Practical Business Python

Lecture 10: Time Series and Forecasting with Python

Iegor Vyshnevskyi

Woosong University

November 23, 2023

Agenda

1. Intro to Forecasting
2. Time Series Data and Forecasting
3. Time Series Forecasting Applications
4. Time Series Components
5. Time Series Forecasting Methods
6. Forecasting: Practical use
7. Forecasting: Practicum
8. In-class Assignment

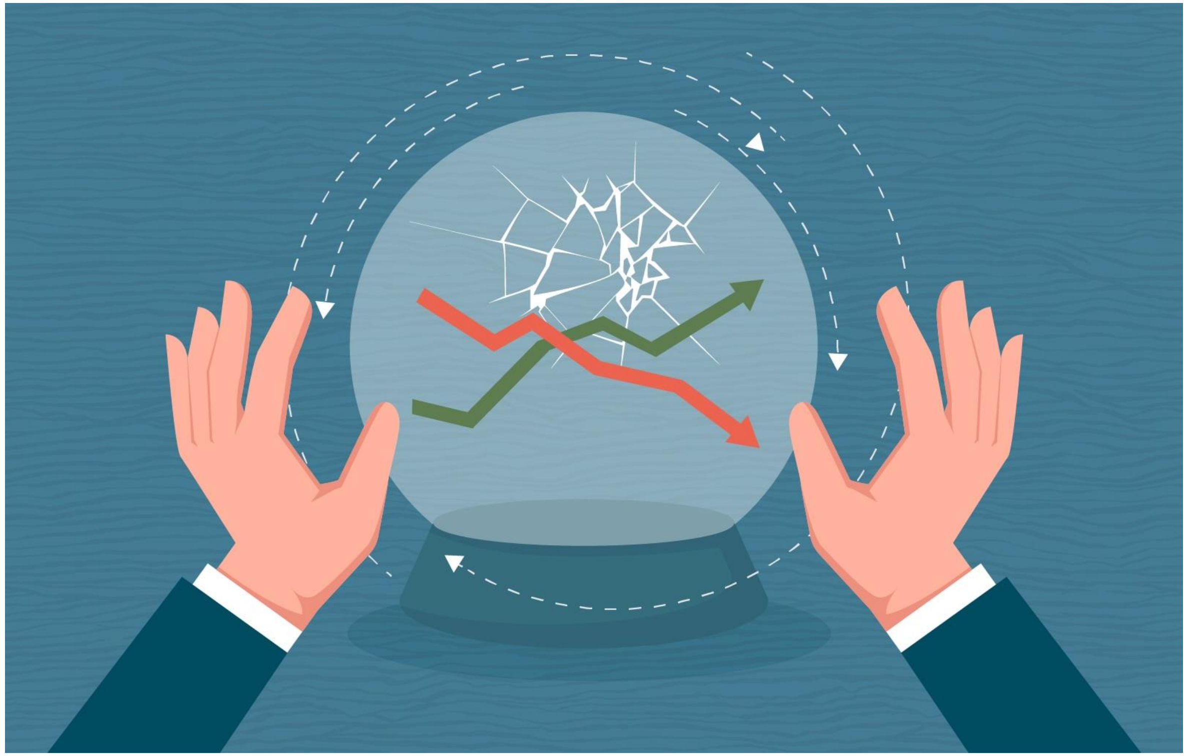
Acknowledgment: Used a number of open sources and materials from the web.

1. Intro to Forecasting

What is ***Forecasting?***

Forecasting is estimating how the sequence of observations will continue into the future.

Forecasting involves making predictions about the future.



Usefulness of Forecasting for Business Analytics

Forecasting is required in many situations:

- deciding whether to build another power generation plant in the next ten years requires forecasts of future demand.
- scheduling staff in a call centre next week requires forecasts of call volumes.
- stocking an inventory requires forecasts of stock requirements.

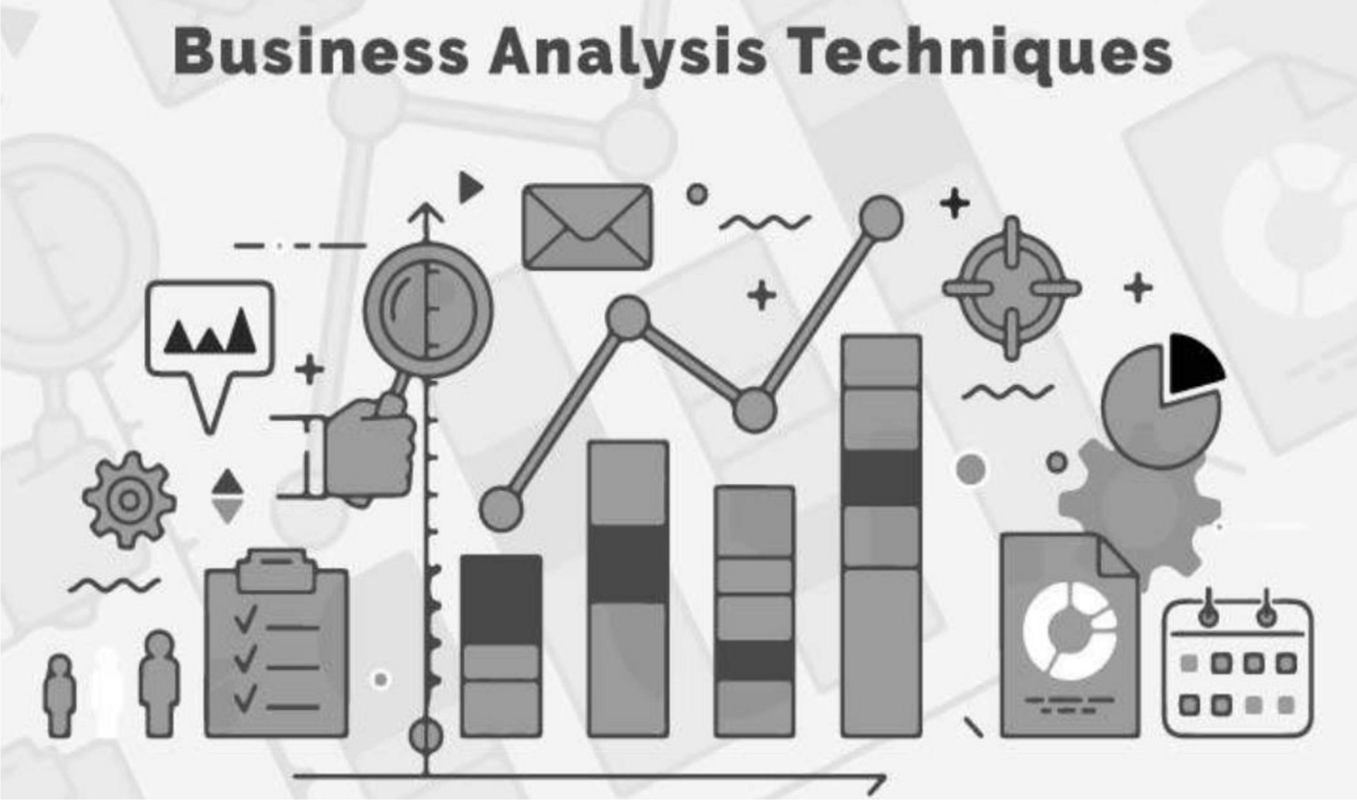
Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for telecommunication routing).

Whatever the circumstances or time horizons involved, forecasting is an important aid to effective and efficient planning.

Usefulness of Forecasting for Business Analytics

Overall, forecasting can be a valuable tool for businesses looking to make data-driven decisions.

However, it should be used in conjunction with other analytical methods to ensure a comprehensive and accurate understanding of business data.



Advantages of using Python for Time Series Forecasting (1)

- ***Rich Ecosystem***: Python has a rich ecosystem of libraries and frameworks that are well-suited for time series forecasting. Notably, libraries such as NumPy, Pandas, and scikit-learn provide powerful tools for data manipulation, analysis, and machine learning.
- ***Time Series Libraries***: Python has dedicated libraries for time series analysis and forecasting, such as Statsmodels, Pytimetk and Prophet. These libraries provide specialized functions and models tailored to the characteristics of time series data.
- ***Machine Learning Libraries***: Python's machine learning libraries, like scikit-learn and TensorFlow, offer a wide range of algorithms that can be applied to time series forecasting. This includes traditional methods like decision trees and ensemble methods, as well as advanced techniques like neural networks.
- ***Data Visualization***: Libraries like Matplotlib, Seaborn and Plotly make it easy to visualize time series data and forecast results. Visualization is crucial for understanding patterns, trends, and the performance of forecasting models.
- ***Community and Documentation***: Python has a large and active community, which means extensive documentation, tutorials, and community support. This is beneficial for both beginners and experienced practitioners working on time series forecasting projects.

Advantages of using Python for Time Series Forecasting (2)

- ***Integration with Other Tools***: Python seamlessly integrates with other tools and technologies. This is crucial for end-to-end data science workflows, where data preprocessing, analysis, forecasting, and deployment might involve different tools.
- ***Flexibility and Customization***: Python provides a high degree of flexibility, allowing practitioners to customize models and algorithms based on specific requirements. This is especially important for dealing with diverse time series data and unique forecasting challenges.
- ***Open Source***: Python is an open-source programming language, and most of the libraries used for time series forecasting are also open source. This not only reduces costs but also allows users to inspect and modify the source code if needed.
- ***Community Packages***: There are numerous community-contributed packages and utilities for time series forecasting in Python. These packages can range from additional forecasting algorithms to specialized tools for feature engineering and model evaluation.
- ***Compatibility with Big Data Tools***: Python can be seamlessly integrated with big data tools and frameworks like Apache Spark. This is important when dealing with large-scale time series data.

Disadvantages of using Python for Time Series Forecasting (1)

- ***Learning Curve***: Python might have a steeper learning curve for individuals who are new to programming or data science. Compared to more specialized tools, beginners might find it challenging to navigate the extensive libraries and frameworks available in Python.
- ***Performance***: Python, in some cases, might not be as fast as languages like C or Java. While this might not be a critical issue for smaller datasets, it can become a concern when dealing with large-scale time series data or when real-time forecasting is required.
- ***GIL (Global Interpreter Lock)***: Python's Global Interpreter Lock can limit the execution of multiple threads in parallel. This might impact the performance of multithreaded applications, although it's less of a concern for time series forecasting where often computation is not highly parallelizable.
- ***Memory Consumption***: Python can be memory-intensive, especially when working with large datasets. This might lead to challenges when working with extensive time series data, and users may need to optimize their code or resort to more memory-efficient tools.
- ***Limited GUI Tools***: Python has fewer graphical user interface (GUI) tools for time series forecasting compared to some specialized software. This might be a disadvantage for users who prefer or are accustomed to working with GUI-based applications.

Disadvantages of using Python for Time Series Forecasting (2)

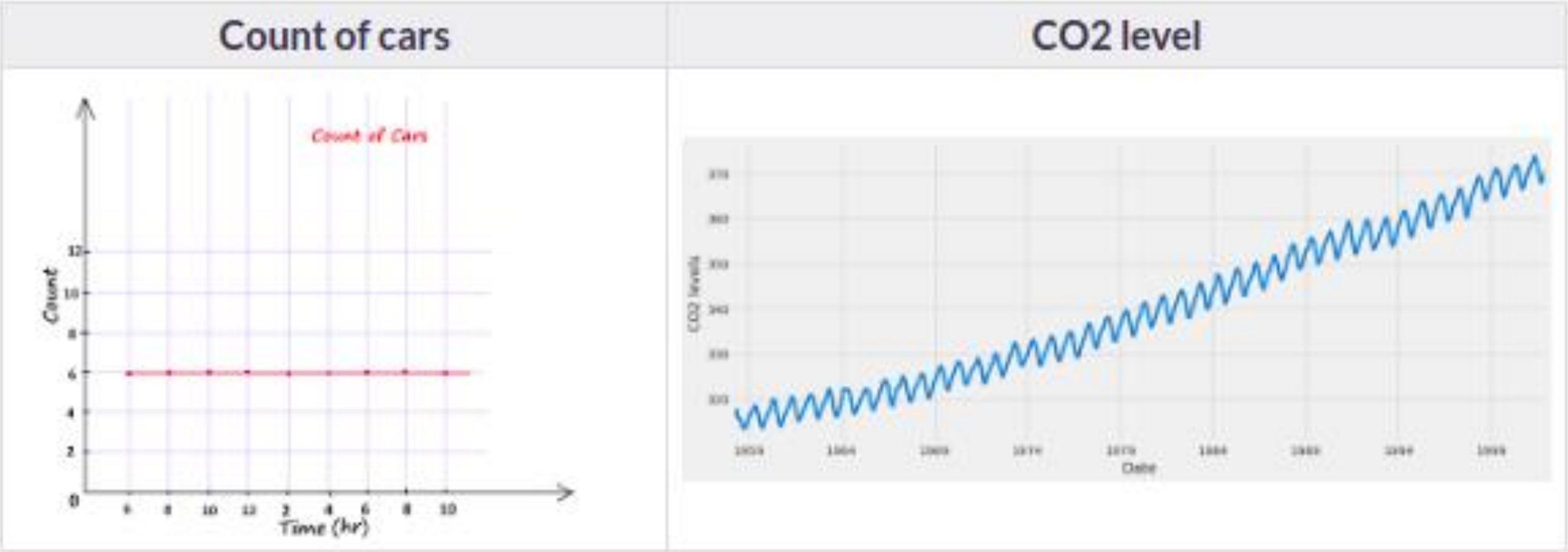
- ***Fragmentation of Libraries***: While the extensive ecosystem is a strength, it can also lead to fragmentation. There might be multiple libraries for similar tasks, and users need to choose the right ones for their specific use case. This can be overwhelming, especially for newcomers.
- ***Reliance on Third-Party Libraries***: The Python ecosystem relies heavily on third-party libraries. While this usually is an advantage, it can be a disadvantage if a critical library becomes deprecated or loses support. This dependency can potentially impact the sustainability of projects.
- ***Ease of Use for Non-Programmers***: For individuals who are not familiar with programming, Python might not be as user-friendly as point-and-click tools that are designed for non-programmers. This can be a drawback for users who prioritize ease of use over customization.
- ***Real-Time Processing Challenges***: While Python is suitable for many time series forecasting tasks, it may face challenges when real-time processing* is a critical requirement. Specialized tools or languages might be preferred in situations where the lowest possible latency is crucial.

*Real-time processing is a method of processing data at a near-instant rate, requiring a constant flow of data intake and output to maintain real-time insights.

2. Time Series Data and Forecasting

What is Time Series Data?

- Which of the following do you believe is an example of a time series? Even if you don't know, make a guess.



- Time Series is data that is accumulated over time and is dependent on it. The count of automobiles is independent of time in this case, hence it is not a time series. Because the CO_2 level increases over time, it is a time series.

Time Series Data

- Series of data observed over time.
  - E.g.: Daily Samsung stock prices, monthly rainfall in Seoul,...
- Not all data that have time values or date values as its features can be considered as a time series data.

Time Series Forecasting

- *Time Series Forecasting* is the method of ***exploring and analyzing time-series data*** recorded or collected over a set period of time.
- This technique is used to ***forecast values and make future predictions***.

Any data fit for time series forecasting should consist of observations over a ***regular, continuous interval***.

3. Time Series Forecasting Applications

Examples of Forecasting Analysis

- Time series forecasting is used in ***stock price prediction*** to predict the closing price of the stock on each given day.
- E-Commerce and retail companies use forecasting to ***predict sales and units sold*** for different products.
- ***Weather prediction*** is another application that can be done using time series forecasting.
- It is used by government departments to predict ***a state's population***, at any particular region, or the nation as a whole.

4. Time Series Components

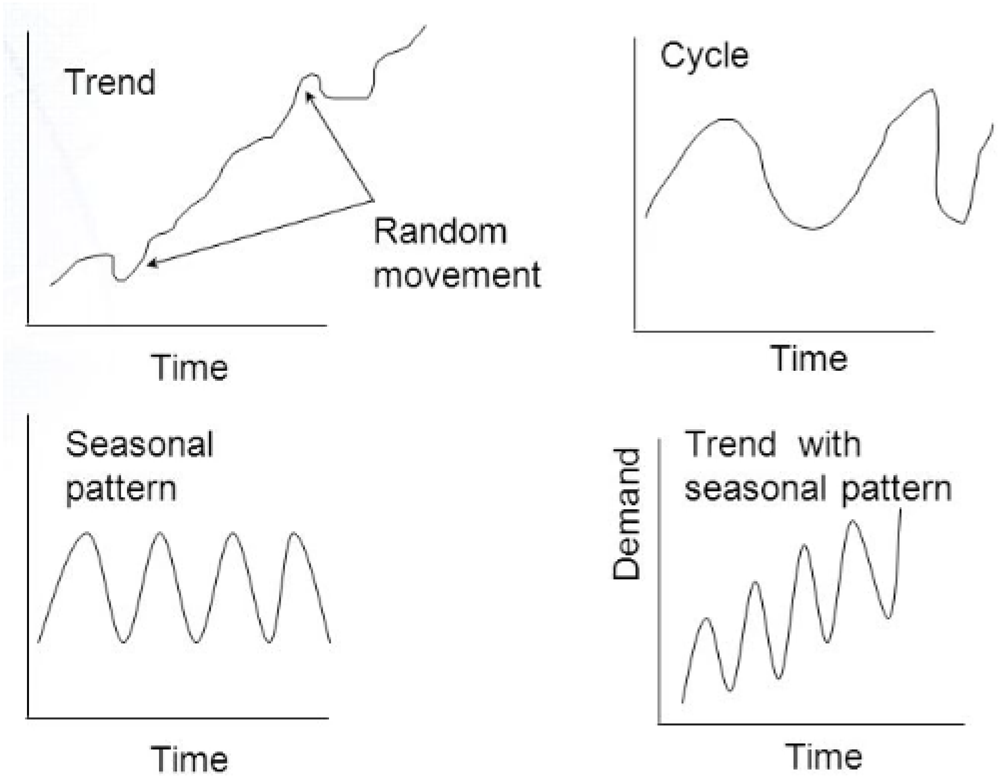
Components

- To use time-series data and develop a model, you *need to understand the patterns in the data over time*.

- These patterns are classified into four components, which are:
  - ***Trend:*** It represents the gradual change in the time series data. The trend pattern depicts long-term growth or decline.
  - ***Level:*** It refers to the baseline values for the series data if it were a straight line.
  - ***Seasonality:*** It represents the short-term patterns that occur within a single unit of time and repeats indefinitely.
  - ***Cyclic:*** A pattern exists where the data exhibits rises and falls that are not of fixed period (duration usually of at least 2 years)
  - ***Noise:*** It represents irregular variations and is purely random. These fluctuations are unforeseen, unpredictable, and cannot be explained by the model.

Seasonal vs cyclic

- Differences between seasonal and cyclic patterns:
  - Seasonal pattern constant length vs. cyclic pattern variable length
  - Average length of cycle longer than length of seasonal pattern
  - Magnitude of cycle more variable than magnitude of seasonal pattern
- The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.



5. Time Series Forecasting Methods

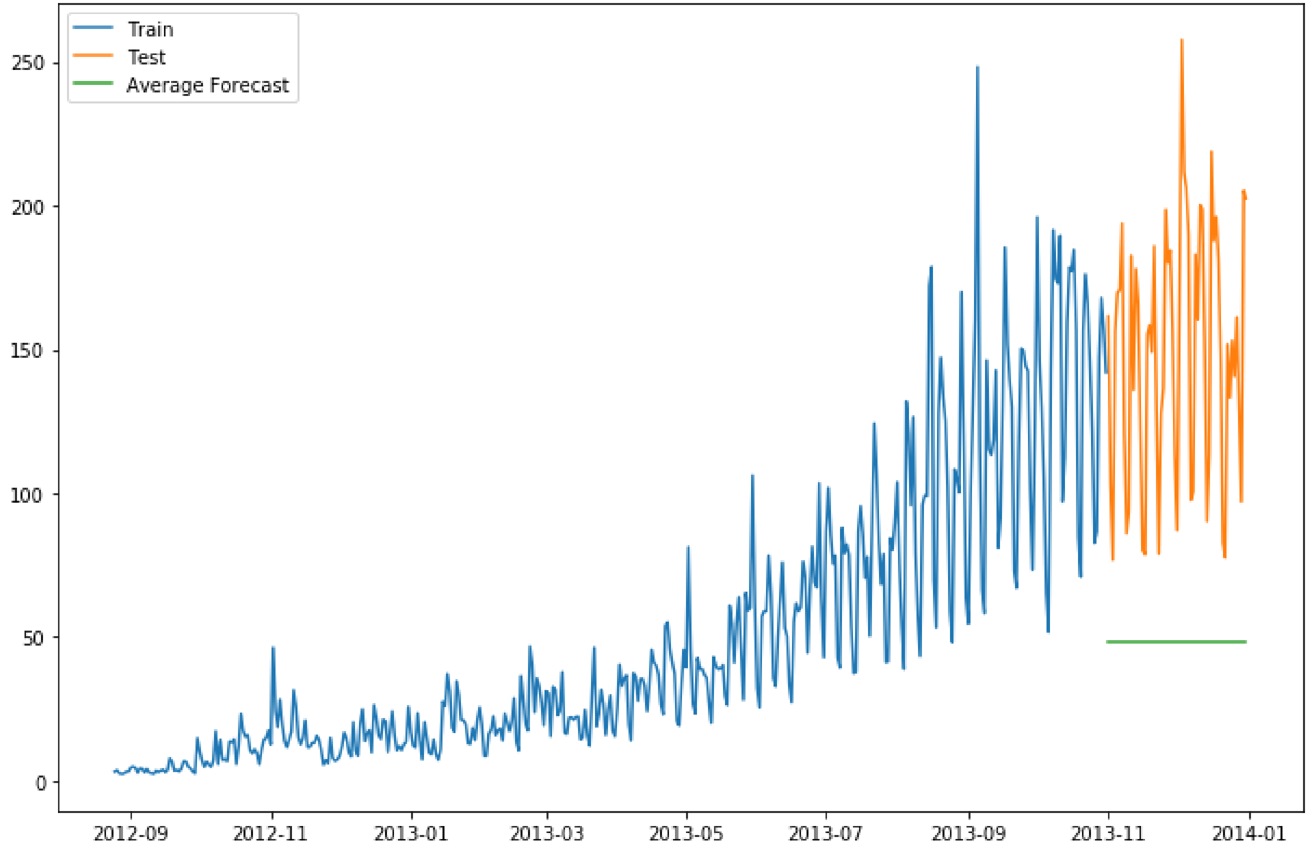
Methods: Naive Approach

- We assume that the next predicted point is identical to the last observed point.
- As a result, we may anticipate a straight horizontal line as the forecast.

1	Period	Actual Sales	Naïve Forecast	% Difference
2	1/1/2012	$11,367.91		0.00%
3	2/1/2012	$12,367.47	$11,367.91	8.79%
4	3/1/2012	$15,433.65	$12,367.47	24.79%
5	4/1/2012	$16,235.66	$15,433.65	5.20%
6	5/1/2012	$23,041.25	$16,235.66	41.92%

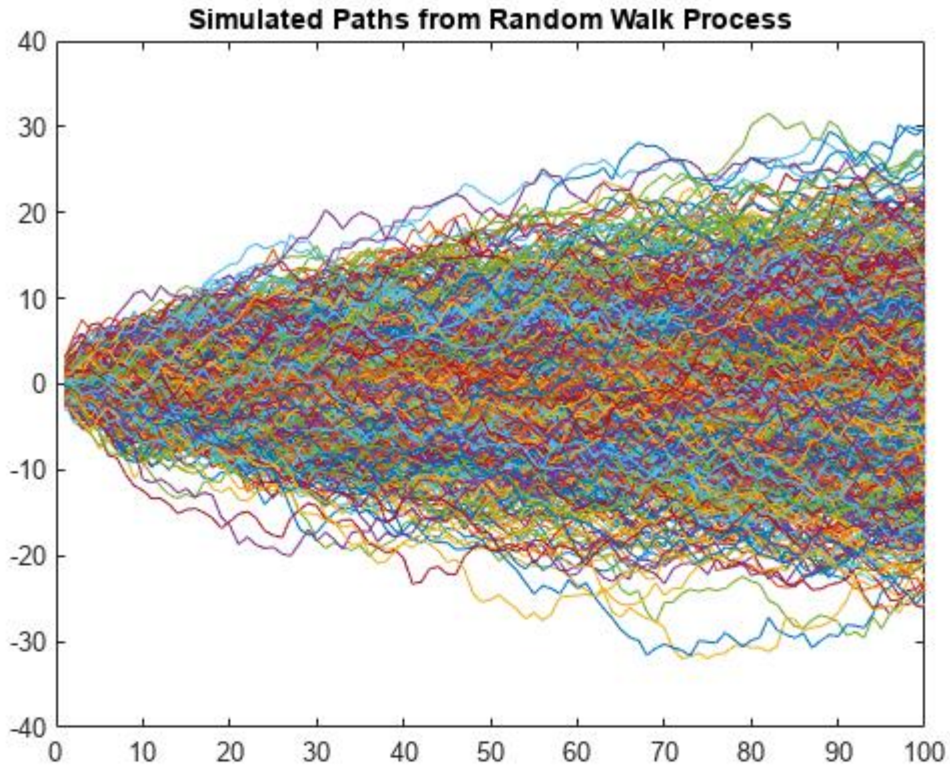
Methods: Simple Average

- The Simple Average approach is a forecasting technique that forecasts the predicted value as the average of all previously observed points. Very sensitive to outliers .



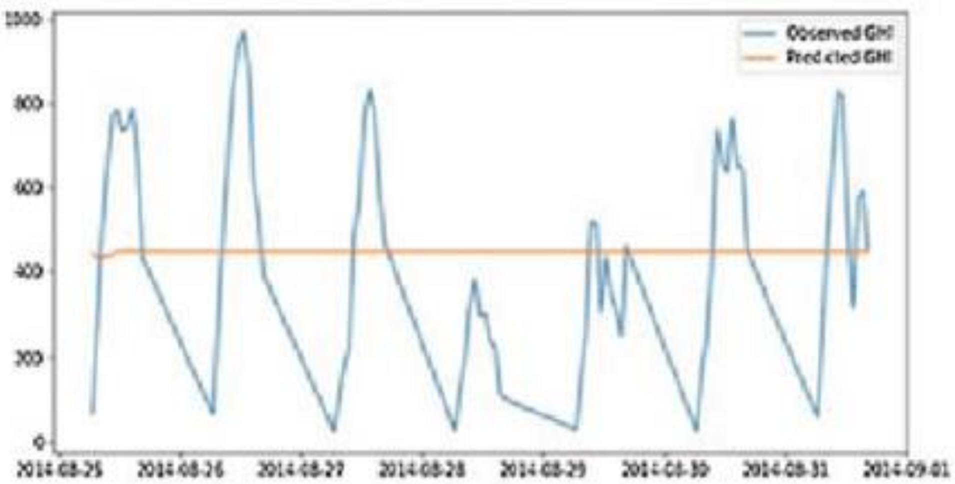
Methods: ARIMA Model

- ARIMA stands for Autoregressive Integrated Moving Average.
- It is a combination of the Autoregressive (AR) and Moving Average (MR) model.
- The AR model forecast corresponds to a linear combination of past values of the variable. The moving average model forecast corresponds to a linear combination of past forecast errors. The “I” represents the data values that are replaced by the difference between their values and the previous values.

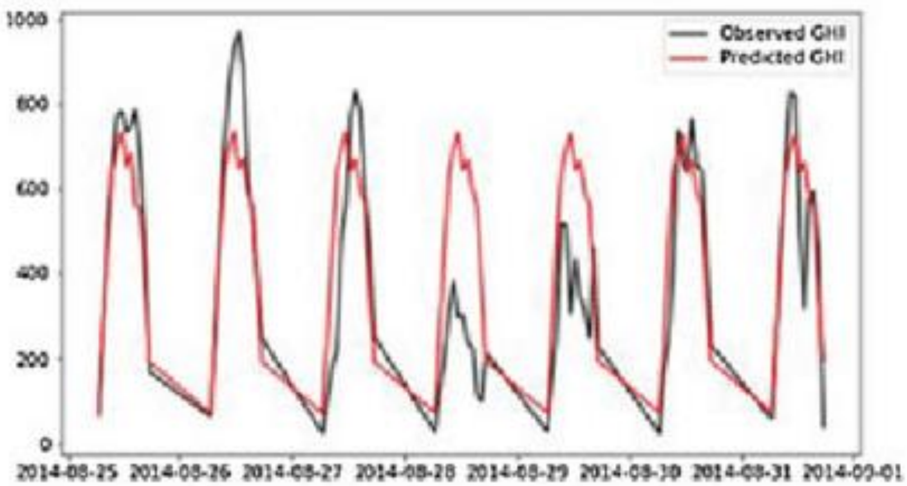


Methods: SARIMA Model

- SARIMA stands for Seasonal Autoregressive Integrated Moving Average.
- It extends the ARIMA model by adding a linear combination of seasonal past values and forecast errors.



ARIMA 7 Day Ahead Prediction



SARIMA 7 Day Ahead Prediction

Methods: Simple Exponential Smoothing

- In this method, we give greater weight to current data than to ones from the distant past.
- As time passes, the weights decline exponentially (the amount that decreases is proportional to the size of the weights), with the smallest weights associated with the oldest data.

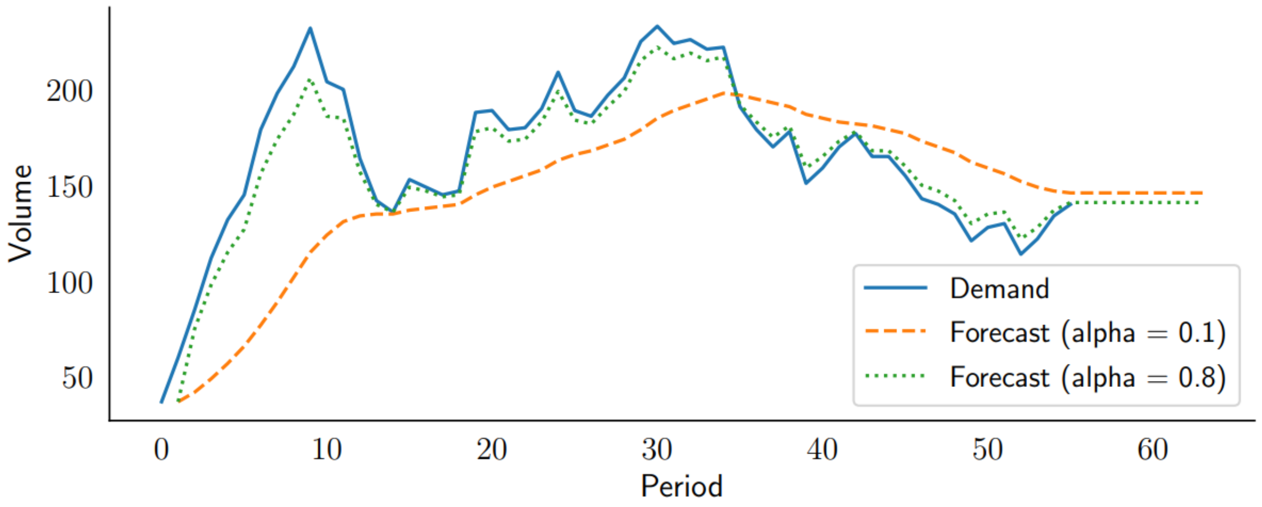
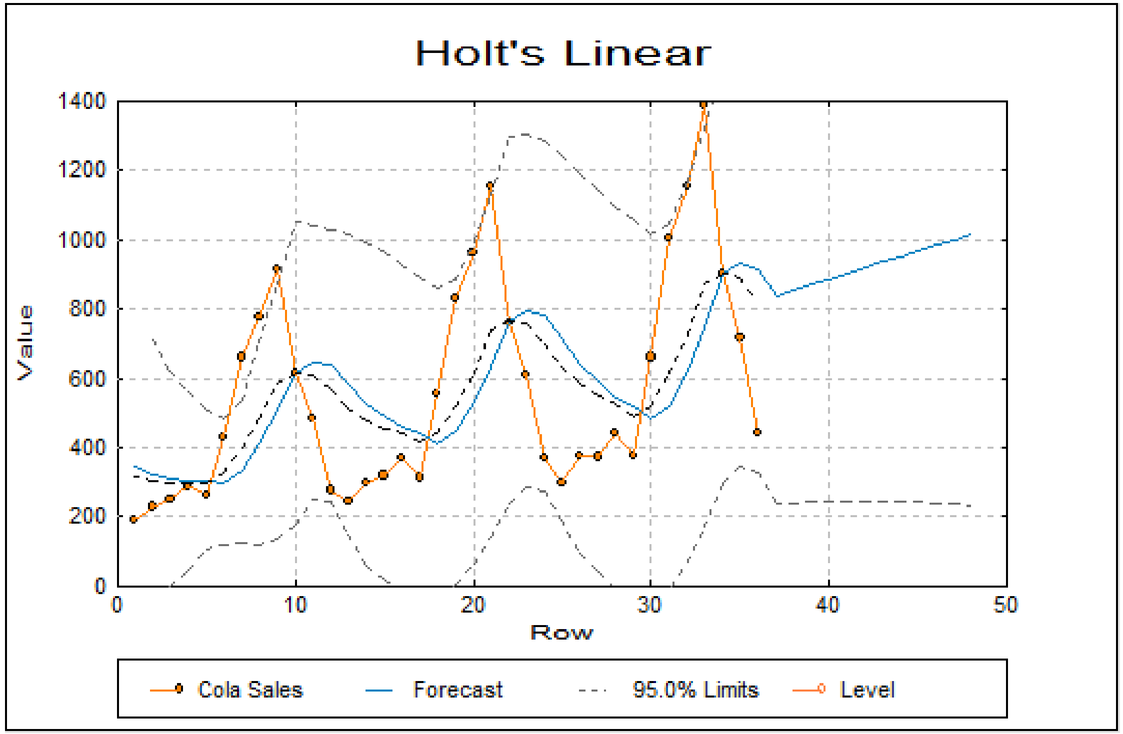


Figure 3.2: Simple smoothing

This strategy is comparable to the naïve approach if we give the whole weight to the last observed value. As a result, the naïve method is likewise a basic exponential smoothing method in which the full weight is assigned to the most recently recorded value.

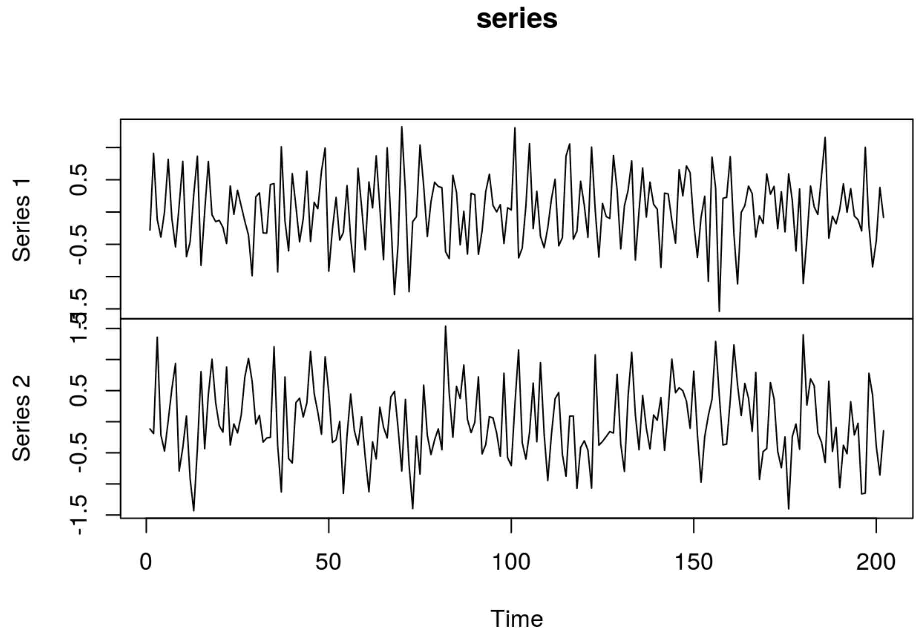
Methods: Holt's Linear Trend Model

- It is an expansion of simple exponential smoothing that allows for trend forecasting of data.
- This approach takes into account the dataset's trend. In this technique, the forecast function is a function of level and trend.



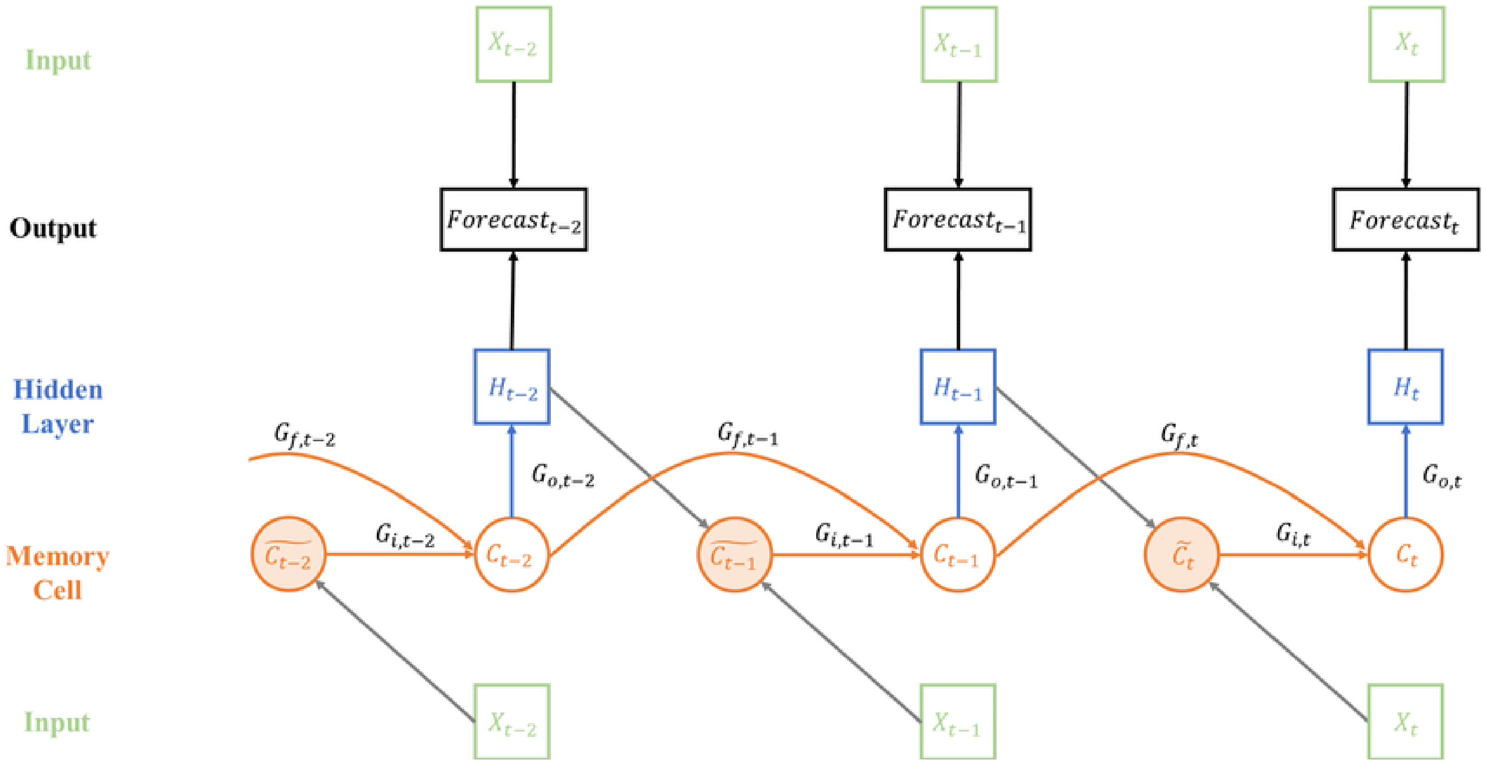
Methods: VAR

- The Vector Autoregression (VAR) method models the next step in each time series using an AR model.
- The VAR model is useful when you are interested in predicting multiple time series variables using a single model.



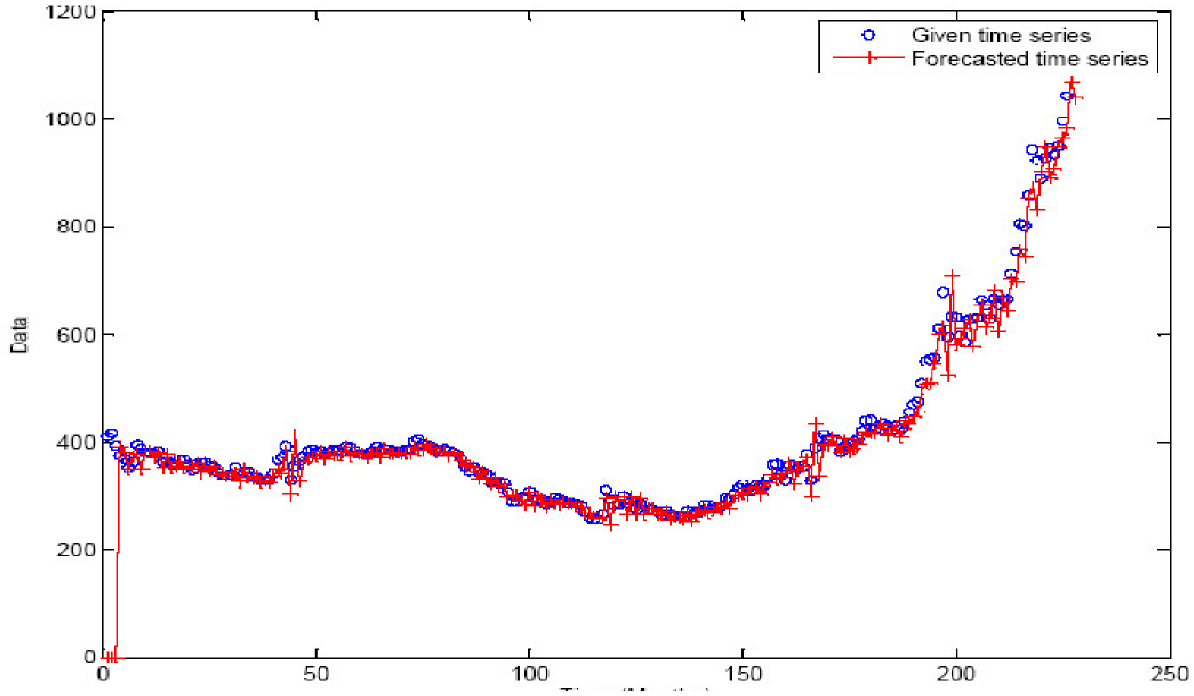
Methods: LSTM

- The Long Short Term Memory network or LSTM is a special kind of recurrent neural network that deals with long-term dependencies.
- It can remember information from past data and is capable of learning order dependence in sequence prediction problems.



Methods: GARCH

- The Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models, most popular time series models used for forecasting conditional volatility.
- These models are conditional heteroskedastic as they take into account the conditional variance in a time series.
- GARCH models are one of the most widely used models for forecasting financial risk measures like VaR and Conditional VaR in financial risk modelling and management.



6. Forecasting: Practical use

Please open the script

7. Forecasting: Practicum

You're going to be working with some retail data from “**Sales.csv**”. The data set contains information retail sales of automotive parts, accessory, and tire stores.

The dataset contains 2 columns:

DATE: monthly data

SALES: sales in Millions of USD,

Seasonally Adjusted

sales ···

	DATE	SALES
0	1/1/1992	3311
1	2/1/1992	3360
2	3/1/1992	3445
3	4/1/1992	3415
4	5/1/1992	3510
...	...	...
369	10/1/2022	10543
370	11/1/2022	10608
371	12/1/2022	10655
372	1/1/2023	10648
373	2/1/2023	10837

374 rows × 2 columns

8. In-class Assignment