A Performance-Based Comparison of Time Series Models for Champagne Sales Prediction

[1*]Harpinder Kaur [2]Atendra Singh Yadav

[1*]Research Scholar, Department of Mathematics, Guru Kashi University, Talwandi Sabo, Punjab, India.

[2]Department of Mathematics, Guru Kashi University, Talwandi Sabo, Punjab, India.

ABSTRACT

Time Series Analysis (TSA) is a critical analytical tool for analyzing data points gathered or captured at predetermined time intervals. It has proven vital in a variety of sectors, particularly financial services, environmental studies, and medicine. It emphasizes the need to handle missing values, smoothing data, and converting non-stationary data to stationary formats. The conducted research dives into the complicated procedures and strategies used in TSA, covering an extensive variety of models including Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), and Seasonal Autoregressive Integrated Moving Average (SARIMA). Enterprises may improve decision-making, streamline procedures, and predict potential patterns by employing strong TSA. The research paper worked on a case study highlighting SARIMA's superior performance in circumstances where seasonality is a substantial factor, making it a preferred choice over ARIMA for TS forecasting involving seasonal data. The study investigates and compares the efficiency of the ARIMA and SARIMA approaches for TSA to evaluate the model's effectiveness using performance metrics like Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). SARIMA outperforms ARIMA in standings of all performance metrics because of its ability to explicitly model seasonality in TS data.

Keywords: ARIMA, ARIMAX, MAE, RMSE, TSA.

I. INTRODUCTION

TSA is an analytical technique for analyzing data points gathered or documented over time at predetermined intervals. This sort of data frequently displays patterns like trends, Seasonal (SS) consequences, or cyclic conduct, making TSA critical for comprehending previous behaviors and determining potential values in the future. Unlike other types of data analysis, TSA takes into account the sequence and structure of time, allowing for a more in-depth investigation of the time-dependent relationships and variations in the data (Ahmad et al., 2001). The value of TSA rests in its capacity to find significant information from temporal data. Individuals and companies can make informed choices by discovering connections and trends over time (Ullrich, 2021; A & A, 2023). Businesses, for example, employ TSA to forecast sales, regulate inventories, and allocate resources more effectively (Al-Douri et al., 2018; Al-turaiki et al., 2021). Similarly, meteorologists use it to predict weather patterns, while economists use it to analyze market movements and financial cycles. TSA's vast utility makes it valuable in a variety of disciplines (Alharbi & Csala, 2022; Andipara, 2022; Sediakgotla et al., 2019). TSA assists in forecasting disease outbreaks in healthcare and forecasting the demand and supply of energy, and it aids in the performance of the systems in real-time applications (Arumugam & Natarajan, 2023; Bora & Hazarika, 2023; Bhandari et al., 2017). With the increase in time-stamped data, TSA plays an even more significant role in making data-driven decisions and forecasting future fluctuations in an ever more intricate world (Case et al., 2019; Patil, 2021; Dahiya et al., 2024).

TSA offers an extensive framework to comprehend temporal dynamics, improve predictive features, and make data-driven choices in a variety of real-world situations. Being able to address difficulties in constantly changing situations demonstrates its importance in research and practice (Doreswamy et al., 2020; Meer & Deborah, 2024). TSA aids in finding long-term trends and cycles in data, helping companies comprehend historical habits and forecast future movements. One of the most important uses of TSA is projecting future values from past measurements. The TSA can identify abnormalities or variations from expected patterns (Permatasari et al., 2018; Deretić et al., 2022). This is highly beneficial for identifying fraud, monitoring industrial equipment, and guaranteeing system dependability (Chodakowska, 2021; Mehmood et al., 2023). Corporations can plan for periodic changes, such as peak sales season, increases in demand for electricity, or outbreaks of illness, by examining SS fluctuations. Cross-correlation methodologies and models such as ARIMAX enable researchers to investigate the impact caused by external influences (exogenous variables) on time-dependent practices, providing beneficial insights for optimization and intervention tactics (Saleem Latteef Mohammed, Khan & Albreem, 2019).

Three critical factors to be considered while forecasting in TSA are mentioned below.

- Seasonality: Seasonality refers to variations in TS data that occur regularly, such as daily, monthly, or annually. It is an important aspect of TSA since it helps uncover anticipated trends that occur due to SS impacts (Mubasher Hassan & Mirza, 2021; Zhang et al., 2022). For example, air conditioner sales rise throughout the summer. Seasonality analytics enable firms to enhance resource allocation (García- Ferrer, 2003). Manufacturers, for example, can arrange production schedules based on predicted SS demand, guaranteeing appropriate inventory while avoiding excess (Imai et al., 2015).
- Trend: In TSA, the trend factor refers to the movement or direction of data across time, which indicates an underpinning growth, decline, or steadiness (Samal et al., 2019; Sediakgotla et al., 2019). It aids in the identification of regular trends that are not instigated by seasonality, thereby offering information regarding the general course of the variable under study. Trends can be linear or nonlinear and are prejudiced by factors like economic growth, technological advancements, or demographic changes (Khan et al., 2020; Sarvaiya & Ramchandani, 2022). Identifying the trend factor is critical for long- term predictions, mounting strategies, and comprehending the larger context of TS data (Jain et al., 2018; Smith et al., 2024).
- Unexpected events: The unexpected events component in TSA refers to outliers created by unusual occurrences that disturb normal data trends, such as natural catastrophes, financial turmoil, or technical breakdowns. These incidents are independent of seasonality, and they can have significant consequences on predictive model accuracy (Somboonsak, 2019; Lee et al., 2021). Identifying and evaluating such events is critical for detecting anomalies, increasing model strength, and designing mitigation methods. Integrating this factor improves the dependability of TSA, allowing for improved risk management and readiness to cope with future uncertainties (Mahsin et al., 2012; Shelatkar et al., 2020).

II. STATE-OF-THE-ART

Ray & Bhattacharyya, 2020 investigated the statistical analysis of food grain in India from 1950 to 2019. TSA parametric regression models such as ARIMA, and ARIMAX were used to represent the pattern of the country's key food grains. Multiple goodness-of-fit metrics, including RMSE, MAPE, MAE, and R-squared values, were developed to identify the best-fitting replicas. Utilizing the best-fitting models, it was determined that the availability of rice (70.05 kg/year), wheat (70.73 kg/year), and total food grain (182.96 kg/year) will decline in 2021 when compared to this year. Pereira da Veiga et al., 2024 examined ARIMA for forecasting five vital TS having an impact on Brazil's healthcare sectors throughout the 2000-2020 economic crisis. The results demonstrate the ARIMA model's (1, 0, 2), (2, 2, 1), (0, 1, 2), (1, 1, 2), and (2, 2, 1) viability in successfully estimating health-related TS, with accuracy above 95%. **Kozitsin & Katser**, 2021 stated that the suggested procedure has been contrasted with two pseudo-online algorithms and the non-learning ARIMA model. The findings reveal that all of the methods were highly accurate, but their computational complexity and system stability varied greatly, favoring the created online ARIMA-based algorithm. The suggested algorithm's capacity to locate irregularities was proven by its third-place finish in the Numenta Anomaly Benchmark competition. **Botelho et al., 2021** Developed the ARIMA model for predicting unreported cases of Hansen disease during the COVID-19 epidemic in Brazil, Tocantins, and Palmas. The research report used an ecologically sound TS analysis of infection signals from 2001 to 2020, which exhausted the ARIMA model in Palmas. Data were collected from NIIS, and population estimates were prepared by the Brazilian Institute of Geography and Statistics. The ARIMA model (4, 0, 3) yields the lowest values for the two evaluated information criteria and the best-fit data, with AIC = 431.30 and BIC = 462.28 at a significance level of 0.05. **Alrweili & Fawzy, 2022** suggested a hybrid ARIMA-ANN model for TS prediction which brings together the ARIMA and ANN. The hybrid ARIMA-ANN model is flexible enough to handle two types of TS: linear and nonlinear. The Crude Oil (petroleum) Monthly Price - Saudi Riyal per Barrel TS data was gathered between July 2001 and May 2021, encompassing 239 observations. The first 215 observations are used as a training series, and the last 24 observations are used as a testing series. The accuracy metrics (MSE, MAPE, and MAE) of the hybrid ARIMA- ANN approach have been juxtaposed with those of the ARIMA and ANN methods. The results reveal that the hybrid ARIMA-ANN approach significantly improves (MSE, MAPE, and MAE) values. **Singh & Gupta, 2022** claimed that pandemic-relevant function frameworks such as Gaussian, gamma, and logistic are special examples of the GSIR model. The GSIR approach addresses time-varying factors that arise throughout time, and enclosed expressions are provided for all system waves. Although the GSIR paradigm can be applied to modeling any pandemic, it is most suited for undertaking a data-driven prediction analysis of the COVID-19 pandemic. Using the predicted model, the study is being conducted on COVID-19 data from various nations, including the United States, India, Brazil, and many more throughout the world. Additionally, because the GSIR model is a progressive theory of the SIR model, it produces better results when compared to the latter. **Pokhrel & Adhikari**, **2023** inspected the ARIMA (1,1,7) and ARIMAX (1,1,7) models to augment paddy production projections in Nepal from 1975 to 2023. The ARIMA was first used to anticipate paddy production. To mend precision, the accessibility of agricultural land was added as an EV to the ARIMAX. Unlike the ARIMA model, which anticipated paddy production of 5787.64 metric tons per hectare in 2022, the ARIMAX model forecasted 5681.17 tons per hectare. **Pereira da Veiga et al., 2024** The ARIMA model, known for its statistical congruence with several linear modeling techniques, has proven effective in a variety of sectors. The research investigates the use of the ARIMA model to forecast numerous crucial economic time frames that had a significant impact on Brazil's public and private healthcare sectors during the 2000–2020 financial crisis. The paper presents a complete overview of the ARIMA implementation procedure, emphasizing the need for exact prediction for managers seeking to reduce financial irregularities and avoid a lack of resources. The results demonstrate the ARIMA model's (1, 0, 2), (2, 2, 1), (0, 1, 2), (1, 1, 2), and (2, 2, 1) viability in properly projecting detrimental to health TS, with accuracy above 95% for the economic variables examined. These findings have important practical consequences for healthcare management and decision-makers. The study gives beneficial perspectives for thinking strategically in the healthcare sector. **Djimasbe et al., 2024** This work satisfactorily constructed and deployed a realistic ARIMAX model to anticipate electric consumption at Kotoka International Airport, illustrating the importance of functional and ecological variables in determining energy demand. The ARIMAX model had an R-squared score of 0.899 and an Adjusted R-squared of 0.878, showing great accuracy and dependability. Additionally, the model's effectiveness indicators, such as MAE, RMSE, and MAPE, revealed its capacity to give accurate projections, which is critical for efficient energy utilization at the airport. Considering its successes, this study recognizes some limitations, such as uncontrolled variability caused by minor procedural details, outside factors, and data resolution. The subsequent research should seek to incorporate other variables, such as particular airport events and external economic factors, to strengthen the model's robustness and prediction power. **Hamiane et al.**, 2024 Analysed three strategies for predicting GDP: LSTM, ARIMA, and a hybrid strategy. The study gathered quarterly GDP data from the FRED from 1947 to 2022. In forecasting GDP, LSTM (MSE=0.010, RMSE=0.104, MAE=0.077, R2=0.96) and ARIMA (MSE=0.095, RMSE=0.309, MAE=0.286, R2=0.75), were outperformed by the Hybrid model (MSE=0.0018, RMSE=0.043, MAE=0.028, R2=0.99).

III. MODELS OF TSA

TSA uses a variety of models to evaluate and forecast data over time while capturing trends, seasonality, and abnormalities.

ARIMA (AutoRegressive Integrated Moving Average)

The ARIMA model is a popular statistical technique for evaluating and projecting TS data. It consists of three key components: Autoregression (AR), which uses the relationship between an observation and its previous values; Integration (I), which makes the data stationary by differentiating it; and Moving Average (MA), which models the dependency between an observation and the residual errors from a moving average model applied to observations with lags (Singh & Gupta, 2020; Oyeleye et al., 2022). ARIMA is best suited for TS data with non-periodic trends and patterns, making it useful for short- to medium-term forecasting in fields such as finance, economics, and weather prediction (Nayak & Naik, 2021).

Each of the aforementioned components is explicitly included in the model as a parameter. The parameters are replaced with integer values (ARIMA(p,d,q)) to quickly identify the ARIMA model in use. The parameters for the ARIMA model are defined as follows:

- *p:* The lag order, or p, refers to the number of lag observations in the model.
- *d:* The degree of difference (d) refers to how many times raw observations are compared.
- *q:* Moving average window size (also known as the order of moving average).

ARIMAX (Autoregressive Integrated Moving Average with Explanatory Variable)

The ARIMAX model is a multiple regression model containing one or more AR and/or MA variables. This method is appropriate for determining if data is stationary or non-stationary, multivariate, and contains any type of data pattern, such as level/trend/seasonality/cyclicity. ARIMAX is identical to the ARIMA approach; however, unlike ARIMA, which is exclusively compatible with univariate datasets, ARIMAX is suitable for evaluation with supplemental descriptive variables (multivariate) in categorical and/or numerical format (Olsavszky et al., 2020; Sirisha et al., 2022). This combination makes ARIMAX particularly useful in situations where external influences are significant, such as energy demand forecasting, where weather conditions are essential, or financial modeling, where macroeconomic considerations influence stock prices. ARIMAX provides a robust framework for researching and forecasting time-dependent data influenced by external influences by successfully merging internal and external variables (Todi, 2023; Więcek & Kubek, 2024).

SARIMA (Seasonal Autoregressive Integrated Moving Average)
SARIMA modifies the ARIMA model by compensating for seasonality in TSA. It is beneficial when the data shows recurring or cyclic tendencies over time. The model adds SS components to the ARIMA framework, allowing it to cope with Non-Seasonal (NSS) and SS data as shown in Fig. 1.

SARIMA is represented as:

$$SARIMA(p,d,q)(P,D,Q,s)$$

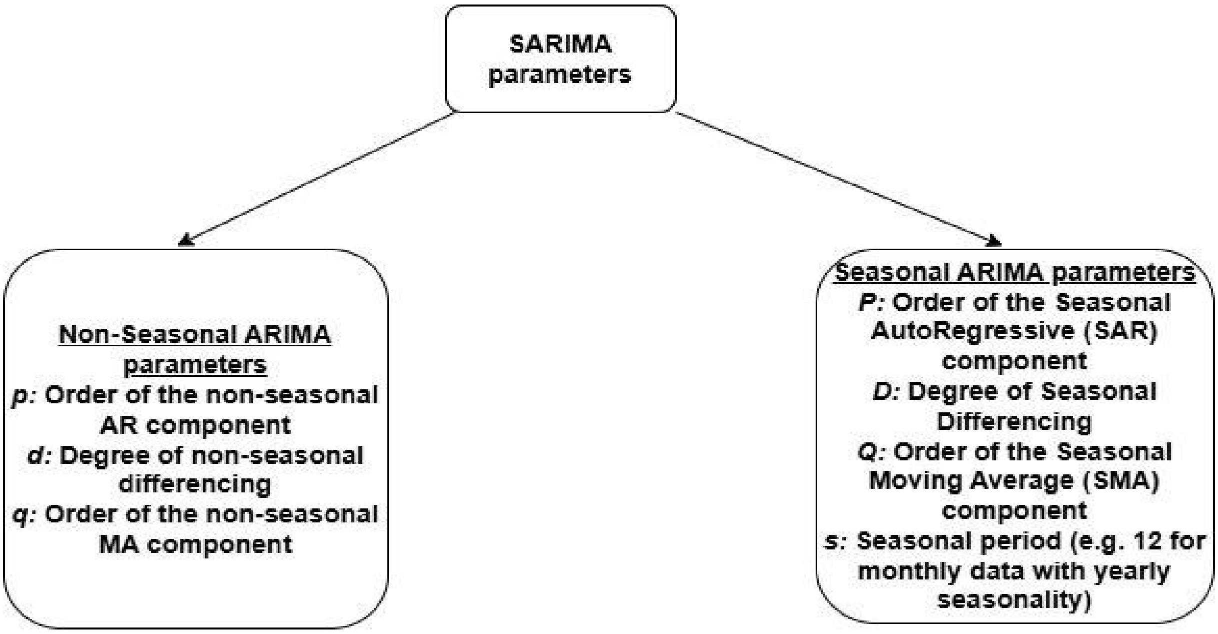

Fig. 1 SARIMA parameters

IV. PERFORMANCE EVALUATION METRICS

Mean Absolute Error (MAE)
In TSA, MAE is a universally used indicator of enactment for assessing ARIMA models. It provides an understandable assessment of the accuracy of the model by measuring the average severity of the errors in projections (MahmoudMohammed H. Alsharif, 2019). MAE is computed as an average of the absolute differences between the actual and predicted values; it is contingent on the scale, therefore being effortless to comprehend in light of the context of the dataset; in contrast to squared error metrics, MAE regulates all errors fairly, without substantially penalizing larger variations, which makes it a robust and metric for appraising ARIMA performance in applications where balanced error appraisal is essential.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|d_t - \hat{d}_t|$$

where

n: refers to the number of observations

$|d_t - \hat{d}_t|$: refers to the difference between actual observed values and predicted values from the ARIMA model

Mean Squared Error (MSE)

MSE computes the average squared difference between predicted and observed values, giving an empirical evaluation of the model's competence. A lower MSE recommends that the model's projections be more accurate. MSE is especially valuable because it penalizes greater inaccuracies more severely, which can aid in identifying prototypes that create infrequent significant discrepancies from observed data (Dama & Sinoquet, 2021). By decreasing MSE, academics and practitioners hope to create TSA models that are dependable and accurate.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(d_t - \hat{d}_t)^2$$

n: refers to the number of observations

$(d_t - \hat{d}_t)^2$: refers to the square of the difference between actual observed values and predicted values from the ARIMA model

Root Mean Squared Error (RMSE)

The RMSE is a prevalent valuation statistic for TSA methods that use ARIMA to examine their predictive ability. RMSE is derived by taking the square root of the average squared discrepancies between the observed and predicted values. Unlike MSE, RMSE is presented in identical units to the data in question, which makes it easier to grasp in practice (Masena, Mahlangu, et al., 2024). It is frequently used in assessments of models, hyperparameter adjustment, and evaluation to guarantee that the ARIMA configuration selected produces precise and trustworthy recommendations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(d_t - \hat{d}_t)^2}$$

n: refers to the number of observations

$(d_t - \hat{d}_t)^2$: refers to the square of the difference between actual observed values and predicted values from the ARIMA model

Mean Absolute Percentage Error (MAPE)

MAPE is derived as the mean of absolute percentage deviations over all data points, resulting in a normalized statistic that is scale-independent. This makes MAPE ideal for assessing the outcome of ARIMA mathematical models spanning diverse datasets or circumstances. A lower MAPE value signifies additional accurate forecasting accuracy, while a MAPE of 0 represents the optimal model (Pramanik et al., 2022). One of the most noticeable benefits of MAPE is its comprehension, which explicitly reveals the average percentage inaccuracy of its forecasts. However, it is cognizant of lower actual numbers, which may inflate the inaccuracy. MAPE is frequently used in aggregation with other measures such as RMSE and MSE to thoroughly judge the performance of ARIMA models in capturing time-dependent patterns and trends.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\mathrm{d}_t - \hat{d}_t}{\mathrm{d}_t}\right| * 100$$

V. CONTRIBUTION AND IMPLEMENTATION

Case Study using ARIMA and SARIMA models

The case study examines the use of TSA to forecast champagne sales. The dataset covers the monthly sales of champagne from 1964 to 1972. The dataset consists of two columns. The first column refers to the month and contains data relating to months, whereas the second column refers to sales data. The dataset's source is provided below.

https://www.kaggle.com/datasets/anupamshah.perrin-freres-monthly-champagne-

sales?resource=download The steps involved in forecasting the sales data are shown below in Fig. 5.

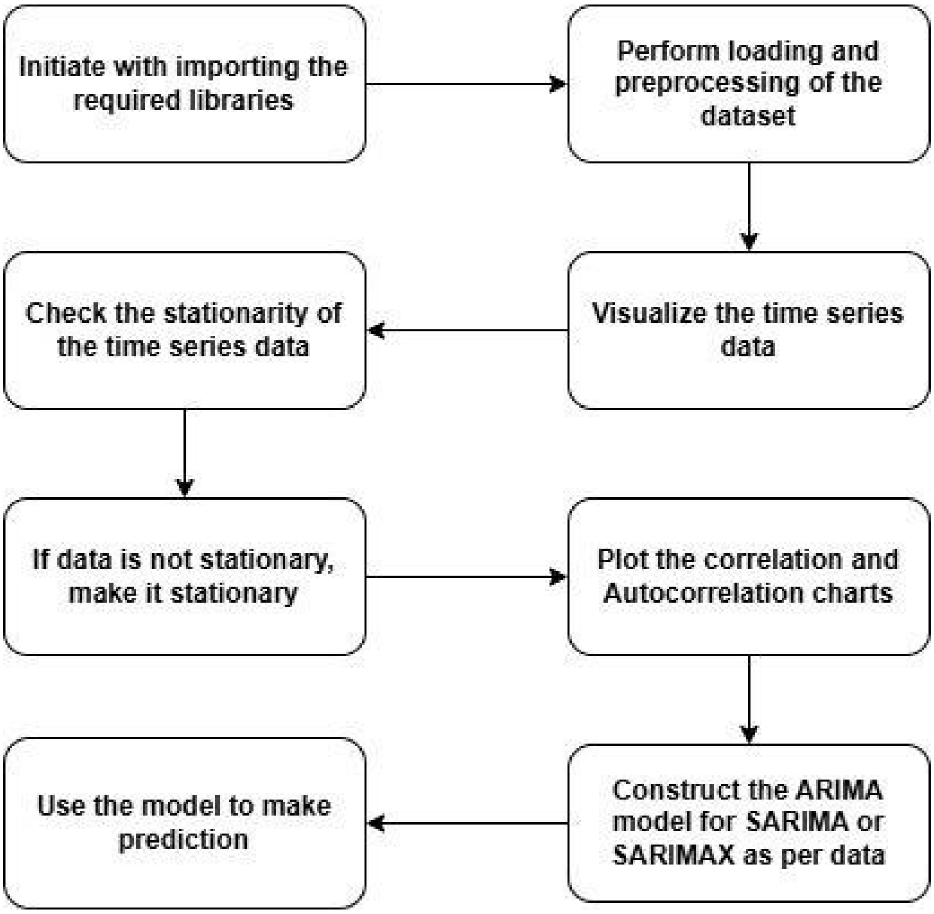

Fig. 5 Procedure for forecasting sales

Decomposition

Decomposition in TSA is an approach for breaking down a TS into its elements to gain insight into its fundamental composition and dynamics. A TS is usually divided into three major components: trend, seasonality, and residuals (noise). The trend aspect shows the data's over time movement or orientation across time, whether it is growing, declining, or stable (Fattah et al., 2018; Dimri et al., 2020). The seasonality factor detects repeated, predictable trends or cycles in the data, such as monthly sales surges or daily temperature swings. The residuals encompass random fluctuations or inexplicable noise that cannot be linked to a trend or seasonality (Goyal & Kundu, 2020).

Augmented Dickey-Fuller (ADF) test

The ADF test is a popular statistical approach in TSA for determining if a series is stationary, which means that its statistical features (mean, variance, and autocorrelation) stay constant across time. Several TS forecasting models, such as ARIMA, rely on stationarity because they presume that the fundamental structure of data remains constant across time (Katambire et al., 2023). The ADF test modifies the standard Dickey-Fuller test by accounting for higher levels of autocorrelation through lagged changes in the TSA, resulting in more robust results. It compares the null hypothesis *(H0)* that the series has a unit root (non-stationary) with the alternative hypothesis *(H1)* that the series is stationary. If the test statistic is smaller than the critical threshold at a predetermined importance threshold (for example, 0.05), the null hypothesis is invalidated, suggesting stationarity (Park & Yang, 2024). Alternatively, the series is not stationary and might demand modifications such as differencing or degenerating to become stationary. The ADF test is essential for preprocessing in TSA, assuring data appropriateness for investigation and projections (Masena, Shongwe, et al., 2024).

Partial Autocorrelation Function (PACF)

The PACF test is a useful technique in TSA for determining the direct link between a TS observation and its lagged values after eliminating the implications of intermediate lags. Unlike the ACF, which takes into account all preceding lags, the PACF separates the correlation at each lag. This makes it especially valuable for determining the proper order of the Autoregressive *(p)* component in models such as ARIMA. The PACF figure depicts the partial correlation coefficients for various lag values, and a considerable drop-off after a given lag indicates the maximum number of lags that can be used in the autoregressive term (Ensafi et al., 2022).

Differencing

Differencing is an approach in TSA that converts a non-stationary TS into a stationary one by eliminating patterns and seasonality. A stationary TS has a consistent mean, variance, and autocorrelation throughout time, which is required for many models for forecasting such as ARIMA. Differencing calculates the difference between successive observations, resulting in a new series that reflects modifications instead of the original values.

Firstly, load and pre-process the dataset as shown in Fig. 6.

	Month	Perrin Freres monthly champagne sales millions ?64-?72
97	1972-02	3564.0
64	1969-05	4968.0
82	1970-11	9842.0

Fig. 6 Glimpse of Dataset comprising 2 columns

Rename the column name *"Perrin Freres monthly champagne sales millions"* with a perceptive name *"Sales"* as shown in Fig. 7.

	Month	Sales
0	1964-01	2815.0
1	1964-02	2672.0
2	1964-03	2755.0

Fig. 7 Glimpse of the dataset after renaming the second column as *"Sales"*

Check for the NULL rows in the dataset. The *'Month'* column has 1 NULL row and the *'Sales'* column has 2 NULL rows as shown in Fig. 8.

	Month	Sales
103	1972-08	1413.0
104	1972-09	5877.0
105	NaN	NaN
106	Perrin Freres monthly champagne sales millions...	NaN

Fig. 8 Records having NULL values in the dataset

Drop rows contain a NULL value for *'sales'* or *'Month'*.

Convert the *'Month'* column type to a Datetime object as shown in Fig. 9.

	Month	Sales
0	1964-01-01	2815.0
1	1964-02-01	2672.0
2	1964-03-01	2755.0
3	1964-04-01	2721.0

Fig. 9 Conversion of *'Month'* column type to DateTime object

The line graph in Fig. 10 depicts Champagne Sales (in Millions) from 1964 to 1972, including monthly sales trends. The sales data shows significant seasonal trends, with peaks occurring annually, demonstrating higher sales during specific months, most likely due to holiday seasons or celebrations. The overall pattern indicates a gradual growth in sales over time, with substantial surges near the end of each year. Notably, the largest sales were around the year 1970, totaling nearly 13,000 million copies. The repeating pattern of rapid spikes followed by drops suggests demand seasonality.

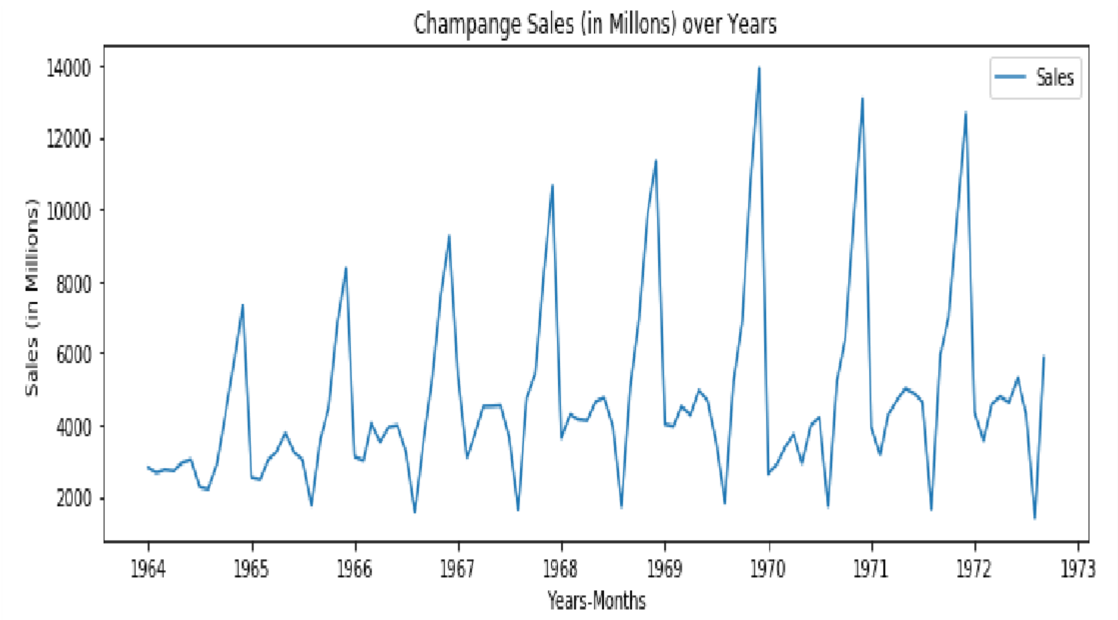

Fig. 10 Champagne sales over the years as per dataset

Check seasonality by plotting for 2 or 3 years of monthly sales. The time range from 1964 to 1966 has been considered as shown in Fig. 11. The line graph in Fig. 11 depicts the Monthly Sales TS from 1964 to 1966, indicating sales changes during these two years. The data demonstrates a seasonal trend, with substantial surges near the end of each year, especially in late 1964 and late 1965, when sales peaked at roughly 8,200 units. A large increasing trend is seen from mid-1964 to the end of the year, followed by a severe decrease at the beginning of 1965. The sales pattern is cyclical, with periods of rise and drop occurring annually, indicating a strong seasonal component in the TS.

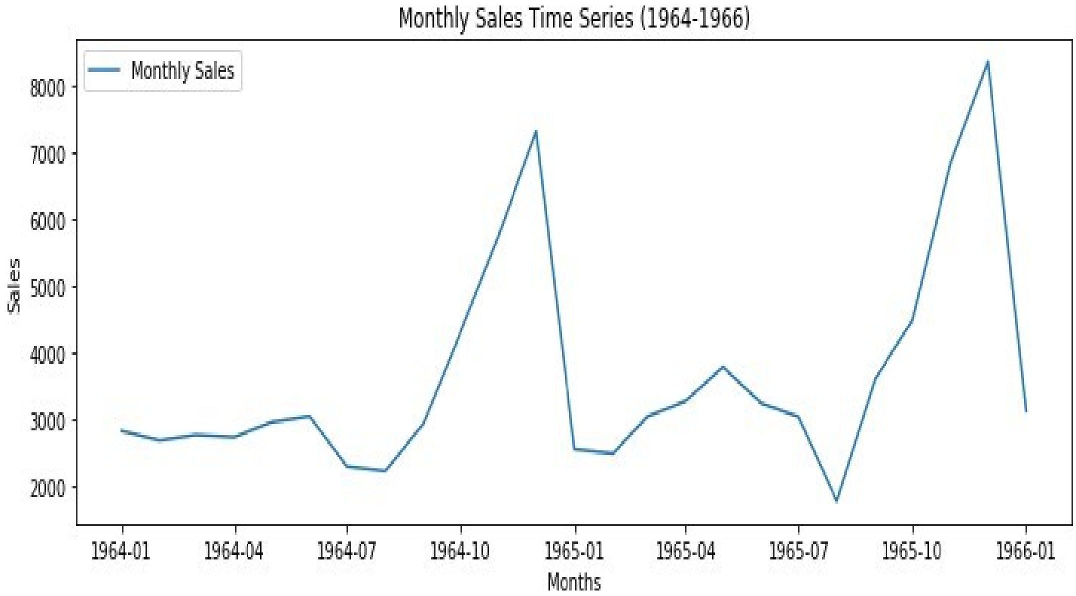

Fig. 11 Monthly Sales TS from 1964 to 1966

Select data again for the desired date range from 1966 to 1968 as shown in Fig. 12. The line graph in Fig. 12 depicts the Monthly Sales TS from 1966 to 1968, displaying changes in sales throughout this period. The figures show a strong seasonal pattern, with substantial surges after each year, especially in late 1966 and late 1967, when sales reached around 10,500 units. The sales trend follows a cyclical pattern of rise and decrease, with large gains near the end of each year and significant drops at the start of the following year. Mid-year sales are often lower, while peak sales are found in the last months of each year, showing a strong seasonal effect, most likely affected by holiday seasons or special events.

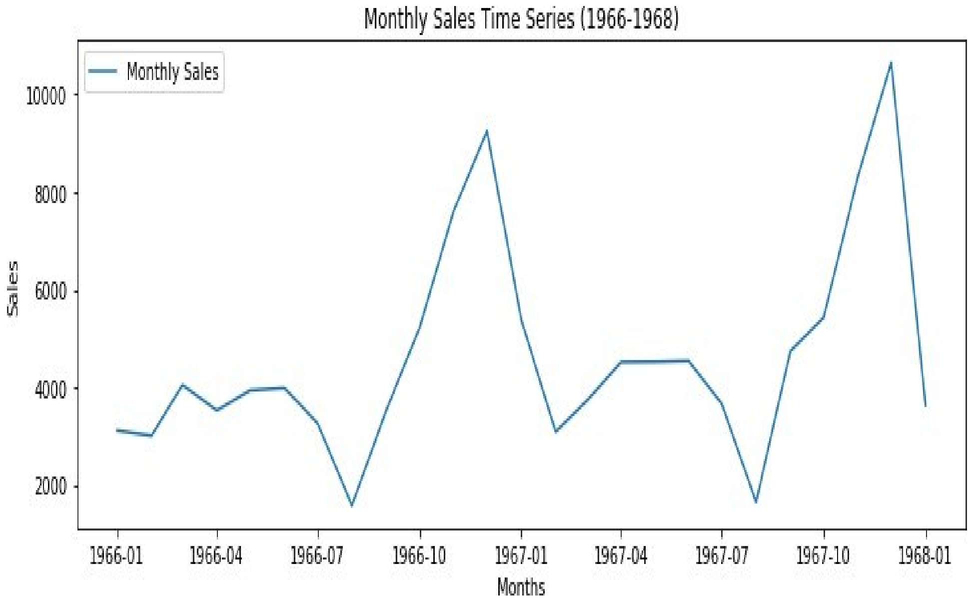

Fig. 12 Monthly Sales TS from 1966 to 1968

Fig. 13 represents a TS decomposition plot, which is used to analyze the underlying patterns in a TS dataset.

- Observed Series (Sales - First Plot): This plot depicts the original TS data for sales over time (about 1964–1972). There are noticeable variations, with peaks appearing at regular intervals, indicating a seasonal pattern.
- Trend Component (Second Plot - Trend): This plot depicts the long-term flow of the data. The sales appear to be increasing over time; however, certain periods show modest dips.
- Seasonal Component (Third Plot - Seasonal): This component detects repeated patterns over a set time. Peaks in the seasonal component confirm that sales fluctuate periodically, most likely on an annual or quarterly basis.
- Residual Component (Fourth Plot - Resid): The residual represents noise or random changes that cannot be explained by the trend or seasonal components. The points are randomly distributed about zero, showing that the model has successfully separated the trend and seasonal trends.

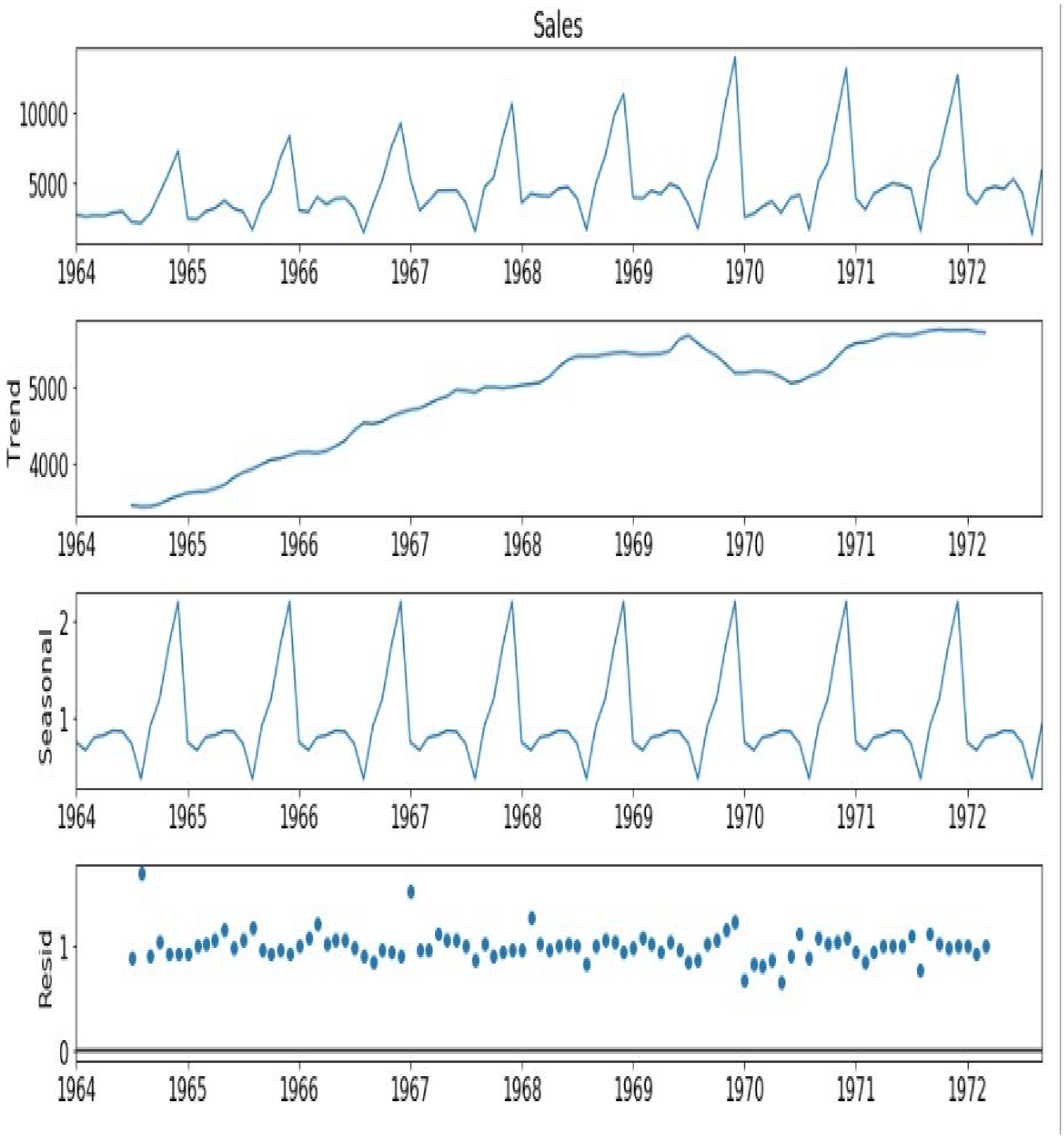

Fig. 13 Components after decomposition

The research conducted an Augmented Dickey-Fuller (ADF) test to check if the data is stationary or not.

Make the TS data stationary applying the Differencing technique as shown in Fig. 14.

Month	Sales
1964-01-01	2815.0
1964-02-01	2672.0
1964-03-01	2755.0
1964-04-01	2721.0
1964-05-01	2946.0
...	...
1972-05-01	4618.0
1972-06-01	5312.0
1972-07-01	4298.0
1972-08-01	1413.0
1972-09-01	5877.0

105 rows × 1 columns

Fig. 14 Data obtained after applying the Differencing technique

Shift the data by 1 time unit as shown in Fig. 15.

```
Month
1964-01-01       NaN
1964-02-01    2815.0
1964-03-01    2672.0
1964-04-01    2755.0
1964-05-01    2721.0
               ...
1972-05-01    4788.0
1972-06-01    4618.0
1972-07-01    5312.0
1972-08-01    4298.0
1972-09-01    1413.0
Name: Sales, Length: 105, dtype: float64
```

Fig. 15 Data after shifting 1-time unit

Add one new column by taking the difference.

Take a *shift(12)* as a year has a 12-month cycle based on SS information.

View the first 20 rows, since 12 rows will have a *"NaN"* value as shown in Fig. 16.

Month	Sales	Sales_first_difference	Seasonal_first_difference
1964-01-01	2815.0	NaN	NaN
1964-02-01	2672.0	-143.0	NaN
1964-03-01	2755.0	83.0	NaN
1964-04-01	2721.0	-34.0	NaN
1964-05-01	2946.0	225.0	NaN
1964-06-01	3036.0	90.0	NaN
1964-07-01	2282.0	-754.0	NaN
1964-08-01	2212.0	-70.0	NaN
1964-09-01	2922.0	710.0	NaN
1964-10-01	4301.0	1379.0	NaN
1964-11-01	5764.0	1463.0	NaN
1964-12-01	7312.0	1548.0	NaN
1965-01-01	2541.0	-4771.0	-274.0
1965-02-01	2475.0	-66.0	-197.0
1965-03-01	3031.0	556.0	276.0
1965-04-01	3266.0	235.0	545.0
1965-05-01	3776.0	510.0	830.0
1965-06-01	3230.0	-546.0	194.0
1965-07-01	3028.0	-202.0	746.0
1965-08-01	1759.0	-1269.0	-453.0

Fig. 16 First 20 rows with 12 rows having *"NaN"* value

Check ADF-Test again on newly shifted sales information. The obtained results are mentioned below.

ADF Test Statistic: -
7.626619157213164 p-value:
2.060579696813685e-11
Lags Used: 0
Number of Observations Used: 92

The strong evidence is obtained against the null hypothesis *(H_0)*. Data has no unit root and is stationary. So, the null hypothesis is rejected. Figure 17 depicts a TS plot of the seasonal first difference of sales data from 1965 to 1973, revealing variations in sales after removing seasonal tendencies. Seasonal differencing is used to stabilize the mean and remove repeating trends, resulting in more stationary data for analysis. The Y-axis depicts sales di sparities in millions, while the X-axis shows time in years and months. The data demonstrates noteworthy volati lity, particularly dramatic spikes in 1968, 1970, and 1971, indicating significant changes or external forces influ encing sales. The modified series appears more stable, making it appropriate for additional statistical modeling, such as ARIMA forecasting.

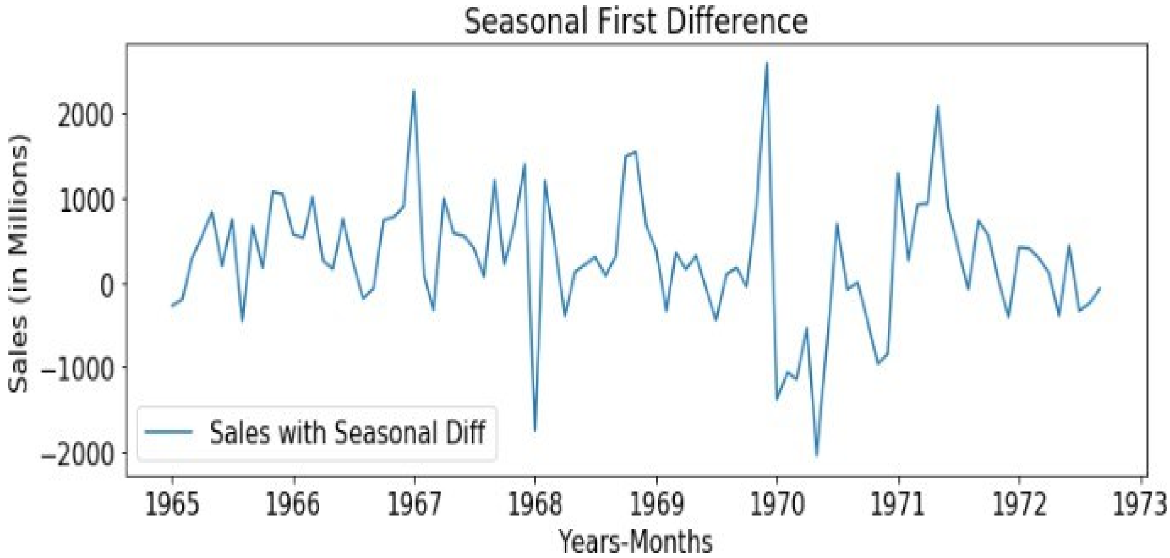

Fig. 17 Graph for Sales with the SS first difference

Fig. 18 displays two diagnostic plots: the ACF on the top and the PACF on the bottom, both used for analyzing TS data. The ACF plot shows how current observations are correlated with past observations (lags), while the PACF plot measures the direct effect of each lag after removing the influence of intermediate lags. Significant spikes outside the blue confidence intervals (usually representing a 95% confidence level) indicate statistically significant correlations. In the ACF plot, there are strong correlations at several lags, suggesting the presence of persistent seasonal patterns. The PACF plot also exhibits notable spikes, which may indicate the appropriate lag order for AR terms in a TS model. These plots are typically used to identify the structure of ARIMA models, where the ACF helps determine the MA component and the PACF suggests the AR component.

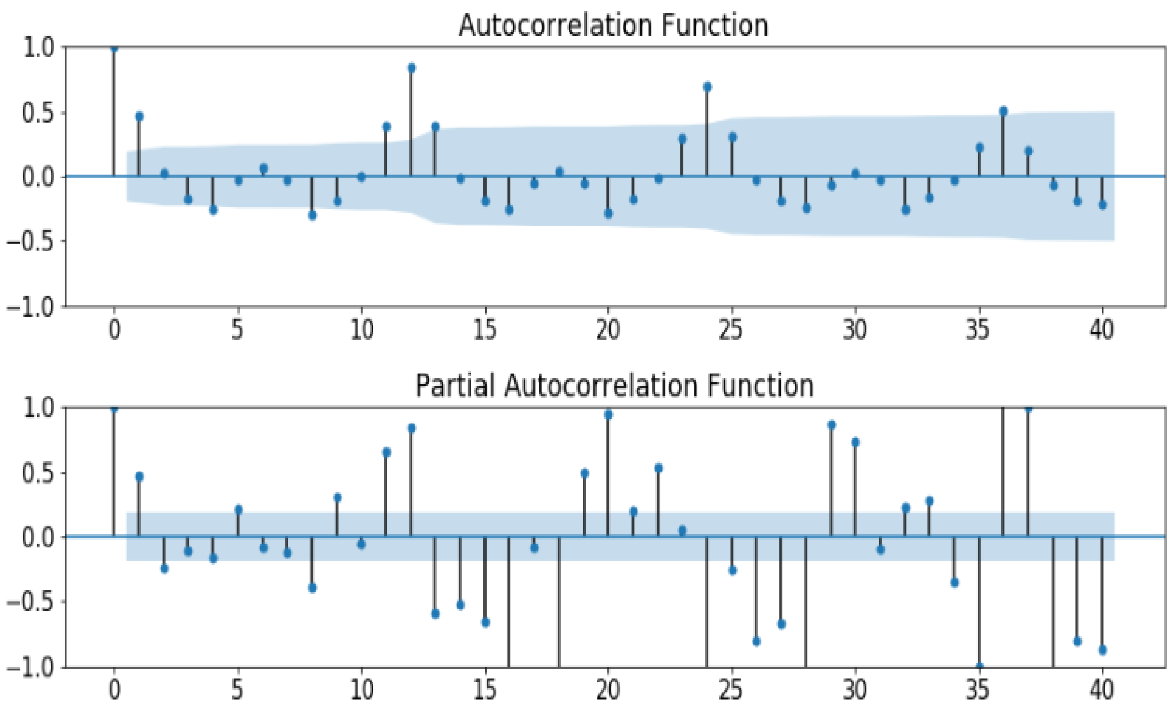

Fig. 18 Graph illustrating ACF and PACF for actual Sales data

Fig. 19 displays two plots: the ACF on the top and the PACF on the bottom. The blue-shaded region represents the confidence intervals; values outside this range are statistically significant. The initial spike at lag 0 is always

1, as it represents a perfect correlation with itself. The gradual decline in ACF and the significant spikes in PACF help determine the order of AR and MA components in TS modeling, such as ARIMA.

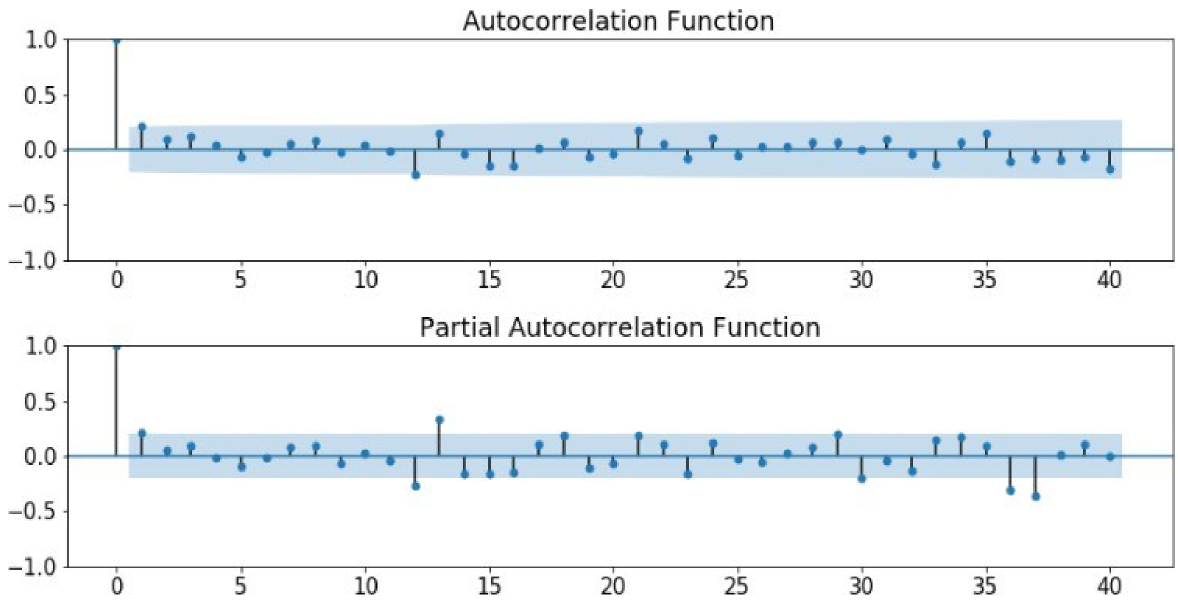

Fig. 19 Graph of ACF and PACF for Seasonal first difference

Fig. 20 exhibits a TS decomposition plot that shows the seasonal first difference of a dataset. The decomposition is divided into four subplots: the first plot represents the original differenced TS, which shows fluctuations over time; the second plot depicts the trend component, which captures long-term upward or downward movements; the third plot highlights the seasonal component, which reveals repeating patterns at regular intervals; and the final plot depicts the residuals, which represent the remaining variations after accounting for trend and seasonality. This decomposition aids in identifying the underlying patterns in the data, which are critical for TSA and analysis.

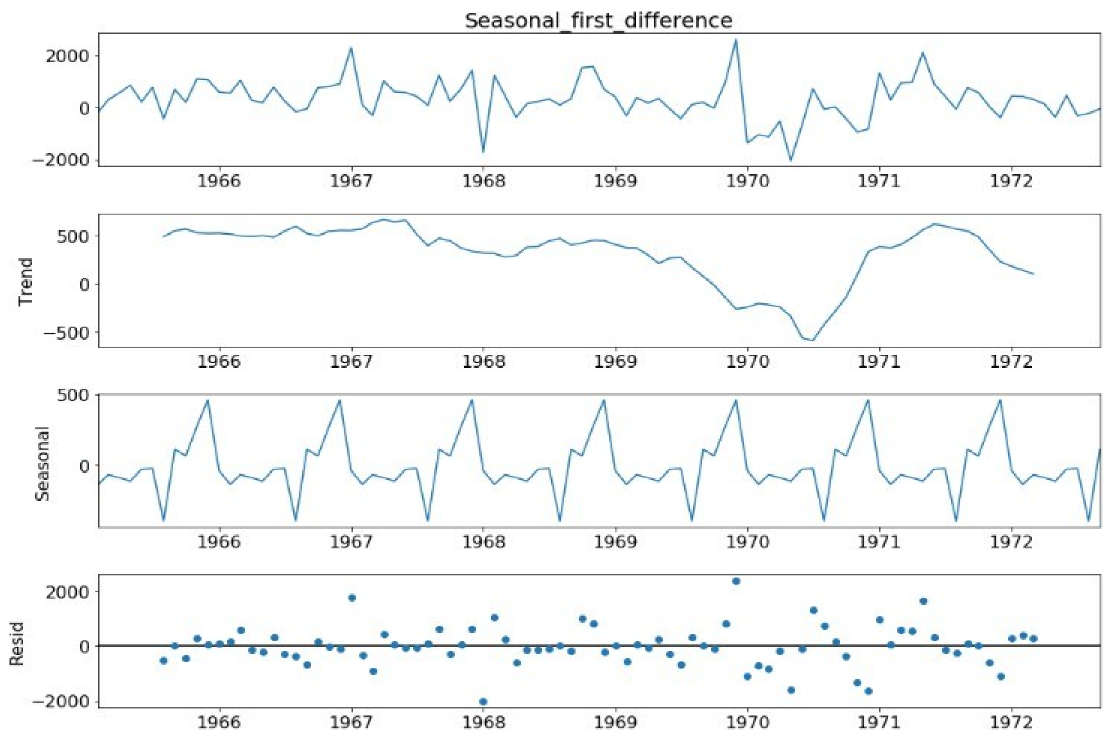

Fig. 20 Decomposition of the TS data

***Fitting ARIMA* -** ARIMA models are a type of linear model that uses historical data to estimate future values. The model is instantiated and fitted on the training dataset. Check the shape and forecast the model. The forecasted model is compared by plotting the actual values together. Fig. 21 depicts a TS plot comparing actual sales data (in blue) with anticipated values (in orange) across time, from 1964 to 1972. The sales statistics show a clear seasonal pattern, with peaks and troughs occurring at regular intervals, indicating periodic swings in demand. The anticipated figures come near the conclusion of the period (about 1971-1972) and roughly reflect the overall pattern of actual sales, with occasional variances. This implies that the forecasting model covers seasonality and trends to some extent, but may contain flaws or residual errors. Such a picture is important for assessing the performance of prediction models in TSA.

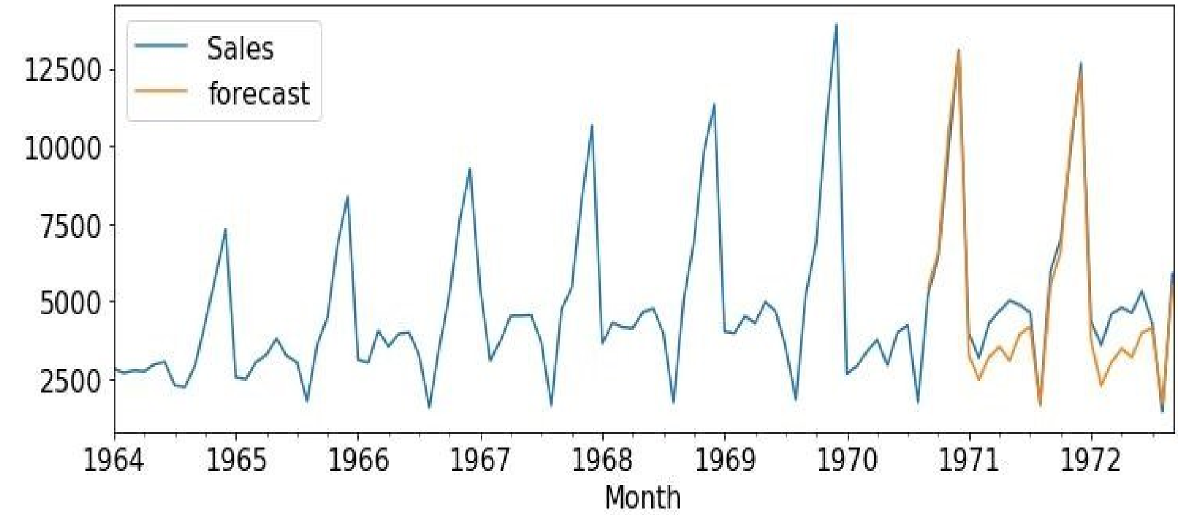

Fig. 21 Graph depicts the actual values and the forecasted values

Fitting SARIMA – Initialize the SARIMA model. Forecast the sales value using SARIMA. The forecasted model is compared by plotting the actual values together. Fig. 22 depicts a TS plot comparing actual sales data (in blue) with a SARIMA forecast (in orange) from 1964 to 1972. The sales data exhibit strong seasonality, with recurring peaks and troughs indicating cyclical trends. The SARIMA forecast appears in the later years (around 1971-1972) and closely follows the actual sales trend, capturing the seasonal fluctuations and overall movement of the data. The alignment between the forecast and actual values suggests that the SARIMA model is effective in modeling both the trend and seasonality, making it a suitable choice for TSA in this context.

Fig. 22 shows that SARIMA captures the seasonality accurately.

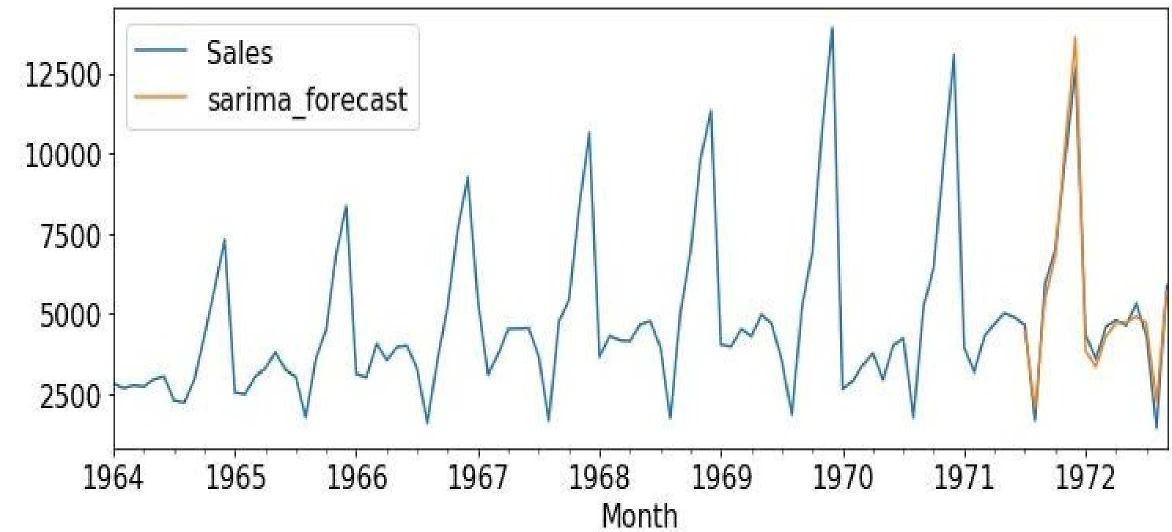

Fig. 22 Graph depicts SARIMA captures the seasonality accurately

The SARIMA model's graph of sales and forecast shows fewer fluctuations, indicating that the approach is more efficient at detecting the patterns and seasonality found in TS data, resulting in smoother and more reliable projections. SARIMA includes seasonal components, allowing it to change estimates based on repeating trends in the data. This lowers the differences between actual and anticipated values. Less fluctuation indicates that SARIMA reduces the impact of inconsistencies in the data, focusing instead on its fundamental trend and seasonal pattern.

The ARIMA graph in Fig. 21 frequently varies considerably from the SARIMA graph in Fig. 22 when contrasted to the actual data because ARIMA does not specifically account for seasonality, making it difficult. ARIMA only considers NSS trends and short-term dependencies, relying on differencing, AR, and MA elements to describe the underlying structure. However, when the data shows strong and consistent SS trends, ARIMA may overfit or underfit some components of the data, leading to greater oscillations and variations from the true value.

In contrast, SARIMA uses particular SS components—P, D, and Q—to directly simulate the periodicity. These SS keywords moderate repeated patterns while efficiently retaining the data's cyclical character. As an outcome, SARIMA produces more dependable and precise estimates, minimizing needless swings and matching with true SS variations in the TS. This distinction highlights SARIMA's usefulness in datasets with strong seasonality, as it incorporates SS and NSS components to create a holistic model.

VI. Calculating performance metrics for ARIMA and SARIMA

Performance indicators in TSA are critical for determining the forecasting model's accuracy and reliability. Metrics like as MAE, MSE, RMSE, and MAPE quantify prediction errors, making it easier to scrutinize models and choose the best-fit strategy for the data. They provide guidance for model adjustment by detecting underperformance and overfitting concerns. Correct metrics guarantee that estimates match real-world requirements, which improves decision-making. The pseudocode for calculating the performance metrics is mentioned below.

```
Load dataset
        data = pd.read_csv('champagne_sales.csv', parse_dates=['Month'], index_col='Month')
Ensure monthly frequency
        data = data.asfreq('MS')
        sales = data['Sales']
Split the dataset for training and testingt
        train_size = int(len(sales) * 0.8)
        train, test = sales[:train_size], sales[train_size:]
Provide specifications for ARIMA
        arima_model = ARIMA(train, order=(1, 1, 1))
        arima_fit = arima_model.fit()
        arima_forecast = arima_fit.forecast(steps=len(test))
Provide specifications for SARIMA
        sarima_model = SARIMAX(train, order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))
        sarima_fit = sarima_model.fit()
        sarima_forecast = sarima_fit.forecast(steps=len(test))
Calculate performance evaluation metrics
        metrics = {}
        for model_name, forecast in [('ARIMA', arima_forecast), ('SARIMA', sarima_forecast)]:
            mae = mean_absolute_error(test, forecast)
            mse = mean_squared_error(test, forecast)
            rmse = np.sqrt(mse)
            mape = np.mean(np.abs((test - forecast) / test)) * 100
            metrics[model_name] = {'MAE': mae, 'MSE': mse, 'RMSE': rmse, 'MAPE': mape}
Obtain the results
        for model, result in metrics.items():
            print(f"{model} Performance Metrics:")
            for metric, value in result.items():
                print(f"{metric}: {value:.4f}")
            print()
```

The calculated values of MAE, MSE, RMSE, and MAPE for ARIMA and SARIMA are shown in Table 2 below.

Table 2. Performance metrics values for ARIMA and SARIMA

Model	MAE	MSE	RMSE	MAPE
ARIMA	2368.7857	9112756.7836	3018.7343	67.0088
SARIMA	1024.0119	1410454.7128	1187.6257	22.0977

The low MAE (1024.0119) for SARIMA compared to 2368.7857 for ARIMA implies that the predictions made by the model are more accurate on average. A lower MAE indicates superior precision, thus rendering it more trustworthy for predicting or decision-making. The low MSE (1410454.7128) for SARIMA compared to 9112756.7836 for ARIMA suggests that the model's predictions have fewer squared deviations from the actual values. MSE increases the influence of bigger errors owing to squaring, making it especially useful for penalizing substantial outliers. The low RMSE (1187.6257) for SARIMA compared to 3018.7343 for ARIMA shows lower average variances between anticipated and actual values. Because RMSE is expressed in the identical unit as the original data, it gives a clear indication of error size. The low MAPE score (22.0977) for SARIMA compared to 67.0088 for ARIMA implies that the model's predictions diverge fewer from actual values in percentage terms.

VII. CONCLUSION

The research conducted shows that, while ARIMA is a strong model for assessing and forecasting TS based entirely on inherent trends, it has limits when external influences substantially impact the objective variable. ARIMAX, on the other hand, overcomes this constraint by adding exogenous variables, resulting in improved forecast accuracy in complicated, real-world settings. Prospective studies can investigate hybrid models that integrate ARIMAX with machine learning approaches to improve predicting capacities in dynamic contexts.

The research demonstrated SARIMA's ability to model short-term trends and recurrent seasonal patterns, resulting in more exact predictions. Performance indicators indicate much lower values for SARIMA than ARIMA. This error reduction demonstrates SARIMA's outstanding forecasting performance, making it a top choice for TS analysis with seasonality.

REFERENCES

A, A. K., & A, S. T. (2023). *Comprehensive Prediction of Stock Prices Using Time Series, Comprehensive Prediction of Stock Prices Using Time Series, Statistical, Machine Learning, and Deep Learning Models Statistical, Machine Learning, and Deep Learning Models Comprehensive Prediction*.

Ahmad, S., Khan, I. H., & Parida, B. P. (2001). Performance of stochastic approaches for forecasting river water quality. *Water Research*, *35*(18), 4261–4266. https://doi.org/10.1016/S0043-1354(01)00167-1

Al-Douri, Y. K., Hamodi, H., & Lundberg, J. (2018). Time series forecasting using a two-level multi-objective genetic algorithm: A case study of maintenance cost data for tunnel fans. *Algorithms*, *11*(8). https://doi.org/10.3390/a11080123

Alharbi, F. R., & Csala, D. (2022). A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. *Inventions*, *7*(4). https://doi.org/10.3390/inventions7040094

Alrweili, H., & Fawzy, H. (2022). Forecasting Crude Oil Prices Using an ARIMA-ANN Hybrid Model. *Journal of Statistics Applications and Probability*, *11*(3), 845–855. https://doi.org/10.18576/jsap/110308

Al-turaiki, I., Almutlaq, F., & Alrasheed, H. (2021). *Empirical Evaluation of Alternative Time-Series Models for COVID-19 Forecasting in Saudi Arabia*.

Andipara, R. (2022). *Applying ARIMA-GARCH models for time series analysis on Seasonal and Nonseasonal datasets Part A : Seasonal Dataset : Perrin Freres Monthly Champagne Sales*. *June*, 0–39.

Arumugam, V., & Natarajan, V. (2023). Time Series Modeling and Forecasting Using Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average Models. *Instrumentation Mesure Metrologie*, *22*(4), 161–168. https://doi.org/10.18280/i2m.220404

Bhandari, S., Bergmann, N., Jurdak, R., & Kusy, B. (2017). Time series data analysis of wireless sensor network measurements of temperature. *Sensors (Switzerland)*, *17*(6). https://doi.org/10.3390/s17061221

Bora, S., & Hazarika, A. (2023). Rainfall time series forecasting using ARIMA model. *2023 International Conference on Artificial Intelligence and Applications, ICAIA 2023 and Alliance Technology Conference, ATCON-1 2023 - Proceeding*, *April*. https://doi.org/10.1109/ICAIA57370.2023.10169493

Botelho, G. M., Henrique, A., Oliveira, M. De, Monteiro, L. D., Gabriel, D., Franco, D. B., & Silva, C. (2021). *Application of the ARIMA Model to Predict Under-Reporting of New Cases of Hansen's Disease during the COVID-19 Pandemic in a Municipality of the Amazon Region*

Case, T., Korea, S., Alsharif, M. H., & Younes, M. K. (2019). *SS symmetry Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation :* 1–17. https://doi.org/10.3390/sym11020240

Dahiya, P., Kumar, M., Manhas, S., Saini, A., Yadav, S. K., Sirohi, S., Kamboj, M., Khichar, M. L., Mishra, E. P., Sharma, V., Kour, V., & Fayezizadeh, M. R. (2024). Time series study of climate variables utilising a seasonal ARIMA technique for the Indian states of Punjab and Haryana. *Discover Applied Sciences*,
6(12). https://doi.org/10.1007/s42452-024-06380-5

Dama, F., & Sinoquet, C. (2021). *Time Series Analysis and Modeling to Forecast: a Survey*. 1–76. http://arxiv.org/abs/2104.00164

Deretić, N., Stanimirović, D., Al Awadh, M., Vujanović, N., & Djukić, A. (2022). SARIMA Modelling Approach for Forecasting of Traffic Accidents. *Sustainability (Switzerland)*, *14*(8). https://doi.org/10.3390/su14084403

Dimri, T., Ahmad, S., & Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. *Journal of Earth System Science*, *129*(1). https://doi.org/10.1007/s12040-020-01408-x

Djimasbe, R., Gyamfi, S., Iweh, C. D., & Ribar, B. N. (2024). Development of an ARIMAX model for forecasting airport electricity consumption in Accra-Ghana: The role of weather and air passenger traffic. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *9*(July), 100691. https://doi.org/10.1016/j.prime.2024.100691

Dong, Y., Li, S., & Gong, X. (2017). *Time Series Analysis: An application of ARIMA model in stock price forecasting*. *29*(Iemss), 703–710. https://doi.org/10.2991/iemss-17.2017.140

Doreswamy, Harish Kumar, K. S., & Gad, I. (2020). Time Series Analysis for Prediction of PM 2.5 Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Model on Taiwan Air Quality Monitoring Network Data. *Journal of Computational and Theoretical Nanoscience*, *17*(9), 3964–3969. https://doi.org/10.1166/jctn.2020.8997

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1–9. https://doi.org/10.1177/1847979018808673

García-Ferrer, A. (2003). A Course in Time Series Analysis, In *International Journal of Forecasting* (Vol. 19, Issue 3, pp. 527–530). https://doi.org/10.1016/s0169-2070(03)00023-2

Goyal, A., & Kundu, A. (2020). ARIMA and Indian Stock Market Forecasting. *Xi'an Jianzhu Keji Daxue Xuebao/Journal of Xi'an University of Architecture & Technology.*, *12*(October), 60–70. https://doi.org/10.37896/JXAT12(s).03/807

Hamel, A. A. (2022). Time series Forecasting Using ARIMA model. *Research Gate*, *February*.

Hamiane, S., Ghanou, Y., Khalifi, H., & Telmem, M. (2024). Comparative Analysis of LSTM, ARIMA, and Hybrid Models for Forecasting Future GDP. *Ingenierie Des Systemes d'Information*, *29*(3), 853–861. https://doi.org/10.18280/isi.290306

Haydier, E., Albarwari, N., & Ali, T. (2023). The Comparison Between VAR and ARIMAX Time Series Models in Forecasting. *Iraqi Journal of Statistical Sciences*, *20*(2), 249–262. https://doi.org/10.33899/iqjoss.2023.181260

Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental Research*, *142*, 319–327. https://doi.org/10.1016/j.envres.2015.06.040

Jain, P. K., Quamer, W., & Pamula, R. (2018). Electricity consumption forecasting using time series analysis. In *Communications in Computer and Information Science* (Vol. 906, Issue April 2018). Springer Singapore. https://doi.org/10.1007/978-981-13-1813-9_33

Katambire, V. N., Musabe, R., Uwitonze, A., & Mukanyiligira, D. (2023). Forecasting the Traffic Flow by Using ARIMA and LSTM Models: Case of Muhima Junction. *Forecasting*, *5*(4), 616–628. https://doi.org/10.3390/forecast5040034

Kaur, J., Parmar, K. S., & Singh, S. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research*, *30*(8), 19617–19641. https://doi.org/10.1007/s11356-023-25148-9

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet*, *15*(8), 1–31. https://doi.org/10.3390/fi15080255

Kozitsin, V., & Katser, I. (2021). *Applied sciences Online Forecasting and Anomaly Detection Based on the ARIMA Model*.

Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, *7*(3), 1–9. https://doi.org/10.1007/s12544-015- 0170-8

Lee, D. H., Kim, Y. S., Koh, Y. Y., Song, K. Y., & Chang, I. H. (2021). *Forecasting COVID-19 Confirmed Cases Using Empirical Data Analysis in Korea*. 1–17.

Mahsin, M., Akhter, Y., & Begum, M. (2012). Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series Analysis. *Journal of Mathematical Modelling and Application*, *1*(5), 67–73.

Masena, T. E., Mahlangu, S. L., & Shongwe, S. C. (2024). Time Series Perspective on the Sustainability of the South African Food and Beverage Sector. *Sustainability (Switzerland)*, *16*(22). https://doi.org/10.3390/su16229746

Masena, T. E., Shongwe, S. C., & Yeganeh, A. (2024). Quantifying Loss to the Economy Using Interrupted Time Series Models: An Application to the Wholesale and Retail Sales Industries in South Africa. *Economies*, *12*(9). https://doi.org/10.3390/economies12090249

Meer, A., & Deborah, A. S. (2024). *Eye Disease Detection in Retinal Images using Deep Transfer Learning Techniques*. *1*, 1–12. https://www.kaggle.com/datasets/gunavenkatdoddi/eye-

Mehmood, M. S., Zafar, Z., Sajjad, M., Hussain, S., Zhai, S., & Qin, Y. (2023). Time Series Analyses and Forecasting of Surface Urban Heat Island Intensity Using ARIMA Model in Punjab, Pakistan. *Land*, *12*(1). https://doi.org/10.3390/land12010142

Mubasher Hassan, M., & Mirza, T. (2021). Using Time Series Forecasting for Analysis of GDP Growth in India. *International Journal of Education and Management Engineering*, *11*(3), 40–49. https://doi.org/10.5815/ijeme.2021.03.05

Nayak, J., & Naik, B. (2021). *Intelligent system for COVID-19 prognosis : a state-of-the-art survey*. 2908–2938.

Olsavszky, V., Dosius, M., Vladescu, C., & Benecke, J. (2020). *Time Series Analysis and Forecasting with Automated Machine Learning on a National ICD-10 Database*. 16–18.

Oyeleye, M., Chen, T., Titarenko, S., & Antoniou, G. (2022). A Predictive Analysis of Heart Rates Using Machine Learning Techniques. *International Journal of Environmental Research and Public Health*, *19*(4). https://doi.org/10.3390/ijerph19042417

Park, M. J., & Yang, H. S. (2024). Comparative Study of Time Series Analysis Algorithms Suitable for Short-Term Forecasting in Implementing Demand Response Based on AMI. *Sensors*, *24*(22). https://doi.org/10.3390/s24227205

Patil, R. (2021). *Time Series Analysis and Stock Price Forecasting using Machine Learning Techniques MSc Research Project Data Analytics Rajat Patil Student ID : x19194048 School of Computing National College of Ireland Supervisor : Jorge Basilio*. *February*. https://doi.org/10.1994/Rajat/AI

Pereira da Veiga, C., Pereira da Veiga, C. R., Girotto, F. M., Marconatto, D. A. B., & Su, Z. (2024). Implementation of the ARIMA model for prediction of economic variables: evidence from the health sector in Brazil. *Humanities and Social Sciences Communications*, *11*(1), 1–11. https://doi.org/10.1057/s41599-024-03023-3

Permatasari, C. I., Sutopo, W., & Hisjam, M. (2018). Sales forecasting newspaper with ARIMA: A case study. *AIP Conference Proceedings*, *1931*(February). https://doi.org/10.1063/1.5024076

Pokhrel, A., & Adhikari, R. (2023). Leveraging Exogenous Insights: A Comparative Forecast of Paddy Production in Nepal Using ARIMA and ARIMAX Models. *Economic Review of Nepal*, *6*(1), 52–69. https://doi.org/10.3126/ern.v6i1.67970

Pramanik, A., Sultana, S., & Rahman, M. S. (2022). Time Series Analysis and Forecasting of Monkeypox Disease Using ARIMA and SARIMA Model. *2022 13th International Conference on Computing Communication and Networking Technologies, ICCCNT 2022*, *October*, 1–7. https://doi.org/10.1109/ICCCNT54827.2022.9984345

Ray, S., & Bhattacharyya, B. (2020). Statistical modeling and forecasting of arima and arimax models for food grains production and net availability of India. *Journal of Experimental Biology and Agricultural Sciences*, *8*(3), 296–309. https://doi.org/10.18006/2020.8(3).296.309

Ritik Vimal, M. (2020). Time Series Analysis: ForecastingWith Sarimax Model and Stationarity Concept.

JETIREJ06034 Journal of Emerging Technologies and Innovative Research, *7*(12), 156–161. www.jetir.org

Saleem Latteef Mohammed, Khan, I., & Albreem, MahmoudMohammed H. Alsharif, S. K. G. (2019). SS Symmetry Robust Hybrid Beamforming Scheme for Wireless Networks. *Mdpi*, *2*(2019), 18.

Samal, K. K. R., Babu, K. S., Das, S. K., & Acharaya, A. (2019). Time series based air pollution forecasting using SARIMA and prophet model. *ACM International Conference Proceeding Series*, 80–85. https://doi.org/10.1145/3355402.3355417

Sarvaiya, D., & Ramchandani, D. (2022). Time Series Analysis and Forecasting of Gold Price using ARIMA and LSTM Model. *International Journal for Research in Applied Science and Engineering Technology*, *10*(9), 168–173. https://doi.org/10.22214/ijraset.2022.46555

Sediakgotla, K., Molefe, W., & Shangodoyin, D. K. (2019). Time Series Modelling of Monthly Average Temperature in Gaborone-Botswana. *African Journal of Applied Statistics*, *6*(2), 689–710. https://doi.org/10.16929/ajas/2019.689.237

Shelatkar, T., Tondale, S., Yadav, S., & Ahir, S. (2020). Web Traffic Time Series Forecasting using ARIMA and LSTM RNN. *ITM Web of Conferences*, *32*, 03017. https://doi.org/10.1051/itmconf/20203203017

Singh, P., & Gupta, A. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information*. *January*.

Singh, P., & Gupta, A. (2022). Generalized SIR (GSIR) epidemic model: An improved framework for the predictive monitoring of COVID-19 pandemic. *ISA Transactions*, *124*, 31–40. https://doi.org/10.1016/j.isatra.2021.02.016

Sirisha, U. M., Belavagi, M. C., & Attigeri, G. (2022). Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison. *IEEE Access*, *10*, 124715–124727. https://doi.org/10.1109/ACCESS.2022.3224938

Smith, D. C., Thumm, E. B., Anderson, J., Kissler, K., Reed, S. M., Centi, S. M., Staley, A. W., Hernandez, T. L., & Barton, A. J. (2024). Sudden Shift to Telehealth in COVID-19: A Retrospective Cohort Study of Disparities in Use of Telehealth for Prenatal Care in a Large Midwifery Service. *Journal of Midwifery and Women's Health*, *69*(4), 522–530. https://doi.org/10.1111/jmwh.13601

Somboonsak, P. (2019). Forecasting Dengue Fever Epidemics using ARIMA Model. *ACM International Conference Proceeding Series*, 144–150. https://doi.org/10.1145/3375959.3375970