Task-1 Presentation

K.R Mahaadevan

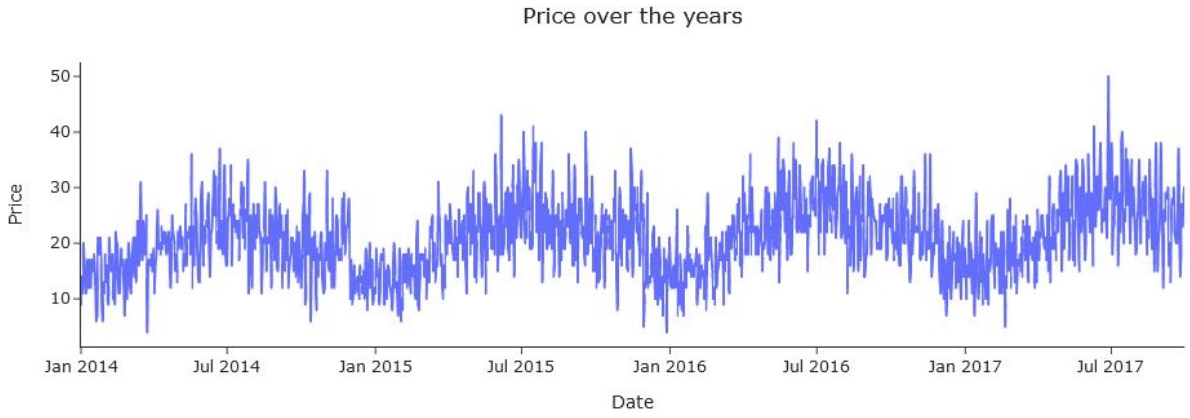
The objective of the case study was to design a time series forecasting model to predict values for the next three months based on the provided historical data

Preliminary Analysis

Firstly,the data was read using the **pandas** library . Then the date column was converted to a datetime object for further visualization

The data was then plotted to gain a initial understanding.

The plot shows a time series visualization of price fluctuations over a span of several years, ranging from January 2014 to mid-2017.

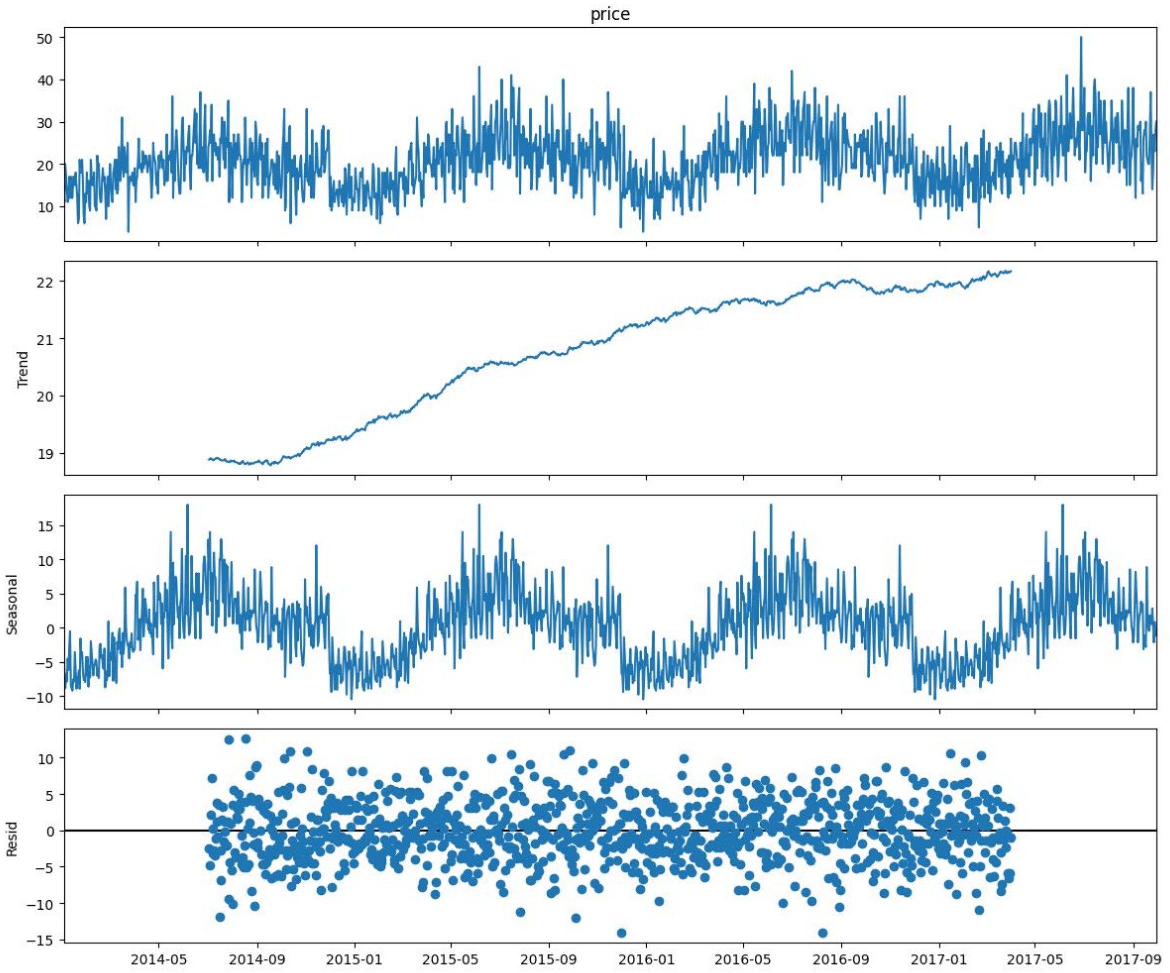


We can observe clear seasonal patterns , with prices fluctuating consistently within a specific range.

Furthermore,a seasonal decomposition of the time series data using the additive model was performed.

In the plot, we can observe that the price exhibits a strong seasonal pattern, with periodic fluctuations occurring over the course of a year suggesting ‘**yearly seasonality**’.

We can also observe that the trend appears to be gradually increasing over the time period.



Stationarity

To perform further analysis , it is important that the data is stationary.

This is validated using the **Augmented Dickey-Fuller(ADF) test.**

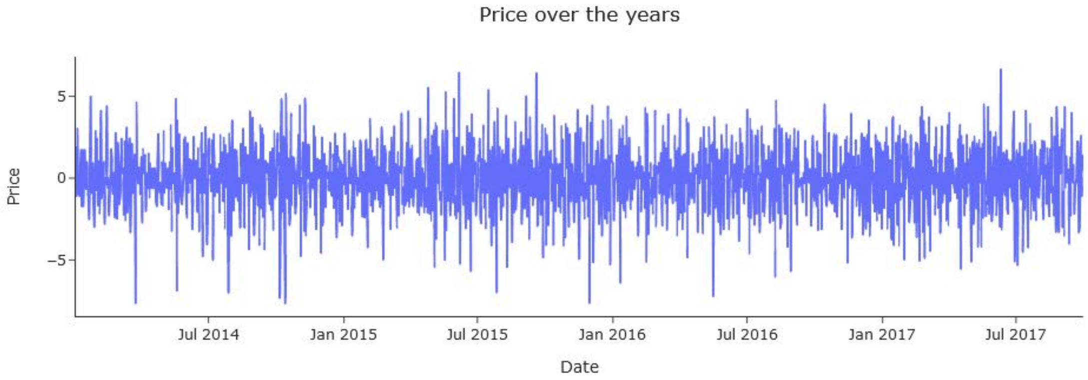
The ADF test compares the null hypothesis that a unit root is present in a time series i.e. non-stationary to the alternative hypothesis, which is it is stationarity

On running the test,we observe the p-value to be 0.144 which confirms the non-stationarity of the data.

Augmented Dickey-Fuller (ADF) Test

```
from statsmodels.tsa.stattools import adfuller
adftest=adfuller(data['price'])
print('P-Value',adftest[1])
```

```
P-Value 0.14495269098248537
```



Firstly, the variance of the series was stabilized by applying the Box-Cox transformation.

Checking for Stationarity

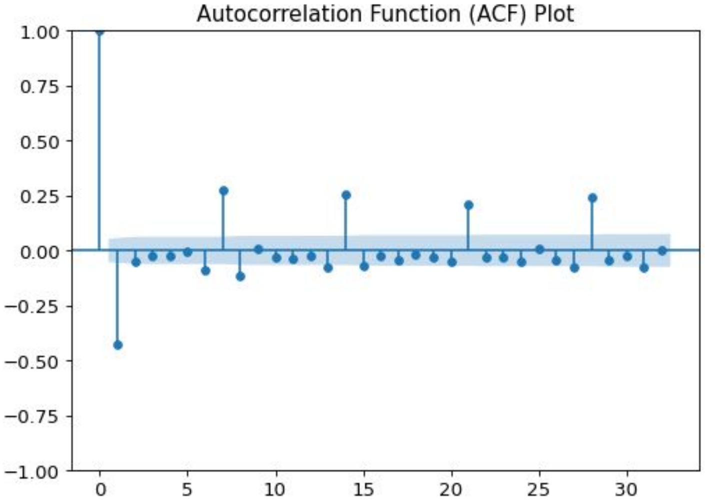
```
result = adfuller(data["price_diff"])
print("ADF Statistic:", result[0])
print(f"P-Value: {result[1]:.40f}")
print("Critical Values:", result[4])
```

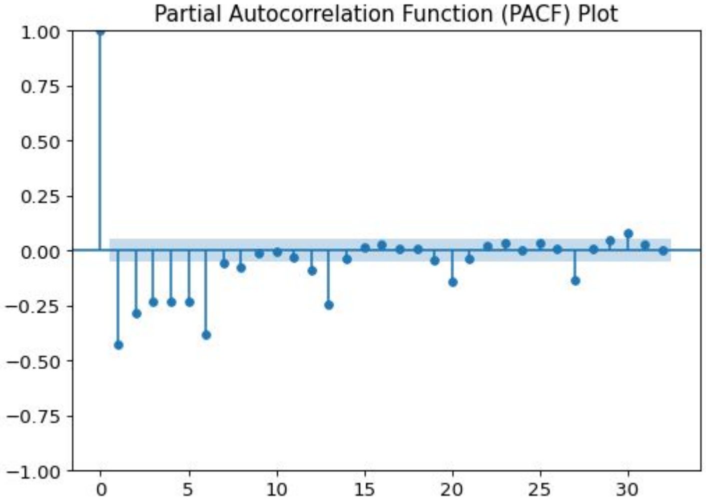
```
ADF Statistic: -13.169341695553152
P-Value: 0.0000000000000000000000012597698695168914
```

Next, to make the data stationary, first-order differencing was applied.

The Augmented Dickey-Fuller (ADF) test was then conducted again to verify stationarity

ACF and PACF Plots



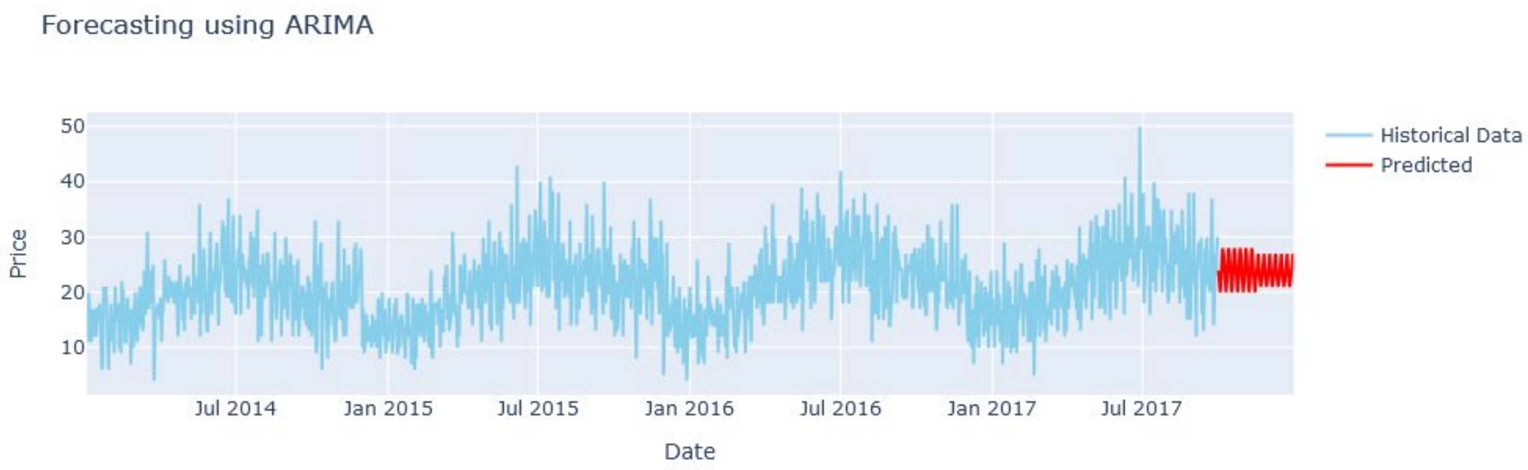


As observed earlier in the plots, the seasonality in the data follows a yearly pattern. To account for this seasonality, we incorporate ARIMA with Fourier terms as exogenous variables. This approach was chosen because the period component in the auto.arima function has a computational limit of 350 and other packages are computationally intensive

```
                                    SARIMAX Results
==========================================================================================
Dep. Variable:                          y   No. Observations:                          13
Model:                    SARIMAX(3, 1, 3)   Log Likelihood                     -2513.9
Date:                     Mon, 09 Dec 2024   AIC                                 5041.9
Time:                             05:29:00   BIC                                 5078.5
Sample:                         01-02-2014   HQIC                                5055.6
                              - 09-30-2017
Covariance Type:                       opg
```

The best model was ARIMA(3,1,3(0,0,0)

Forecasting the data



After finalizing the model, the prices for the next three months from the test1.csv file were predicted. Note that the scaled values were transformed back to their original scale.

Holt Winters Model

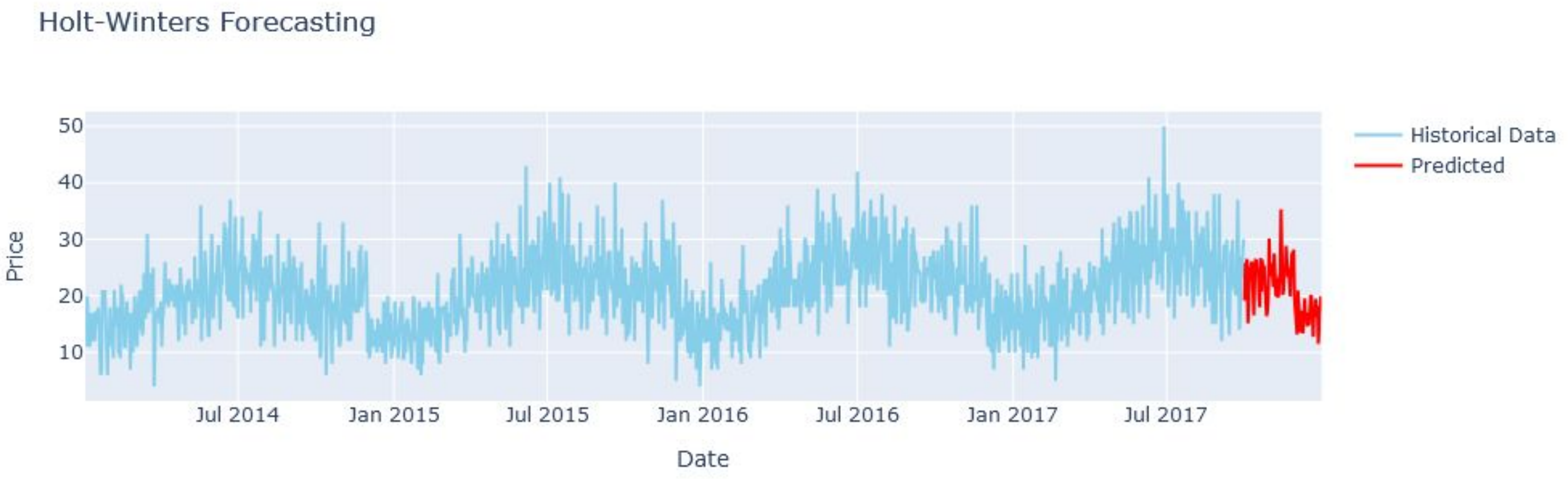
To compare the forecasts of the previous model, analysis was further done using the Holt Winters Model.

Holt Winters' forecast was primarily used as it captures both the trend and seasonality of the time series.

We observe that the Holt-Winters (HW) model allows the seasonal period to be set to 365 and has a lower AIC value compared to the SARIMAX model

```
                       ExponentialSmoothing Model Results
================================================================================
Dep. Variable:                price_boxcox   No. Observations:              1368
Model:                ExponentialSmoothing   SSE                        2586.168
Optimized:                            True   AIC                        1609.180
Trend:                            Additive   BIC                        3535.768
Seasonal:                         Additive   AICC                       1886.312
Seasonal Periods:                      365   Date:              Mon, 09 Dec 2024
Box-Cox:                             False   Time:                      05:25:30
Box-Cox Coeff.:                       None
================================================================================
```

Forecasting the data



As done earlier , the prices for the next three months from the test1.csv file were also predicted using the Holt Winters Model. Note that the scaled values were transformed back to their original scale.

From the plot, we observe that the HW model predicts the values with greater accuracy compared to the SARIMAX model. Based on the plot and the AIC of the models, the forecasted values from the HW model were selected and filled into the test1.csv file.