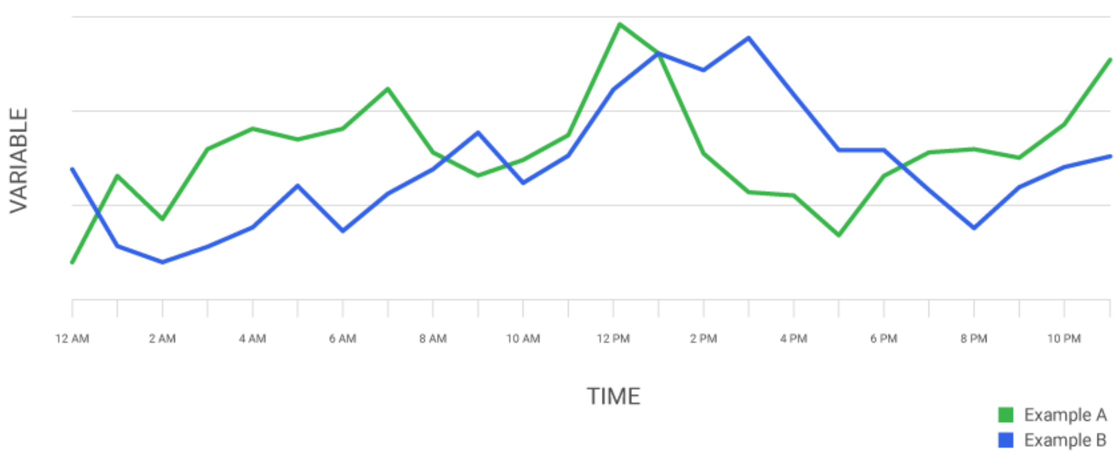
1- Introduction to Time Series Forecasting

By: Eng. Esraa Madhi

Introduction

Time series data is data collected on the same subject at different points in time. This could involve tracking sales figures or monitoring patient health statistics over a period.



There are other types of data as well, such as:
Cross-sectional data: This is data captured at a single moment and can encompass details about customers, products, sales, and more.

In practice, datasets often combine elements of both time series and cross-sectional data. Take, for instance, a scenario where you're evaluating product sales on a specific day; your analysis would be cross-sectional. However, if you shift your focus to the sales trajectory of a single product over the last five years, you're engaging in time-series analysis.

The goals differ when you're working with time-series data versus when you're dealing with cross-sectional data. Nevertheless, it's common for real-world datasets to integrate aspects of both, offering a more comprehensive view.

What is Time Series Forecasting?

Time series forecasting	Traditional machine learning
Time series forecasting models are	Traditional machine learning models

designed to handle data where observations are dependent on time, capturing trends and seasonal patterns for future predictions.	typically assume data points are independent and do not inherently account for time, focusing on finding patterns across features for prediction or classification.
Predicting tomorrow's stock price based on the past 60 days of prices.	Predicting whether an email is spam or not based on its content and sender.
Prediction model: • Input: time interval • Output: target variable	Prediction model: • Input: features • Output: target variable

---

Methods for Forecasting Time Series Data:

1. Classical / Statistical Models: Moving Averages, Exponential Smoothing, ARIMA or SARIMA
2. Machine Learning: Linear Regression, XGBoost or Random Forest.
3. Deep Learning: RNN, LSTM

This lesson is focused on the first two methods: classical/statistical models and machine learning. Deep learning methods are out-of-scope for this tutorial.

Components of Time Series

1. **Trend:** is a pattern in data that shows the movement of a series to relatively higher or lower values over a long period of time. In other words, a trend is observed when there is an increasing or decreasing slope in the time series.

There are different types of trends, such as:

- **Uptrend**: When the data from a time series analysis generally moves upward, it's considered to be in an uptrend.
- **Downtrend**: Conversely, if the analysis reveals a general downward movement in the data, this is known as a downtrend.
- **Horizontal or Stationary Trend**: When the time series data does not display any clear upward or downward pattern, it's described as horizontal or stationary.

To see the complexity behind linear visualization we can decompose the data. The function called **seasonal_decompose** within the **statsmodels** package can help us to decompose the data into its components/show patterns — trend, seasonality and residual components of time series

Note:

- Decomposition is a statistical job that involves breaking down Time Series data into many components or identifying seasonality and trend from a series of data.
- seasonal_decompose function uses **moving averages** method to estimate the trend.

2. **Seasonality**: The pattern that repeats over regular intervals due to seasonal factors, such as monthly or quarterly fluctuations influenced by weather, holidays, or other cyclical events.

   ex: Retail sales often show an increase during the holiday season every year, typically starting from Black Friday through to the end-of-year holidays.

3. **Cyclical:** These are fluctuations that occur over non-fixed periods, often tied to economic or business cycles, which can last longer than a year and are not necessarily seasonal.

   ex: Over a period of several years, the real estate market might experience a boom with rising property values, followed by a bust where prices fall or stagnate. This cycle is not tied to a specific season or annual pattern; instead, it is related to longer-term economic conditions, interest rates, and consumer confidence.

4. **Irregular (or Random):** These are unpredictable, sporadic fluctuations that don't follow a pattern or cycle. This component includes random or one-off events, often referred to as "noise" in the data.

   ex: A sudden and unexpected event, such as a natural disaster, may cause a spike in insurance claims. This irregularity is not part of a pattern or cycle and is typically unpredictable.

In time series analysis, these components can be modeled separately to understand underlying patterns and to make forecasts. Some models, like the additive model, assume that these components are added together to make up the time series. Others, like the multiplicative model, assume that these components are multiplied together. Identifying and understanding these components is crucial for accurate analysis and forecasting in time series data.

Simple Moving Average (SMA)
A simple moving average tells us the unweighted mean of the previous K data points(lags) . The more the value of K the more smooth is the curve, but increasing K decreases accuracy. If the data points are p_1, p_2, . . . , p_n then we calculate the simple moving average.

Note: Lagging a time series means to shift its values forward one or more time steps, or equivalently, to shift the times in its index backward one or more steps.

The MA method is based on the idea that the current value of a time series is a function of the average of the values of previous periods.

Raw data is often noisy, and identifying meaningful patterns or trends requires smoothing the data. MA helps smooth the time series by averaging out the effects of noise, which makes it easier to identify and forecast underlying patterns.

Advantages:

1. Simple to implement: MA is a straightforward method that requires only a few lines of code to implement.
2. Good for short-term forecasting: MA is suitable for **short-term** forecasting, such as predicting sales for the next few weeks or months.
3. Smoothes out the noise: MA can help remove noise from the time series data, making it easier to identify underlying patterns or trends.

Disadvantages:

1. Ignores long-term trends: MA is not suitable for long-term forecasting as it only considers past values and does not take into account other factors that may affect the time series in the long run.
2. May require tuning: The window size and the type of MA used may need to be adjusted to achieve the best performance.

Exponential Smoothing

Exponential smoothing of time series data assigns exponentially decreasing weights for newest to oldest observations. The older the data, the less weight the data is given, whereas newer data is given more weight

---

Machine Learning:

Machine learning is an alternative way of modeling time-series data for forecasting. In this method, we extract features from the date to add to our "X variable" and the value of the time-series is "y variable"

Resources:

- https://pub.towardsai.net/10-top-time-series-courses-to-proficient-this-important-data-science-skills-565212d29a0
- https://levelup.gitconnected.com/13-guided-time-series-projects-to-build-your-portfolio-491d959f62af
- https://pub.towardsai.net/5-free-practical-kaggle-notebook-to-get-started-with-time-series-analysis-7d73ba0b1d07
- https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b
- https://www.kaggle.com/learn/time-series
- https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/
-