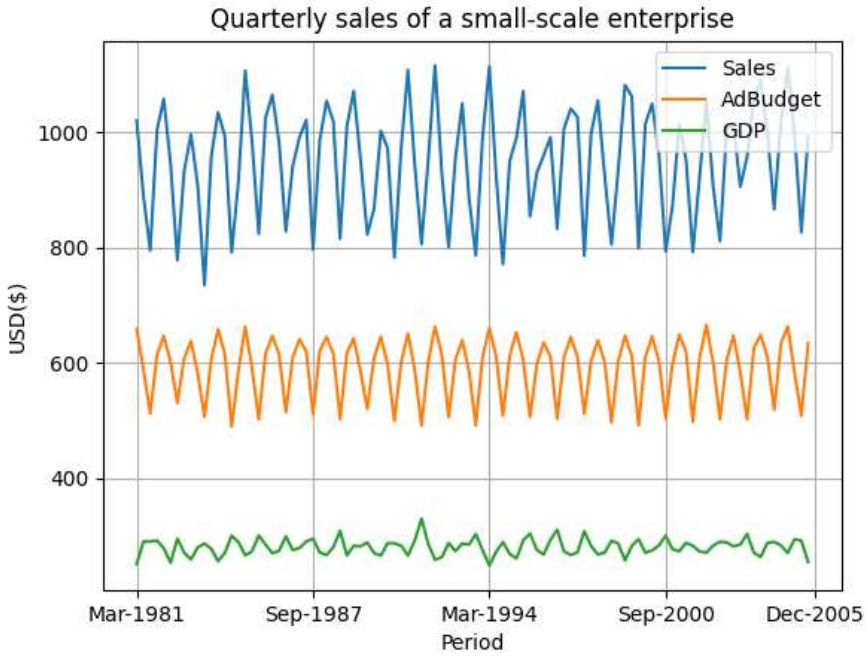
George Washington University

Timer Series Analysis and Modeling – Assignment 1

Rajkumar Conjeevaram Mohan

09/14/2022

Q1. The plot of quarterly sales data



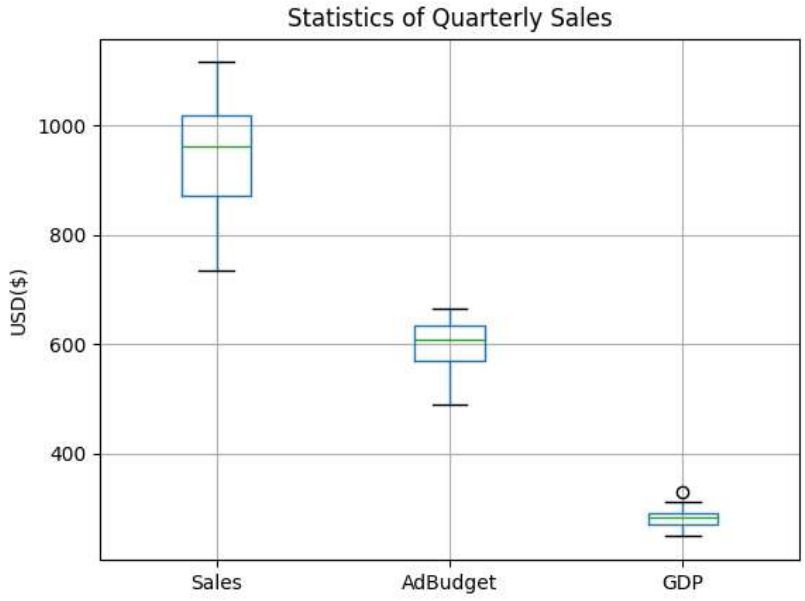
Q2. Time-series statistics

The Sales mean is 948.74 and has a variance of 9653.49 with a standard deviation of 98.25 Median: 960.65

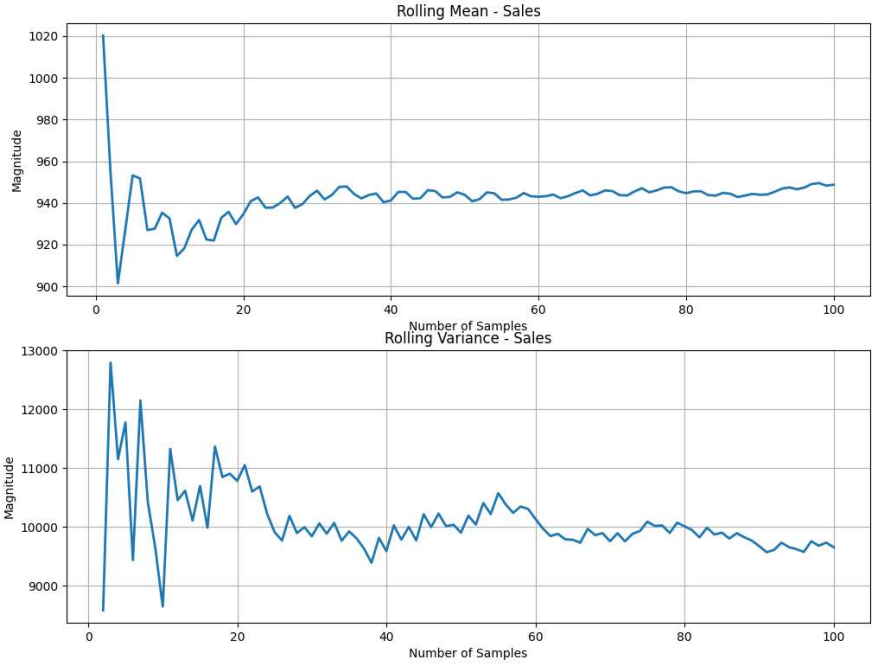
The AdBudget mean is 591.93 and has a variance of 2953.1 with a standard deviation of 54.34 Median: 608.5

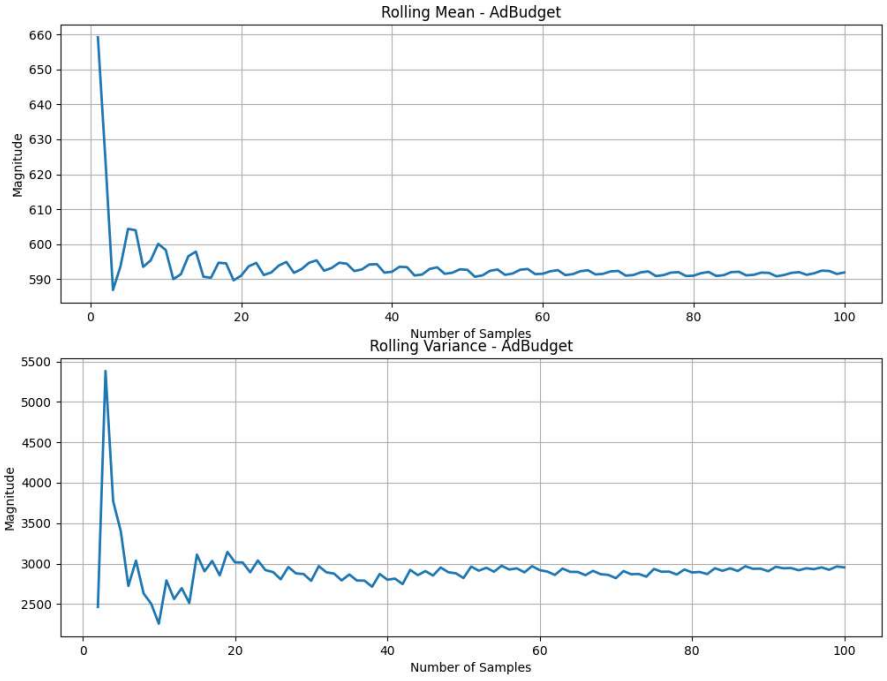
The GDP mean is 281.18 and has a variance of 206.51 with a standard deviation of 14.37 Median: 282.6

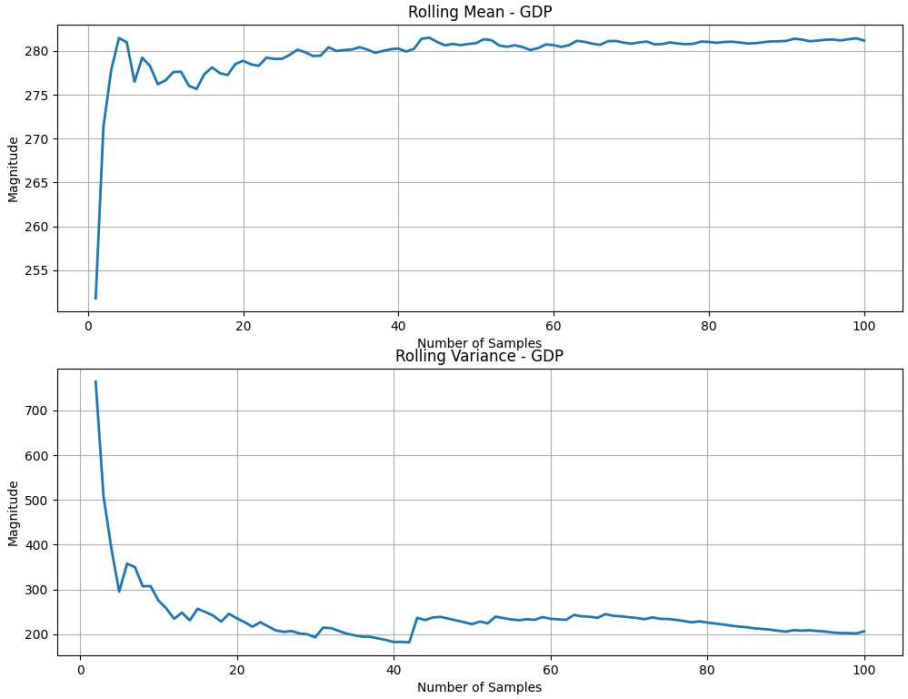
Boxplot of Sales, AdBudget, and GDP (not part of any question)



Q3. Rolling mean and variance for Sales, AdBudget, and GDP







Q4. Write down your observation about the plot of the mean and variance in the previous step. Is Sales, GDP, and AdBudget stationary or not? Explain why.

Although Sales data shows some sign of fluctuation initially, overall, it appears that all three data - Sales, AdBudget, and GDP are stationary, given the rolling mean and the variance stabilizes as more samples are included.

Q5.

ADF results on **Sales**:
ADF Statistic: -3.262755
p-value: 0.016628
Critical Values:
1%: -3.505
5%: -2.894
10%: -2.584

ADF results on **AdBudget**:
ADF Statistic: -2.758605
p-value: 0.064434
Critical Values:
1%: -3.504
5%: -2.894
10%: -2.584

ADF results on **GDP**:
ADF Statistic: -3.227577
p-value: 0.018443
Critical Values:
1%: -3.504
5%: -2.894
10%: -2.584

At a confidence interval of 95%, the Sales and the GPD appear stationary since their p-values are less than the significance threshold of 0.05. In the case of AdBudget, the p-value is greater than the significance threshold, which doesn't allow us to reject the null hypothesis. In other words, AdBudget is non-stationary with 95% confidence.

As with a 95% confidence level, ADF-test results do reinforce visual observation from the previous step for time series except for AdBudget, which has a p-value of 0.064434 that is above the significance threshold of 0.05.

At a confidence interval of 99%, all three data - Sales, AdBudget, and GDP would be non-stationary given the p-values for all three data are beyond the significance threshold of 0.01. This means we fail to reject the null hypothesis, which states for the ADF test that the series has a unit root. In other words, the data is non-stationary.

Q6.

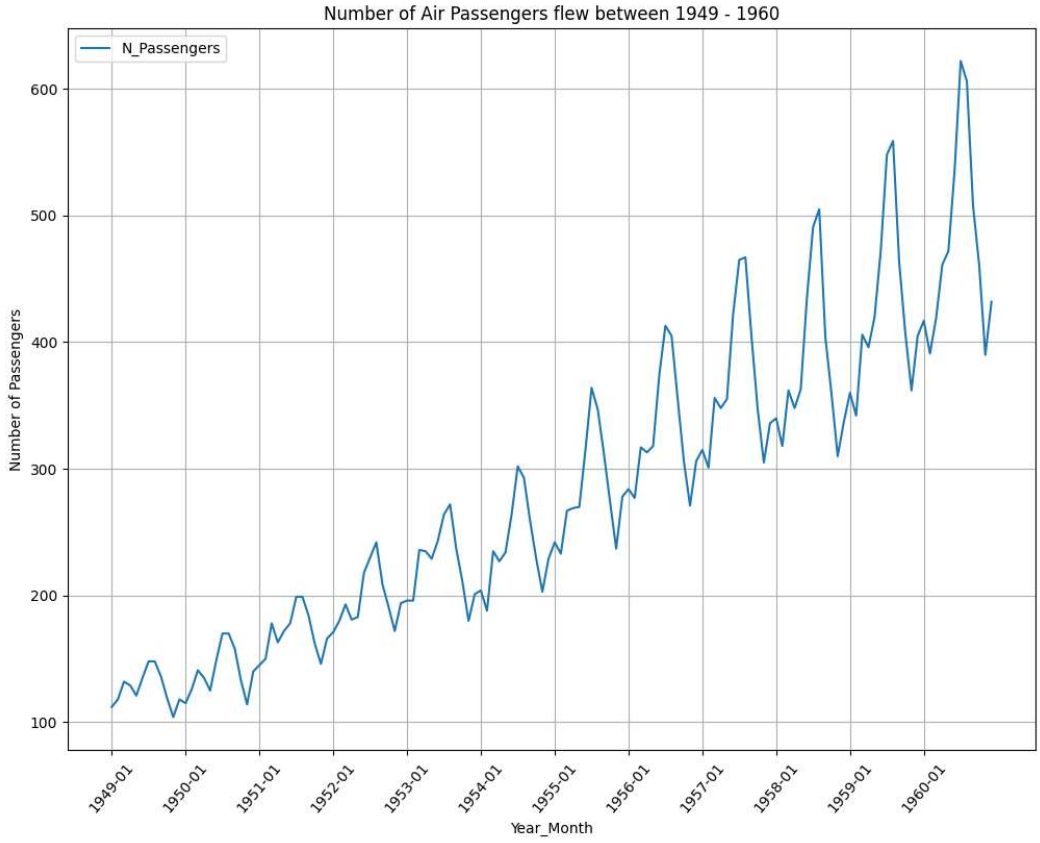
KPSS results on **Sales**:
Results of KPSS Test:
Test Statistic 0.305544
p-value 0.100000
Lags Used 19.000000
Critical Value (10%) 0.347000
Critical Value (5%) 0.463000
Critical Value (2.5%) 0.574000
Critical Value (1%) 0.739000

KPSS results on **AdBudget**:
Results of KPSS Test:
Test Statistic 0.087946
p-value 0.100000
Lags Used 14.000000
Critical Value (10%) 0.347000
Critical Value (5%) 0.463000
Critical Value (2.5%) 0.574000
Critical Value (1%) 0.739000

KPSS results on **GDP**:
Results of KPSS Test:
Test Statistic 0.319751
p-value 0.100000
Lags Used 42.000000
Critical Value (10%) 0.347000
Critical Value (5%) 0.463000
Critical Value (2.5%) 0.574000
Critical Value (1%) 0.739000

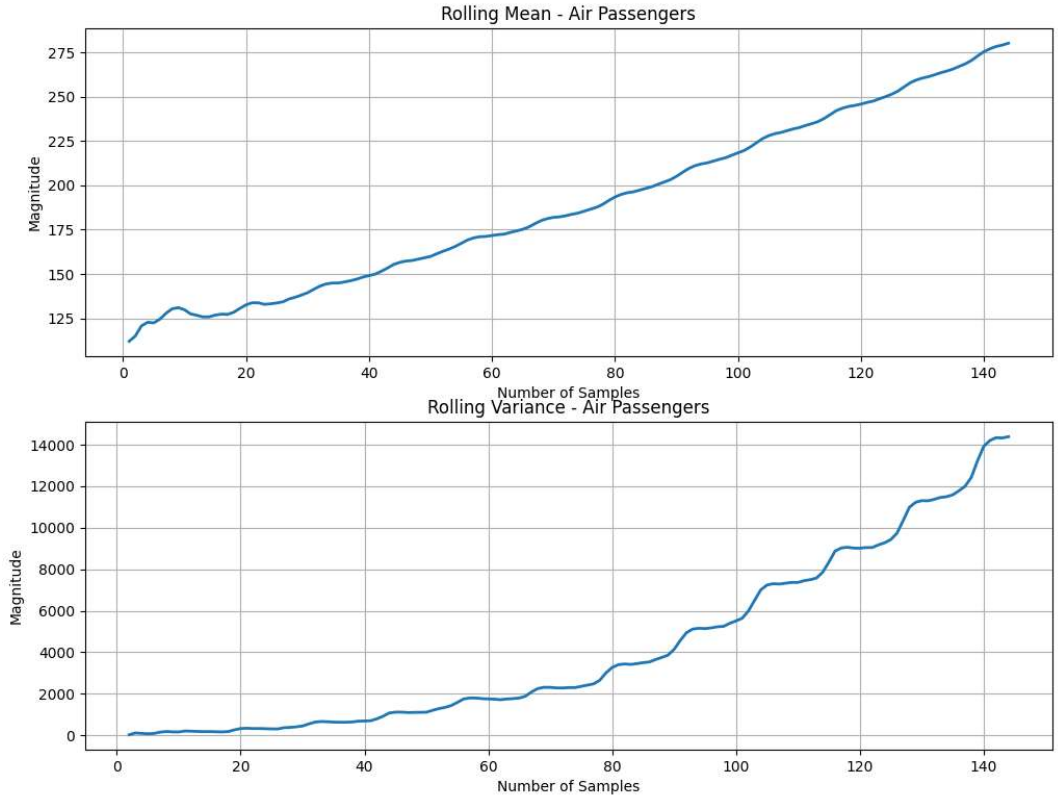
At 95% confidence or above, the KPSS test confirms that all three time-series Sales, AdBudget, and GDP are stationary since their p-values (0.1) are above the significance threshold. Hence, we fail to reject the null hypothesis, which states that the time series is stationary.

Q7.1 Plot the number of air passengers time-series.



Q7.2 Time-series statistics
Number of air passengers travelled each month on average would be 280.3 and has a variance of 14391.92 with a standard deviation of 119.97 Median: 265.5

Q7.3 Rolling mean and variance



Q7.4 Write down your observation about the plot of the mean and variance in the previous step. Is Air Passengers data stationary or not? Explain why?

The rolling mean and the variance of the Air Passengers time-series data do not appear to stabilize, instead, continue to increase. Hence the data is not stationary.

Two causes for this pattern:

a. Trend that is causing the mean to keep increasing
b. The height of the seasonal pattern (from the base) each year, or in other words, the number of air passengers growing each year is causing the variance to grow.

Q7.5

ADF results on the **number of air passengers**:
ADF Statistic: 0.815369
p-value: 0.991880
Critical Values:
1%: -3.482
5%: -2.884
10%: -2.579

The p-value of the ADF test result is higher than the significance threshold for 95% confidence interval and above. Hence, we fail to reject the null hypothesis, which states the time-series is non-stationary. This result is intuitive and aligns with the visual interpretation from steps 7.1 and 7.2 showing the plot of the time-series itself and the rolling mean & variance respectively.

Q7.6

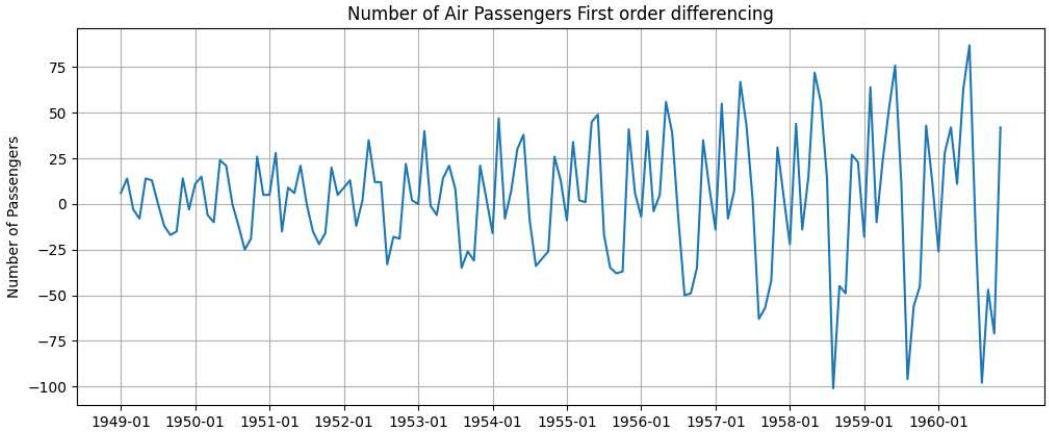
KPSS results on the **number of air passengers**:
Results of KPSS Test:

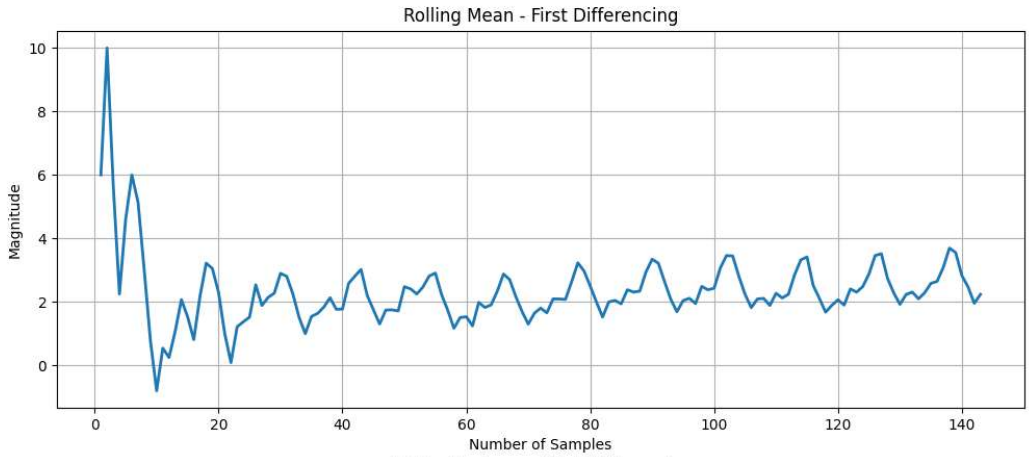
Test Statistic	1.651312
p-value	0.010000
Lags Used	8.000000
Critical Value (10%)	0.347000
Critical Value (5%)	0.463000
Critical Value (2.5%)	0.574000
Critical Value (1%)	0.739000

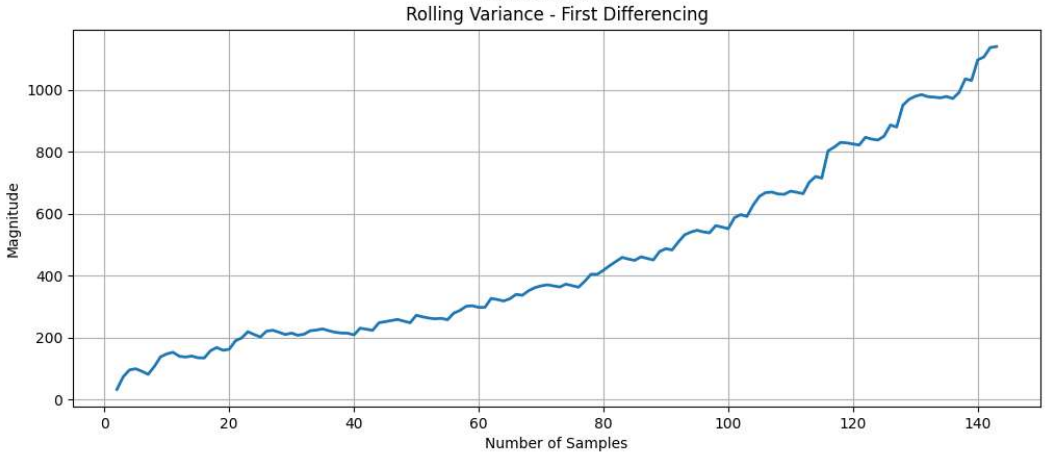
For a confidence level of 95% or above:
My conclusion is that as per KPSS test results, the time series is non-stationary as we reject the null hypothesis in favor of the alternate hypothesis. In the KPSS test, the null hypothesis and the alternate hypothesis are the vice-versa of ADF. Here the null hypothesis states the data is stationary whereas the alternate hypothesis signifies the time series is a non-stationary signal.

Q8.

a. 1st order non-seasonal differencing

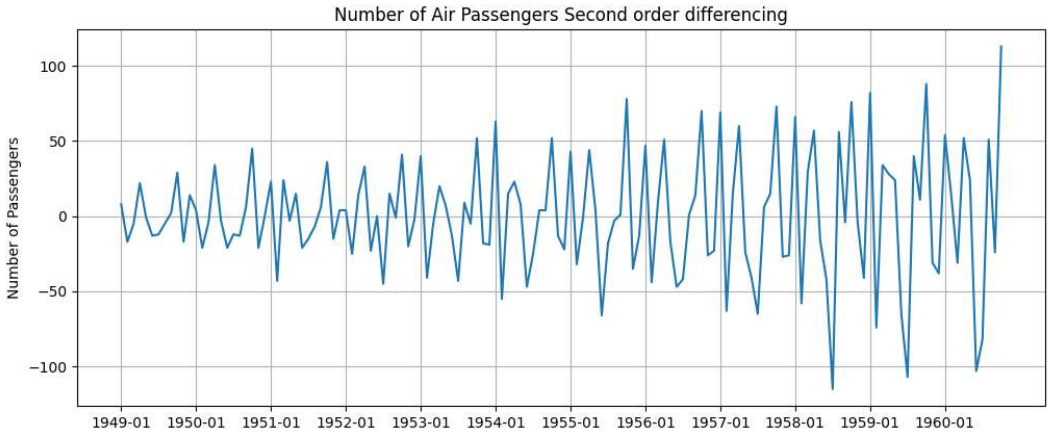


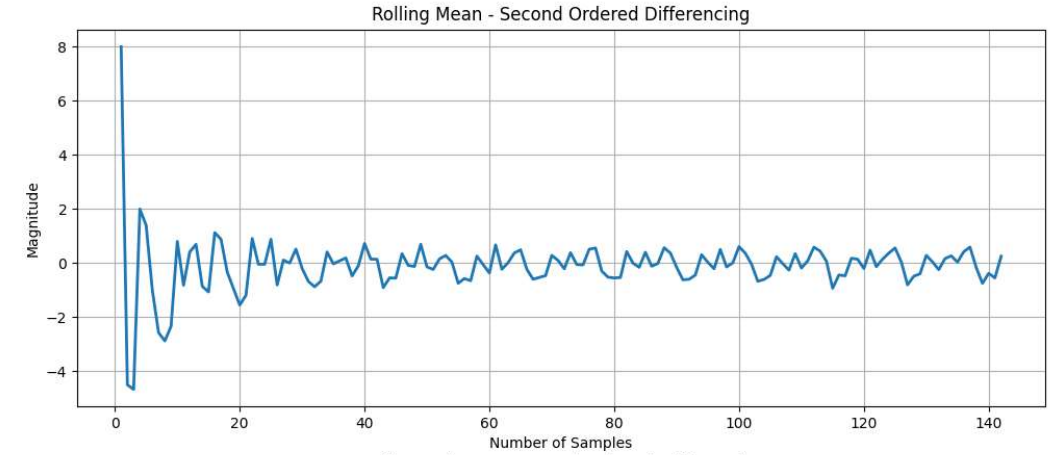


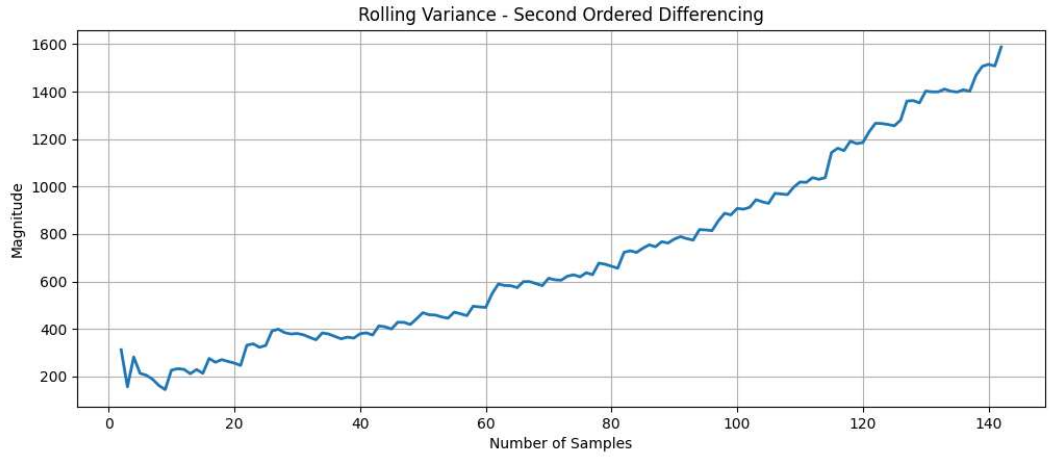


The trend is eradicated, but the seasonality exists. The rolling mean appears to fluctuate, which confirms the time-series has seasonality.

b. 2nd order differencing

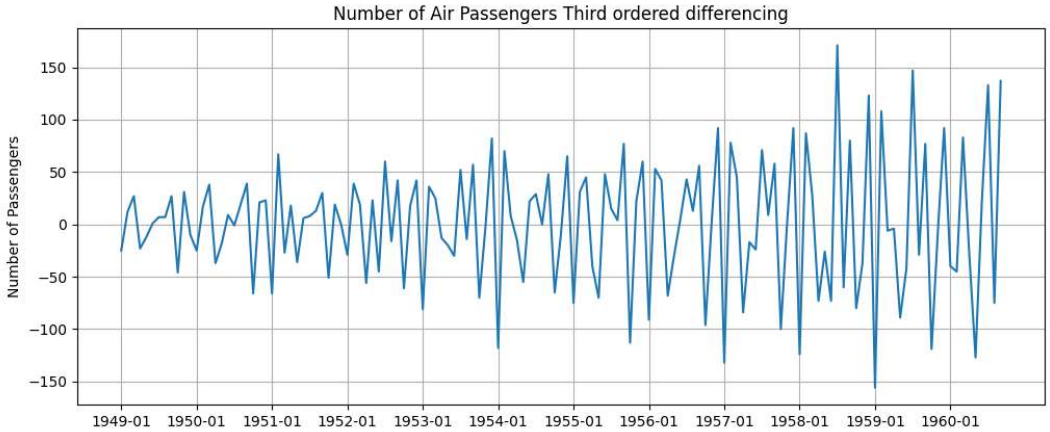


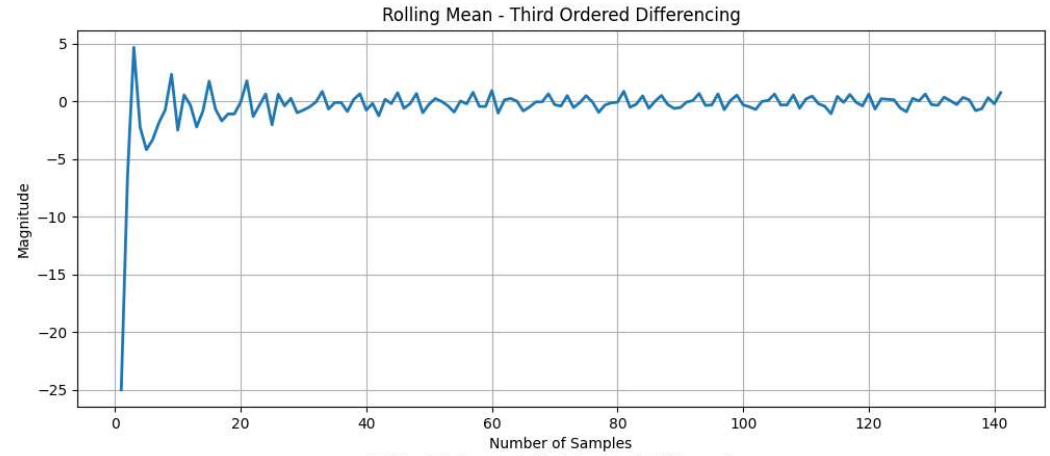


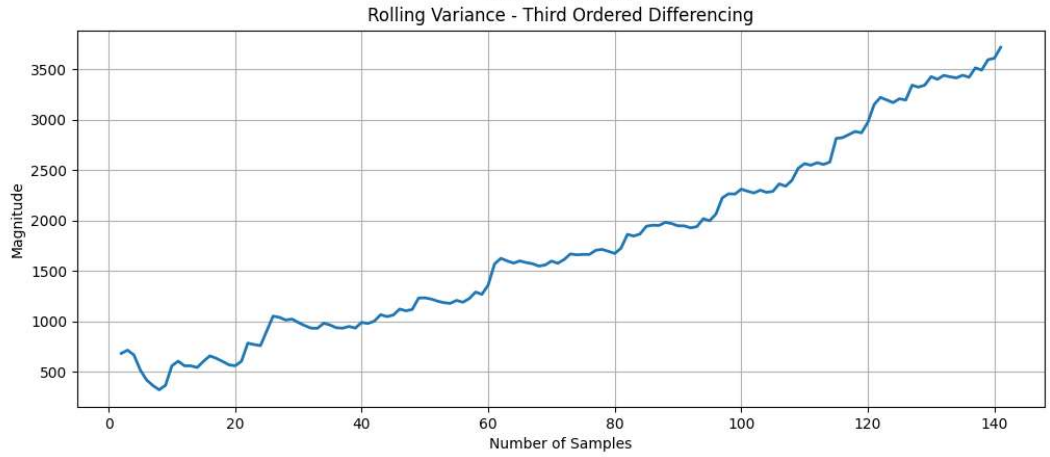


The rolling mean shows signs of beginning to stabilize after the second order differencing. However, the variance continues to grow as is intuitive from the first subplot that the data on the y-axis does expand over time.

c. 3rd order differencing

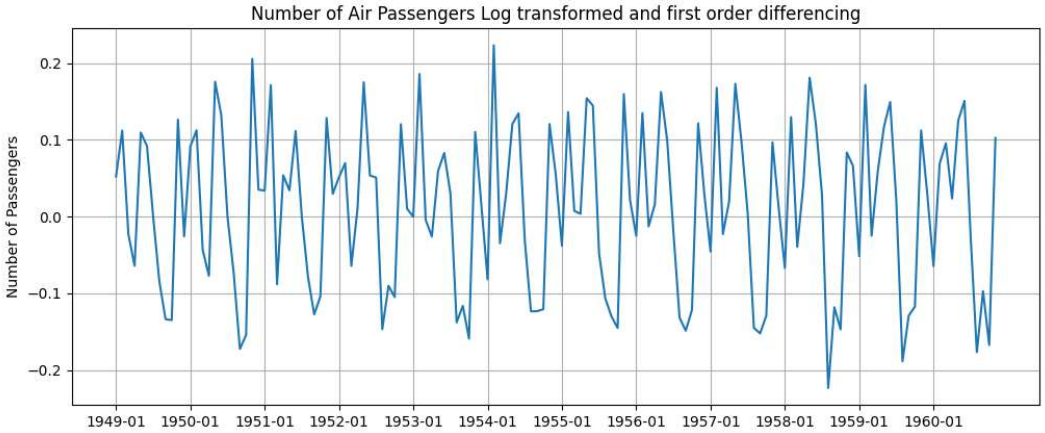


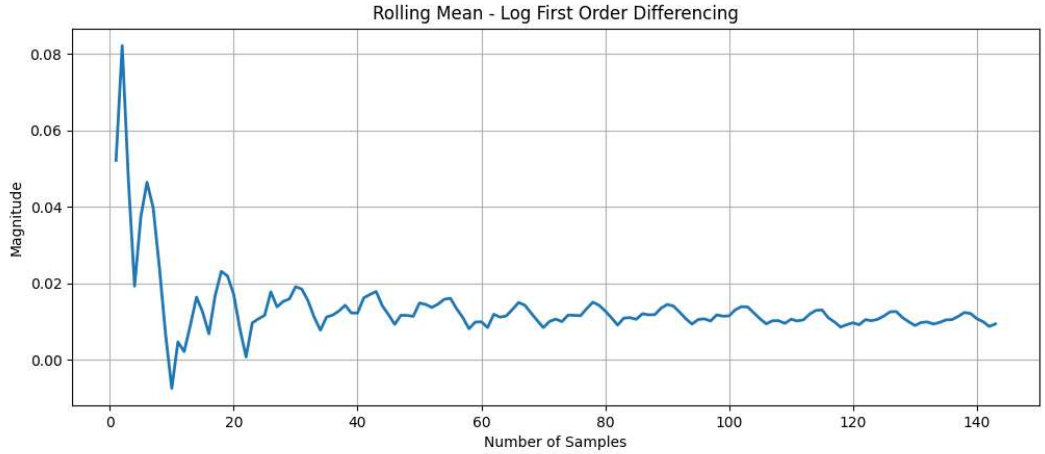


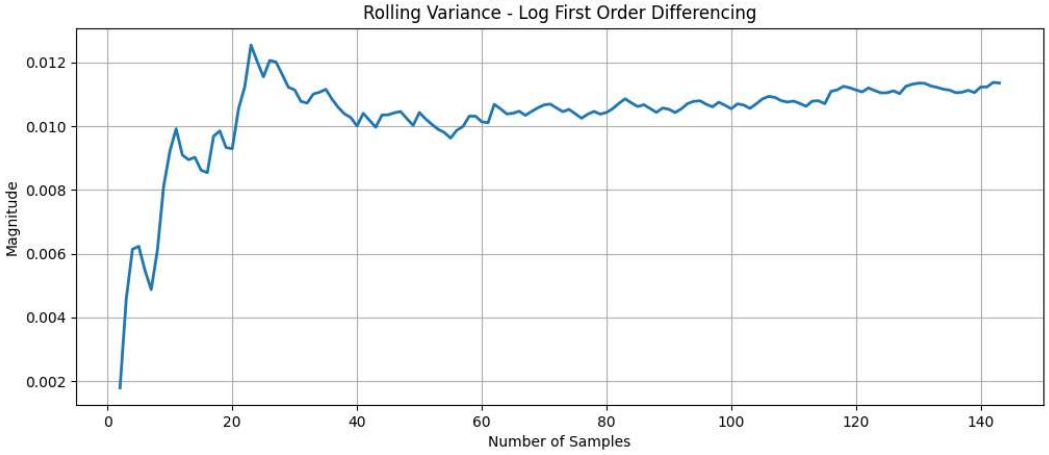


The rolling mean appears to have stabilized after the third-order differencing. However, higher order differencing has no effect on the seasonality that let the rolling variance grow.

d. Differencing following the log transformation







ADF results on the log-transformed and differenced **number of air passengers**:
ADF Statistic: -2.717135
p-value: 0.071120
Critical Values:
1%: -3.483
5%: -2.884
10%: -2.579

KPSS results on the number of the log-transformed and 1st order differenced **air passengers time-series**:
Results of KPSS Test:

Test Statistic	0.038304
p-value	0.100000
Lags Used	6.000000
Critical Value (10%)	0.347000
Critical Value (5%)	0.463000
Critical Value (2.5%)	0.574000
Critical Value (1%)	0.739000

KPSS does confirm that the time series is stationary.