First Year (Semester-2) Research Assignment on

Time Series Analysis of Marine Fish Sales

in partial fulfilment of the requirement for the successful completion of semester 2 of MSc Big Data Analytics

Submitted By

23-PBD-010

Kotia Divyakumar Ashok

(Semester – II MSc. BDA)

Under the supervision of

Dr. Pravida Raja

2023-2024

Department of Computer Sciences (MSc. BDA)

St. Xavier’s College (Autonomous) Ahmedabad – 380009

DECLARATION

I, the undersigned solemnly declare that the research assignment *Time Series Analysis of Marine Fish Sales* is based on my work carried out during the course of our study under the supervision of *Dr. Pravida Raja*. I assert the statements made and conclusions drawn are an outcome of my research work. I further certify that

- The work contained in the report is original and has been done by me under the general supervision of my supervisor.
- The work has not been submitted to any other Institution for any other degree / diploma / certificate in this university or any other University of India or abroad.
- We have followed the guidelines provided by the department in writing the report.

Kotia Divyakumar Ashok

23-PBD-010

MSc. BDA (Big Data Analytics)

St. Xavier's College (Autonomous), Ahmedabad

TABLE OF CONTENTS

Section No.	Title of section	Page Number
1	Abstract	4
2	Introduction	5
3	Review of Literature	6
4	Objective, Data & Methodology	7
5	Data and Data Analysis	8
6	Findings and Conclusions	12
7	References	13

1. ABSTRACT

This study examines monthly sales data of marine fish from a fishing vessel in Porbandar district, Gujarat, spanning 2015 to 2022. It aims to discern local-level trends and patterns in marine fish sales to inform fisheries management and policy formulation. Key findings reveal seasonal fluctuations, with October exhibiting peak sales and February showing the lowest. The R-squared value of 0.369 suggests a moderate influence of linear trend components on sales variability. The study underscores the importance of advanced time series analysis for understanding marine fish sales dynamics, contributing to sustainable resource management and policy decisions in the fishing industry.

Keywords: Marine fish sales, Time series analysis, Economic development, Sustainability

2. INTRODUCTION

Understanding changes in fisheries over time is crucial for ensuring the sustainable use of marine resources, which are essential for food security and economic development. For instance, tuna stocks play a vital role in providing nutrition, income, and jobs in many countries. Predicting these changes helps policymakers make better decisions, but it can be challenging due to various factors influencing fisheries [6]. The global demand for food has surged with the growing population, leading to increased fish consumption, which rose from 9.0 kg per capita in 1961 to 20.5 kg in 2018. India, the world's second-largest fish producer, plays a pivotal role in providing nutritional security, food security, and employment[1]. With over 14.5 million people engaged in the fisheries sector, fish production significantly contributes to India's economy. Gujarat, endowed with a vast coastline and an extensive Exclusive Economic Zone (EEZ), is a major player in marine fish production, with substantial contributions to the country's total fish production [4].

Fish production forecasts are crucial to meet India's future fish demand, with Gujarat being a prominent marine fish production hub [4]. Boasting a coastline of 1600 km and an expansive continental shelf, Gujarat offers abundant fishing opportunities [4]. With nearly 37,980 registered marine boats and over 5.59 lakh fishers (more than 1 lakh families) dependent on fishing, Gujarat's fisheries sector is integral to the state's economy. Porbandar district, with its 106 km Arabian Sea coastline, is a significant contributor to Gujarat's fish catch, primarily through marine fishing activities of the "Kharva" community[5][8]. This study focuses on analysing time series data related to marine fish sales, particularly collected from Porbandar district, Gujarat, to understand trends and patterns in marine fish sales dynamics at the local level. Understanding marine ecosystem interactions with human activities is crucial for fishing-dependent community resilience amid environmental changes. This study aims to contribute to managing marine resources by analyzing marine fish sales data. Subsequent sections will review literature, outline objectives, describe methodology, and present findings to inform fisheries management and policies.

3. REVIEW OF LITERATURE

The reviewed studies provide insights into fisheries time series analysis, focusing on methods for sales or production estimation, model selection, trend prediction, and time series decomposition, contributing to sustainable fisheries management.

Shikhova et al. (2020) delve into revenue forecasting for fish processing enterprises, proposing a regression model that incorporates environmental temperature dynamics. By considering the impact of temperature fluctuations on demand for fish products, their approach aids in optimizing production planning and procurement decisions amidst market volatility. [6]

In contrast, Handayani et al. (2020) introduce the SARIMA model for predicting catfish sales, emphasizing its utility in capturing seasonal sales patterns. Their study underscores the importance of informed decision-making and flexible operations in micro cooperatives, highlighting the practical application of advanced forecasting techniques in the fisheries sector. [7]

Solanki et al. (1998) analyse seasonal trends in fishery resources off Gujarat's coast, utilizing thermal imagery to identify potential fishing grounds. Their findings provide valuable insights for fisheries management and resource allocation in the region, emphasizing the significance of environmental factors in shaping fishing patterns. [4]

Anuja et al. (2017) investigates trends in marine fish production in Tamil Nadu, comparing the forecasting accuracy of ARIMA and regression models. Their study highlights the importance of sustainable fishing practices and suggests potential avenues for improving prediction accuracy using advanced methods. Additionally, they acknowledge the limitations of linear models like ARIMA and advocate for the exploration of non-linear methods to capture complex relationships in fisheries data. [1]

Collectively, these studies contribute to a comprehensive understanding of fish sales and production dynamics, offering valuable insights for informed decision-making and sustainable resource management in the fisheries sector.

4. OBJECTIVE, DATA & METHODOLOGY

Objective: The primary of this study is to examine and predict the monthly sales data of Marine Fish from 2015 to 2022. The objective is to understand how sales change over time, providing valuable information about the marine fish market. By doing so, this research aims to help with fisheries management and policy-making, supporting the development of strategies for using marine resources sustainably.

Data: The data used for this study belong to the sales (in INR) of marine fish caught over a period of 2015 to 2022 by one of the fishing vessels of one of the fishermen of Porbandar district. The data was converted to the .csv format by using the Google lens and was pre-processed using the Python and Excel.

Methodology:

The methodology performed in this study involved the utilization of the R programming language for data analysis and visualization. First of all, data was imported and the time series was plotted to gain initial insights into the sales pattern of Marine Fish over the specified period. After that following methods were performed:

1. **Autocorrelation Analysis:** Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were generated to assess the correlation between observations at different time lags, aiding in identifying potential seasonal and trend components within the data.
2. **Time Series Decomposition:** Time series decomposition was performed to extract the trend and seasonal components, providing insights into the underlying patterns and variations in the sales data.
3. **Seasonal Index Calculation:** Seasonal indices were calculated to quantify the seasonal variation in sales data. It shows to impact of each month on the overall sales pattern.
4. **Least Square Trend estimation:** Trendline was fitted using the method of Least squares and the estimated values of sales were obtained.

5. DATA AND DATA ANALYSIS

	Jan	Feb	Mar	Apr	May	Jun	Jul
1	117730.0	140832.5	181632.5	113061.5	143101.0	275182.5	123740.0
2	298367.5	281775.0	258609.5	180257.0	185687.0	215592.0	83626.5
3	266104.0	149929.0	221976.5	245192.5	309446.0	183954.0	269967.5
4	267493.5	176432.0	392811.0	423425.5	458064.5	441704.0	284542.5
5	373883.0	171045.0	269568.5	312162.0	397966.0	342753.0	457034.0
6	523098.5	387657.0	323620.0	254364.0	481405.5		

	Aug	Sep	Oct	Nov	Dec
1	159032.5	216095.0	437628.0	266464.0	103112.0
2	359655.0	466050.0	285996.5	143336.0	147200.0
3	366083.0	321813.0	237852.0	204871.0	309545.0
4	273524.0	178865.5	415341.0	424609.0	421153.5
5	381967.0	447750.0	327694.5	273509.0	322154.0
6					

fig 1. Overview of data

Fig 1. shows the general overview of the data which spans from January 2015 to December 2022. Also, for some months specifically for June – August for every year no data was captured as during those months there is restriction in the fishing activities due to not suitable weather conditions. So, such months were removed the during the preprocessing stage

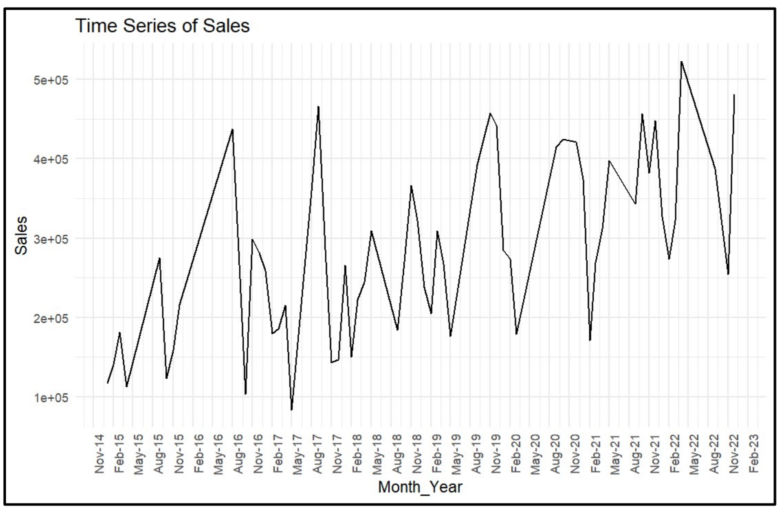


fig 2. Time Series plot of the sales data

From fig. 1 we can say that there exists trend as well as seasonality component in the timeseries. We will also consider ACF plot to comment on the seasonality.

ACF and PACF Plots

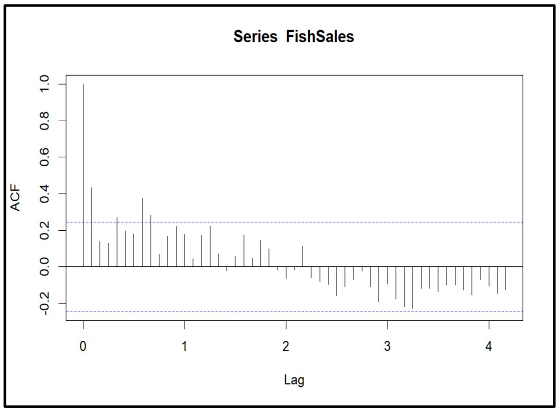


fig. 3 ACF plot

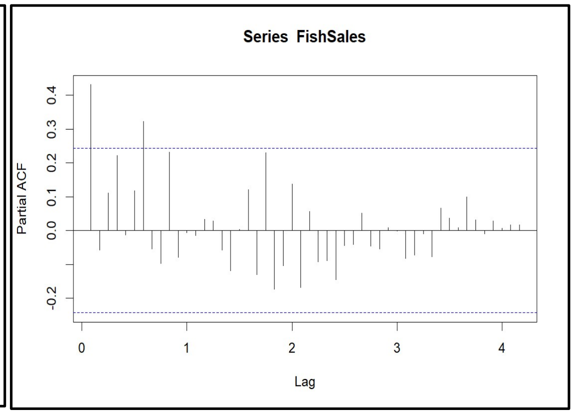


fig. 4 PACF plot

From the fig 3., we can see that the ACF plot is in the oscillating manner which suggests the presence of seasonality in our timeseries. But from ACF and PACF plot it is difficult to conclude anything about the presence of AR or MA components as the lags are not in the way as they are when there is certainty of presence of AR or/and MA components. Now, since we know that there exists seasonality in the timeseries, we will decompose the time series into its components.

Time Series Decomposition

Every time series contains mainly 4 components namely:

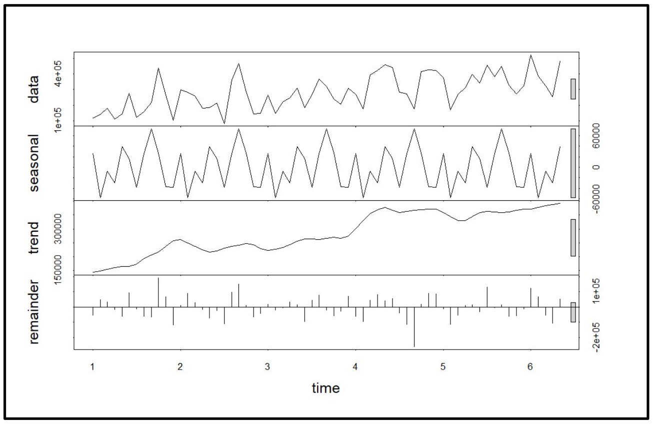
Trend (T): A trend is a long-term represents a growth or a decline of a time series over an extended period of time.

Seasonal component (S): This term of seasonality is used for time series defined at time intervals which are fractions of a year. It is a pattern of change that repeats itself from year to year.

Cyclical component (C): Changes in sometimes show a wave trend, which shows the possible existence of periodicity with longer intervals.

Irregular component (ε): This is a part of a time series represented by residuals, after the above-mentioned components have been removed.

It can be expressed as: $y_t = T_t + S_t + C_t + \varepsilon_t$



From the fig. 5) we can visualize every components of the time series and from that we can say that good amount of seasonal as well as trend component is present in our time series data.

fig. 5 Decomposition of Time Series

Seasonal index

	Jan	Feb	Mar	Apr	May
1	21653.424	-90807.373	-4596.555	-1346.612	47256.768
2	21653.424	-90807.373	-4596.555	-1346.612	47256.768
3	21653.424	-90807.373	-4596.555	-1346.612	47256.768
4	21653.424	-90807.373	-4596.555	-1346.612	47256.768
5	21653.424	-90807.373	-4596.555	-1346.612	47256.768
6	21653.424	-90807.373	-4596.555	-1346.612	47256.768

	Jun	Jul	Aug	Sep	Oct
1	3111.570	-33442.290	25392.968	40215.268	52642.218
2	3111.570	-33442.290	25392.968	40215.268	52642.218
3	3111.570	-33442.290	25392.968	40215.268	52642.218
4	3111.570	-33442.290	25392.968	40215.268	52642.218
5	3111.570	-33442.290	25392.968	40215.268	52642.218
6					

	Nov	Dec
1	-29699.107	-30380.279
2	-29699.107	-30380.279
3	-29699.107	-30380.279
4	-29699.107	-30380.279
5	-29699.107	-30380.279
6		

fig. 6 Seasonal index values

Fig 6. Shows the seasonal indices for each month. The magnitude of seasonal index shows the degree of deviation from the average or baseline level. Months having the positive seasonal index suggest that there is an increase in sales in those months compared to the average while the negative values suggest the decrease.

So, we can say that on an average October month shows the most increase in the Marine fish sales compared to the average sales and similarly February shows the maximum decrease in the sales.

Least square trend estimation

Now we will try to find the estimated values of trend using the method of least squares.

Least squares method is a fundamental technique used in time series analysis to estimate the relationship between variables and make predictions based on historical data. In the context of time series analysis, the least squares method is commonly employed to identify and quantify trends by fitting a straight line to the observed data points.

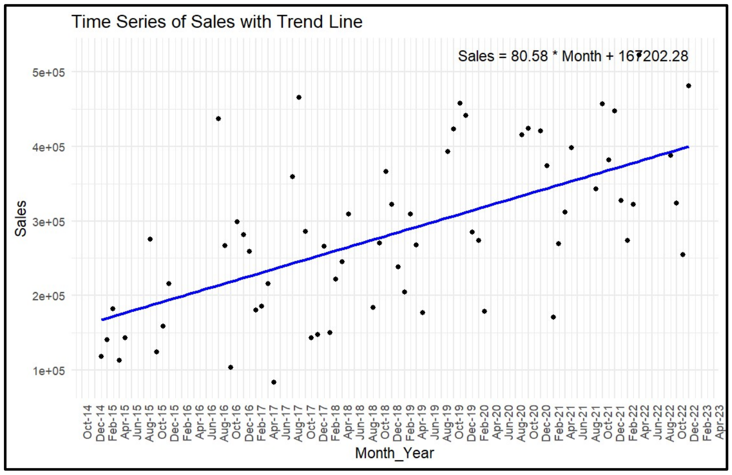


fig. 7 Fitted trendline using method of least square

As shown in fig 7. we can see that a straight line is fitted to the observed data points. From that we can conclude that the slope of the fitted line is positive which suggests that there is positive trend signifies that over a period of time the sales of fish are likely to increase. The equation of the fitted trendline is obtained as:

$$Sales = 80.58 \times Month + 167202.28$$

Month_Year	Original_Values	Estimated_Values	Residuals
2015-01-01	117730	167202	-49472
2015-02-01	140833	169700	-28868
2015-03-01	181633	171956	9676
2015-04-01	113062	174454	-61393
2015-05-01	143101	176871	-33770
2015-09-01	275183	186782	88400
2015-10-01	123740	189200	-65460
2015-11-01	159033	191698	-32665
2015-12-01	216095	194115	21980

fig. 8 Estimated values of the sales

Considering this trend equation, the estimated values of sales and the corresponding residuals are as mentioned in fig. 8. Also, according to the trendline equation the predicted value for the next 4 periods were also obtained as mentioned in fig. 9

	Month_Year	Predicted_Sales
1	2022-12-02	400229.3
2	2023-01-02	402727.2
3	2023-02-02	405225.1
4	2023-03-02	407481.2

fig. 9 Predicted values

6. FINDINGS AND CONCLUSIONS

By decomposing the timeseries and calculating the seasonal indices, it was found that the magnitude of the seasonal index was maximum for the October month (52642.218). So, we can say that on an average October month shows the most increase in the Marine fish sales compared to the average sales and similarly February shows the least seasonal index with negative values (-90807.373) suggesting the month having the maximum decrease in the fish sales.

The R-squared value representing the goodness of fit of the trendline was obtained to be around 0.369, which suggest that the component of the linear trend is not much dominant in terms of the sales. It suggests that around 36.9 % of the variability in the marine fish sales can be explained by the linear trend component. Thus, the remaining variability could be explained by other components such as seasonal component.

For the sales of Marine fish, the appropriated time series model cannot be decided easily from the ACF and PACF plots so we need to look for more advanced TS models like SARIMA.

In conclusion, looking ahead, there is a need to address several areas for further enhancement in this research. These areas include developing strategies to handle missing data within the time series, exploring the integration of advanced time series models, and incorporating additional environmental factors such as sea surface temperature and wind speed into the analysis.

7. REFERENCES

1. Anuja, A., N.R. Kumar, V.S. Bharti, & Yadav, V. (2017). Trends in marine fish production in Tamil Nadu using regression and autoregressive integrated moving average (ARIMA) model. *Journal of Applied and Natural Science*, 653-657.

2. Boruah, B. B., Dutta, A., Hazarika, B. B., & Roy, P. (2020). Forecasting Marine and Total Production in India Using ARIMA Models. *Indian Journal of Economics and Business, 19*.

3. Gianpaolo Coro, S. L. (2016). Analysing and forecasting fisheries time series: purse seine in Indian Ocean as a case study. *ICES Journal of Marine Science*.

4. H U Solanki, M. R. (1998). Seasonal trends in the fishery resources off Gujarat: Salient observation using NOAA AVHRR. *Indian Journal of Marine Sciences*, 438-442.

5. Jazreen Deboo, R. G. (2020, November 2). *Porbandar Fisheries & Impact of Covid-19.* Retrieved from https://iimb.ac.in/: https://iimb.ac.in/turn_turn/porbandar-fisheries-impact-covid-19.php

6. Oksana Shikhova, M. S. (2020). Application of methods of modeling and forecasting for planning of operating results of fish processing enterprises. *BIO Web of Conferences 17, 00157 (2020)*

7. Nandia Rani Woro Tri Handayani1, M. M. (2020). Forecasting of Catfish Sales by Time Series. *84 Jurnal Buana Informatika, Volume 11*, 84-93.

8. Commisionaer of Fisheries, Agriculture, Farmers Welfare & Co-operation Department, Government of Gujarat. (2021-22). *Statistical Information.*