

**NAME – SHUBHAM KENDAL**

**REGISTRATION NO – 12301314**

**COURSE – M.Sc. (Statistics and Data Analytics)**

**SECTION – G3304**

**ROLL NO – RG3304A04**

**TOPIC – Detection and Modelling of Seasonal Patterns in Time Series.**

**SUBMISSION DATE – 2nd April , 2024.**

* **Time Series Analysis:-**

Time series analysis is a specific way of analysing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

**Examples for Time Series Analysis:**

Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms. Likewise, time series analysis is ideal for forecasting weather changes, helping meteorologists predict everything from tomorrow’s weather report to future years of climate change. Examples of time series analysis in action include:

* Weather data
* Rainfall measurements
* Temperature readings
* Heart rate monitoring (EKG)
* Brain monitoring (EEG)
* Quarterly sales
* Stock prices
* Automated stock trading
* Industry forecasts
* Interest rates

**Components of Time Series Analysis:-**

The various forces at work, affecting the values of a phenomenon in a time series, can be broadly classified into the following four categories, commonly known as the Components of a time series, some or all of which are present ( in a given time series) in varying degrees.

1. Long-term Movement or Secular trend.
2. Periodic Changes or Short-term Fluctuations.
3. Seasonal Variations, and **(ii)** Cyclic Variation
4. Random or Irregular Movements.

**Our main focus will be on the Detection and modelling of seasonal patterns in a time series.**

* **Seasonal Variations:**

These variations in a time series are due to the rhythmic forces which operate in a regular and periodic manner over a span of less than a year, i.e., during a period of 12 months and have the same or almost same pattern year after year. Thus Seasonal variations in a time series will be there if the data are recorded quarterly (every three months), monthly, weekly, daily, hourly and so on.

Most of economic time series are influenced by seasonal swings, e.g, prices, production and consumption of commodities; sales ans profits in a departmental stores; bank clearings and bank deposites, all are affected by seasonal variations. The seasonal variations may be attributed to the following two causes:

1. **Those resulting from natural forces.** As the name suggests, the various seasons or weather conditions and climatic changes play an important role in seasonal movements. For instance, the sale of umbrellas pick up very fast in rainy seasons; the demand for electric fans goes up in summer seasons; the sales of ice and ice cream increases very much in summer – all are being affected by natural forces, viz., weather or seasons.
2. **Those resulting from man-made conventions.** These variations in a time series within a period of 12 months are due to habits, fashions, customs and conventions of the people in the society. For instance, the sale of jewellery and ornaments goes up in marriages; the sales and profits in departmental stores goes up considerably during marriages, and festivals like Diwali, Dussehra, Christmas, etc. Such variations operate in a regular spasmodic manner and recur year after year.

The main objective of the measurement of seasonal variations is to isolate them from the trend and study their effects. A study of seasonal variations is very useful to businessman, producers, sales manager, etc., in planning future operations and in formulation of policy decisions regarding programmmes. Thus, to understand the beahvious of the phenomenon in a time series properly, the time series data must be adjusted for seasonal variations. This is done by isolating them from trend and other components by dividing the given time series values (yt) by the seasonal variations (St). This technique is called de-seasonalisation of data.

Now, we will discuss about the technique called de-seasonalisation of data. Here we will measure the seasonal variations in a time series data. There are diffrent ways to measure seasonal variations which are discussed below:

1. **Method of Simple Averages**:-

The method of simple averages is a basic technique used to estimate seasonality in time series data. It includes the following steps :

* **Calculate the Average** : First, you take the historical data for each time period within a season. Then, you calculate the average value for each time period across multiple seasons.
* **Subtract the Average** : Next, you subtract the average value calculated for each time period from the actual data values for that time period. This gives you the seasonal component of the time series data.
* **Smooth the Data** : Finally, you can use the seasonal component to adjust the original data, effectively removing the seasonal fluctuations and leaving behind the underlying trend and random components.

Now, The Merits of Simple Average method are :

* **Simplicity** : The method of simple average is straight forward and easy to implement, making it accessible to users within minimal statistical expertise.
* **Interpretability** : The result obtained from this method are easy to understand and interpret, making it suitable for communication findings to a non-technical audience.

And, The Demerits of Simple Average method are :

* **Lack of Accuracy** : This method assumes that the seasonal patterns remain constant over time, which may not always be the case. Seasonal patterns can evolve or change due to the various factors, leading to inaccuracies in the estimation.
* **Sensitive to Outliers** : The method can be sensitive to outliers or irregularities in the data, which can disort the calculated averages and affect the accuracy of the seasonal component.

1. **Ratio to Trend method**:-

The Ratio to Trend method is a technique used to estimate seasonality in time series data. So, the brief explanation is given as,

1. **Calculate the Trend :** First, you need to estimate the underlying trend in the time series data. This can be done using various methods such as moving averages, linear regression, or exponential smoothing.
2. **Calculate the Ratios :** Next, you divide each observed data point by the corresponding value of the trend component. This gives you a set of ratios representing the relationship between the actual data and the trend at each point in time.
3. **Calculate Seasonal Indices :** Once, you have the ratios, you calculate the average ratio for each season. These average ratios represent the seasonal indices, which indicate how much higher or lower the data tends to be during each season compared to the underlying trend.
4. **Adjust the Data :** Finally, you adjust the original data by dividing each observation by the seasonal index corresponding to its season. This effectively removes the seasonal component, leaving behind the trend and irregular components of the time series data.

Now, let’s discuss the merits of this method :

1. **Flexibility :** The Ratio to Trend method allows for detection of changing seasonal patterns over time. Unlike the simple average method, it does not assume that seasonal pattern remains constant, making it more adaptable to evolving data.
2. **Robustness :** By dividing the data by trend component, this method helps to stabilize the seasonal indices, reducing the influence of outliers or irregularities in the data.

And, The Demerits of this method are :

1. **Complexity :** The Ratio to Trend method requires the estimation of underlying trend, which can be challenging, especially for noisy or volatile time series data.
2. **Subjectivity :** The choice of trend estimation method can impact the results of the analysis. Different methods may yield different trend estimates, leading to the variations in the calculated seasonal indices.
3. **Ratio to Moving average method:-**

The Ratio to Moving Average (RMA) method is a technique used in time series analysis to handle seasonality. The RMA method involves dividing each observation in the time series by the corresponding value of moving average. The moving average is typically calculated over a certain number of periods to smooth out random fluctuations.

The primary purpose of using the RMA method is to remove or minimize the effect of seasonality from the time series data, allowing for better analysis of underlying trends or pattern.

Now, let’s discuss its merits which are given as,

1. **Seasonal Adjustment :** The RMA method effectively adjusts time series data for seasonal variations, making it easier to identify underlying trends or patterns.
2. **Simplicity :** It’s a relatively simple technique to implement and understand, making it accessible to analysts and practitioners with varying levels of statistical expertise.

And, the demerits of this method are –

1. **Data Requirements :** RMA requires a sufficient amount of historical data to compute reliable moving averages, which might be a limitation in cases where data availability is limited.
2. **Loss of Information :** The process of dividing each observation by the moving averages can lead to a loss of information, particularly if the seasonal component is not accurately captured by the moving average.
3. **Link relative method:-**

The Link Relative method is another approach to handle the seasonality. The link relative method involves calculating the ratio of each observation to the corresponding observation from the previous season or period. This ratio is used to adjust for seasonal variations in the data.

So, Merits of this method are given as;

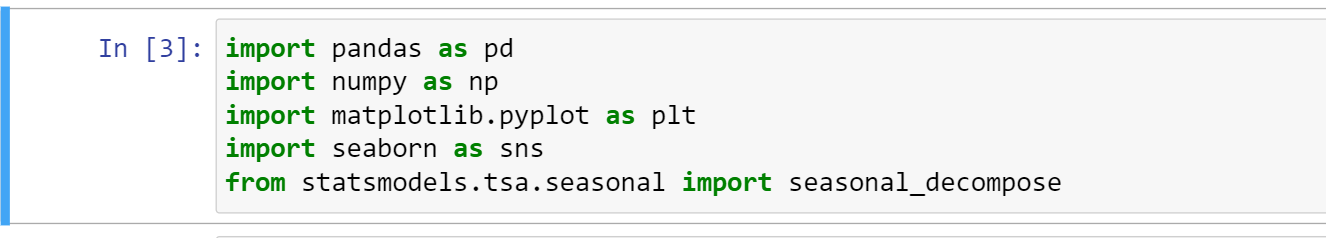
1. **Seasonal Adjustment :** The link relative method effectively adjusts the time series data for seasonal variations by comparing each observation to the corresponding observation from the previous season.
2. **Interpretability :** The results obtained from the link relative method are often intuitive and easy to interpret, as they provide a direct comparison to previous periods.
3. **Less Data Requirement :** Compared to some other methods, the link relative method may require less historical data, particularly if the seasonal patterns are consistent over time.

And, the demerits of this method are –

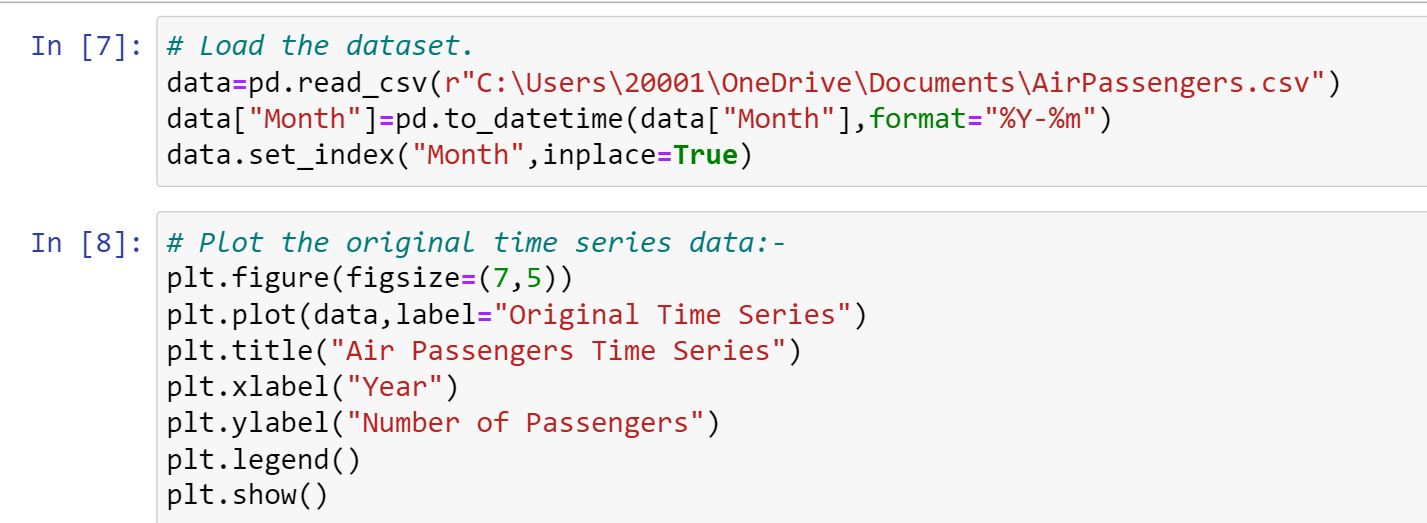
1. **Assumption of Stationarity :** Similar to other seasonal adjustment methods, the link relative method assumes stationarity in the time series data, which ,might not always hold true, especially in cases where there are underlying structural changes over time.
2. **Less Flexibility :** Compared to methods like Ratio to Moving Average, the link relative method may offer less flexibility in adjusting for different types of seasonality or incorporating additional factors into the analysis.

**Now we will see how to detect the seasonality in our data and model the data in Python.**

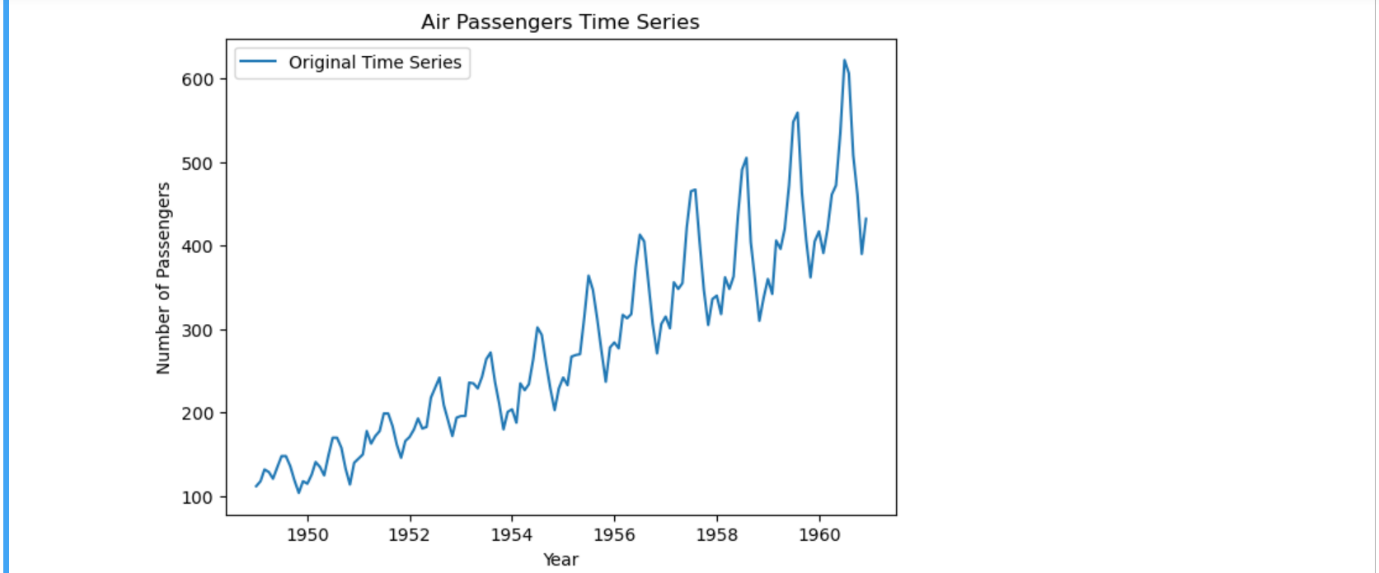
At first, we will import all required Python modules like Pandas, NumPy, Matplotlib and Seaborn etc.



Now we will load the time series dataset which we have taken from the Kaggle named as “AirPassengers.csv” then we will visualize the raw data.

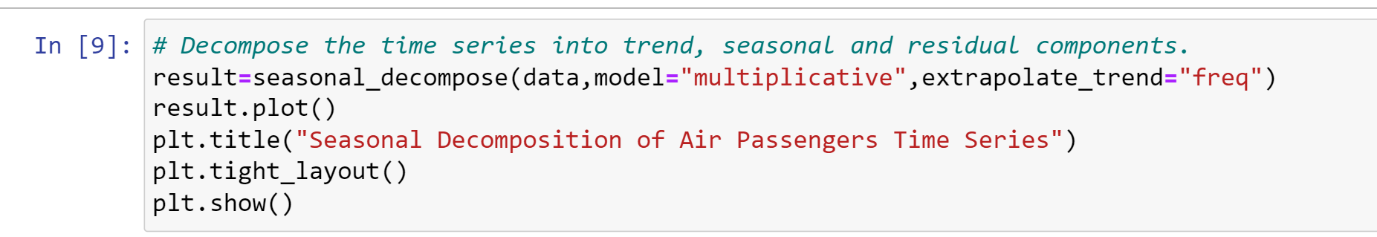
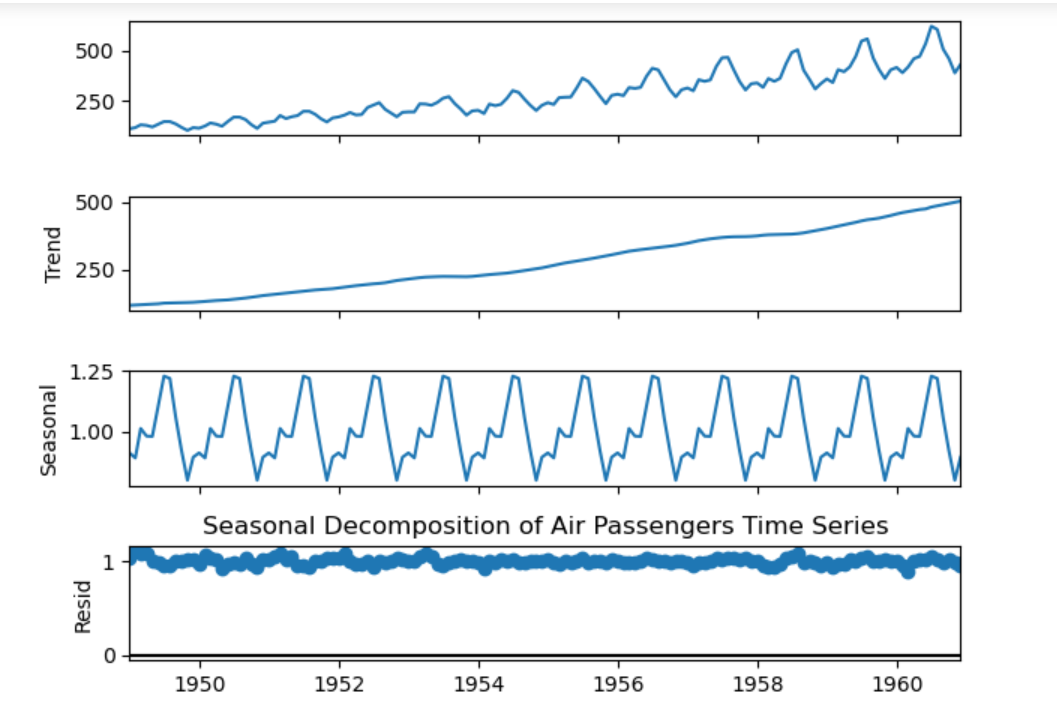


Output:-



As we have already got the time series plot, now we will decompose it to the trend, seasonal and residual components. To do this we need to specify some of the parameters of seasonal decompose function which are listed below:

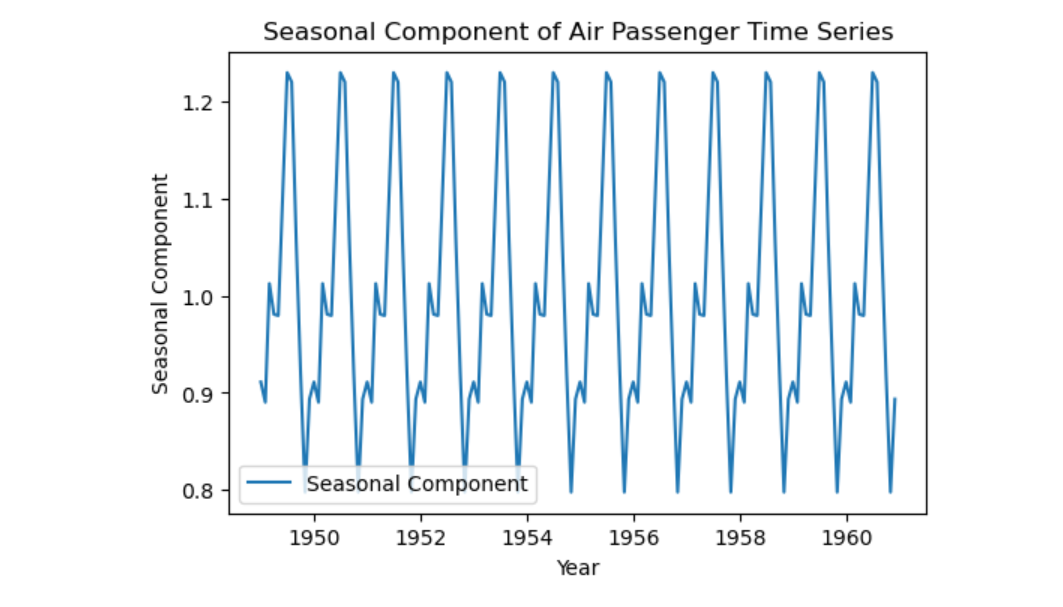
* **data:** This parameter represents the time series data that we want to decompose which is should be in a pandas Data Frame or Series with a datetime index.
* **model:** This parameter specifies the type of decomposition to be performed which can take two values “additive” or “multiplicative”. Here we will use “multiplicative” model as we can see the amplitude of seasonal component is relatively constant across different levels of the time series.
* **extrapolate\_trend:** This parameter controls whether to extrapolate the trend component to cover missing values at the end of the time series. Here we will sit it to “freq” means that the trend component is extrapolated using the frequency of the time series. Extrapolate the trend is useful when there are missing values at the end of the time series.

Output:

Now we will visualize the only seasonal component by extracting it from the decomposition results.

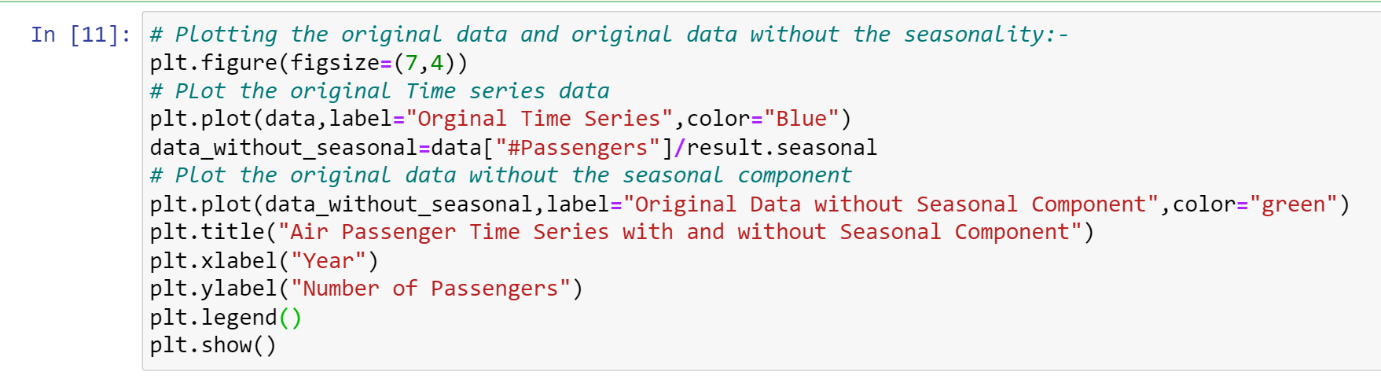


Output:

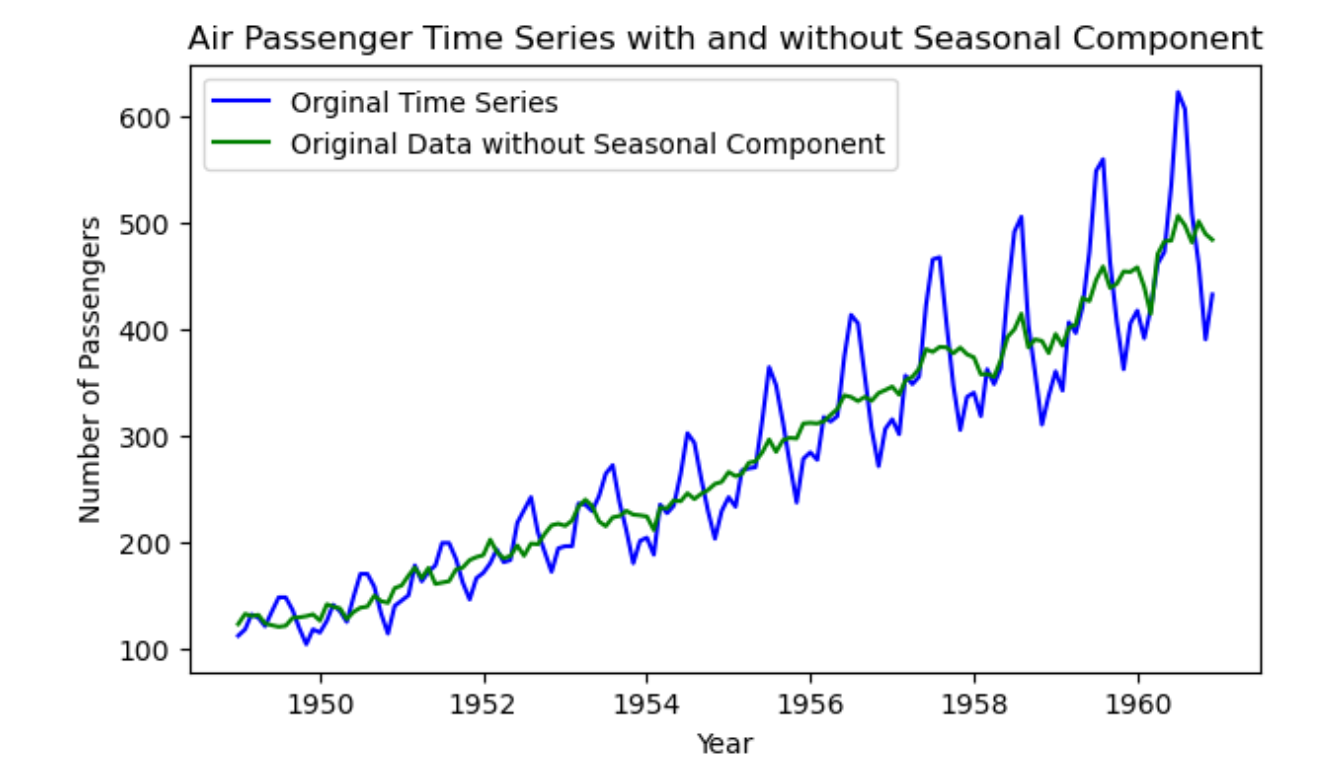


Removing seasonality from the data:

To use a time series data for various purposes including model training it is required to have a seasonality free time series data. Here we will visualize how organised it will look after removing the seasonality.



Output:



From the plot we can see that after removing seasonality the time series data became very organized which required for model training for any further purposes.

Conclusion:

We can conclude that seasonality detection and remove it from the data is very important step before proceed to the model training phase. Seasonality can degrade the performance of the predictive model which may lead to wrong forecast.

**Bibliography:-**

* S.C. Gupta, V.K Kapoor. Fundamentals of Applied Statistics. Sultan Chand & Sons, 1977 Edition.
* Dataset is taken from the website – <http://www.kaggle.com/datasets/chirag19/air-passengers/download?datasetVersionNumber=1>
* Python codes are run in the -<http://localhost:8888/notebooks/Untitled.ipynb?kernel_name=python3>

**Thank You !!!!!**