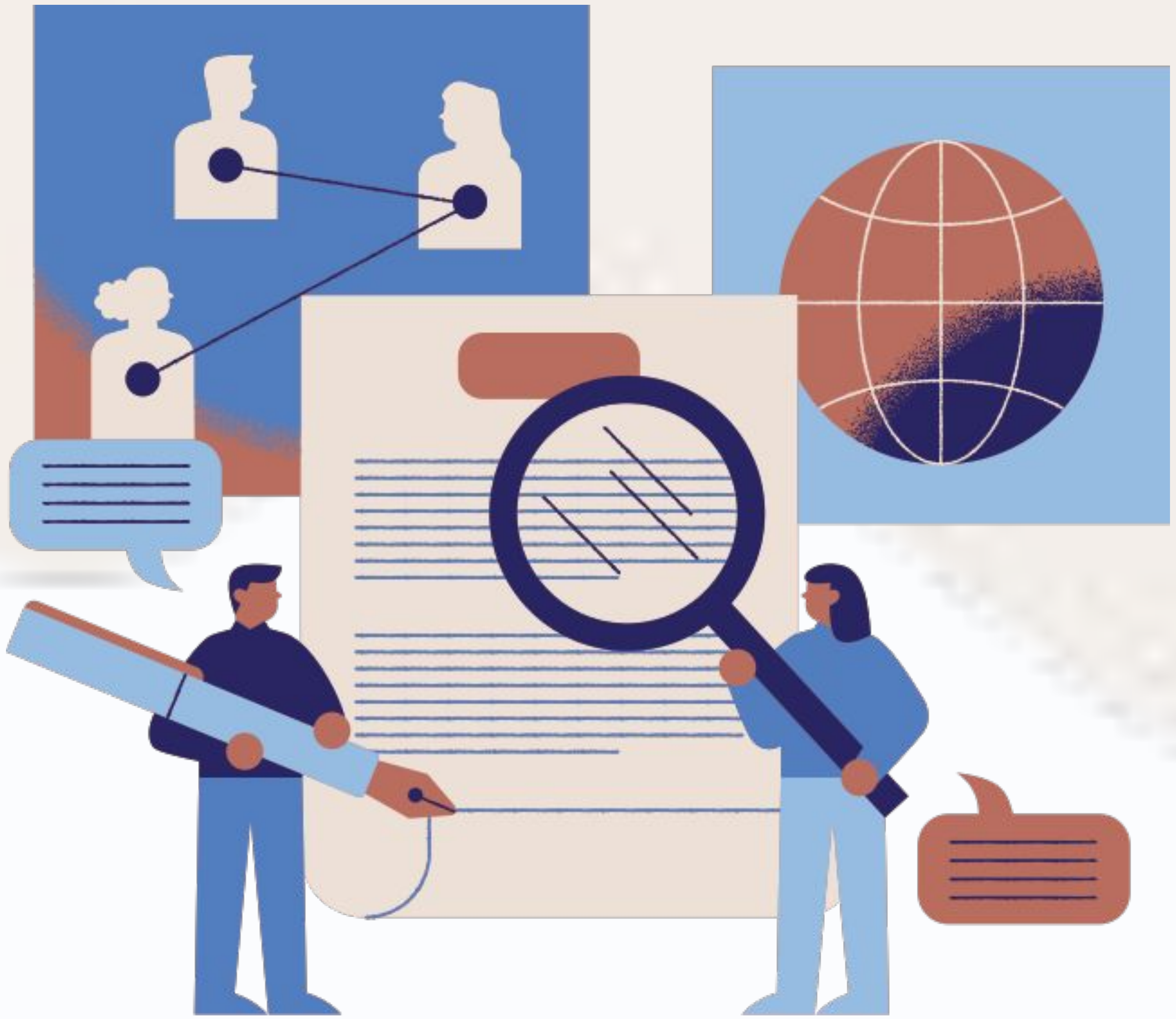
|| JAI SRI GURUDEV ||

Sri Adichunchanagiri Shikshana Trust®

SJB INSTITUTE OF TECHNOLOGY

AN AUTONOMOUS INSTITUTE UNDER VISVESVARAYA TECHNOLOGICAL UNIVERSITY



Course: Exploratory Data Analytics

Course Code: 23CSE422

Mrs. Shilpashree S,
Assistant Professor
Department of CS & E

Module-4

Understanding the time series dataset

Definition: A time series is a collection of observations recorded sequentially over time.

Key Elements of Time Series

- ***Collection of Observations: Multiple data points, not just a single measurement.***
- ***Sequential in Time: Data is ordered by time—each observation has a time stamp or time-based index.***

Time series data captures temporal patterns, allowing us to:

- ***Track trends over time***
- ***Make forecasts***
- ***Detect anomalies***

Understanding the time series dataset

Examples of Time Series Data in Real Life

🛍️ ***Retail Sales Trends: Analyzing total sales over days, months, or seasons.***

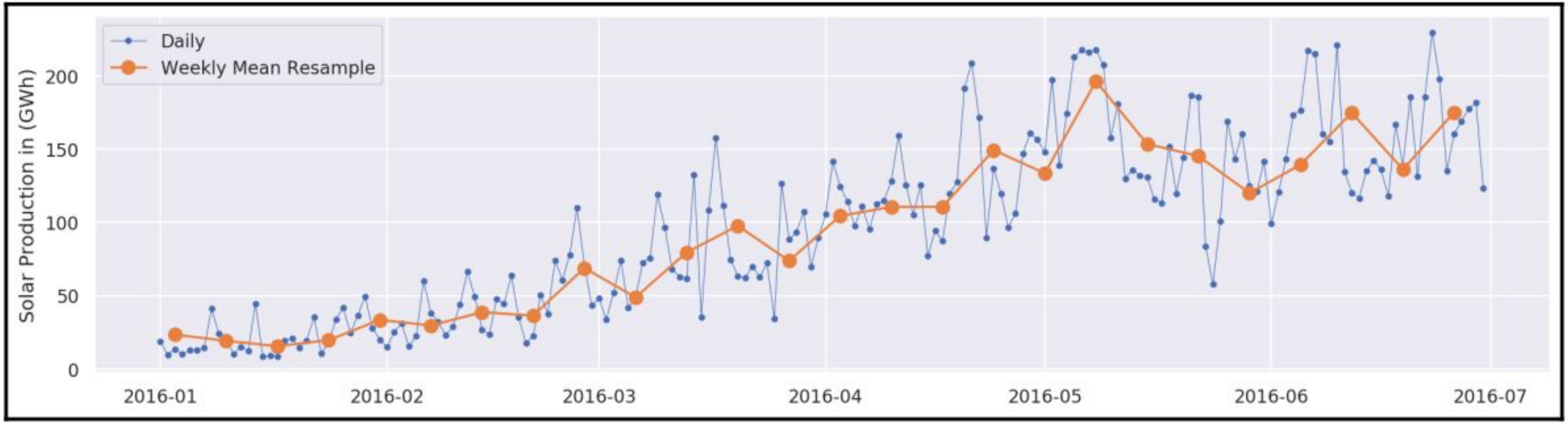
📈 ***Stock Market Analysis: Understanding patterns and fluctuations in stock prices over time.***

🌤️ ***Weather Forecasting: Predicting temperature changes across months and seasons.***

🩺 ***Medical Monitoring: Tracking heart rate patterns to observe effects of medications.***

📰 ***Subscriber Growth Analysis: Measuring increase or decrease in subscribers year over year.***

Understanding the time series dataset



Solar Energy and Electricity Usage (Jan–Jun 2016)
The chart shows solar energy production (in GWh) and electricity consumption on a daily and weekly basis.

Fundamentals of TSA

In order to understand the time series dataset, let's **randomly** generate a **normalized dataset**:

1. **Generating a "Zero Mean" Time Series:** We can generate the dataset using the **numpy** library

Check the code snippet & Output given here:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

zero_mean_series = np.random.normal(loc=0.0, scale=1., size=50)
print(zero_mean_series)
```

```
[ 1.34203177 -0.1955064  -0.34609476  0.50451503  1.3049901  -0.9321934
  0.62887145  0.0205069  -0.59912361  1.35901266 -0.24349178  1.87055621
  0.70027315  0.75123089 -0.93358856  0.15629235 -0.82295271  0.31264348
 -0.20340048  1.13491894 -2.56840318 -0.36250103  0.98282982 -0.73117821
 -0.19881863  0.66063834 -1.62415512 -1.93481907  0.85412951  0.31556022
  0.75982472 -0.76656602  1.37314653  0.17849685  1.25962592  2.15471396
 -0.7792634   0.56528432 -0.84644363  0.73818031 -0.01247721  2.30925747
 -0.32218973 -0.36374626 -1.12339489 -0.32626342 -0.5940559  -0.47641825
 -1.47557415  0.83616162]
```

Fundamentals of TSA

What is a Zero Mean Time Series?

A zero mean time series is a series where the average (mean) value of all observations is zero.

Example

If a time series has values like [-2, 1, 0, 3, -2], then the mean = 0, so it's a zero mean series.

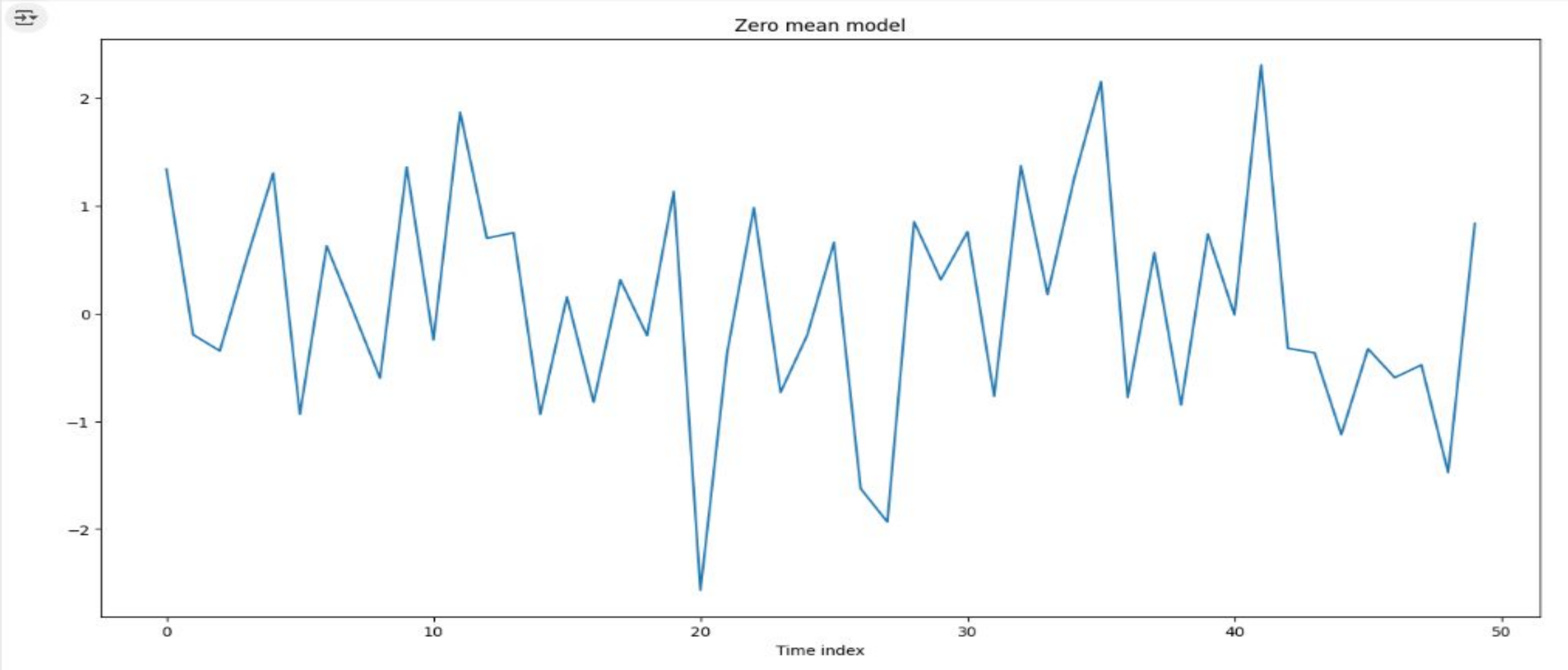
```
zero_mean_series = np.random.normal(loc=0.0, scale=1.0, size=50)
```

- loc=0.0: Mean of the distribution (centered at 0)
- scale=1.0: Standard deviation (spread of data)
- size=50: Number of values in the series

Fundamentals of TSA

2. Next, we are going to use the **seaborn** library to plot the time series data.

```
plt.figure(figsize=(16, 8))
g = sns.lineplot(data=zero_mean_series)
g.set_title('Zero mean model')
g.set_xlabel('Time index')
plt.show()
```



Fundamentals of TSA

3. Creating a "Random Walk" Time Series (Create more complex time series): We can perform a **cumulative sum** over the list and then **plot the data** using a time series plot. The plot gives more interesting results.

Check the following code snippet & Output:

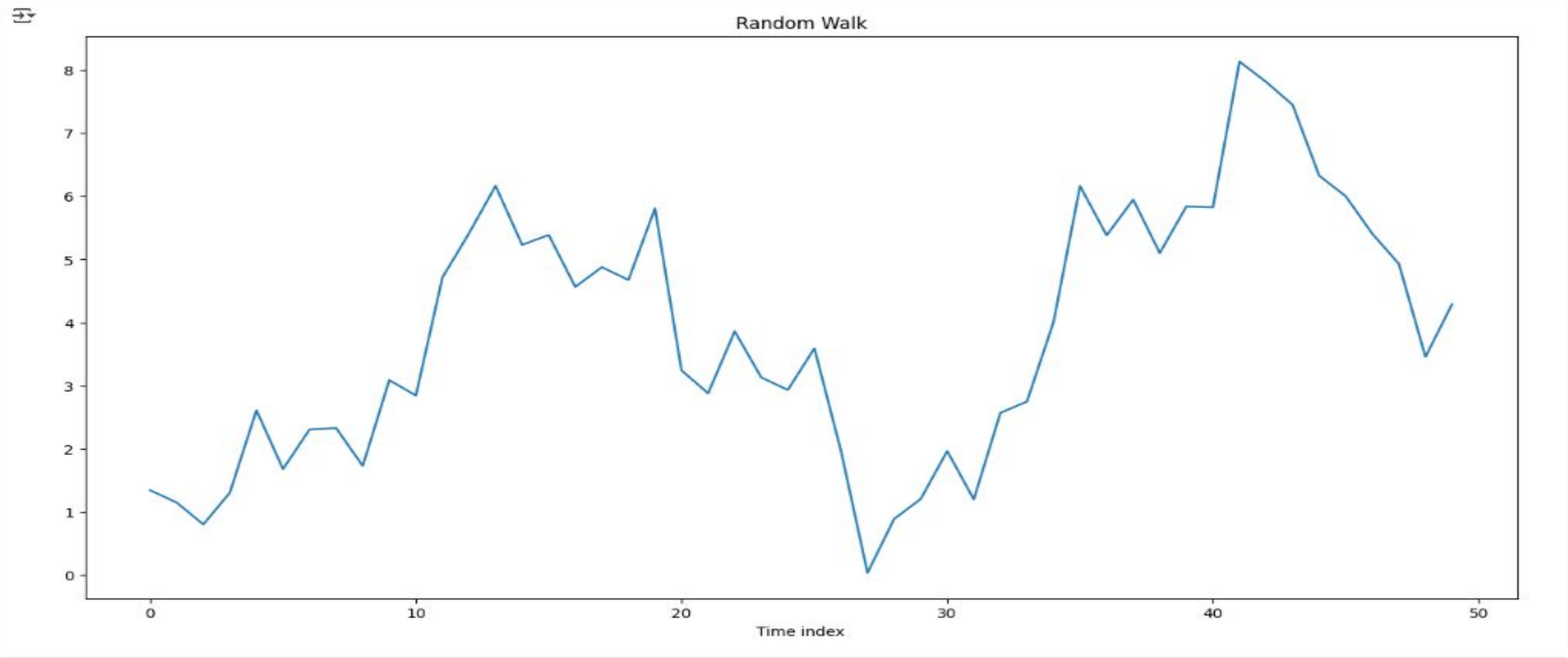
```
random_walk = np.cumsum(zero_mean_series)
print(random_walk)
```

```
[1.34203177 1.14652537 0.80043062 1.30494565 2.60993575 1.67774234
 2.3066138  2.32712069 1.72799709 3.08700975 2.84351797 4.71407418
 5.41434733 6.16557822 5.23198966 5.38828202 4.56532931 4.87797279
 4.67457231 5.80949125 3.24108807 2.87858704 3.86141686 3.13023865
 2.93142002 3.59205836 1.96790324 0.03308417 0.88721368 1.2027739
 1.96259861 1.1960326  2.56917913 2.74767598 4.0073019  6.16201586
 5.38275246 5.94803678 5.10159315 5.83977346 5.82729625 8.13655372
 7.81436399 7.45061773 6.32722283 6.00095941 5.40690351 4.93048526
 3.45491111 4.29107272]
```

Fundamentals of TSA

4. Plotting the "Random Walk" Time Series: Now, we get an interesting graph that shows the change in values over time.
Check the following code snippet & Output:

```
plt.figure(figsize=(16, 8))
g = sns.lineplot(data=random_walk)
g.set_title('Random Walk')
g.set_xlabel('Time index')
plt.show()
```



- The **"random walk"** is a fundamental concept in time series analysis.
- It's a process where the **next value** is a **random step away** from the current value, making it **inherently difficult** to predict.
- Many real-world time series, like stock prices, can exhibit characteristics of a random walk in the short term.

Univariate time series

 Defnition: When we capture a sequence of observations for the **same/ one variable** over a particular duration of time, the series is referred to as **univariate time series.**

Simple Example: Hourly Temperature in Your Room

✔Imagine you have a thermometer in your room, and you decide to record the **temperature (univariate)** every hour for the next 12 hours.

Here's what your recordings might look like:

- **Hour 1 (9:00 AM):** 25°C
- **Hour 2 (10:00 AM):** 26°C
- **Hour 3 (11:00 AM):** 27°C
- **Hour 4 (12:00 PM):** 28°C
- **Hour 5 (1:00 PM):** 28.5°C
- **Hour 6 (2:00 PM):** 29°C
- **Hour 7 (3:00 PM):** 28.8°C
- **Hour 8 (4:00 PM):** 28°C
- **Hour 9 (5:00 PM):** 27.5°C
- **Hour 10 (6:00 PM):** 27°C
- **Hour 11 (7:00 PM):** 26.5°C
- **Hour 12 (8:00 PM):** 26°C

Characteristics of time series data

1. **Trend:**

- **Definition:** "Observing a trend means that the average measurement values seem either to decrease or increase over time.“

- A trend is the general direction in which the time series is moving over a longer period. It's the underlying "slope" of the data.

- **Example:**

  Increasing Trend: The sales of electric vehicles are generally increasing year after year. If you plot the annual sales of EVs, you'd likely see an upward trend.

  Decreasing Trend: The sales of traditional film cameras have been generally decreasing over the past two decades. A plot of their annual sales would likely show a downward trend.

- **Visual:** Imagine a line graph that, overall, is going uphill (increasing trend) or downhill (decreasing trend), even if it has some smaller ups and downs along the way.

Characteristics of time series data

2. Outliers:

Definition: "Time series data may contain a notable amount of outliers. These outliers can be noted when plotted on a graph.“

Outliers are data points that are significantly different from the other values in the time series. They are the "odd ones out."

Example: Imagine you're tracking the daily temperature in Bengaluru, and one day the temperature suddenly spikes to 45°C when the usual range is 30-35°C. This unusually high temperature would be an outlier. It could be due to a heatwave or even a data recording error.

Visual: On a line graph, outliers would appear as points that are far away from the general pattern of the line – either much higher or much lower than the surrounding points.

Characteristics of time series data

3. Seasonality:

- **Definition:** "Some data in time series tends to repeat over a certain interval in some patterns. We refer to such repeating patterns as seasonality.“
- Seasonality refers to patterns that repeat at fixed intervals, like daily, weekly, monthly, or yearly. These patterns are often related to calendar cycles or other predictable events.
- **Example:**

  (i)Yearly Seasonality: Sales of winter clothing tend to peak in the colder months (November-February) and dip in the summer months. This pattern repeats every year.

  (ii)Weekly Seasonality: Website traffic might be higher on weekdays and lower on weekends, with this pattern repeating every week.

  (iii)Daily Seasonality: Electricity consumption often peaks during the day when businesses and households are active and is lower at night. This pattern repeats every day.
- **Visual:** Imagine a line graph with regular, repeating humps or dips at consistent intervals.

Characteristics of time series data

4. Abrupt Changes:

- **Definition:** "Sometimes, there is an uneven change in time series data. We refer to such uneven changes as abrupt changes. Observing abrupt changes in time series is essential as it reveals essential underlying phenomena.“
- Abrupt changes are sudden, significant shifts in the level or trend of the time series. They are often caused by unexpected events.
- **Example:**

  (i)A sudden drop in airline bookings after a major travel restriction is announced.

  (ii)A sharp increase in the stock price of a company after a successful product launch.

  (iii)A sudden decrease in water levels in a reservoir due to a prolonged drought.
- **Visual:** On a line graph, an abrupt change would look like a sharp, almost vertical jump or drop in the line.

Characteristics of time series data

5. Constant Variance (Stationarity in Variance):

- **Definition:** "Some series tend to follow constant variance over time. Hence, it is essential to look at the time series data and see whether or not the data exhibits constant variance over time."
- Constant variance means that the spread or the "wiggliness" of the data around its average level remains roughly the same over the entire time period. If the variance is not constant, the spread of the data changes over time.
- **Example:**

  (i)Constant Variance: The daily temperature in a relatively stable climate might fluctuate within a consistent range throughout the year.

  (ii)Non-Constant Variance: The price of a volatile stock might have small fluctuations during periods of stability but experience much larger swings during times of economic uncertainty.
- **Visual:** Imagine comparing two time series plots. One might have the line consistently staying within a certain vertical range, while the other might show the line having small ups and downs at the beginning but then much larger ups and downs later on.