Business Forecasting

ADIA Course

Day 1 – Session 1

Dr. Tanujit Chakraborty

Sorbonne University, Abu Dhabi, UAE

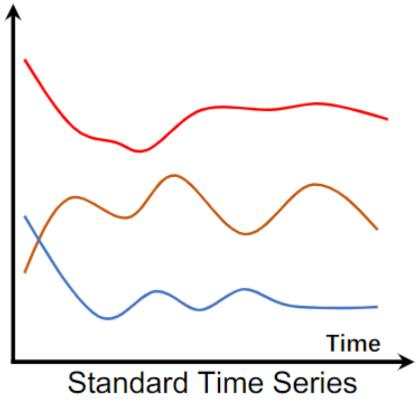
Sorbonne Centre for Artificial Intelligence, Paris, France

Research Areas: Time Series Forecasting, Machine Learning, Econometrics, Health Data Science

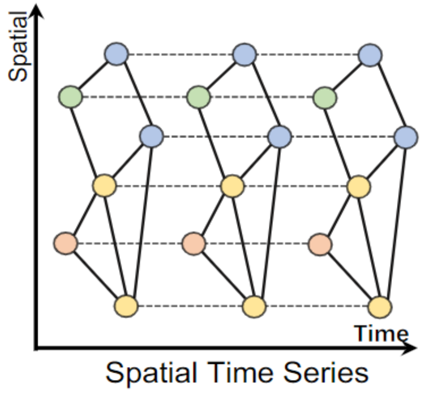
June 30, 2025

Space, Time, and Forecasting

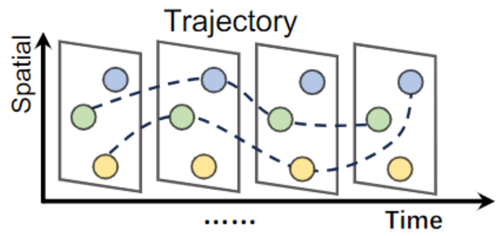
Time series is a set of observations, each one being recorded at a specific time, e.g., Annual GDP of UAE, iPhone Sales figure, etc.



A spatial time series is a type of data that combines spatial (cross-sectional) and temporal dimensions, capturing how measurements or observations vary across different locations over time, e.g., Air quality of Abu Dhabi.



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



Time Series is omnipresent

- Time series data is a specialized form of data that plays a vital role in various fields, including economics, finance, climate science, healthcare, and many others.

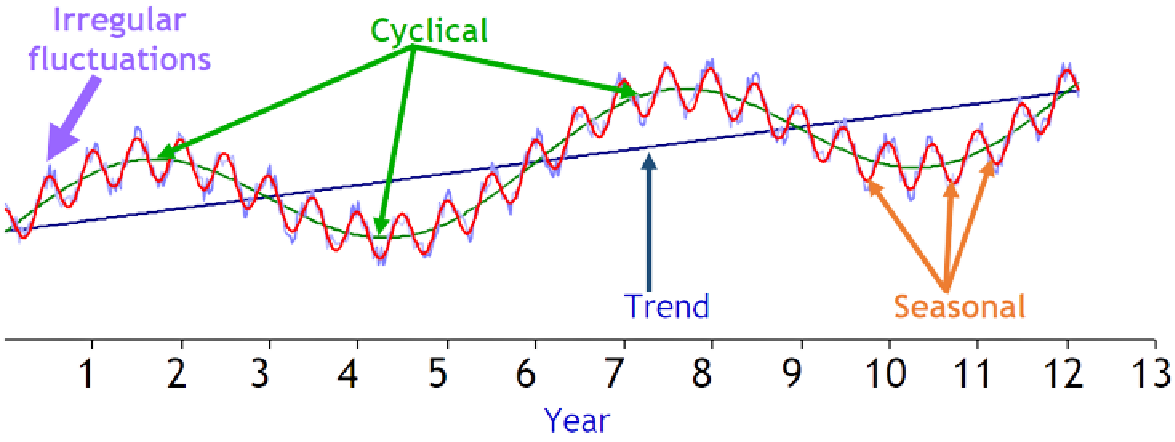
- A forecast is a scientifically justified assertion about the possible states of an object in the future

Time Series is everywhere

- Epidemiology: Epidemic/Flu/Covid-19 cases observed over some time period.
- Economics: Stock prices, unemployment rates, inflation rates, etc.
- Earth and Environmental Sciences: Daily Sea Surface Temperature, Southern Oscillation Index, Seismic Waves, Air Quality Index, Global Warming, etc.
- Astronomy: Sunspot numbers, Luminosity of stars, etc.
- Demography: Population series, Birthrates, Mortality rates, etc.
- Medical Science: Blood pressure, Blood oxygen level, Sugar level, etc.
- Business: Product demand, Sales, Market share, etc.

Time Series Components

- Trend (T_t) : pattern exists when there is a long-term increase or decrease in the data.
- Seasonal (S_t) : pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).
- Cyclic (C_t) : pattern exists when data exhibit rises and falls that are not of fixed period (duration usually of at least 2 years).



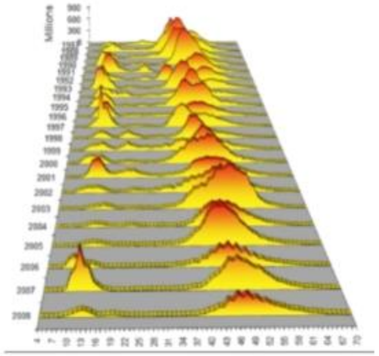
Shape of Time Series Data

- We usually think that the data is one-dimensional.
- It only consists of the time and the data associated with it (Temp.).

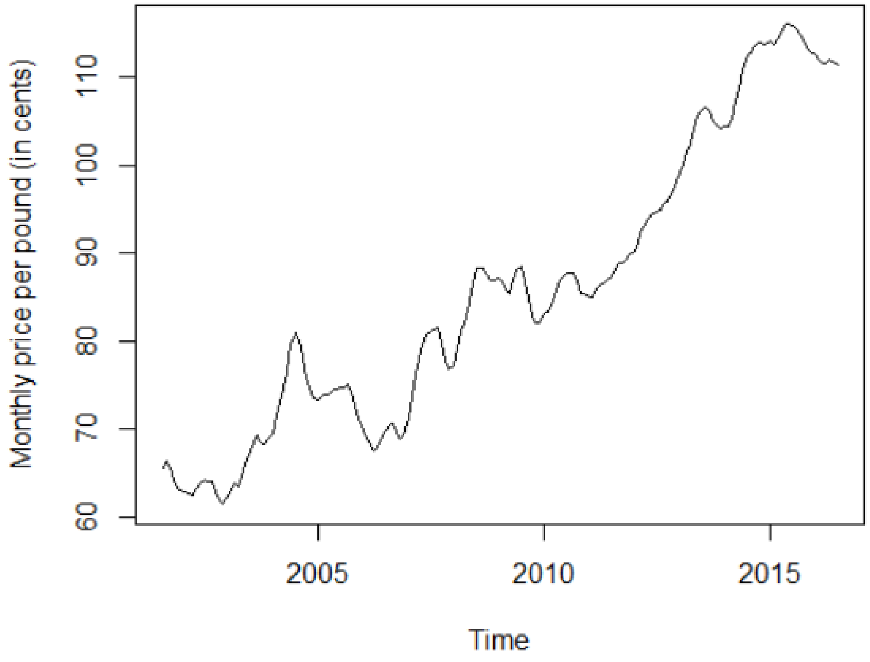
	New York City
1990-01-01	1
1990-01-02	5
1990-01-03	9
1990-01-04	13

- But it can be Multidimensional.

	New York City	London	Tokyo	Paris
1990-01-01	1	2	3	4
1990-01-02	5	6	7	8
1990-01-03	9	10	11	12
1990-01-04	13	14	15	16



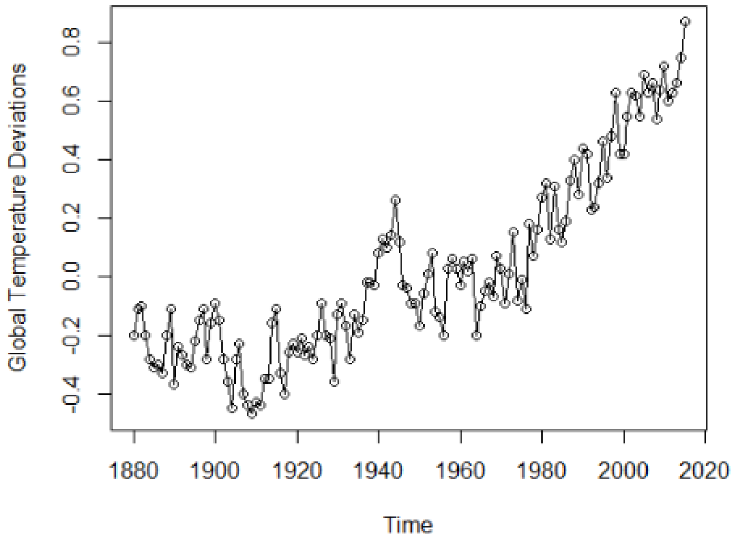
Examples of Time Series data

<table>
<tr><th>Time series data plot</th><th>Observations</th></tr>
<tr><td>
</td><td>

- Chicken price in the US (Economics)
- Response variable is continuous.
- Overall, the price is increasing (upward trend). The trend is more or less linear.
- Possibly no significant periodic fluctuation, i.e., possibly no seasonal or cyclical components.
- Removing the trend may lead to a series that has a more or less constant mean.
- It may be cross-correlated with the Inflation time series.

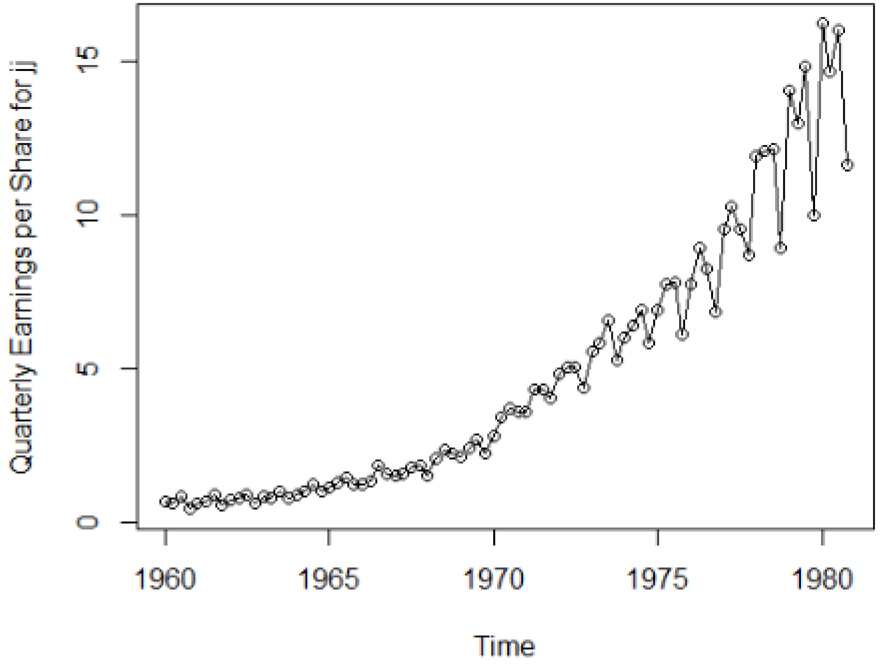
</td></tr>
</table>

Examples of Time Series data



- Global temperature anomaly (avg of 1951-80 as base) (Earth Sciences)
- Response variable is continuous.
- There is no trend until around 1935. Then there is a sharp increase which levels off around 1970.
- After 1970, global temperature is clearly increasing until now, i.e., there is an upward trend. The trend is more or less linear.
- Although there is always some local fluctuation, the importance is more on the upward trend.
- Removing the trend may lead to a series that has a more or less constant mean.
- Glacial sediment data or the data on particulate matter over the past 100 years may be used as proxies.

Examples of Time Series data

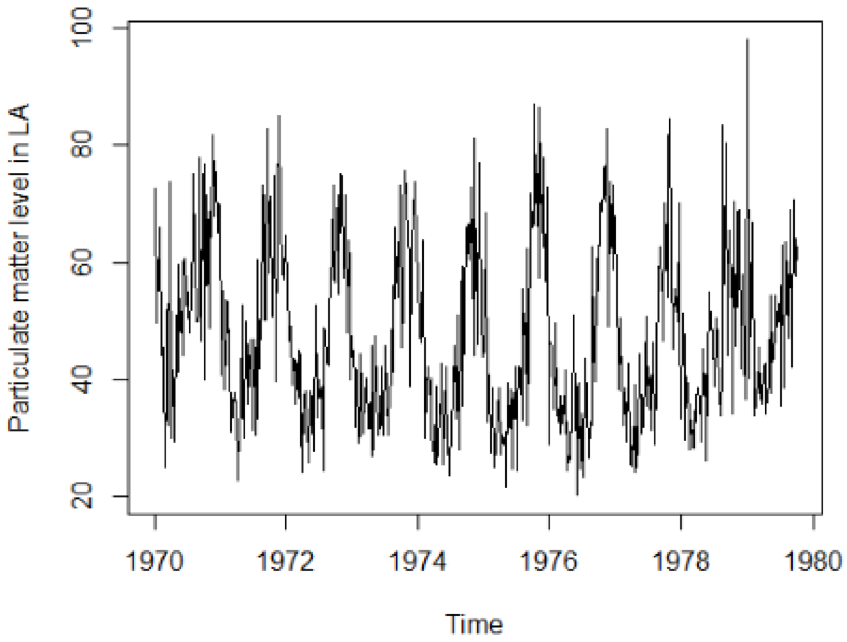
<table>
<tr><th>Time series data plot</th><th>Observations</th></tr>
<tr><td>


</td><td>

- J&J share price (Economics)
- Response variable is continuous.
- There is an increasing quadratic trend.
- There may be periodic fluctuations (with a period of 1 year – seasonal fluctuations) towards the end.
- Removing the trend and seasonality may lead to a series that has a more or less constant mean.

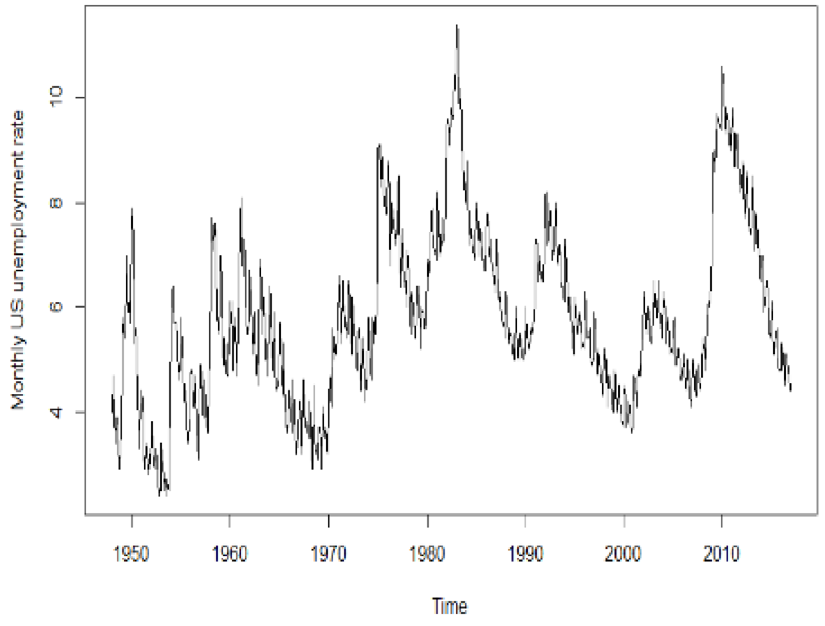
</td></tr>
</table>

Examples of Time Series data



- Particulate matter (say PM-10) in LA (Environmental Science)
- Response variable is continuous.
- The series is mostly periodic, with a possible period of 1 year, i.e., there are seasonal fluctuations.
- The level is at highest during the winter and lowest during the summer. Hence, it is cross-correlated with the temperature.
- Sine-cosine curves may be useful in explaining the structure.
- One may check the usage of Fast Fourier/Wavelet transformation on this data before modeling.

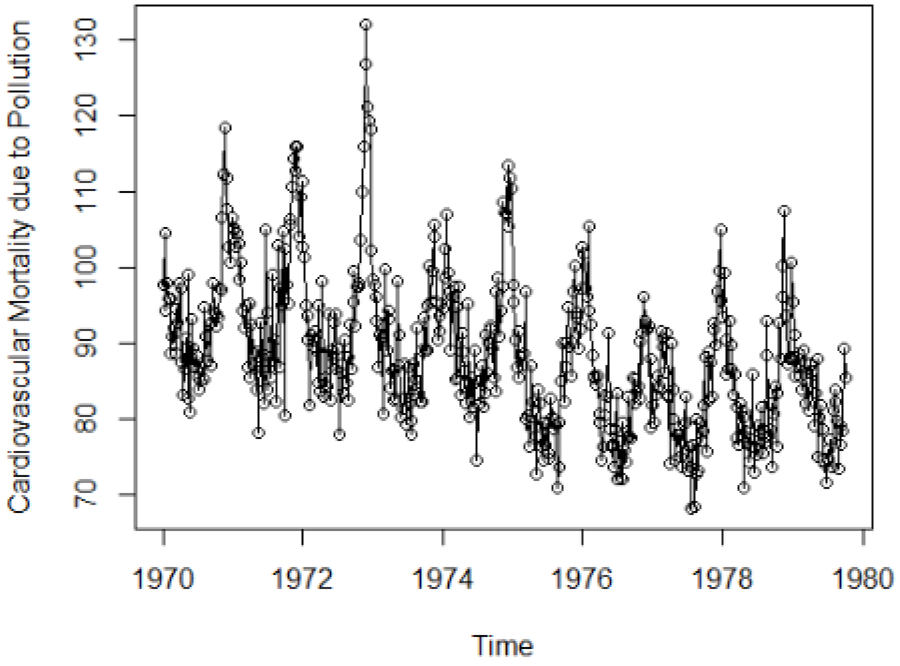
Examples of Time Series data

<table>
<tr><th>Time series data plot</th><th>Observations</th></tr>
<tr><td>
</td><td>

- Unemployment Rate in US (Economics)
- Response variable is continuous.
- Mostly random fluctuations except during 1982 (monetary policy for reducing inflation) and 2008 (Global Financial Crisis).
- Models with only random factors may be ideal.
- To explore the 1982/2008 case, need to consider the series only around those periods. For those, the previous remark may not be true.
- Recent series will show an increase during the waves of Covid-19.

</td></tr>
</table>

Examples of Time Series data

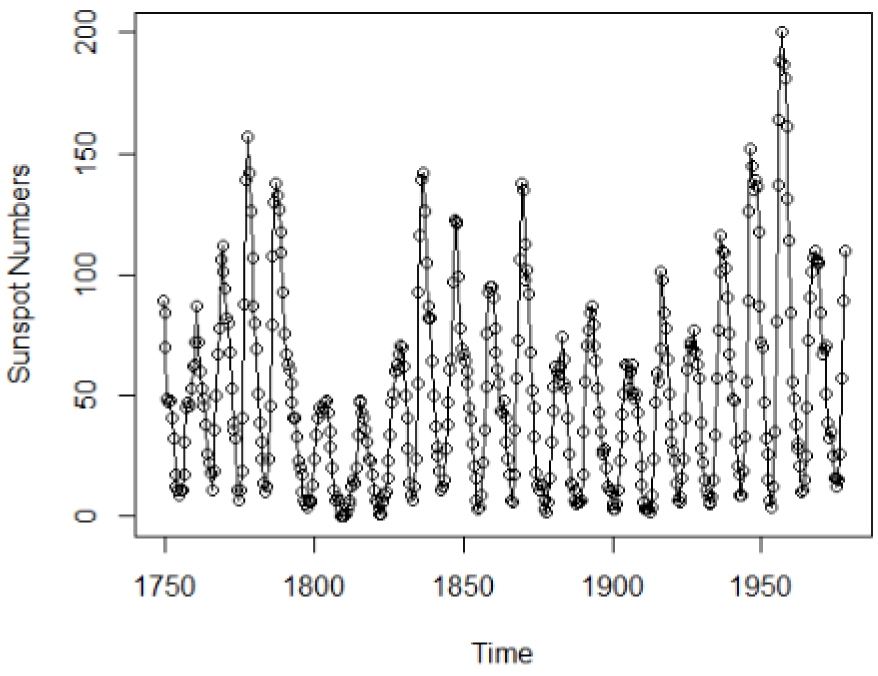


- Cardiovascular mortality in LA (Demography)
- Response variable is continuous.
- The series is mostly periodic, but there is a gradually decreasing trend.
- Mostly explainable by the previous series. A gradual decreasing trend may be due to better medications.
- One may check the usage of Fast Fourier/Wavelet transformation on this data before modeling.

Examples of Time Series data

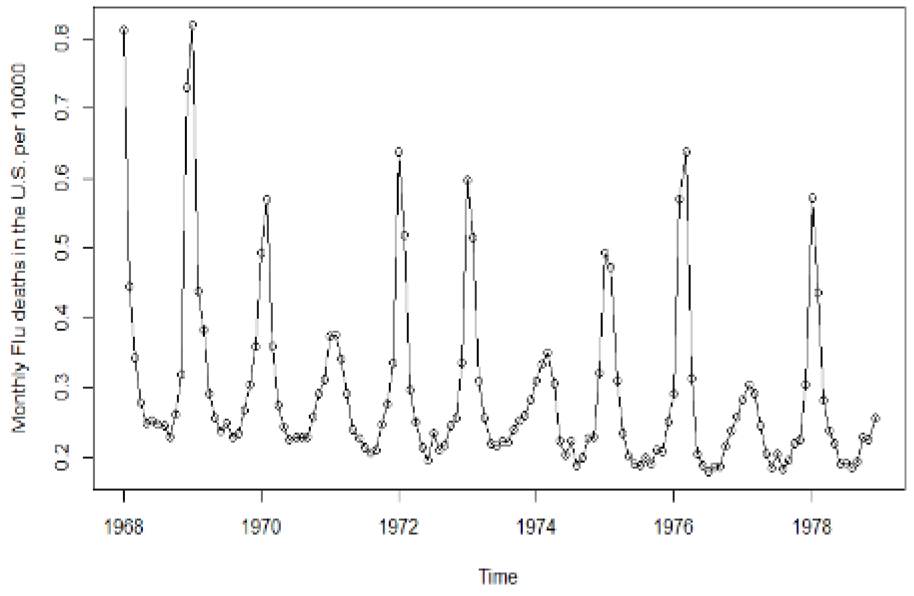
Time series data plot	Observations
	• Sensex for last 5 years from Google Finance website (Economics) • Response variable is continuous. • Trend is linear here. Hence, a snapshot of a series may depict different patterns. • The dip is due to the impact of Covid. To understand this dip, we should consider zooming around the dip.

Examples of Time Series data



- Sunspot numbers (Astronomy)
- Response variable is discrete.
- The series is mostly periodic with a possible period of 11 years (Solar cycle), i.e., there are cyclical fluctuations.
- There is very little randomness. Fluctuations are mostly systematic.
- Sine-cosine curves may be useful in explaining the basic structure.

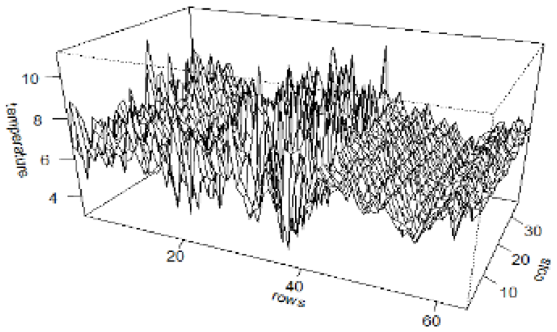
Examples of Time Series data

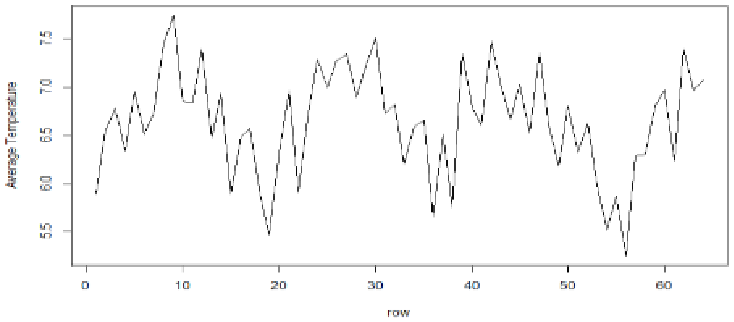


- Flu deaths in US (Epidemiology)
- Response variable is discrete.
- The series is mostly periodic with a possible period of 1 year, i.e., there are seasonal fluctuations.
- The level is at highest during the winter and lowest during the summer. Hence, it is cross-correlated with the temperature, which is natural.
- Sine-cosine curves may be useful in explaining the structure.

Examples of Spatial Time Series data

<table>
<tr><th>Spatial-time series data plot</th><th>Observations</th></tr>
<tr><td>





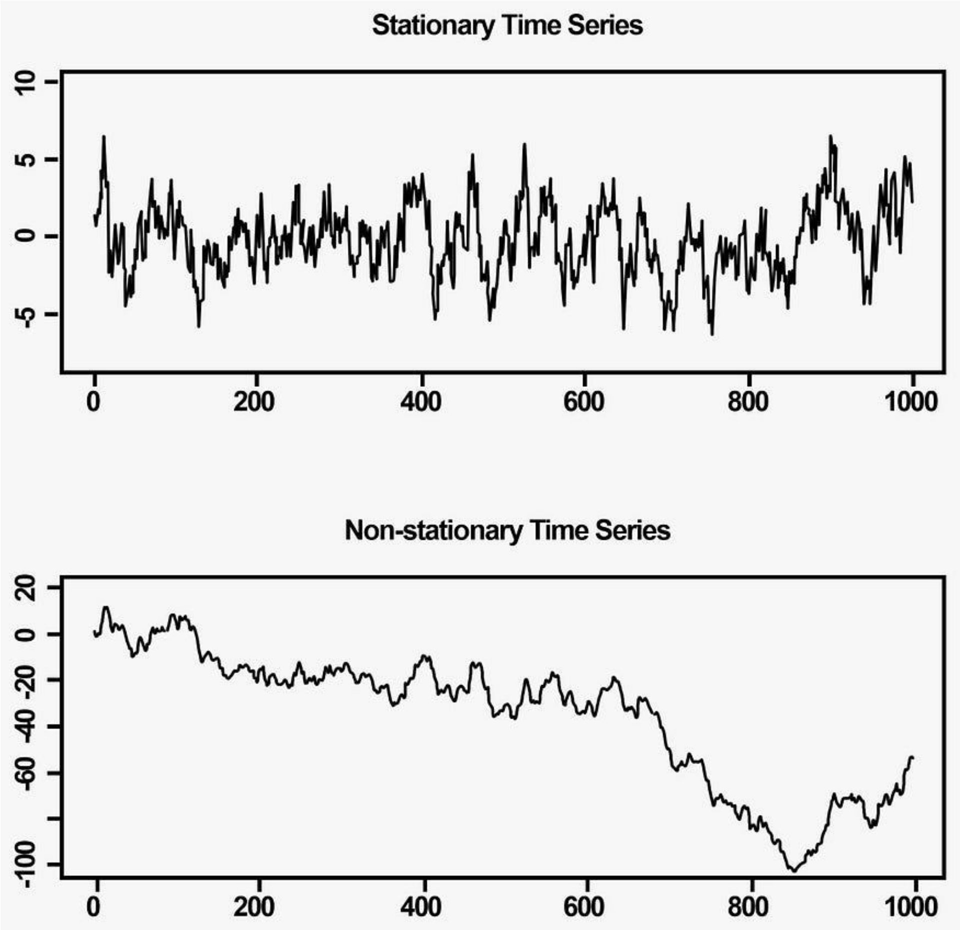
</td><td>

- Soil Temperature in a field (Earth Science)
- Spatial data and hence the index is a two-dimensional grid.
- Response variable is continuous.
- The second series is the average of the first series over columns.
- There is a seasonal pattern.
- Removing seasonality will result in a random series.
- Sine-cosine curves may be useful to explain the systemic variation.

</td></tr>
</table>

Time Series Decomposition

- Decomposition : $Y_t = f\,(T_t; S_t; C_t; I_t)$, where Y_t is data at period t and I_t is irregular component at period t.
- Additive decomposition: : $Y_t = T_t + S_t + C_t + I_t$
- Multiplicative decomposition: $Y_t = T_t * S_t * C_t * I_t$
- A stationary series is roughly horizontal, constant variance and no patterns predictable in the long-term.



Stationary Time Series

Stationary Series:

- A series free from trend and seasonal patterns
- A series exhibits only random fluctuations around mean
- A stationary time series exhibits similar statistical behavior in time and this is often characterized by a constant probability distribution in time

Unit root test Augmented Dickey Fuller Test (ADF) :

- Checks whether any specific patterns exists in the series H0: data is non-stationary
- H1: data is stationary
- A small p-value suggest data is stationary

Steps to follow for handling Time series data

When we have time series data, these are the following steps to draw inferences about the process that generates the observed series:

- **Step 1:** Plot the data.
- **Step 2:**Identify if there is any trend and/or periodic fluctuations (seasonal or cyclical).
- **Step 3:** Remove the trend and periodic components and get the residuals that are stationary.
- **Step 4:** Model the stationary residuals using some time series models (e.g., AR, MA, ARMA, etc.). Use these models to forecast the residuals.
- **Step 5:** Then add the trend and periodic components to get an appropriate model and to perform forecasting.

What people think I forecast?

When I go to any university, and I tell people that my job is time series forecasting and machine learning, usually one of two things happens:

- ...like, weather forecasting?
  - Lots of domain knowledge and specialized models exist
  - We leave it to meteorologists
- ...so, can you predict the stock market and we all get rich?
  - I'll tell you how, and we're all going to be rich!
  - Try it on your own risk!

What I forecast?

- Epidemic time series (e.g., dengue, malaria, hepatitis, etc.)
- Sales forecasting in the supply chain, retail at pharmacy companies
- Forecasting in climate
  - Air quality
  - El Nino
  - Seismic events
- Key macroeconomic variables (inflation, unemployment, exchange rate, etc.)
- ...

Can these be forecasted?

1. daily electricity demand in 3 days' time
2. Google stock price tomorrow
3. Google stock price in 6 months' time
4. maximum temperature tomorrow
5. total sales of drugs in pharmacies next month

Something is easy to forecast if:

1. we have a good understanding of the factors that contribute to it
2. there is a lot of data available
3. the future is somewhat similar to the past
   - ID assumption: samples are identically distributed
4. the forecasts cannot affect the thing we are trying to forecast.
   - self-fulfilling prophecies (election polls)
   - controlled systems
   - Big bull effect in stock markets / bitcoin prices

Past of Forecasting

- In ancient Babylon, forecasters would foretell the future based on the distribution of maggots in a rotten sheep's liver.
- Beginning in 800 BC, a priestess known as the Oracle of Delphi would answer questions about the future at the Temple of Apollo on Greece's Mount Parnassus.

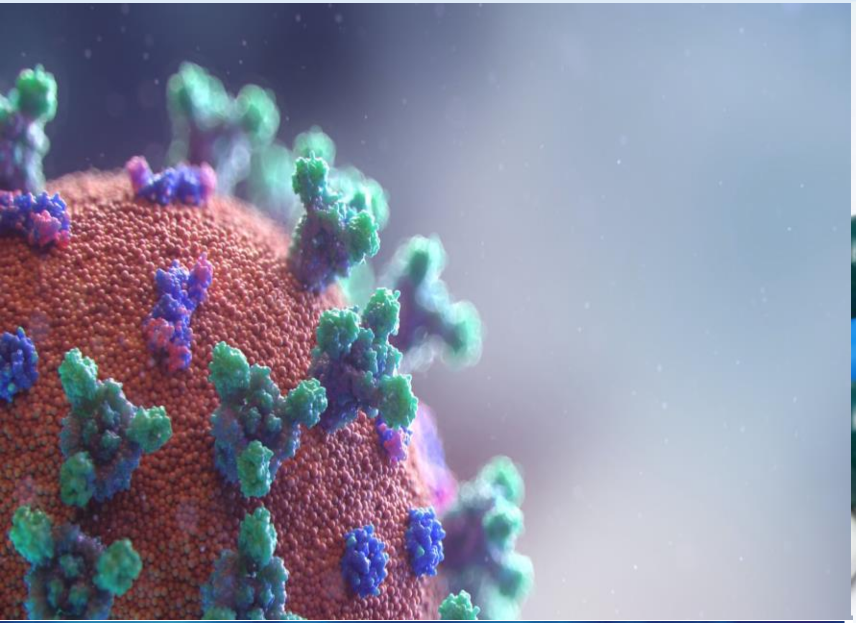
Forecasters are to blame!

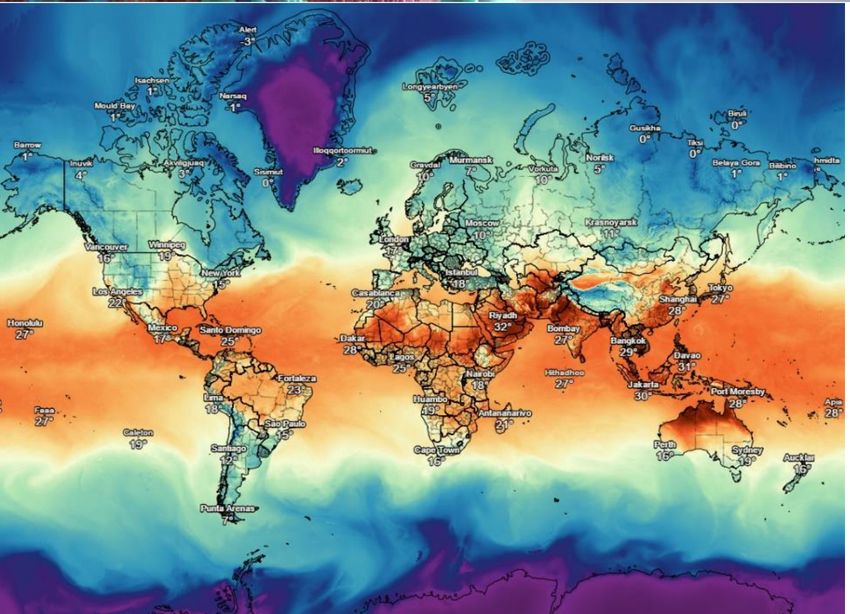
- Forecasters had a tougher time under the emperor Constantius, who issued a decree in AD357 forbidding anyone “to consult a soothsayer, a mathematician, or a forecaster -- May curiosity to foretell the future be silenced forever.”

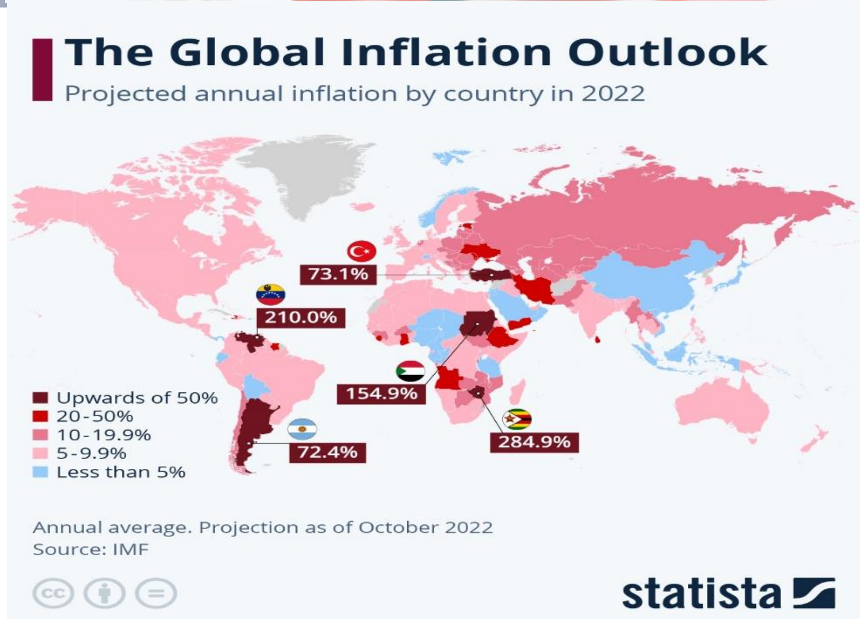
- News report on 16 August 2006: A Russian woman is suing weather forecasters for wrecking her holiday. A court in Uljanovsk heard that Alyona Gabitova had been promised 28 degrees and sunshine when she planned a camping trip to a local nature reserve, newspaper Nowyje Iswestija said.

- But it did nothing but pour with rain the whole time, leaving her with a cold. Gabitova has asked the court to order the weather service to pay the cost of her travel.

What can we forecast?



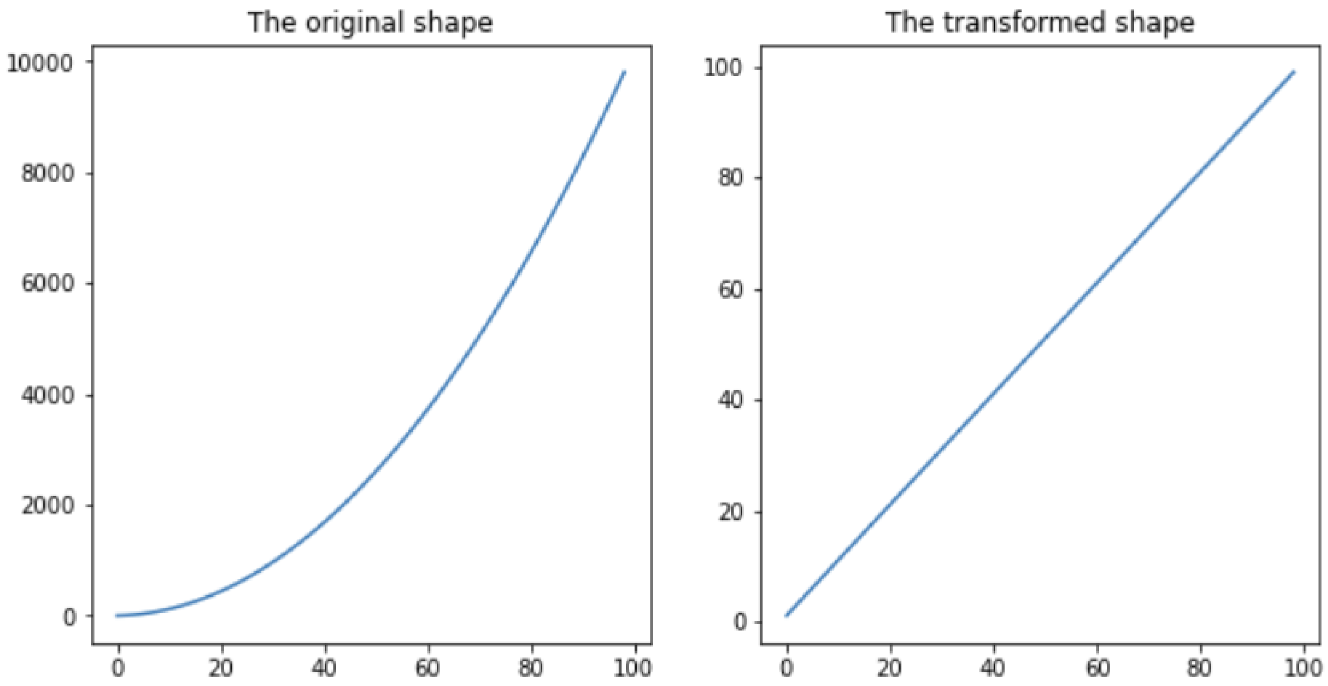




Transformation in Time Series

Power Transformation

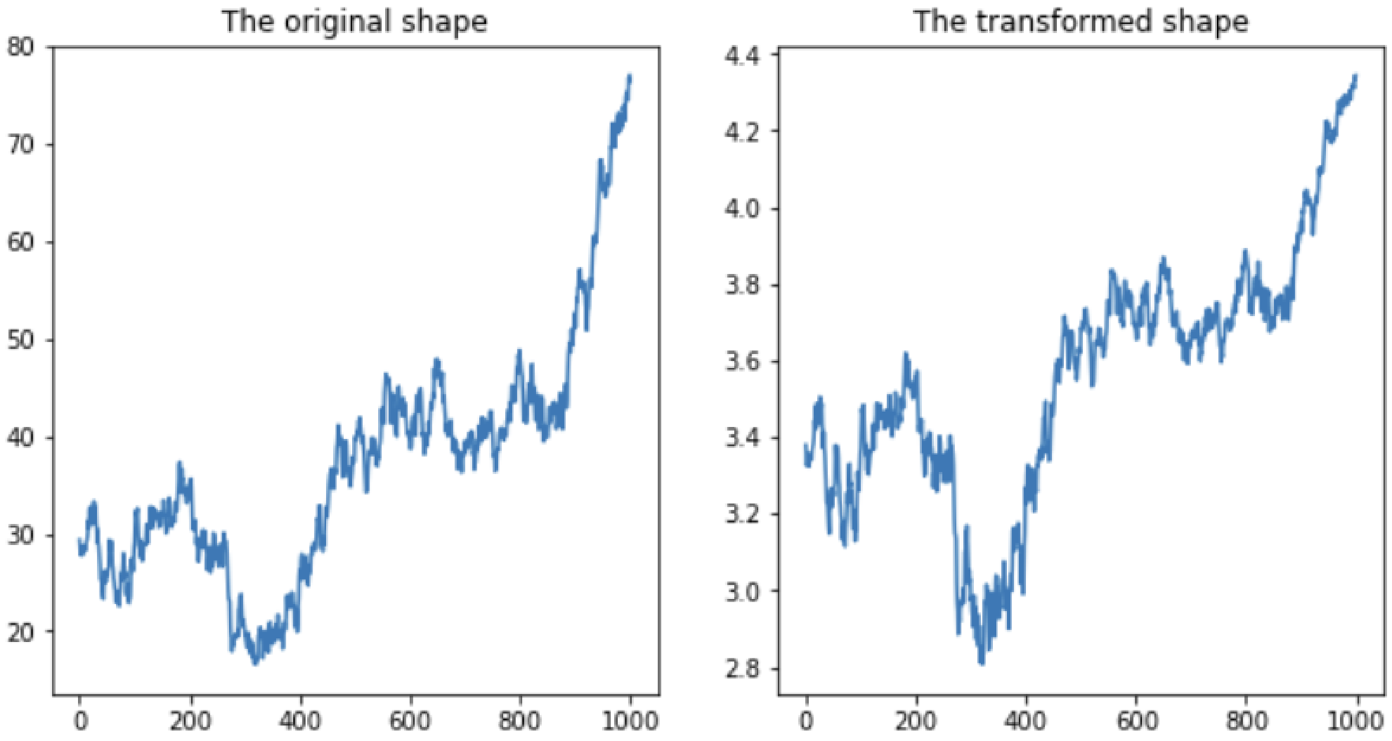
- This means, we will take the power of **some** value for each values on the time-series. Here, we will **need** to understand the nature of the data first before applying any transformation.



- We have done this because **some model returns better results when the "form is linear"**. Thus, that time we want to transform them in to some linear form. And for that you need to understand the "nature" of the data.

Log Transformation

- This is the fundamental transformation of all.
- This can be used to "squash" the data in the smaller range.
- This can be the **default** transformation too!
- It has the frequent application in *finance*.



Box-cox Transformation

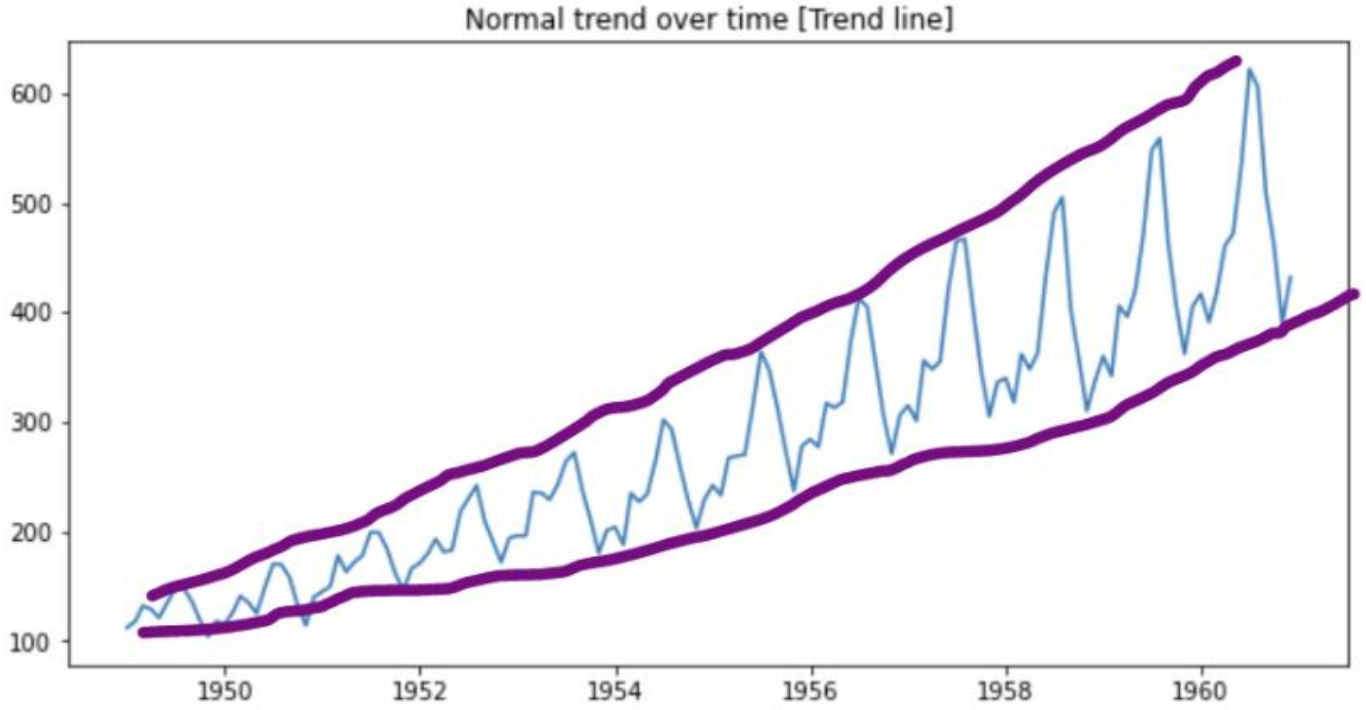
- This is the **generalization** of the 'Power' and 'Log' transforamtion.
- It is used to achieve **'normality'** in the *non-normal* variables.
- Unifies the power and log transforms.
- λ is chosen automatically by the *boxcox*( ) function in *Scipy*
- "Estimating Box-Cox power transformation parameter via goodness of fit tests".

$$y'(t) = \frac{y(t)^{\lambda} - 1}{\lambda} \qquad if\ \lambda \neq 0$$

$$y'(t) = \log y(t) \qquad if\ \lambda = 0$$

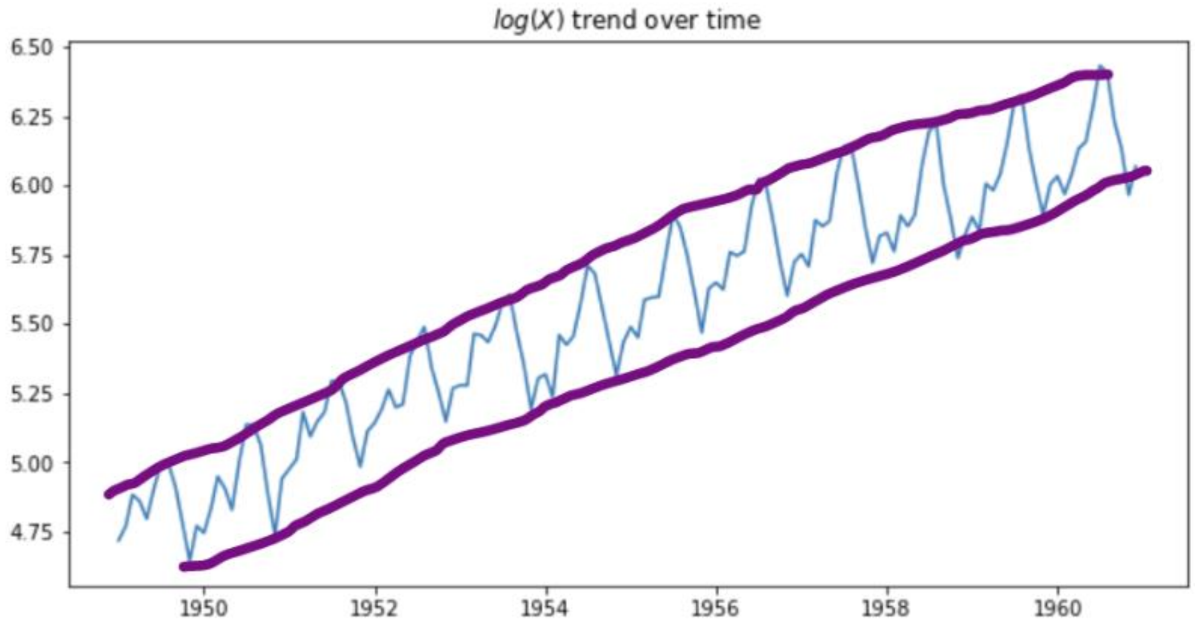
Since, $\lim_{\lambda \to 0} \frac{\lambda^{\lambda} - 1}{\lambda} = \ln x$

Air Passengers data



- Here, we can see that the trend is **increasing** and **too** the magnitude.
- There is the **seasonality**: On each year of some specific months the amount rises and then dips
- The amount is increasing i.e., the seasonality is same but the magnitude is increasing over time: *which the model has to consider*.

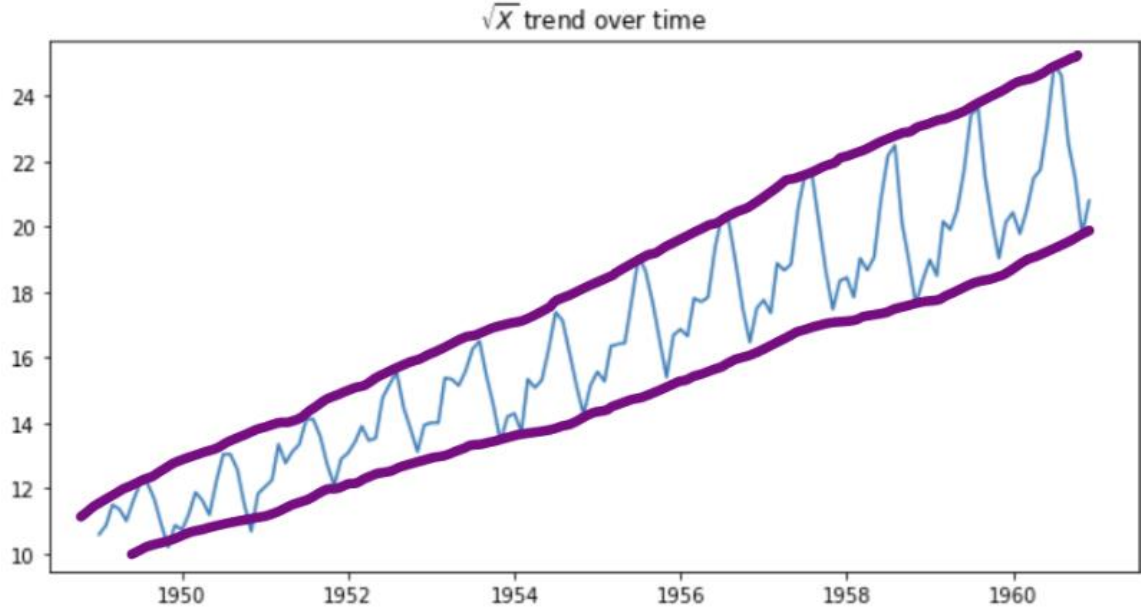
Transformation on Air Passengers data



It is more **even on the both sides?** like, the magnitude is not changing much as it did in the normal data.

Other Transformation: Fourier, Wavelet

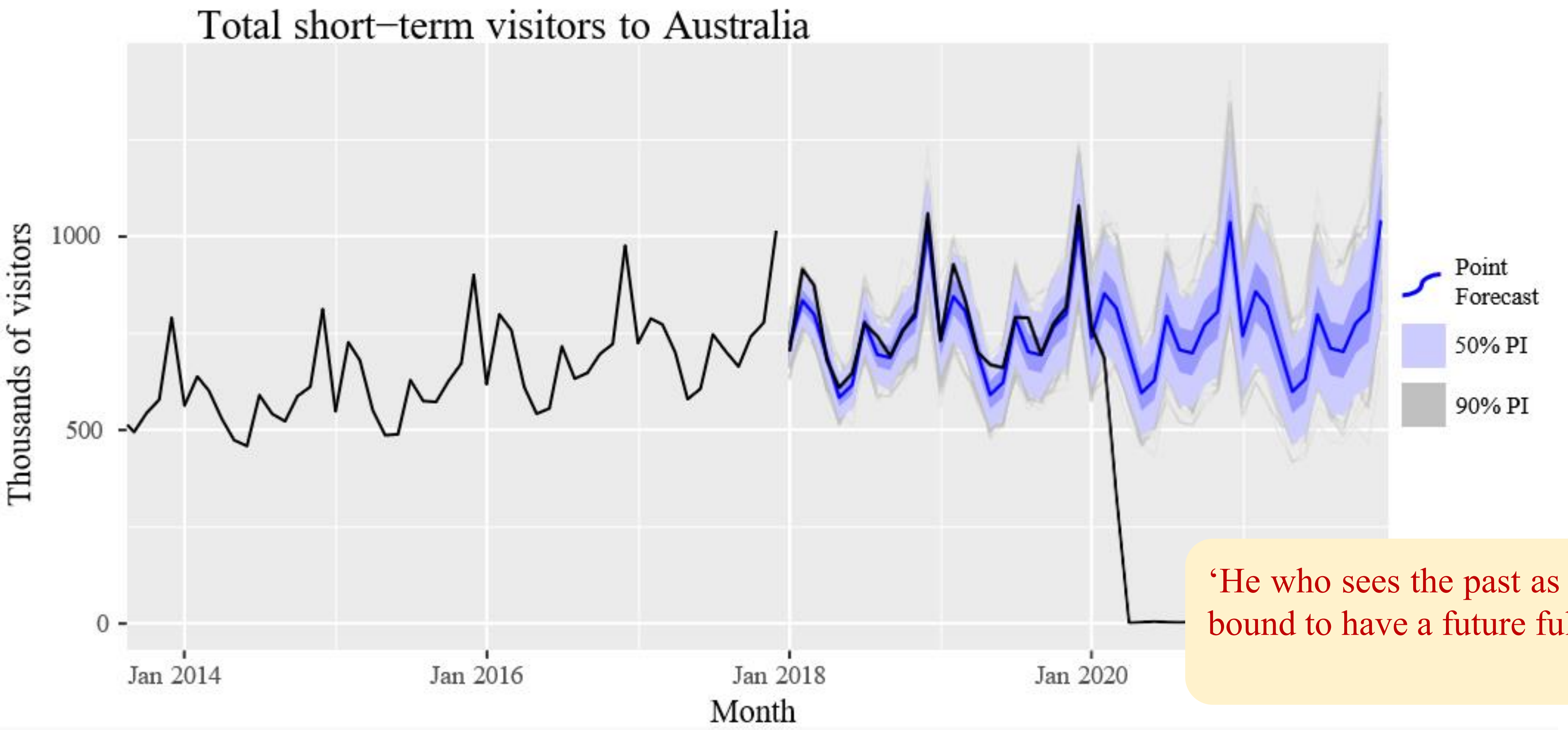
It still has some increase in the magnitude but **not much as the original data**.



Forecasting

Uncertainty and Forecasting

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



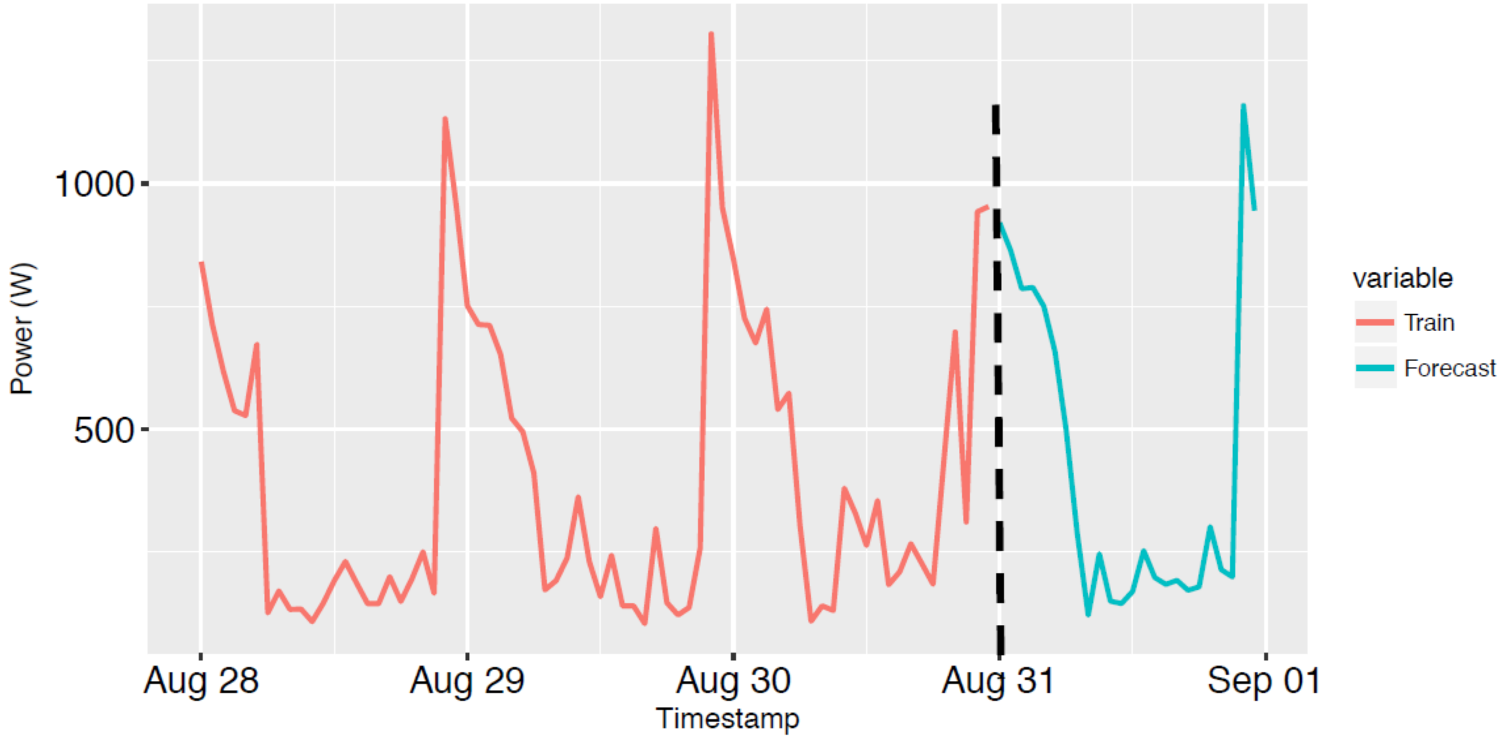
'He who sees the past as surprise-free is bound to have a future full of surprise.'
- Amos Tversky

Forecasting: Assumptions

SORBONNE UNIVERSITÉ

- **Time series Forecasting:** Data collected at regular intervals of time (e.g., Electricity Forecasting).

- **Assumptions:** (a) Historical information is available;
  (b) Past patterns will continue in the future.



Reference

HAPPY FORECASTING

"A good forecaster is not smarter than everyone else, he merely has his ignorance better organised."