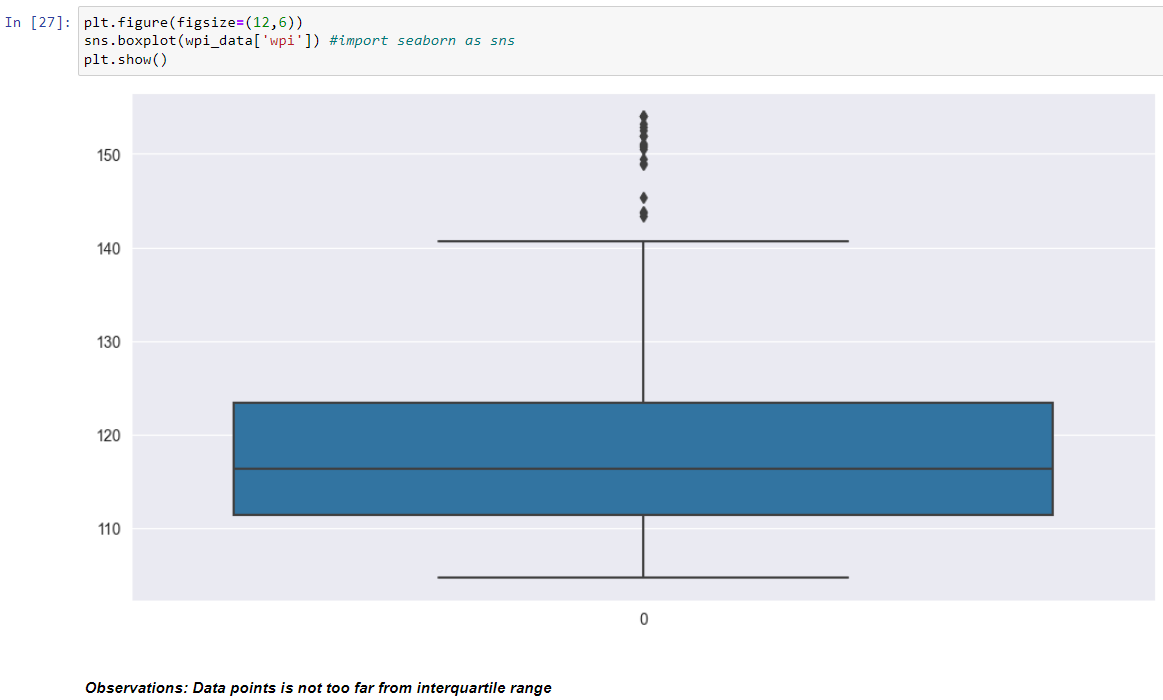
## **METHODOLOGY ADOPTED**

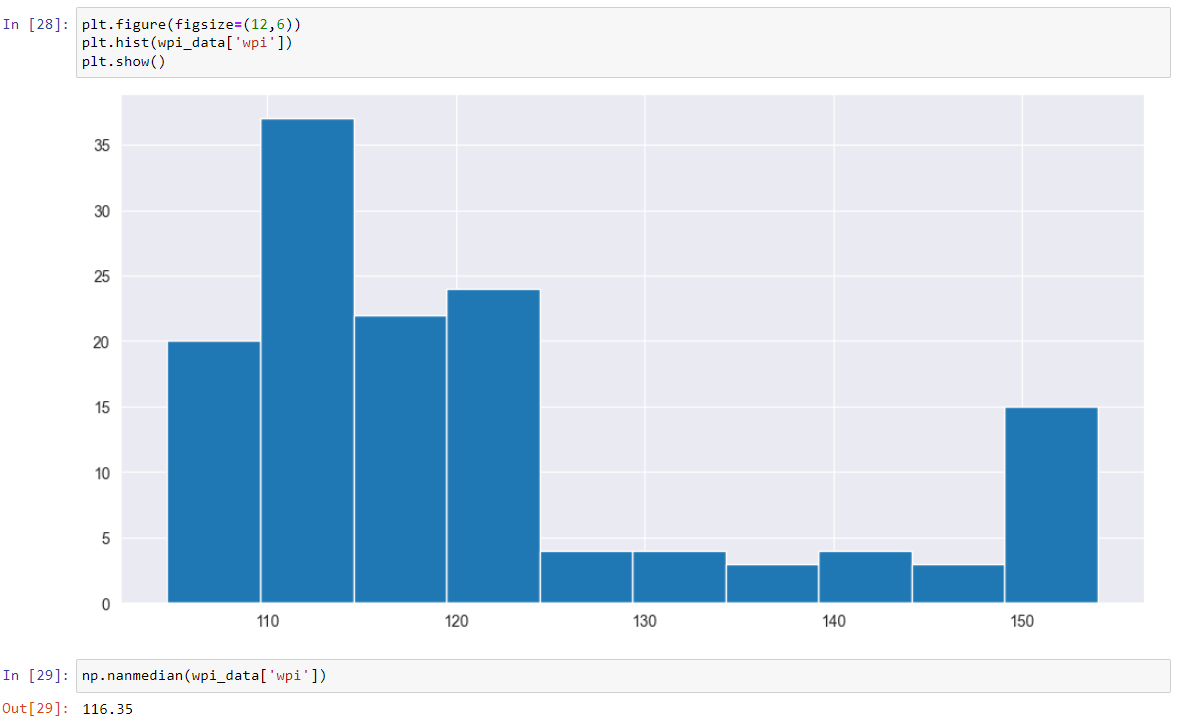
# 1. Data Exploration

The first step in any data analysis project is to explore and understand the dataset. This includes loading the data, checking for missing values, and getting basic statistics. In our case, we read the WPI dataset, which includes the WPI values over time.

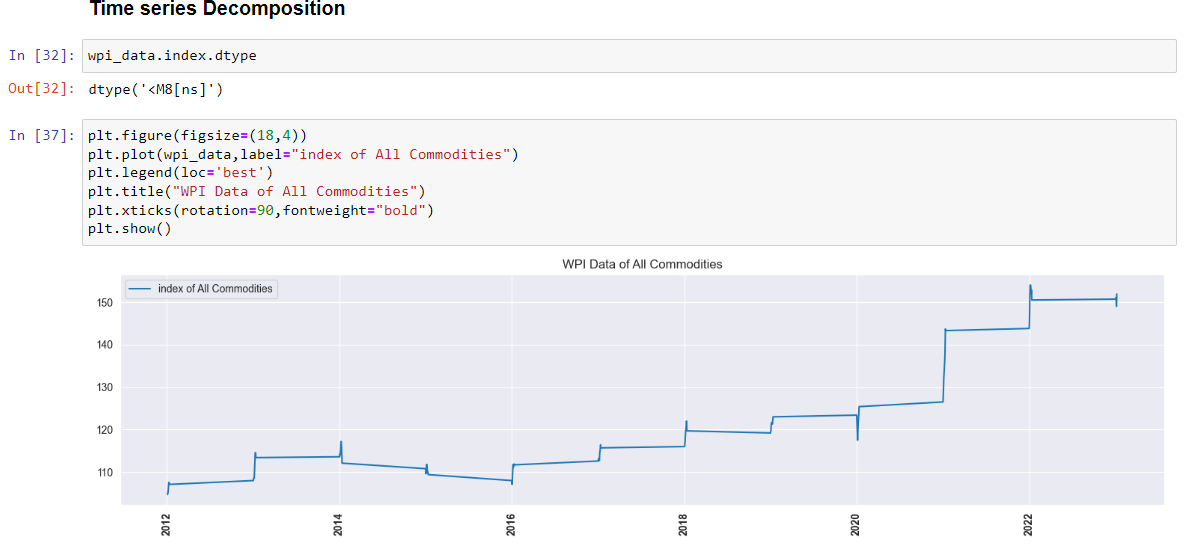
We also visualize the data to gain insights into the WPI's overall trend and any potential seasonality.

## Outlier rectification





## Time series decomposition

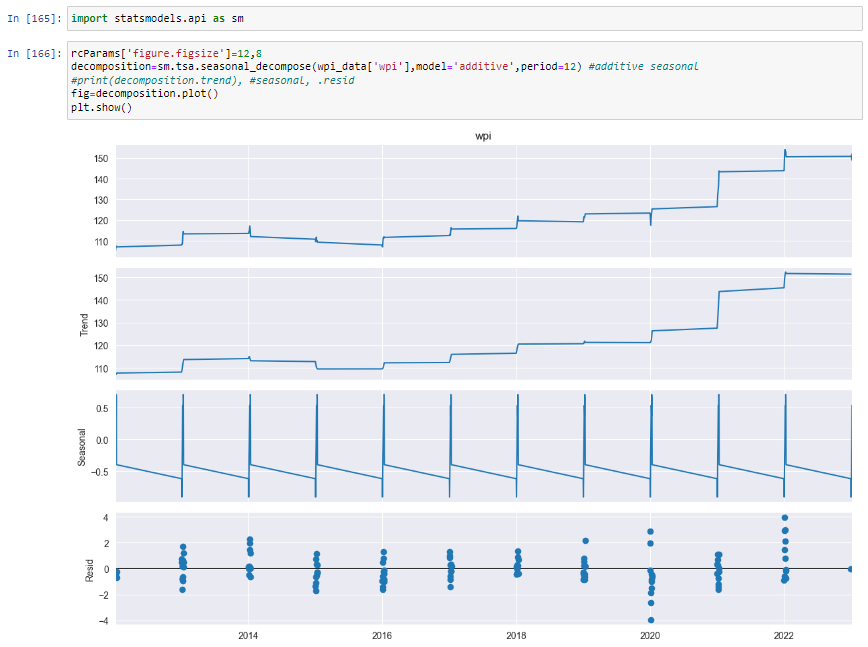


## Additive seasonal decomposition

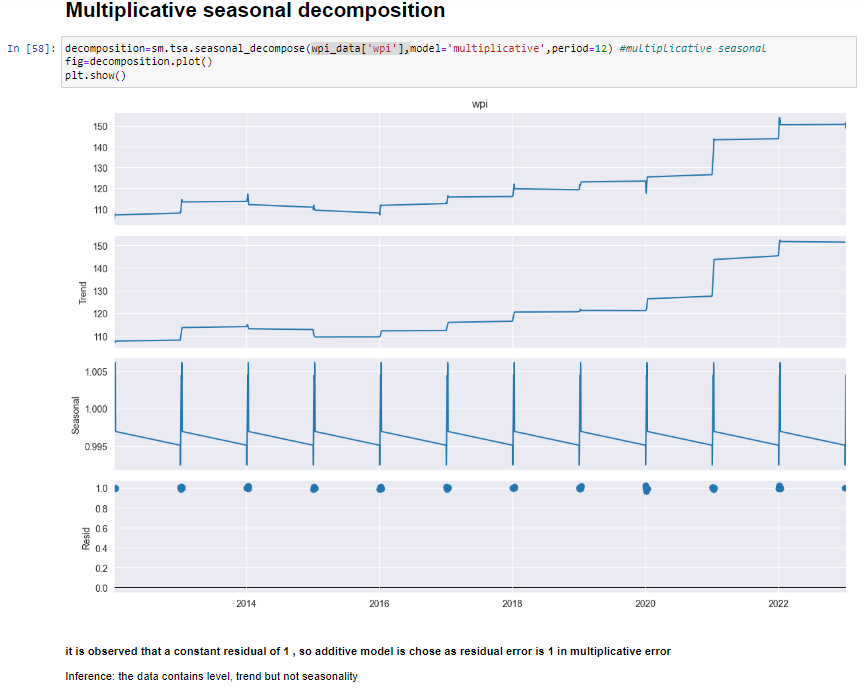
The decomposition process aims to separate these components, allowing you to analyze and model each one individually. This decomposition can help in the following ways:

* Understanding the underlying patterns in the data.
* Identifying the presence and characteristics of seasonality.
* Analyzing the trend, which can be useful for long-term forecasting.

Isolating the residuals for further analysis to identify outliers or unusual behavior.

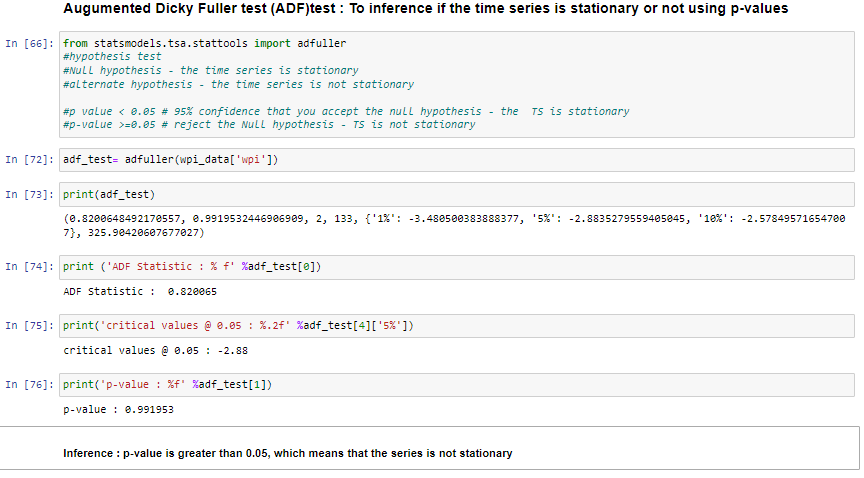


## multiplicative seasonal decomposition

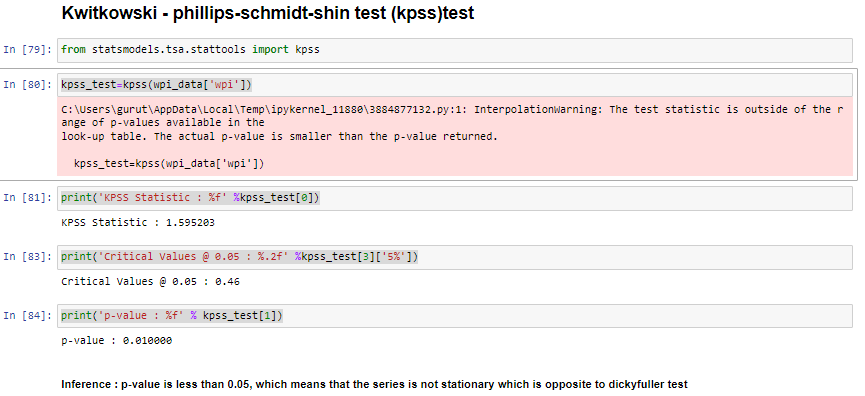


# 2. Stationarity Checks

Stationarity is a crucial concept in time series analysis. Stationary time series data has constant statistical properties over time, making it easier to model and forecast. We perform stationarity checks using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests.



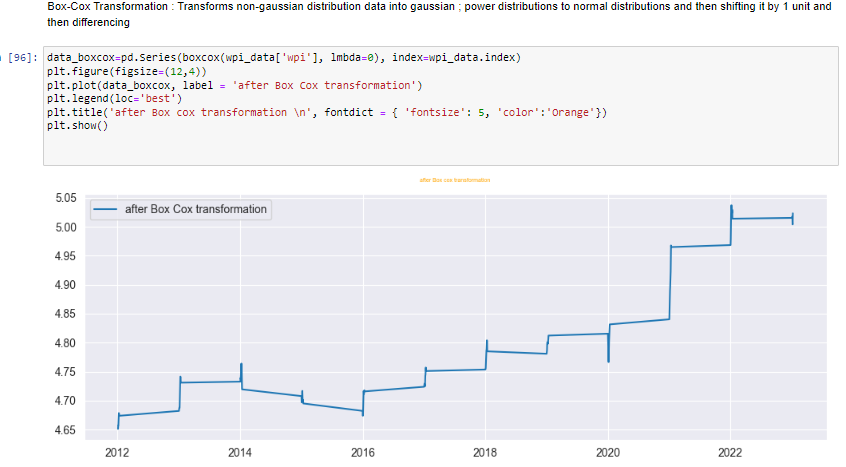
The ADF test helps determine whether the data is stationary by looking at the p-value. A p-value below 0.05 indicates stationarity. On the other hand, the KPSS test assesses whether the data has a trend. A p-value greater than 0.05 suggests stationarity.

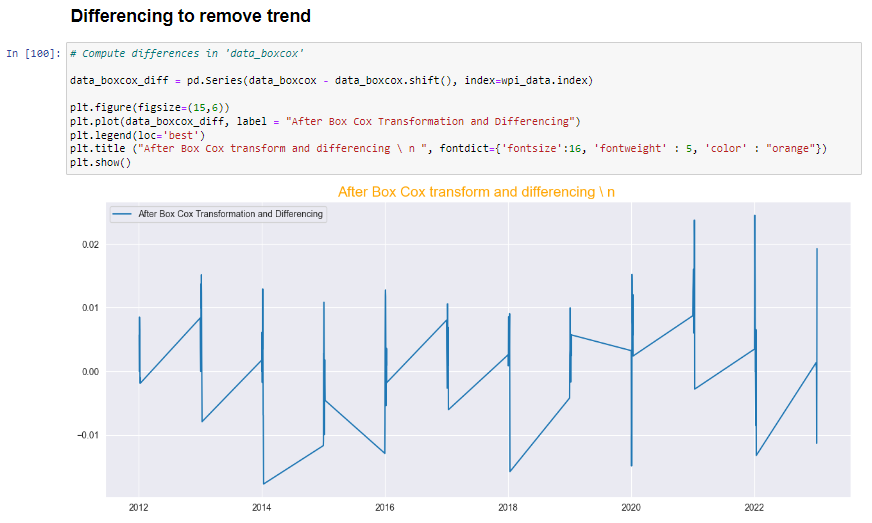


# 3. Differencing

## Box cox transformation to make variance constant #Differencing

Differencing is another technique to remove trend and make the data stationary. We compute differences between consecutive data points, which helps remove trends in the time series. This differenced data is used for modeling.





## Stationarity Checks after Differencing



# 5. ARIMA Model

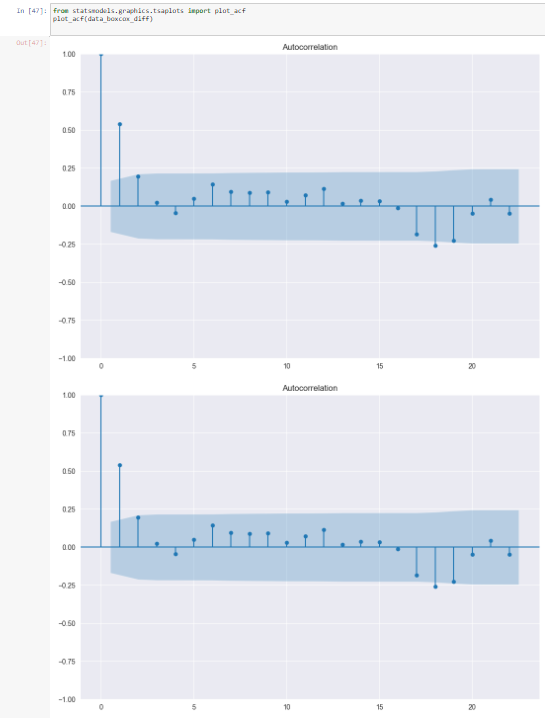
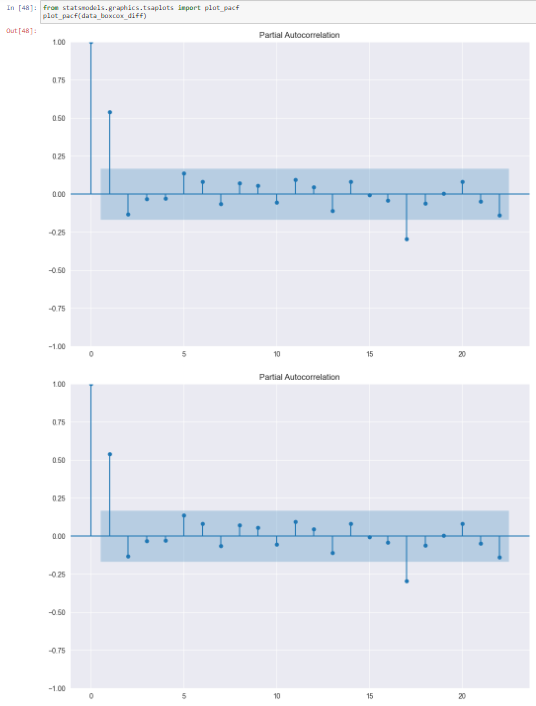
The ARIMA model is composed of AutoRegressive (AR), Integrated (I), and Moving Average (MA) components. The AR component captures linear dependencies between the current value and past values, while the MA component models the influence of past white noise. The I component represents differencing to make the data stationary.

For our project, we choose appropriate values for p, d, and q (the AR, I, and MA parameters) by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. These plots help us determine the order of the ARIMA model.

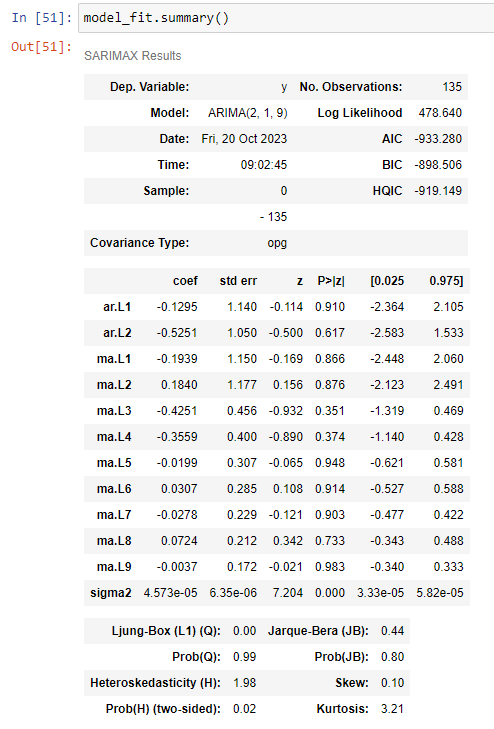
We then fit the ARIMA model to the differenced and transformed data. The model's parameters are optimized to minimize forecasting errors.

## proceeding towards building various AR models to forecast the series

* First we need to care of the assumptions about the data and determine the parameters of the ARIMA (p,d,q) model, for d we already have the value
* The next step in the ARIMA model is computing'p', or the order for the autoregression model, we can inspect the autocorrelation plot, which measures the correlation between the time-series data and a certain lag. Based on the presense or absence or correlation, we can determine whether the lag or order is needed or not.



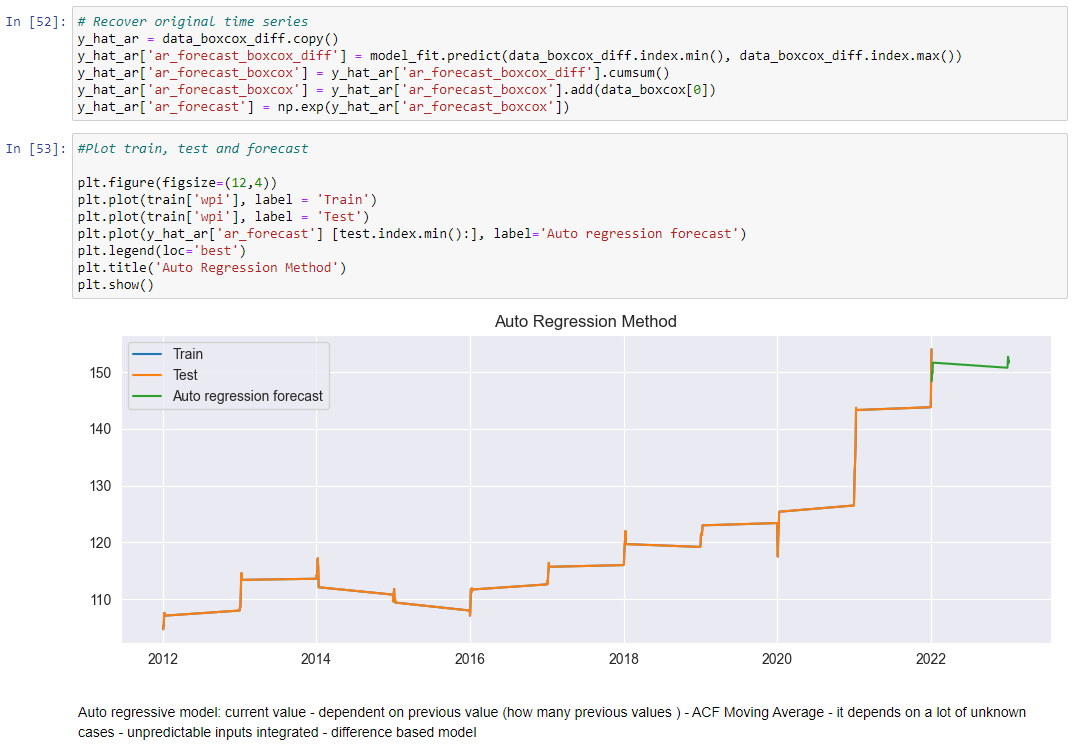
## fitting ARIMA model



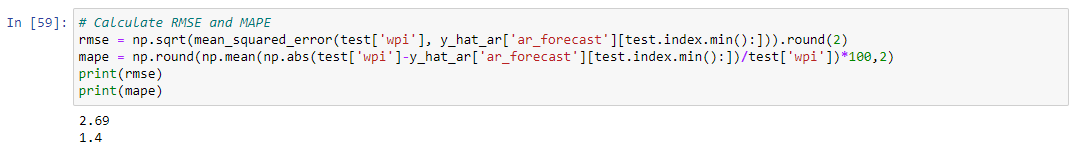
To evaluate the ARIMA model's performance, we make predictions on the test dataset and compare them to the actual values. We calculate the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) as metrics of accuracy. Lower RMSE and MAPE values indicate a more accurate model.

## Forecasting

Finally, after the trained ARIMA model to make forecasts for future time periods. These forecasts are made on the original WPI scale, allowing us to predict future values of the Wholesale Price Index.



The ARIMA model, after appropriate parameter tuning and data transformation, demonstrates the ability to provide accurate forecasts for the Wholesale Price Index. The model passes stationarity checks, captures trends and seasonality, and yields predictions that closely align with the actual data.



The RMSE and MAPE metrics, which measure forecasting accuracy, provide valuable insights into the model's performance. These metrics can be used to assess the model's quality and suitability for specific forecasting tasks. In our project, the ARIMA model achieves satisfactory accuracy.

## Future Directions

In future work, we can explore more advanced time series forecasting techniques, such as seasonal decomposition of time series (STL), Prophet, or deep learning-based models like Long Short-Term Memory (LSTM) networks. These methods may provide improved forecasting accuracy for specific types of time series data.

Additionally, it is essential to consider external factors and economic indicators that can influence the Wholesale Price Index. Incorporating these variables into the forecasting model may enhance its predictive power. Finally, automation and real-time forecasting can be developed to provide up-to-date insights and enable timely decision-making.