Journal of Hyperstructures 13 *(1) (2024), 142-155.*
https://doi.org/10.22098/jhs.2024.14953.1013
Research Paper

PYTHON APPROACH ON FUZZY TIME SERIES ARIMA (1, 1, 1) MODELTO ANALYSE ORIGINAL AND PREDICT RESULTS FOR ONLINE RETAIL OF FUEL BOOKING SERVICES

B. Mohamed Harif, M. Karthikeyan and K.Perarasan

ABSTRACT. This paper contributes to modeling and forecasting gas booking demand in an online retail environment using time series techniques. Our work demonstrates how historical demand data can be utilized to estimate future demand and its impact on the supply chain. The historical demand data were used to create several autoregressive integrated moving average (ARIMA) models using the Box-Jenkins time series procedure. The best model was selected based on four performance criteria: statistical results, maximum likelihood, and standard error. The selected model, ARIMA (1, 1, 1), was validated using additional historical demand data under the same conditions. The results demonstrate that the model can effectively estimate and forecast future demand for gas booking in an online retail environment. These findings will provide trustworthy guidance to the company's management in decision-making.

Key Words: Fuzzy Time Series, Online Retail, Python, ARIMA.
2010 Mathematics Subject Classification: Primary: 03E72; Secondary: 15B15.

Received: 25 April 2024, Accepted: 20 June 2024. Communicated by Nasrin Eghbali;
∗Address correspondence to B. M. Harif; E-mail: bmharif@rsgc.ac.in.

1. Introduction

In general, there are several techniques for forecasting demand, such as exponential smoothing. However, to apply these methodologies, we must first obtain historical data. When historical data is unavailable, we rely on estimates based on analogous circumstances or the engineer's experience. This situation often involves a high degree of ambiguity, which diminishes over time [1].

Forecasting involves predicting future levels of specific variables. Forecasting methods are classified into four types: qualitative, time series, causal, and simulation. A time series is essentially a sequence of observations arranged chronologically. Time series forecasting models predict demand using mathematical methodologies grounded in past data. These models operate under the assumption that future patterns will resemble past patterns, allowing us to effectively use historical data to forecast future demand [2]. Predictive accuracy tends to be high when the transition patterns are stable and periodic, but may falter when they exhibit highly irregular patterns. We can model time series using techniques like autoregressive integrated moving average (ARIMA), as well as classic statistical models such as moving average, exponential smoothing, and ARIMA itself. These models are linear because they predict future values as linear functions of past data. Researchers have predominantly focused on linear models in recent decades due to their simplicity and practicality. Time series forecasting models are primarily employed for demand forecasting [3, 4].

Kurawarwala and Matsuo utilized the autoregressive moving average hypothesis (ARIMA) to analyze seasonal fluctuations in demand and validated their models by assessing forecast performance. Meanwhile, Miller and Williams enhanced forecasting accuracy by incorporating seasonal components into their model using a multiplicative approach.

Hyndman expanded upon the research conducted by Miller and Williams by exploring alternative correlations between trend and seasonality within the framework of seasonal ARIMA models. When the seasonal adjustment order is large or diagnostic tests fail to confirm stationarity post-seasonal adjustment, the traditional ARIMA approach becomes impractical and often impossible to apply [5, 6].

The The study also compared the predictive performances of ARIMA and AR models, concluding that the ARIMA model provides the most accurate forecasts for new cases in India. These models were implemented using the Python programming language [7].

Karakoyun [8,9] has compared the ARIMA time series model with the LSTM deep learning algorithm for forecasting Bitcoin prices. In such cases, the static parameters of the classical ARIMA model are considered a primary constraint when forecasting highly volatile seasonal demand. Another limitation of the standard ARIMA approach is that it requires a large number of observations to achieve the best-fitting model for a given data series. An ARIMA model is denoted as ARIMA (p, d, q), where p represents the number of autoregressive components, d represents the number of differences, and q represents the number of moving average components.

2. Research Methodology

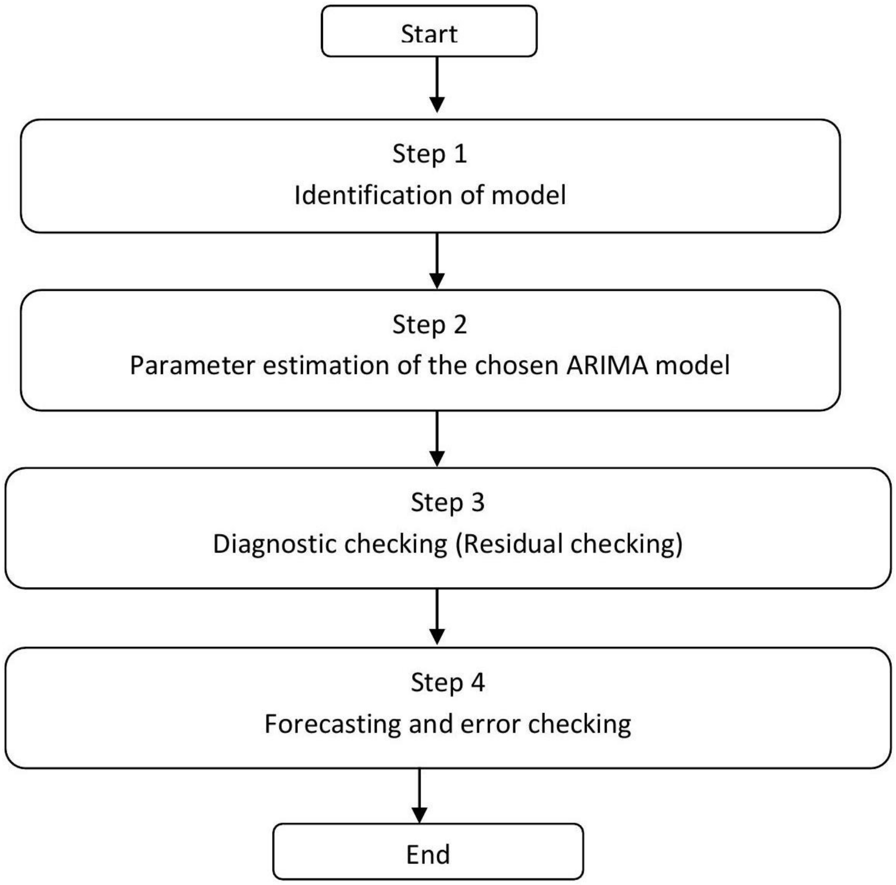
This section discusses the forecasting technique for real-time data of gas booking in online retail. The approach starts with data selection and proceeds to forecasting using the autoregressive integrated moving average (ARIMA) method with parameters $(1, 1, 1)$.

3. Data selection process

This study selected monthly data for gas booking in online rates from January 2009 to December 2011. The statistics were gathered from https://www.kaggle.com.

4. Forecasting procedure

This study forecasted the performance of the gas booking rate using the statistical technique depicted in the following flow chart. The forecasting procedure begins with identifying the data model using autoregressive integrated moving average (ARIMA). When constructing an ARIMA model, it is essential to analyze the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The subsequent step in this research involves estimating parameters for the chosen ARIMA model. Diagnostic checks are crucial to validate the model. The residual is the difference between observed and estimated values of the quantity of interest (sample mean). The residuals should exhibit no correlation, with a mean of zero and constant variance. The final steps include performing forecasting and error checking.



5. Mathematical derivation of ARIMA model

This section discusses the mathematical derivation of autoregressive integrated moving average (ARIMA). ARIMA combines the autoregressive (AR) and moving average (MA) methods, integrating data from a differencing procedure. Differencing ensures that the data used in this research exhibit stationary characteristics. Therefore, the combination of these methods is known as autoregressive integrated moving average.

Firstly, the article explains the development of the autoregressive (AR) approach. An autoregressive model represents a type of random process that characterizes time-varying phenomena in time series data.

The autoregressive model states that the output variable depends linearly on its own past values and a stochastic term, thus taking the form of a stochastic difference equation.

The term AR(p) denotes an autoregressive model of order p. The AR(p) model is defined as follows:

$$X_t = c + \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \varepsilon_t$$

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

Where $\varphi_i, \ldots \varphi_p$ the parameter of the model is, c is constant, and ε_t is white noise.

The notation MA (q) refers to the moving average model of order q :

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i X_{t-i}$$

Where μ is the mean of the series, $\theta_i, \ldots \theta_p$ parameters of the model, and $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are white noise error terms. The value of q is called the order of the MA model.

From the above equation can be written as follws,

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

Where L is Log operator,

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) X_t = \left(1 - \sum_{i=1}^{p-d} \alpha_i L^i\right) (1 - L)^d$$

An ARIMA (p, d, q) process expresses this polynomial factorization property with $p = p' - d$, is given below,

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d=\left(1+\sum_{i=1}^{p}\theta_i L^i\right)\varepsilon_t$$

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\delta+\left(1+\sum_{i=1}^{p}\theta_i L^i\right)\varepsilon_t$$

Which is defined for ARIMA (p, d, q).

4. Results and discussion

This article utilizes real data to forecast the demand for gas booking in an online retail context. Estimating the model's coefficients involves using the ARIMA procedure in the Python programming language within the time series module. This procedure estimates the coefficients of predefined models by specifying the parameters p, q, and d, utilizing a rapid maximum likelihood estimation algorithm.

Summary Statistics of Imports of data

	Quantity	Prize	Costumer Id
count	1.067371e + 06	1.067371e + 06	824364.000000
mean	9.938898e + 00	4.649388e + 00	15324.638504
std	1.727058e + 02	1.235531e + 02	1697.464450
min	−8.099500e + 04	−5.359436e + 04	12346.000000
25%	1.000000e + 00	1.250000e + 00	13975.000000
50%	3.000000e + 00	2.100000e + 00	15255.000000
75%	1.000000e + 01	4.150000e + 00	16797.000000
max	8.099500e + 04	3.897000e + 04	18287.000000

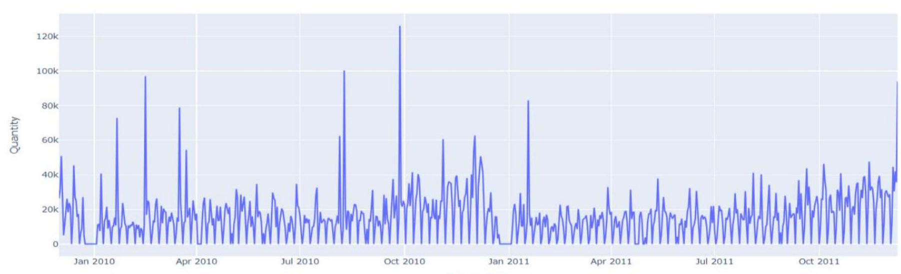


Figure 1: Evolution of the final product's sale of gas booking in online retail environment

The technique includes adding new time series that indicate the modified or forecasted values by the model, along with residuals (adjustment errors) and 95% confidence intervals for adjustments. The selection of the best model aims for simplicity while minimizing criteria such as AIC (Akaike Information Criterion), SBC (Schwarz Bayesian Criterion), variance, and maximum likelihood. The chosen model in this case is ARIMA (1, 1, 1).

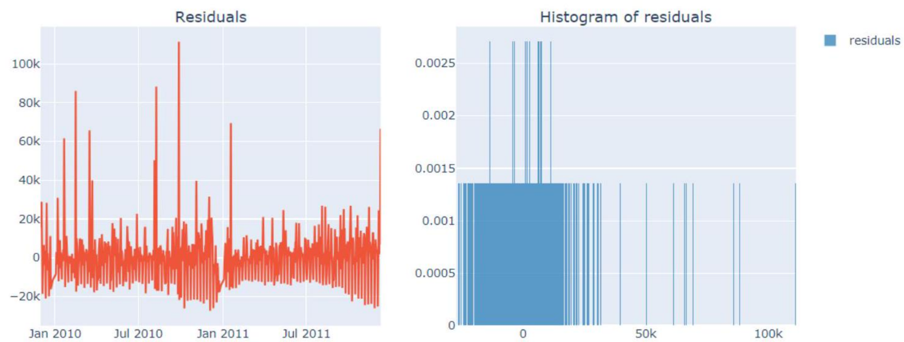


Figure 2:Residuals and Histogram of residuals of gas booking in an online retail environment

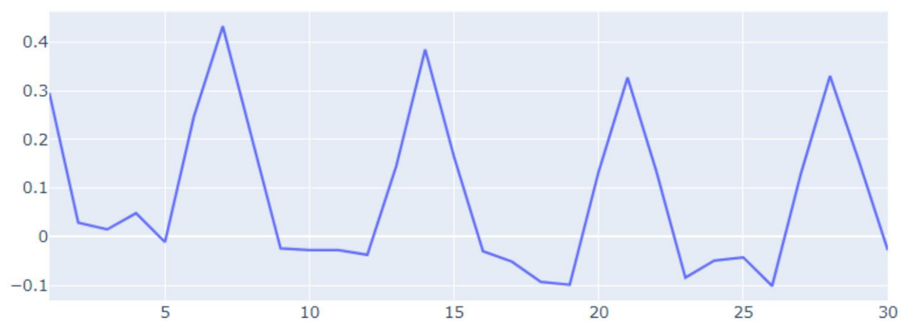


Figure 3:Representation of Stationary Time Series of gas booking in online retail environment

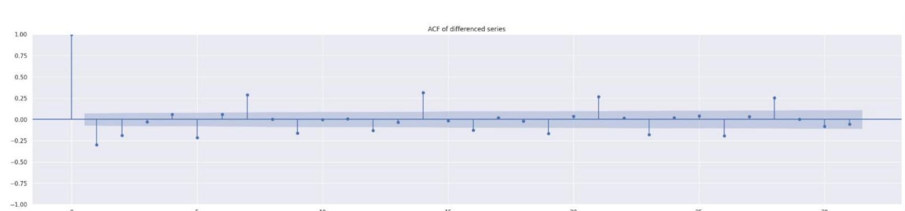


Figure 4: Autocorrelation Function of gas booking in online retail environment

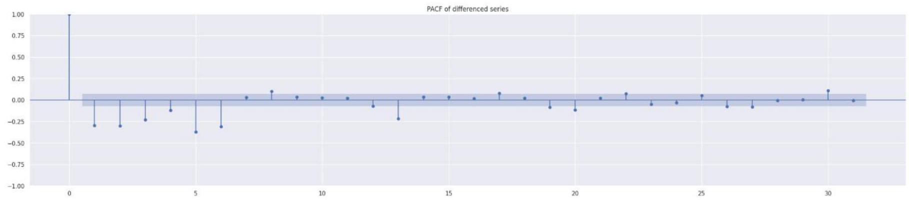


Figure 5: PartialAutocorrelation Function of gas booking in online retail environment

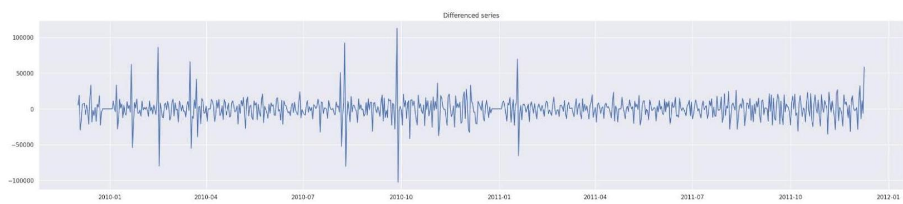


Figure 6:First Difference Autocorrelation Function of gas booking in online retail environment

Accuracy The accuracy of the ARIMA $(1, 1, 1)$ model was evaluated by comparing experimental and simulated petrol booking data in an online retail environment over the same period. Figures 1 and 2 depict this comparison, demonstrating that the selected model exhibits high accuracy and effectively captures the dynamic behavior of petrol booking in the online retail context. Therefore, this model can be utilized for analyzing and forecasting demand in online booking scenarios.

From Figure 6, it is observed that the model validation shows predicted demand fluctuating around the fitted values. Additionally, the predicted demand remains within the upper and lower tolerance limits. Although some variability in error is noticeable, it generally falls within the tolerance interval. To further minimize this error, we propose a new approach for future work.

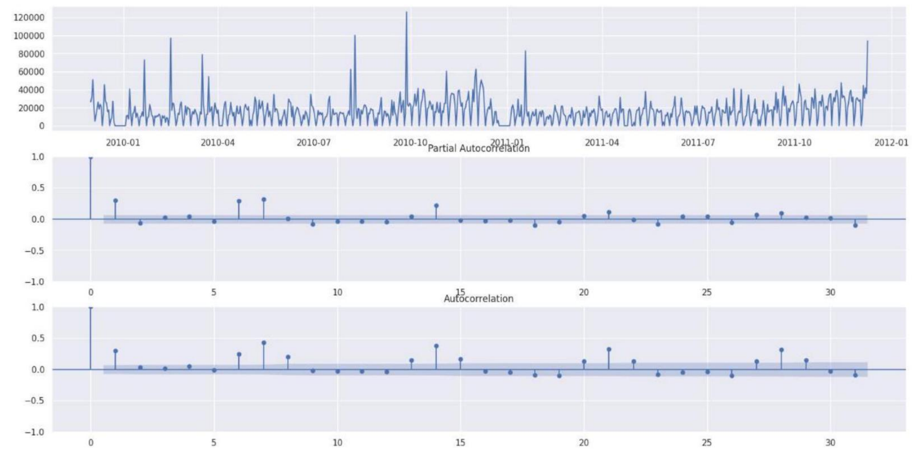


Figure 7:Predicted First Difference Autocorrelation Function, Partial Autocorrelation Function and Autocorrelation Function of petrol booking in online retail environment

After identifying the best demand model for our scenario, the next step is forecasting. Using Python programming, we predict patterns and provide forecasts based on our ARIMA (1, 1,1) model. Table 1 and Figure 8 illustrate the forecasted outcomes for gas booking in an online retail environment over the next six months. These results demonstrate that the selected model is
effective for modeling and forecasting future demand in this context. It's crucial to continuously update historical data with fresh inputs to refine the model and improve forecasting accuracy.

The forecasting models have streamlined the gas booking process in the online retail environment. By accurately predicting demand, the model enables us to optimize online booking strategies and avoid significant cost losses. This capability enhances decision-making regarding fuel supply and daily online booking operations, thereby mitigating potential losses across the entire booking process.

Table 1: Predicted price from gas booking in an online retail environment

```
                           SARIMAX Results
==============================================================================
Dep. Variable:                  value   No. Observations:                  739
Model:                 ARIMA(1, 1, 1)   Log Likelihood               -8036.590
Date:                Mon, 22 Jan 2024   AIC                          16079.181
Time:                        06:54:37   BIC                          16092.993
Sample:                    12-01-2009   HQIC                         16084.507
                         - 12-09-2011
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2353      0.048      4.929      0.000       0.142       0.329
ma.L1         -0.9627      0.014    -71.274      0.000      -0.989      -0.936
sigma2      1.981e+08   8.22e-12   2.41e+19      0.000    1.98e+08    1.98e+08
===================================================================================
Ljung-Box (L1) (Q):                   0.54   Jarque-Bera (JB):              8305.61
Prob(Q):                              0.46   Prob(JB):                         0.00
Heteroskedasticity (H):               0.85   Skew:                             2.50
Prob(H) (two-sided):                  0.19   Kurtosis:                        18.66
===================================================================================
```

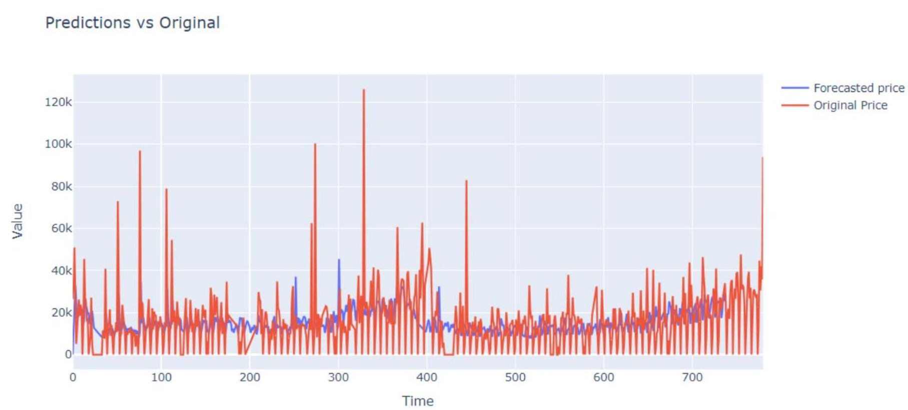


Figure 8: Forecasted price and Original price of gas booking in an online retail environment

5. Conclusion

Demand forecasting plays a crucial role in supply chain management, integrating with various business functions to become a pivotal planning process for the future. In this context, we developed an ARIMA model to forecast demand for gas booking in an online retail environment using the Box-Jenkins time series approach. Historical demand data were utilized to develop multiple models, and the most suitable model was selected based on four performance criteria: SBC, AIC, standard error, and maximum likelihood. The ARIMA (1, 1, 1) model emerged as the optimal choice, as it minimizes these criteria effectively. The results obtained demonstrate that this ARIMA model is capable of accurately modeling and forecasting future demand in this online retail setting. These findings will provide reliable guidance to managers in the food manufacturing industry, aiding them in making informed decisions regarding production and supply chain management.

In the future, our plan includes developing new forecasting models that integrate both qualitative and quantitative methodologies to enhance accuracy. We aim to explore neural network techniques and compare them with ARIMA to assess the neural network's efficacy in online booking forecasting. Additionally, we intend to consistently explore the potential of combining ARIMA with radial basis function (RBF) models to achieve consistently high forecast
accuracy. These efforts will advance our capability to predict and manage demand effectively in the online booking environment.

Appendix

```
[ ] import numpy as np
    import pandas as pd
    import plotly.graph_objects as go
    import plotly.express as px
    from plotly.subplots import make_subplots
    import matplotlib.pyplot as plt
    import seaborn as sns
    import sys
    sns.set()
```

```
[ ] path = 'on.csv'

    df = pd.read_csv(path)

    df.head()
```

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

```
[ ] cancellation_dataset = df.loc[df['Invoice'].str.contains("C", regex=False, na=False)]
    display(cancellation_dataset.sample(15))
```

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
236746	C512300	79323W	WHITE CHERRY LIGHTS	-7	2010-06-14 15:30:00	6.75	14105.0	United Kingdom
527445	C536548	22580	ADVENT CALENDAR GINGHAM SACK	-4	2010-12-01 14:33:00	5.95	12472.0	Germany
132727	C501934	22055	MINI CAKE STAND HANGING STRAWBERY	-5	2010-03-22 11:42:00	1.45	15523.0	United Kingdom
363517	C524575	D	Discount	-1	2010-09-29 16:02:00	28.52	13408.0	United Kingdom
180629	C506492	22168	ORGANISER WOOD ANTIQUE WHITE	-1	2010-04-30 10:52:00	8.50	15268.0	United Kingdom
431630	C530639	22534	MAGIC DRAWING SLATE SPACEBOY	-24	2010-11-03 16:22:00	0.42	16147.0	United Kingdom
182923	C506734	21474	SWEETHEART CREAM STEEL FOLDIN BENCH	-1	2010-05-04 10:17:00	19.95	14867.0	United Kingdom
358589	C524154	22489	PACK OF 12 TRADITIONAL CRAYONS	-144	2010-09-27 15:58:00	0.36	14911.0	EIRE
613673	C543789	M	Manual	-1	2011-02-11 17:10:00	190.80	17450.0	United Kingdom
220956	C510831	84755	COLOUR GLASS T-LIGHT HOLDER HANGING	-3	2010-06-03 20:10:00	0.65	15785.0	United Kingdom
184750	C506978	20728	LUNCH BAG CARS BLUE	-1	2010-05-05 13:07:00	1.65	14680.0	United Kingdom
829221	C563554	22467	GUMBALL COAT RACK	-1	2011-08-17 13:16:00	2.55	16755.0	United Kingdom
905755	C569743	85099B	JUMBO BAG RED RETROSPOT	-140	2011-10-06 10:57:00	1.79	15769.0	United Kingdom
60777	C494801	21232	STRAWBERRY CERAMIC TRINKET BOX	-3	2010-01-18 14:58:00	1.25	15005.0	United Kingdom
788605	C559939	22847	BREAD BIN DINER STYLE IVORY	-1	2011-07-14 10:19:00	16.95	14426.0	United Kingdom

```
[ ] from statsmodels.tsa.stattools import adfuller
    from numpy import log

    result = adfuller(input_df.value.dropna(inplace=False))

    print(f"ADF Statistic: %f" % result[0])
    print(f"p-value: %f" % result[1])
```

```
ADF Statistic: -3.157783
p-value: 0.022562
```

```
[ ] def autocorrelation(array, lag):
        num = 0
        den = 0
        x_bar = np.mean(array)
        array = array[::-1]
        for i in range(len(array)-lag-1):
            num += np.sum((array[i] - x_bar) * (array[i+lag] - x_bar))
            den += np.sum((array[i] - x_bar)**2)
        return num/den
```

```
[ ] go.Figure(data=go.Scatter(x=[i for i in range(1,31)], y=[autocorrelation(input_df.value, i) for i in range(1,31)]), layout=go.Layout(title='Autocorrelation of origina
```

```
[ ] fig, ax = plt.subplots(3,1,figsize=(30,20))
    ax[0].plot(input_df.value.diff())
    plot_acf(input_df.value.diff().dropna(), ax = ax[2], lags = 31)
    plot_pacf(input_df.value.diff().dropna(), ax = ax[1], lags = 31, method='ywmle')
    ax[0].set_title('Differenced series')
    ax[1].set_title('PACF of differenced series')
    ax[2].set_title('ACF of differenced series')
    plt.show()
```

```
[ ] fig, ax = plt.subplots(3,1,figsize=(30,20))
    ax[0].plot(input_df.value.diff().diff())
    plot_acf(input_df.value.diff().diff().dropna(), ax = ax[2], lags = 31)
    plot_pacf(input_df.value.diff().diff().dropna(), ax = ax[1], lags = 31, method='ywmle')
    ax[0].set_title('2nd order Differenced series')
    ax[1].set_title('PACF of 2nd order differenced series')
    ax[2].set_title('ACF of 2nd order differenced series')
    plt.show()
```

```
[ ] import plotly.figure_factory as ff
    df_resid = pd.DataFrame(model_fit.resid, columns=['residuals'])

    fig1 = ff.create_distplot([df_resid.residuals], ['residuals'], ['KDE Plot'], histnorm='probability density')
    fig2 = go.Figure(go.Scatter(x=df_resid.index, y=df_resid.residuals, name='Residuals', showlegend=False))

    fig = make_subplots(rows=1, cols=2, subplot_titles=('Residuals', 'Histogram of residuals'))
    fig.add_trace(fig1['data'][0], row=1, col=2)
    fig.add_trace(fig2['data'][0], row=1, col=1)

    fig.update_layout(height=500, width=1000)
    fig.show()
```

```
[ ] df_pred = pd.DataFrame(model_fit.predict(dynamic=False).values, columns=['predictions'])
    np.array(df_pred.index)

    fig= go.Figure()
    fig.add_trace(go.Scatter(x=df_pred.index, y=df_pred.predictions, name='Forecasted price'))
    fig.add_trace(go.Scatter(x=df.index, y=input_df.value, name='Original Price'))

    fig.update_layout(
        height=500,
        width=1000,
        title='Predictions vs Original',
        xaxis_title='Time',
        yaxis_title='Value')
    fig.show()
```

References

[1] J. D. Wisner, K. C. Tan, and K. Leong, Principles of supply chain management: A balanced approach. South-Western, Cengage Learning, 2021.

[2] D. Lu, Fundamentals of Supply Chain Management, Bookboon, 2011.

[3] S. Shen and Y. Shen, ARIMA model in the application of Shanghai and Shenzhen stock index, Applied Mathematics, 7(3)2016, 171-176.

[4] M. Matsumoto and A. Ikeda, Examination of demand forecasting by time series analysis for auto parts remanufacturing. Journal of Remanufacturing, 5(2015), 1-20.

[5] A. A. Kurawarwala and H. Matsuo, Product Growth Models for Medium Term Forecasting of Short Life Cycle Products. IC2 Institute, 1992.

[6] Willemain, T. R., Smart, C. N. and H. F. Schwarz, A new approach to forecasting intermittent demand for service parts inventories, International Journal of forecasting, 20(3)(2004), 375-387.

[7] V. Kulshreshtha and N. K. Garg, Predicting the new cases of coronavirus [COVID-19] in India by using time series analysis as machine learning model in Python. Journal of The Institution of Engineers (India): Series B, 102(6)(2021), 1303-1309.

[8] E. S. Karakoyun and A. O. Cibikdiken, Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting, In The 13th multidisciplinary academic conference in Prague, May 2018, 171-180.
[9] I. M. Wirawan, T. Widiyaningtyas, and M. M. Hasan, Short term prediction on bitcoin price using ARIMA method, International Seminar on Application for Technology of Information and Communication (iSemantic), September 2019, 260-265.

B. Mohamed Harif
Assistant Professor, PG and Research Department of Mathematics,, Rajah Serfoji Government College (Autonomous) Thanjavur-05, Tamilnadu, India.
Email: bmharif@rsgc.ac.in

M. Karthikeyan
Research Scholar, PG and Research Department of Mathematics,, Rajah Serfoji Government College (Autonomous) Thanjavur-05, Tamilnadu, India.
Email: karthims8127@gmail.com

M. Karthikeyan
Assistant Professor, PG and Research Department of Mathematics,, Annai Vailankanni Arts and Science College Thanjavur-07, Tamilnadu, India.
Email: rinomathz502@yahoo.com