DSC 275/475: Time Series Analysis and Forecasting (Fall 2022)

HW #1

(Total points: 40)

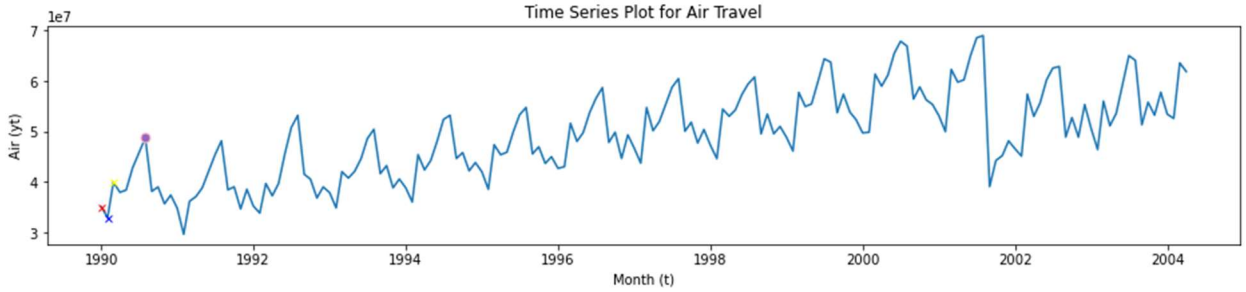
Aradhya Mathur

1. ***(8 pts)*** **The Bureau of Transportation Statistics (BTS) conducted a study to evaluate the impact of Sept 11 attacks (9/11) on U.S. air transportation. The purpose of this study is to provide a greater understanding of the passenger travel behavior patterns of persons travelling by air before and after the event. In order to assess the impact of September 11, BTS took the following approach: Using data before September 11, it forecasted future data (under the assumption of no terrorist attack). Then, BTS compared the forecasted series with the actual data to assess the impact of the event.**

The data is available in the file: *BTS_Air_Rail_Vehicle_Miles.csv*.

a) Is the goal of this study descriptive or predictive? (2 pts)
Descriptive study. As this study focuses on and assess patterns of 9/11 attacks on US transportation. There is no prediction or forecasting involved.

b) **Create a time series plot of the *Air* data, i.e. a plot y_t versus *t, where t=1,2,3* … What would t=1, 2, 3 refer to in the time series? Which time period does t=1 refer to?** (3 pts)



Time periods *t1 = Jan-90, t2=Feb-90, t3= Mar-90*

```
In [33]: #t1, t2 and t3
         travel.Month[0] , travel.Month[1] , travel.Month[2]

Out[33]: (Timestamp('1990-01-01 00:00:00'),
          Timestamp('1990-02-01 00:00:00'),
          Timestamp('1990-03-01 00:00:00'))
```

Converted Jan to 01

c) **What are the values for y_1, y_2 and y_3 in the time series**? (3 pts)
Values of *y1= 35153577, y2=32965187, y3=39993913*

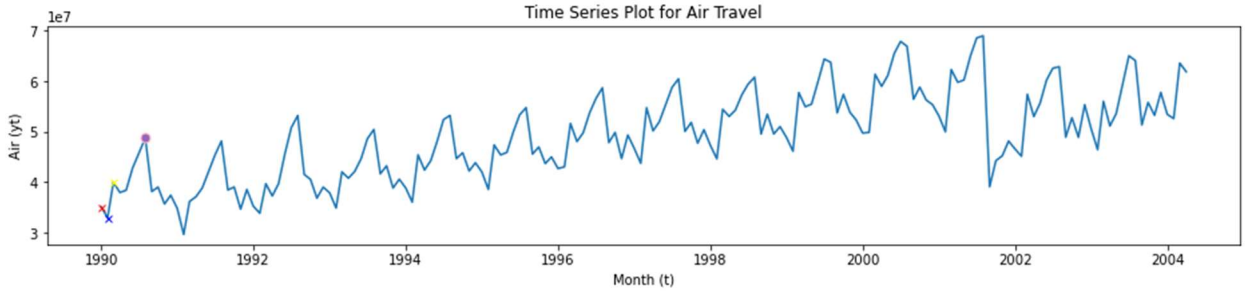
```
In [32]: #y1, y2 and y3
         travel.Air[0] , travel.Air[1] , travel.Air[2]

Out[32]: (35153577, 32965187, 39993913)
```

2. ***(10 pts)*** **In addition to air travel data, two additional time series are also provided in the same data file – Rail and Vehicle travel.**

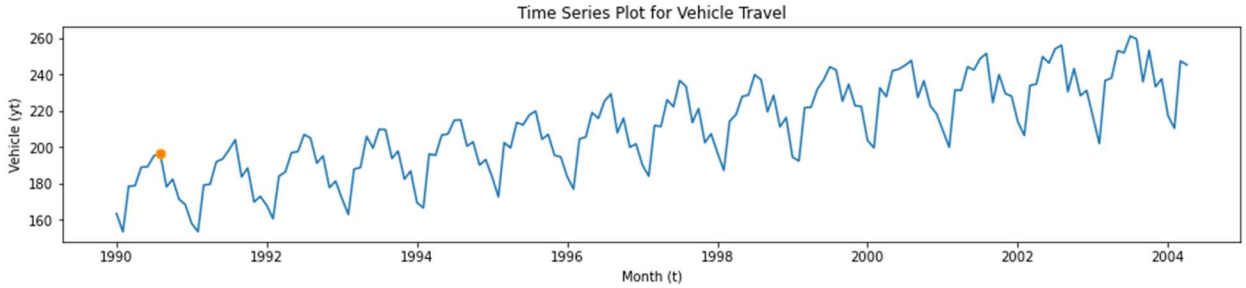
a) **Which of these components appear in the *Air* and *Vehicle* time series: i) Level; ii) Seasonality; iii) Trend; iv) Noise. List for each data set.** (8 pts)

Air-



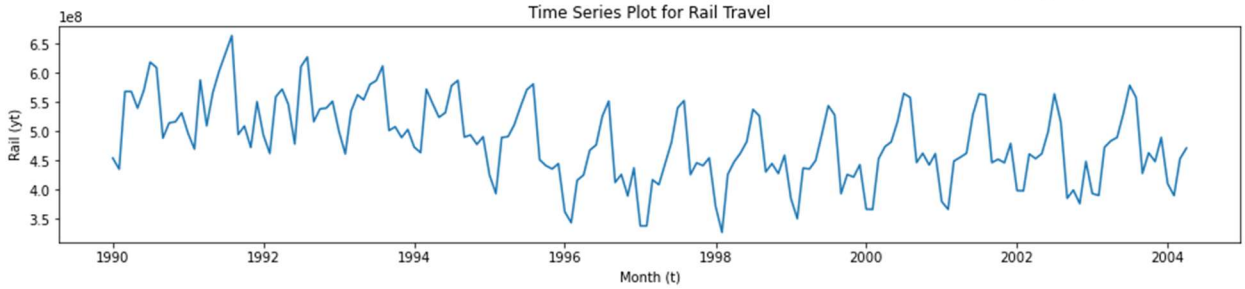
Level: Level appears.
Seasonality: Yes, seasonality can be observed. At eighth month in a year, peak is reached and after that it decreases till the end of year. Travelling is directly related to vacation periods and first peak in March and highest peak in August confirms the same. Additionally, additive form of seasonality can be observed.
Trend: Positive trend noticed. Irregularity can be seen in 2001 where the trend decreases for a year and after that it follows prior positive trend.
Noise: No significant noise detected

Vehicle-



Level: Level is present
Seasonality: Yes, seasonality can be observed. At eighth month in a year, peak is reached and after that it decreases till the last month of the year. Additive seasonality observed.
Trend: Positive trend is easily observed.
Noise: No significant noise observed

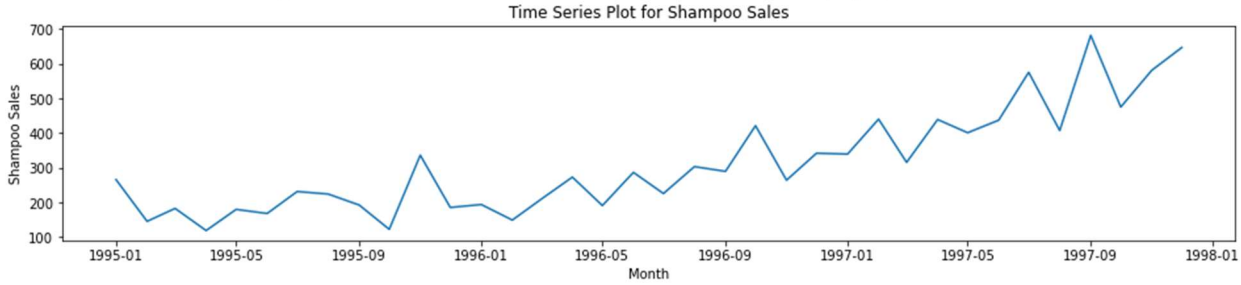
b) **For the Rail data set, describe the trend, i.e. how does the trend vary across the time series?** (2 pts)



Trend: *Irregular trend can be observed in rail dataset. From 1990 to 1998 we can observe a downward trend where it reaches minimum value and from 1998 to 2004 a positive trend can be observed.*

3. *(6 pts)* Forecasting Shampoo Sales: The file *ShampooSales.csv* contains data on the monthly sales of a certain shampoo over a 3-year period.

a) **Create a time series plot of the data. Label the axes.** (2 pts)

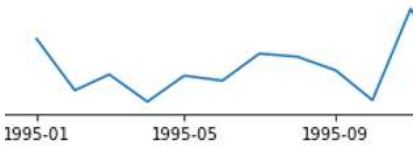


b) Which of the four components (level, trend, seasonality, noise) are present in this series? (4 pts)

Level: Level is present

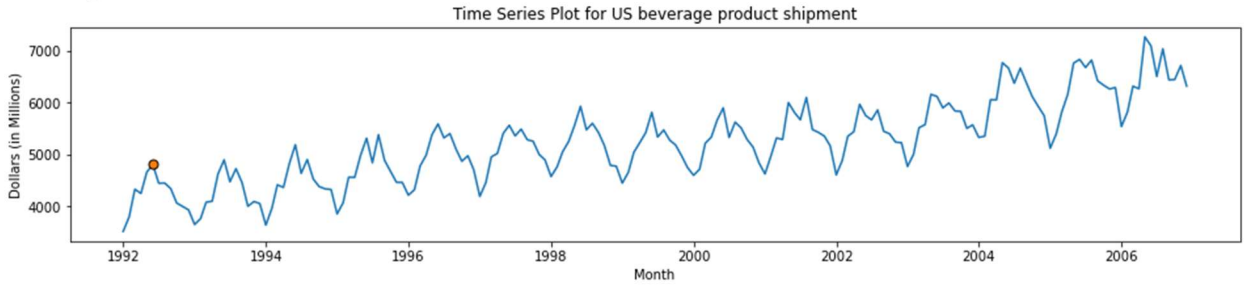
Trend: Positive trend can be observed.

Seasonality: Seasonality observed. Maximum is at start of every year. Additive seasonality observed.



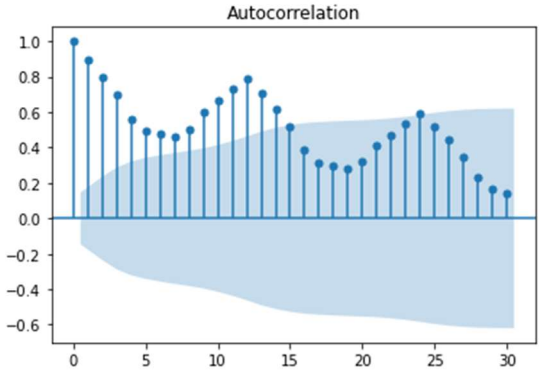
Noise: Noise can not be seen in the graph clearly.

4. ***(6 pts)*** **The file, *Beverages_Shipment_2020.csv*, contains the US beverage product shipments data.**



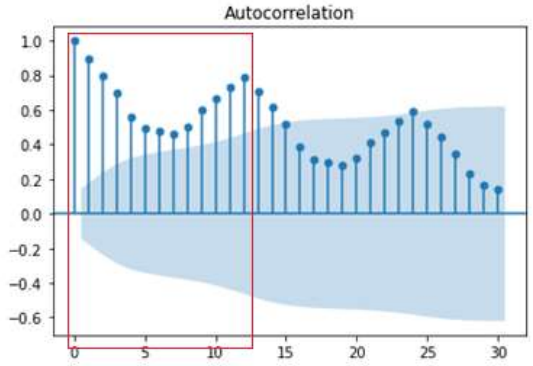
a) **Is there seasonality in this time series?** (2 pts)
Yes, Seasonality can be observed, at sixth month in a year peak is reached and after that it decreases till the end of year. Additive seasonality seen.

b) Find the sample autocorrelation function for this data set. (2 pts)
(For Python, you can use the "plot_acf" function in "statmodels" module. Plot at least 25-30 lags)
Autocorrelation is the linear relation between time series and its lagged values.



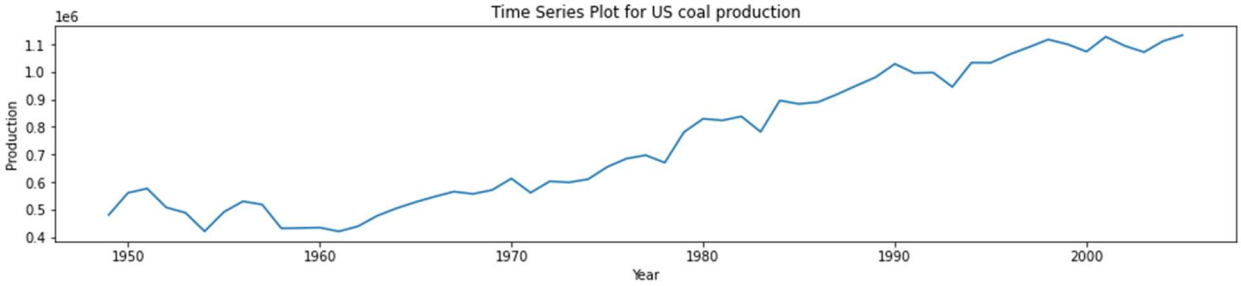
Highest value is 1 which is naturally at start of the function. Seasonality can be observed and lags of trend are reducing.

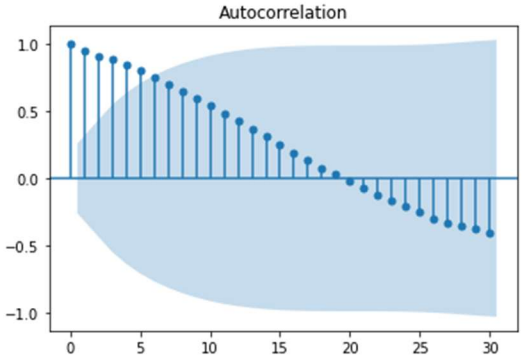
c) **From the autocorrelation plot in (b), what is the seasonal period?** (2 pts)
Seasonal period is 0-11 Lags (12 in total, basically a year)



5. ***(10 pts)*** **Data on US coal production is given in *Coal_Production_US_2020.csv*.**

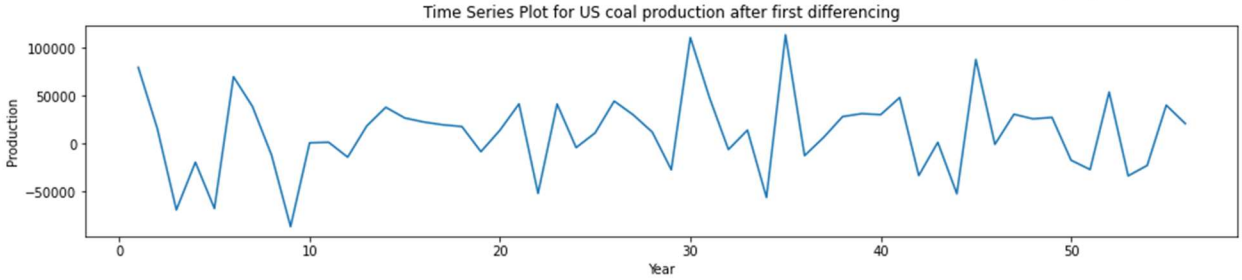
a) **Plot the coal production data and the sample autocorrelation function.** (2 pts)

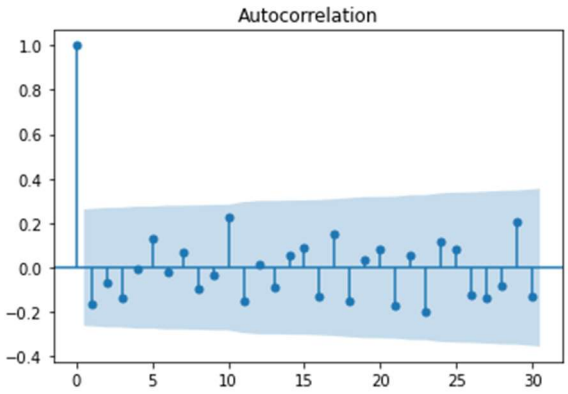




b) **Is the time series stationary or non-stationary?** (2 pts)
Non-Stationary as pattern can be observed in autocorrelation function and a positive trend is visible.

c) **Plot the first difference of the time series and the sample autocorrelation function of the first difference.** (4 pts)





d) **What impact has differencing had on the time series? Comment with respect to presence or absence of stationarity** (2 pts)

Initially it was non-Stationary but after differencing, it became Stationary as trend and seasonality were removed/reduced and no pattern can be observed.