Mastering Time Series Analysis and Forecasting with Python

Bridging Theory and Practice Through Insights, Techniques, and Tools for Effective Time Series Analysis in Python

Sulekha Aloorravi

www.orangeava.com

First published: March 2024

Published by: Orange Education Pvt Ltd, AVA™

Address: 9, Daryaganj, Delhi, 110002, India

275 New North Road Islington Suite 1314 London,
N1 7AA, United Kingdom

ISBN: 978-81-96815-10-3

www.orangeava.com

Dedicated To

My Loving Husband:

Dileep

My Strength and Support System

About the Author

Sulekha Aloorravi is a professional with over 18+ years of experience along with a diverse background and several key roles. She is currently the Vice President in the Banking industry, where she also specializes as a Data Scientist. In addition to her corporate role, Sulekha is also a mentor with Great Learning. Her contributions in the academic field have been recognized and cited.

Her expertise extends into the realm of engineering and data science, with a noted deep understanding of various technologies and systems. This technical proficiency is further exemplified through her work as an author. Sulekha has written "*Metaprogramming with Python*," a guide for programmers on writing reusable code to build smarter applications.

This combination of roles in both the corporate and academic sectors, along with her contributions to the field of programming through her publication, highlights Sulekha Aloorravi's multifaceted expertise and her significant presence in the fields of data science, business management, and technology.

Sulekha is a passionate advocate for the use of data science to solve real-world problems. She has a strong track record of success in identifying and extracting valuable insights from large datasets, which she has then used to improve business processes, optimize data-driven solutions, and make better-informed decisions. In addition to her technical expertise, Sulekha is also a highly effective communicator and collaborator. She has a proven ability to work with cross-functional teams to translate complex data into actionable insights that can be readily understood and adopted by business.

About the Technical Reviewer

Dileep Vuppaladhadiam is an accomplished leader in the field of Artificial Intelligence and Machine Learning (AI/ML) with a remarkable 18-year track record of shaping the industry. His expertise spans a wide spectrum of domains, including solution design, data architecture, data engineering, data science, and the practical application of artificial intelligence and machine learning technologies.

Throughout his career, Dileep has played a pivotal role in deploying cutting-edge AI/ML-based applications, employing rigorous data science methodologies to empower data-driven decision-making. His achievements include the successful implementation of numerous AI/ML solutions, both on-premises and in cloud environments, delivering substantial business value.

Dileep's academic journey is equally impressive, featuring multiple majors encompassing accounting, economics, finance, business administration, data science, artificial intelligence, business analytics, and business intelligence. His commitment to education extends beyond his own studies, as he has also served as a dedicated coach, mentor, and faculty member, inspiring and guiding countless aspiring data scientists.

His professional footprint spans diverse sectors, including Manufacturing and Industrial, Information Technology, Banking, Finance, Retail, Travel and Tourism, and Consulting. Dileep has had the privilege of leading Research and Development, Technical Consultancy, and Support teams, bringing innovative solutions to renowned organizations such as IBM, Capgemini, Barclays, HSBC, HomeCredit, Mindtree, and Infosys.

Currently, Dileep holds the prestigious position of Vice President at Wells Fargo, where he plays a pivotal role in guiding the global enterprise towards achieving significant and sustainable business value. His focus centers on harnessing the power of robust and scalable data engineering and data science approaches tailored to the modern business landscape. Through his leadership, he champions data-driven business and technical transformations, enabling the creation of substantial value on a global scale.

Acknowledgements

I am deeply grateful to my family for their unwavering support and encouragement throughout the journey of writing this book. Their love, understanding, and patience have been my anchor, enabling me to devote countless hours to research, writing, and revision. To my spouse Dileep, whose belief in me never wavered, and to my nephew Sai Sathvik, whose boundless energy and laughter filled my days with joy and inspiration, I owe a debt of gratitude beyond words.

Thank you for your sacrifices, your understanding during late nights and weekends spent at the keyboard, and for being my constant source of motivation. Your love and encouragement have been instrumental in bringing this project to fruition. I am forever grateful for your unwavering support and belief in me.

I extend my sincere gratitude to the publishers for their guidance, expertise, and support throughout the publication process. Their dedication to excellence and commitment to providing valuable resources to readers have been instrumental in bringing this book to successful completion. I am deeply grateful for their collaboration and unwavering belief in the importance of this work.

To the readers, I offer my heartfelt thanks for choosing this book as a resource on your journey to mastering time series analysis with Python. Your interest and engagement inspire me to continue striving for clarity, depth, and relevance in my writing. It is my hope that the knowledge shared within these pages will empower you to tackle complex data challenges with confidence and creativity.

Thank you for your trust, curiosity, and commitment to lifelong learning. Your feedback and insights are invaluable, and I am honored to have the opportunity to contribute to your growth and success in the field of data science.

Preface

Welcome to "*Mastering Time Series Analysis and Forecasting with Python*." In this book, we embark on an exploration of time series analysis, a foundational pillar of data science with far-reaching applications across industries. Whether you are a seasoned data scientist seeking to deepen your understanding of time series methods or a beginner eager to unlock the potential of Python for analyzing sequential data, this book is designed to be your comprehensive guide.

The field of time series analysis encompasses numerous methods, algorithms, and techniques tailored to uncover patterns, trends, and insights within sequential data. Through a hands-on, practical approach, we delve into the fundamental concepts of time series analysis, explore state-of-the-art methodologies, and provide step-by-step tutorials to implement these techniques using Python libraries such as pandas, NumPy, Matplotlib, Statsmodels, and more.

As you journey through the pages of this book, you will learn how to visualize time series data, extract meaningful features, build predictive models, and evaluate their performance. From classical methods like ARIMA and exponential smoothing to modern approaches like machine learning and deep learning, we cover a diverse array of techniques to suit various data scenarios and business requirements.

Each chapter is crafted to provide both theoretical foundations and practical applications, ensuring that you not only understand the underlying principles but also gain the skills to apply them effectively in real-world projects. Along the way, you will encounter Python code examples, illustrative plots, and hands-on exercises to reinforce your learning and deepen your understanding.

This book comprises 9 chapters, each a complete module in itself, serving as your comprehensive guide to mastering time series in Python. Whether you are analyzing financial data, forecasting sales, predicting demand, or studying sensor readings, the techniques presented in this book will equip you with the tools and knowledge to tackle a wide range of time series challenges. I invite you to embark on this journey with me and discover the fascinating world of time series analysis with Python.

Chapter 1 Introduction to Time Series: In this chapter, we embark on a journey to explore the fundamental concepts of time series analysis and its applications across various domains. We begin by introducing the concept of time series data and its significance in analyzing sequential data. Through real-world examples, we demonstrate how time series analysis plays a crucial role in diverse industries such as finance, healthcare, manufacturing, and more. Whether you're a beginner or an experienced practitioner, this chapter serves as an essential primer for understanding the role of time series analysis in modern data science applications.

Chapter 2 Overview of Times Series Libraries in Python: In this chapter, readers will embark on an exploration of popular time series libraries in Python, gaining insights into their features and applications through practical examples and illustrations. Beginning with Pandas, the versatile library provides robust capabilities for handling time series data, followed by an examination of NumPy's role in numerical computing for time series analysis. The chapter further delves into Statsmodels, showcasing its capabilities for statistical modeling and forecasting. Additionally, readers will be introduced to other significant libraries such as Prophet and AutoTS, which offer advanced features for time series forecasting and modeling. Through this comprehensive exploration, readers will acquire the knowledge and tools necessary to harness Python's capabilities for effective time series analysis.

Chapter 3 Visualization of Time Series Data: In this chapter, readers will be introduced to a myriad of time series visualization techniques facilitated by Python libraries. Beginning with an overview of time series visualization libraries in Python, the chapter delves into basic plotting functionalities using matplotlib, providing a foundational understanding of visualizing time series data. Subsequently, readers will explore advanced visualization capabilities using seaborn, enabling them to analyze complex temporal patterns and relationships with ease. The chapter culminates with an exploration of interactive time series visualizations using plotly, empowering readers to create dynamic and interactive plots that facilitate deeper insights into time series data. Through practical examples and step-by-step guidance, readers will gain proficiency in leveraging Python's visualization tools for effective time series analysis and interpretation.

Chapter 4 Exploratory Analysis of Time Series Data: In this comprehensive chapter, readers will embark on a journey through the foundational aspects of time series data analysis. Beginning with the essential tasks of loading and inspecting time series data, readers will gain proficiency in navigating diverse datasets.

Through an exploration of descriptive statistics, readers will uncover key insights into the characteristics and distributions of time series data, laying the groundwork for deeper analysis. The chapter further delves into advanced techniques such as time series decomposition, stationarity analysis, and autocorrelation analysis, equipping readers with the tools to discern underlying patterns and structures within time series data. Additionally, readers will explore the significance of rolling statistics in capturing temporal trends and variability, enabling them to extract meaningful insights from dynamic datasets. With practical examples and illustrative explanations, this chapter serves as an indispensable guide for mastering the fundamental principles of time series data analysis.

Chapter 5 Feature Engineering on Time Series: In this illuminating chapter, readers will delve into the intricate realm of feature engineering for time series data. Beginning with the exploration of univariate feature engineering techniques, readers will learn how to create lag-based features, compute rolling statistics, derive expanding window statistics, and calculate exponential moving averages. Through practical examples and step-by-step explanations, readers will gain a profound understanding of each technique's application and significance in uncovering temporal patterns and relationships within univariate time series data. The chapter further advances into the realm of multivariate feature engineering, where readers will discover how to craft lag-based multivariate features, generate interaction terms-based features, and derive aggregated features. By mastering these techniques, readers will acquire the expertise to extract valuable insights and enhance the predictive power of their time series models, propelling their data science endeavors to new heights.

Chapter 6 Time Series Forecasting – ML Approach Part 1: In this enlightening chapter, readers will embark on a captivating journey through a myriad of time series forecasting techniques and models. Commencing with a deep dive into the intricacies of Autoregressive Integrated Moving Average (ARIMA), readers will unravel the essence of this foundational model and its application in capturing temporal patterns and trends. Subsequently, the chapter illuminates the process of Seasonal Decomposition of Time Series (STL), providing insights into dissecting time series data into its constituent components of trend, seasonality, and noise. Moving forward, readers will delve into the realm of Exponential Smoothing Models, exploring their efficacy in capturing underlying patterns and making accurate predictions. Furthermore, the chapter unveils the prowess of Facebook Prophet, shedding light on its intuitive interface and robust forecasting

capabilities. Lastly, readers will be introduced to Support Vector Machines (SVM), a powerful machine learning algorithm adept at handling non-linear relationships and making accurate predictions in time series data. Through practical examples and insightful explanations, readers will emerge equipped with the knowledge and skills to navigate the intricate landscape of time series forecasting, empowering them to unlock valuable insights and drive informed decision-making in their data-driven endeavors.

Chapter 7 Time Series Forecasting – ML Approach Part 2: In this interesting chapter, readers will embark on a captivating exploration of diverse machine learning algorithms tailored for time series forecasting. Beginning with a comprehensive overview of Hidden Markov Models (HMM), readers will uncover the intricacies of this probabilistic framework and its application in capturing latent states and transitions within time series data. Subsequently, the chapter delves into the realm of Gaussian Processes, revealing their inherent flexibility and ability to model complex relationships in time series data. Moreover, readers will discover the art of developing machine learning-based approaches for time series forecasting, harnessing the predictive power of algorithms such as Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest, and Gradient Boosting. Through practical examples and insightful discussions, readers will gain invaluable insights into the diverse array of machine learning techniques available for time series analysis, empowering them to make informed decisions and extract actionable insights from their data.

Chapter 8 Time Series Forecasting - DL Approach: In this captivating chapter, readers will embark on an enlightening journey into the realm of deep learning for time series forecasting. The exploration begins with a deep dive into Long Short-Term Memory (LSTM) networks, unraveling the inner workings of these powerful recurrent neural networks designed to capture long-term dependencies in sequential data. Through practical examples and insightful discussions, readers will gain a comprehensive understanding of how LSTM networks can be effectively applied to time series forecasting tasks. Additionally, the chapter explores the application of Gated Recurrent Units (GRUs), another variant of recurrent neural networks known for their efficiency in modeling sequential data. Furthermore, readers will delve into the realm of Convolutional Neural Networks (CNNs), uncovering how these versatile architectures can be adapted to extract meaningful features from time series data for forecasting purposes. Armed with this knowledge, readers will be equipped to leverage the full potential of deep

learning techniques for tackling a myriad of time series forecasting challenges with confidence and proficiency.

Chapter 9 Multivariate Time Series, Metrics, and Validation: In this comprehensive chapter, readers will embark on a comprehensive journey through the fundamental concepts and techniques essential for understanding and analyzing time series data. Beginning with the crucial step of loading and inspecting time series data, readers will learn how to effectively explore the structure and characteristics of their datasets. Subsequently, the chapter delves into the realm of descriptive statistics, providing readers with valuable insights into summarizing and understanding the statistical properties of time series data. The exploration continues with an in-depth examination of time series decomposition techniques, enabling readers to disentangle the underlying components such as trend, seasonality, and noise within their data. Furthermore, readers will learn how to perform stationarity analysis to assess the stability of time series data over time, followed by a thorough review of autocorrelation and partial autocorrelation functions to identify temporal dependencies. Finally, the chapter concludes with an exploration of rolling statistics, empowering readers with the ability to analyze trends and patterns in their time series data using moving averages and other rolling window techniques. Through practical examples and step-by-step explanations, readers will gain a solid foundation in time series analysis, equipping them with the knowledge and skills needed to navigate the complexities of real-world time series datasets with confidence and precision.

This book presents a comprehensive and accessible guide to mastering the intricacies of time series data analysis. Through a comprehensive coverage of theoretical concepts, practical examples, and hands-on Python code, this book equips readers with the knowledge and skills needed to navigate the complexities of time series data across diverse domains such as finance, healthcare, manufacturing, and beyond. From the foundational principles of time series analysis to advanced forecasting models and techniques, each chapter offers invaluable insights and actionable strategies for harnessing the power of time series data to derive meaningful insights and drive informed decision-making.

Downloading the code bundles and colored images

Please follow the links or scan the QR codes to download the ***Code Bundles and Images*** of the book:

https://github.com/OrangeAVA/Mastering-Time-Series-Analysis-and-Forecasting-with-Python

The code bundles and images of the book are also hosted on ***https://rebrand.ly/e6e2d8***

In case there's an update to the code, it will be updated on the existing GitHub repository.

Errata

We take immense pride in our work at **Orange Education Pvt Ltd,** and follow best practices to ensure the accuracy of our content to provide an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@orangeava.com

Your support, suggestions, and feedback are highly appreciated.

DID YOU KNOW

Did you know that Orange Education Pvt Ltd offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at **www.orangeava.com** and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at: **info@orangeava.com** for more details.

At **www.orangeava.com**, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on AVA™ Books and eBooks.

PIRACY

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **info@orangeava.com** with a link to the material.

ARE YOU INTERESTED IN AUTHORING WITH US?

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please write to us at **business@orangeava.com**. We are on a journey to help developers and tech professionals to gain insights on the present technological advancements and innovations happening across the globe and build a community that believes Knowledge is best acquired by sharing and learning with others. Please reach out to us to learn what our audience demands and how you can be part of this educational reform. We also welcome ideas from tech experts and help them build learning and development content for their domains.

REVIEWS

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at Orange Education would love to know what you think about our products, and our authors can learn from your feedback. Thank you!

For more information about Orange Education, please visit **www.orangeava.com**.

Table of Contents

Chapter 1
Introduction to Time Series

Time is the most important dimension of our lives and all the changes happen around the world constantly because of the impact of time. As human beings with access to enormous technology, we collect a lot of data over time with or without realizing it. Understanding the data that we collected over time and utilizing the information we derived from it to make things better helps in human evolution and the evolution of technology for the betterment of this world. The concept of time series comes in handy to perform this analysis.

Whether you are a data scientist, data analyst, data engineer, or Python programmer who works on data that deals with time, you will benefit from the content of this book. Let us together explore the value of time series data and learn how to derive more insights throughout this learning journey.

We will begin this chapter by introducing time series, then examine its applications in various domains across the industry such as finance, healthcare, manufacturing, and more. Finally, we will conclude this chapter by discussing methods for preparing time series data.

Structure

In this chapter, the following topics will be covered:

- Overview of Time Series
- Time Series Usage in Different Domains
- Preparation of Time Series Data

Overview of Time Series

Data collected over regular intervals of time is termed time series data. This can be data collected every second on your mobile phone, data collected throughout the day on your smartwatch, or even data collected manually by someone every hour of the day for a continuous period of one month, among other examples. All of this data, collected for specific purposes, can be used to understand patterns and derive insights for one or more use cases. As time goes on, more data gets collected, and the repository of such large data becomes a rich source of information that can be explored and analyzed to make informed decisions.

Time series data also helps in identifying hidden patterns in the data collected over time. It can provide information on trends, such as the upward and downward movements of the price of oil in the market, as well as seasonality, such as the right time to sell gift items in your shop.

The analysis of time series data can be performed by applying various powerful techniques using programming languages such as Python, R, and many more. You can understand the data simply by exploring it and examining its statistical information, such as measures of central tendency. Alternatively, you can apply machine learning and deep learning to the data to identify hidden insights or to predict and forecast the future. This book covers a wide range of these concepts and provides examples of where they can be applied.

Applications of Time Series Across Industries

Time series data is of extensive use in various industry domains. To understand the examples of time series data in various domains, we will be using the UCI machine learning repository's datasets widely throughout this book. You can find these datasets at https://archive.ics.uci.edu/ml/datasets. Let us explore some of the industry domains here.

Usage in Finance and Economics

Finance and Economics are an integral part of the world's economy. Large amounts of data are accumulated in these domains across the industry. The study of the patterns in this data is very crucial for effective financial decision-making in a constantly changing economic landscape. Some of the applications of time series data in finance and economics will be explored in this section.

Stock Market Analysis

Stock markets are an excellent source of time series data. Trades in stock markets happen during business hours every weekday on several investment products across multiple industry sectors. Using the stock market data, we can analyze market conditions, predict future stock prices, make trading decisions, and manage individual investor interests.

An example of daily stock market time series data for Lenevo Group Limited, sourced from https://finance.yahoo.com/, is represented in the following table (see *Figure 1.1*):

Daily high price of LNVGY Ticker for March 2023

Date	High
01-03-2023	18.719999
02-03-2023	18.59
03-03-2023	18.639999
06-03-2023	19.23
07-03-2023	19.200001
08-03-2023	18.98
09-03-2023	18.6
10-03-2023	19.299999
13-03-2023	19
14-03-2023	18.91
15-03-2023	18.379999
16-03-2023	18.450001
17-03-2023	19.290001
20-03-2023	18.65
21-03-2023	18.620001
22-03-2023	18.74
23-03-2023	20.809999

***Figure 1.1**: Time series data of stock prices*

The trend line for the preceding stock market prices can be reviewed in the following chart (see *Figure* 1.2):

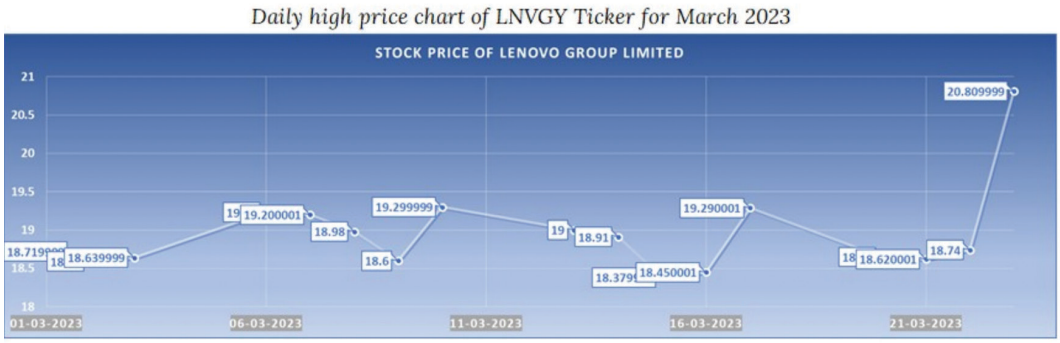

Figure 1.2: *Time series trend of stock prices*

Figure 1.2 demonstrates the behavior of the daily high prices of LNVGY stock over 23 days. A simple analysis of the trend of the stock can explain when the prices were high and when the prices were low during these 23 days.

Market Risk Analysis

Market risk analysis is another type of analysis that can be performed in financial markets to understand the factors that can adversely impact market performance. Time series data is an important factor or input that is used to analyze market risk. Value-at-risk is one of the statistical measures that is majorly used to analyze market risk. Historical market data, such as market prices, spreads, volatilities, interest rates, and more, are used in the analysis. The volatility index, denoted by VIX, is an example of market risk analysis input data. VIX is a volatility index measure calculated based on exchange-specific index option prices, denoting the expectation of volatility for a stock market.

An example of India VIX time series data for NIFTY index option prices is sourced from https://www1.nseindia.com/products/content/equities/indices/historical_vix.htm and is represented in the following table (see *Figure* 1.3):

Daily VIX for India

Date	Open	High	Low	Close	Prev. Close	Change	% Change
01-Mar-23	14.02	14.02	12.275	12.995	14.02	-1.03	-7.31
02-Mar-23	12.995	13.55	11.1	12.97	12.995	-0.03	-0.19
03-Mar-23	12.97	12.97	11.98	12.18	12.97	-0.79	-6.09
06-Mar-23	12.18	12.4725	11.8375	12.2675	12.18	0.09	0.72
08-Mar-23	12.2675	12.8475	11.8975	12.45	12.2675	0.18	1.49
09-Mar-23	12.45	12.8425	11.78	12.725	12.45	0.28	2.21
10-Mar-23	12.725	13.8175	12.725	13.4125	12.725	0.69	5.4
13-Mar-23	13.4125	16.4275	13.4125	16.215	13.4125	2.8	20.89
14-Mar-23	16.215	16.6425	15.0025	16.2175	16.215	0	0.02
15-Mar-23	16.2175	16.505	14.5175	16.295	16.2175	0.08	0.48
16-Mar-23	16.295	17.355	13.95	16.2175	16.295	-0.08	-0.48
17-Mar-23	16.2175	16.2175	14.6	14.7675	16.2175	-1.45	-8.94
20-Mar-23	14.7675	17.22	14.52	16.0075	14.7675	1.24	8.4
21-Mar-23	16.0075	16.035	14.0825	15.0825	16.0075	-0.93	-5.78
22-Mar-23	15.0825	15.0925	14.2425	14.8075	15.0825	-0.28	-1.82
23-Mar-23	14.8075	15.0575	13.9025	14.49	14.8075	-0.32	-2.14

Figure 1.3: *Time series data of India VIX*

India VIX is based on the NIFTY index option prices. The trend lines for the preceding India VIX can be reviewed in the following chart (see *Figure* 1.4):

Daily VIX chart for India

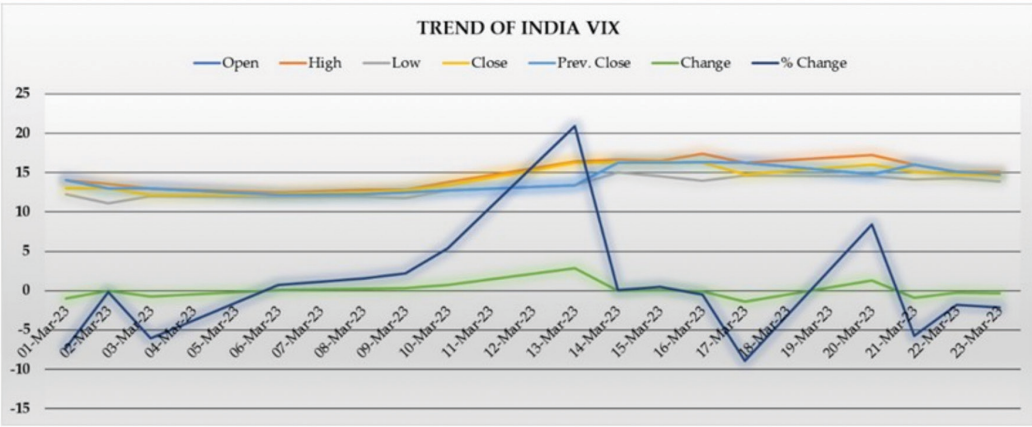

Figure 1.4: *Trend lines of India VIX*

The preceding figure represents the trends of multiple variables of India VIX over 23 days. A trend analysis of each of these variables can help us understand how each variable has performed by itself, as well as with respect to other variables.

Credit Risk Analysis

In finance, lending or loans is a major financial product for many banks and financial institutions. Lending is a process that always poses a threat of credit risks. Credit risk analysis is performed by financial institutions before entering into any lending contract either with customers who are individuals/retailers or with counterparty firms. Credit risk analysis involves various calculations and scorings using historical time series data with multiple variables related to the credit behavior of counterparties. The statistical measures such as the probability of default, loss given default, and exposure at default are derived from the time series data to understand the creditworthiness of a customer or counterparty.

An example of a credit risk dataset generated specifically for this chapter can be seen in *Figure* 1.5:

Loan repayment ledger by date

Date	Loan Repayment Amount (USD
10-01-2022	500
10-02-2022	500
10-03-2022	500
10-04-2022	500
10-05-2022	500
10-06-2022	500
10-07-2022	500
10-08-2022	500
10-09-2022	500
10-10-2022	500
10-11-2022	500
10-12-2022	500
10-02-2023	1020
10-03-2023	500

Figure 1.5: *Loan repayment time series data*

The trend line for the preceding loan repayment ledger can be reviewed in the following chart:

Figure 1.6: *Trend line for loan repayment*

The trend line in *Figure* 1.6 shows the credit behavior of a customer. During the month of January 2023, no repayment was made, and the repayment happened during February 2023 along with a penalty of $20. The credit behavior of a customer becomes an input for the bank while calculating the credit score, thus influencing the lending decision of banks for future loans.

Analysis of Economic Conditions

Time series data can be used to analyze economic indicators such as inflation rates, interest rates, and GDP of a nation, and forecast future values for informed decision-making.

Inflation rate, in general, indicates the percentage increase in the price of goods over time. Similarly, interest rate, in general, denotes the percentage increase of interest on loans borrowed by customers of a financial institution. GDP, or Gross Domestic Product, is a measure with a time dimension that indicates the market value of finished products within a country.

An example of GDP time series data for the United States is sourced from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD and is represented in the following table (see *Figure* 1.7):

Annual GDP of United States

Year	GDP
2011	$ 1,55,99,72,81,23,000.00
2012	$ 1,62,53,97,22,30,000.00
2013	$ 1,68,43,19,09,93,000.00
2014	$ 1,75,50,68,01,74,000.00
2015	$ 1,82,06,02,07,41,000.00
2016	$ 1,86,95,11,08,42,000.00
2017	$ 1,94,77,33,65,49,000.00
2018	$ 2,05,33,05,73,12,000.00
2019	$ 2,13,80,97,61,19,000.00
2020	$ 2,10,60,47,36,13,000.00
2021	$ 2,33,15,08,05,60,000.00

Figure 1.7: *GDP data of the United States*

The preceding data in *Figure 1.7* shows the GDP values of the United States, starting from the year 2011 to 2021. Forecasting the GDP for subsequent years through time series forecasting techniques is one of the use cases of time series data in economics.

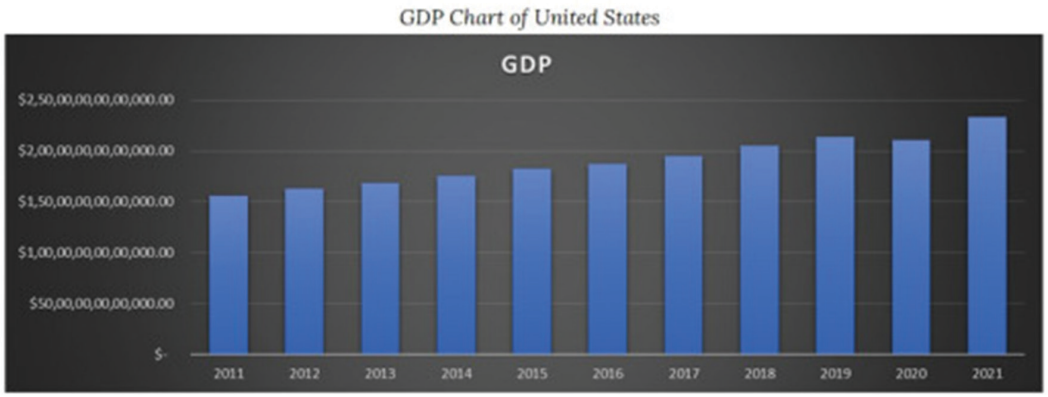

Figure 1.8: *Trend of the GDP of a country*

The year-on-year view of GDP along with a trend of movement can be observed from the preceding chart in *Figure 1.8*. From the chart, it is evident that there is a constant upward trend in GDP, except for the year 2020.

These are some of the applications of time series data in finance. However, there are many more applications, such as investment portfolio management, algorithmic trading, pricing strategies, macroeconomic analysis, and more. With this understanding, let us further explore the usage of time series data in sales and marketing.

Usage in Sales and Marketing

Sales and marketing are domains that are related to each other, as the growth in one drives the growth in the other. There are numerous uses of time series data in these two domains. Some of the applications will be explored in this section.

Forecasting of Retail Sales

Time series data can be used to study the pattern of sales of various products in a supermarket daily and forecast sales for the near future. This helps the management to plan their production and inventory. Sales forecasting can be performed using multiple time series analysis techniques, which will be further studied in subsequent chapters of this book, along with case studies.

Let us now look at an example of time series data that can become an input to sales forecasting. The sample data for this example is downloaded from https://archive.ics.uci.edu/ml/datasets/Sales_Transactions_Dataset_Weekly. This data, as shown in *Figure* 1.9, lists products in rows, the time in weeks in the columns, and the values being weekly sales numbers.

Weekly Sales by Product

Product_Code	W0	W1	W2	W3	W4	W5	W6	W7
P1	11	12	10	8	13	12	14	21
P2	7	6	3	2	7	1	6	3
P3	7	11	8	9	10	8	7	13
P4	12	8	13	5	9	6	9	13
P5	8	5	13	11	6	7	9	14
P6	3	3	2	7	6	3	8	6
P7	4	8	3	7	8	7	2	3

Figure 1.9: *Weekly purchase quantity of products*

Forecasting the purchase quantity for each product is one of the use cases of time series data in sales. Let us look at the weekly purchase trend of each product in *Figure* 1.10.

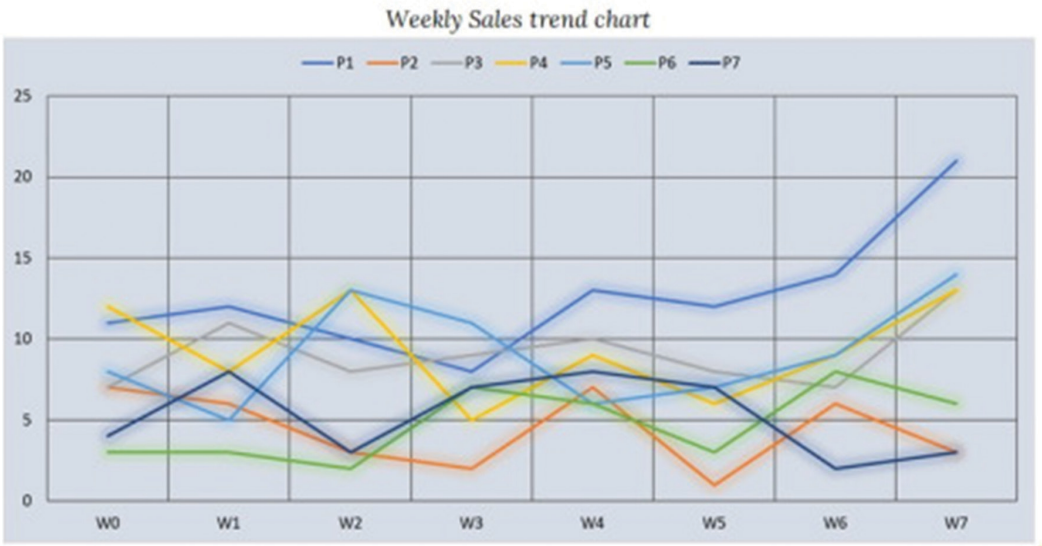

Figure 1.10: *Purchase trend of products*

The purchase quantity of each product in *Figure 1.10* follows a specific pattern. For instance, product P1 has an upward purchase trend over time. This observation can help management to produce or stock more of Product 1 to handle the upcoming demand.

Seasonality Analysis and Planning

Seasonality can be considered as the repetition of a particular behavior at a specific point in time, which is observed as a continuous trend in sales patterns. For example, the sales of air conditioners go up every summer, and the sales of gift articles go up during every festival season such as Christmas to the new year. These are known patterns, but there may also be a lot of unknown seasonal patterns in the sales of products that are not visible evidently by looking at the data. Such patterns can be analyzed by performing a seasonal analysis of the time series data.

An example of time series data depicting the trends of sales quantity of consumer products, which can be used to analyze seasonality is downloaded from https://archive.ics.uci.edu/ml/datasets/online%20retail. A sample of the dataset is presented in *Figure 1.11*:

Retail dataset depicting product sales quantity

StockCo	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 08:26	2.55	17850	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 09:02	2.55	17850	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 09:32	2.55	17850	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	64	01-12-2010 10:19	2.55	17511	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	32	01-12-2010 10:39	2.55	13408	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 10:51	2.55	17850	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	4	01-12-2010 11:21	2.95	15862	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	8	01-12-2010 11:33	2.55	17850	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 12:36	2.95	16552	United Kingdom
85123A	WHITE HANGING HEART T-LIGHT HOLDER	3	01-12-2010 12:43	2.95	14729	United Kingdom

Figure 1.11: *Retail dataset*

The sales quantity in the dataset can be further analyzed to identify seasonal patterns, as represented in *Figure 1.12*:

Retail sales quantity data for the product WHITE HANGING HEART T-LIGHT HOLDER

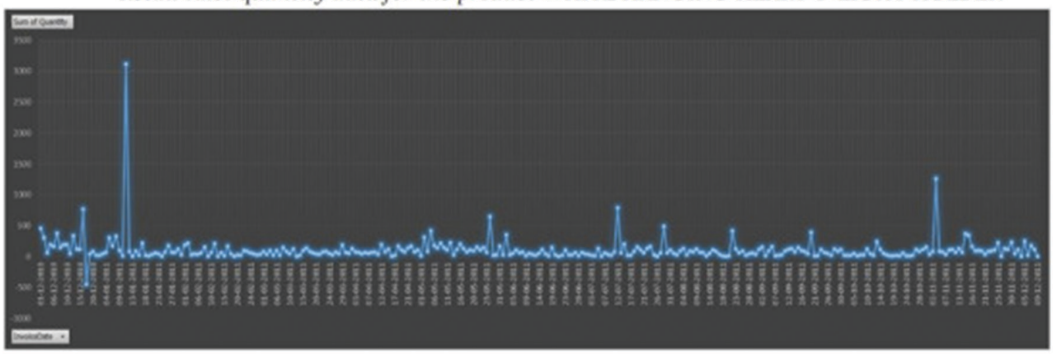

Figure 1.12: *Sales quantity from 2010 to 2011*

The spikes in the data in *Figure 1.12* can be further analyzed in detail for seasonal sales and the reason behind them. Such analysis can help in managing the inventory of the product to handle the spikes in demand.

Inventory Management

Sales forecasting through time series analysis will, in turn, help in managing the stocks in the inventory for specific products that are in demand. Managing inventory efficiently also helps in avoiding excess inventory stocking and the storage costs that will be incurred. Stocking excess inventory in cases of perishable goods such as FMCG (fast-moving consumer goods) like milk, vegetables, baked food, and more, can increase the losses for a retail business.

To explain the benefit of inventory management, let us look at an FMCG inventory and sales dataset generated specifically for this chapter, as shown in *Figure* 1.13. This dataset was generated between 1st January 2023 and 31st January 2023 for four FMCG products: milk, bread, tomatoes, and mangoes. The data is divided into inventory quantity and sales quantity for each day. The inventory quantity denotes the inventory stocked at the beginning of the day, while the sales quantity denotes the sales that happened by the end of the same day.

Sales and Inventory dataset

	Inventory				Sales			
Date	Milk	Bread	Tomatoes	Mangoes	Milk	Bread	Tomatoes	Mangoes
01-Jan-23	100	200	75	80	67	22	18	72
02-Jan-23	100	200	75	80	1	36	66	34
03-Jan-23	100	200	75	80	89	107	30	20
04-Jan-23	100	200	75	80	93	22	74	9
05-Jan-23	100	200	75	80	75	90	16	32
06-Jan-23	100	200	75	80	60	2	14	24
07-Jan-23	100	200	75	80	51	28	14	70
08-Jan-23	100	200	75	80	61	116	53	33
09-Jan-23	100	200	75	80	66	174	37	20
10-Jan-23	100	200	75	80	55	142	5	21
11-Jan-23	100	200	75	80	54	80	9	15
12-Jan-23	100	200	75	80	4	25	52	0
13-Jan-23	100	200	75	80	65	38	24	69
14-Jan-23	100	200	75	80	42	141	35	61
15-Jan-23	100	200	75	80	65	62	61	70
16-Jan-23	100	200	75	80	42	20	22	70
17-Jan-23	100	200	75	80	53	14	30	16
18-Jan-23	100	200	75	80	97	94	31	37
19-Jan-23	100	200	75	80	50	196	19	30
20-Jan-23	100	200	75	80	61	157	8	9
21-Jan-23	100	200	75	80	33	119	69	18
22-Jan-23	100	200	75	80	55	175	0	10
23-Jan-23	100	200	75	80	84	104	15	47
24-Jan-23	100	200	75	80	74	188	48	31
25-Jan-23	100	200	75	80	75	50	3	39
26-Jan-23	100	200	75	80	36	121	37	49
27-Jan-23	100	200	75	80	39	20	35	61
28-Jan-23	100	200	75	80	59	92	52	47
29-Jan-23	100	200	75	80	49	23	57	45
30-Jan-23	100	200	75	80	45	38	24	29
31-Jan-23	100	200	75	80	48	167	25	50

Figure 1.13: *Sales and Inventory of FMCG*

A study of each individual product inventory stock vs. sales is depicted in *Figure 1.14*:

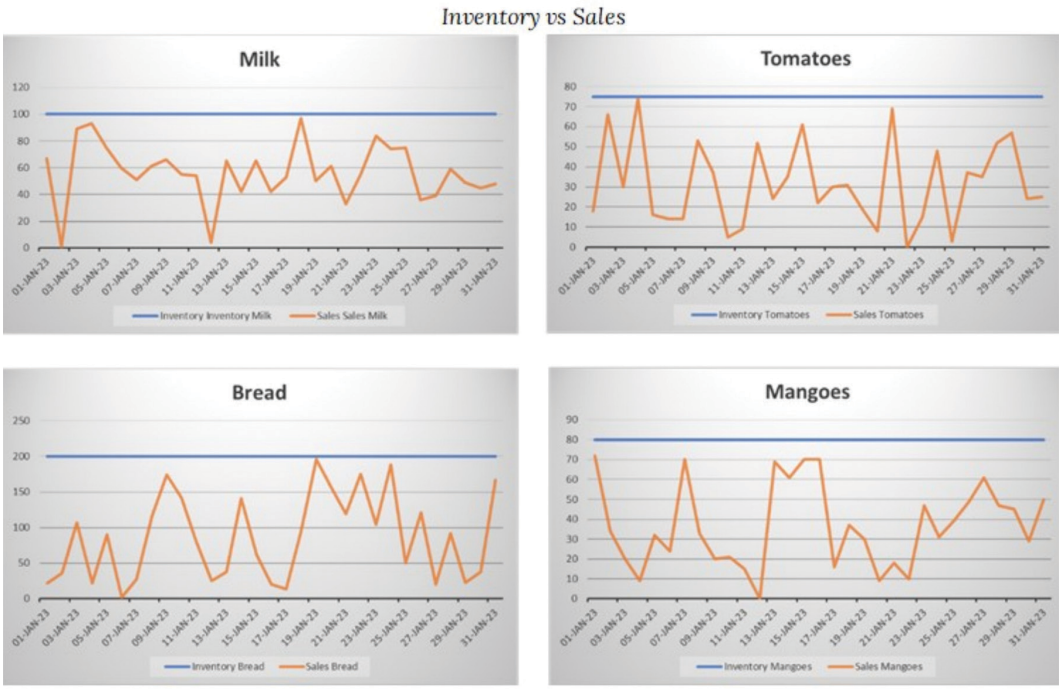

Figure 1.14: *Inventory vs. Sales trend of FMCG products*

Let us look at the trend of one example product from *Figure 1.14*. The inventory quantity of milk for all 31 days of January 2023 has been continuously 100, whereas the actual sales have changed each day, with very low sales on 2nd and 12th January compared to high sales on 4th and 18th January. This analysis gives an idea to decide on what should be the appropriate daily inventory stocking of milk based on the expected demand.

Marketing Campaign Analysis

While marketing new or existing products or services, campaigns are conducted to increase their reach through various channels. For instance, a new broadband service of a telecom provider might conduct a campaign by printing advertisements and delivering them to prospective customers inside a mall or by doing a door-to-door delivery of the advertisement brochures in a particular prospective locality. The other way they might conduct a campaign is by reaching out to prospects through social media platforms or by posting advertisements on websites. The reach of these campaigns can be analyzed through various

analytical methods and tools. The types of data that get collected through these campaigns are time series in nature and help in analyzing various customer behavioral aspects that can help in generating leads through prospects.

To examine an example of campaign time series data, let us look at the bank campaign dataset available at https://data.world/data-society/bank-marketing-data. A sample data from this dataset is presented in *Figure 1.15*. The dataset in *Figure 1.15* has the following input variables: age, job, marital status, education, credit default, loan details, mode of contact, last contacted month and day of the week, last contact duration, number of contacts during the campaign, last contact time gap, campaign outcome, employment variation rate, consumer price index, consumer confidence index, daily indicator, and quarterly indicator.

Banking campaign data

month	may	may	may	may	may	may	may	may
day_of_week	mon	mon	mon	mon	mon	mon	mon	mon
age	56	57	37	40	56	45	59	41
job	housemaid	services	services	admin.	services	services	admin.	blue-collar
marital	married	married	married	married	married	married	married	married
education	basic.4y	high.school	high.school	basic.6y	high.school	basic.9y	professional.course	unknown
default	no	unknown	no	no	no	unknown	no	unknown
housing	no	no	yes	no	no	no	no	no
loan	no	no	no	no	yes	no	no	no
contact	telephone	telephone	telephone	telephone	telephone	telephone	telephone	telephone
duration	261	149	226	151	307	198	139	217
campaign	1	1	1	1	1	1	1	1
pdays	999	999	999	999	999	999	999	999
previous	0	0	0	0	0	0	0	0
poutcome	nonexistent	nonexistent	nonexistent	nonexistent	nonexistent	nonexistent	nonexistent	nonexistent
emp.var.rate	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
cons.price.idx	93.994	93.994	93.994	93.994	93.994	93.994	93.994	93.994
cons.conf.idx	-36.4	-36.4	-36.4	-36.4	-36.4	-36.4	-36.4	-36.4
euribor3m	4.857	4.857	4.857	4.857	4.857	4.857	4.857	4.857
nr.employed	5191	5191	5191	5191	5191	5191	5191	5191
y	no	no	no	no	no	no	no	no

Figure 1.15: *Sample data of a bank campaign*

The input variables in this dataset can be used to perform time series campaign analysis and understand the campaign outcome. This kind of analysis will help in further product development as well as enhancing the reach of existing products to more prospective customers and helps in an effective lead generation too.