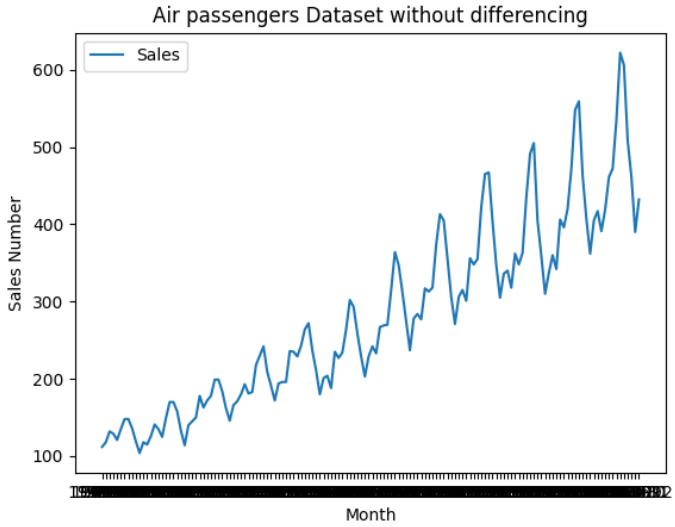
Time Series Modeling & Analysis
Homework 1
Bradley Reardon

1. Write a Python code to display the first 5 rows of the data on the console.

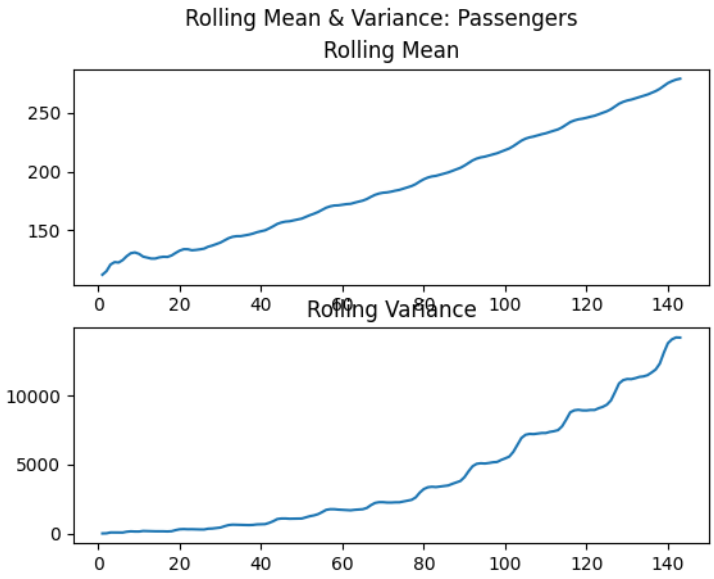
	Month	#Passengers
0	1949-01	112
1	1949-02	118
2	1949-03	132
3	1949-04	129
4	1949-05	121

2. Explore the dataset by plotting the entire dataset, where you can learn more about the data set pattern (trend, seasonality, cyclic, …). Add the label to the horizontal and vertical axis as Month and Sales Number. Add the title as “Air passengers Dataset without differencing”. Add an appropriate legend to your plot. Do you see any trend, seasonality, or cyclical behavior in the plotted dataset? If yes, what is it?



The trend is seasonal since similar change in trend occurs at a set interval of time. The number of passengers sharply increases in the latter half of the year.

3. Plot the rolling mean and rolling variance (the function developed in LAB #1). Is the raw dataset stationary or non-stationary? Explain why.



Based on the rolling mean and variance, the data seems to be non-stationary as there is a consistent increase in the metrics as additional samples are added.

4. Execute an ADF-test on the given dataset and display the ADF-test statistics. Interpret the ADFtest statistics. Is the dataset stationary or not with 95% confidence? Justify your answer.

```
ADF Statistic: 0.815369
p-value: 0.991880
Critical Values:
    1%: -3.482
    5%: -2.884
    10%: -2.579
```

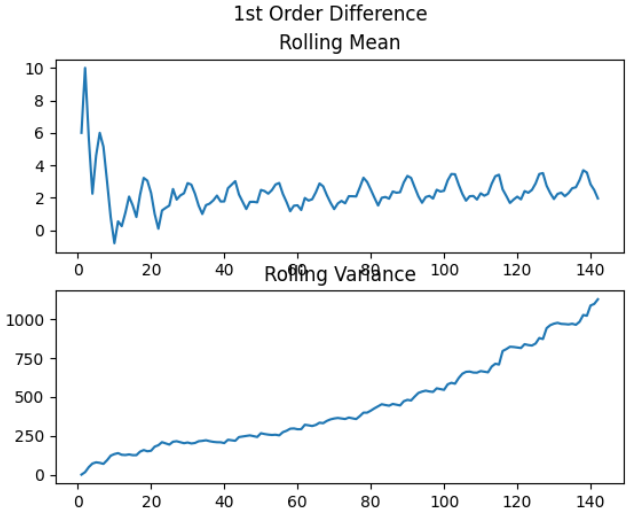
P-value is >0.05 and the test statistic is < the critical value at 5%, so we fail to reject the null hypothesis and assume the data is non-stationary.

5. Execute a KPSS-test on the given dataset and display the KPSS-test statistics. Interpret the KPSStest statistics. Is the dataset stationary or not with 95% confidence? Explain why.

```
Results of KPSS Test:
Test Statistic           1.651312
p-value                  0.010000
LagsUsed                 8.000000
Critical Value (10%)     0.347000
Critical Value (5%)      0.463000
Critical Value (2.5%)    0.574000
Critical Value (1%)      0.739000
dtype: float64
```

The test statistic for Passengers is higher than the critical value givena confidence interval of 95% (critical value 5%), which aligns with the p-value being <0.05. This means we reject the null hypothesis, assuming that the dataset is non-stationary.

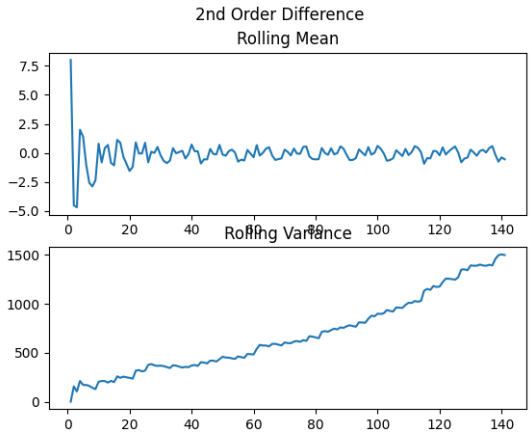
6. Perform a 1st order non-seasonal differencing and repeat step 3 and 4. Is the dataset stationary? Explain why.




```
ADF Statistic: -2.829267
p-value: 0.054213
Critical Values:
    1%: -3.482
    5%: -2.884
    10%: -2.579
```

ADF-test: p-value >0.05, reject null hypothesis; assume non-stationary.
Visual test looks to be non-stationary given the change in the metrics as samples are added.
1st order dataset is non-stationary.

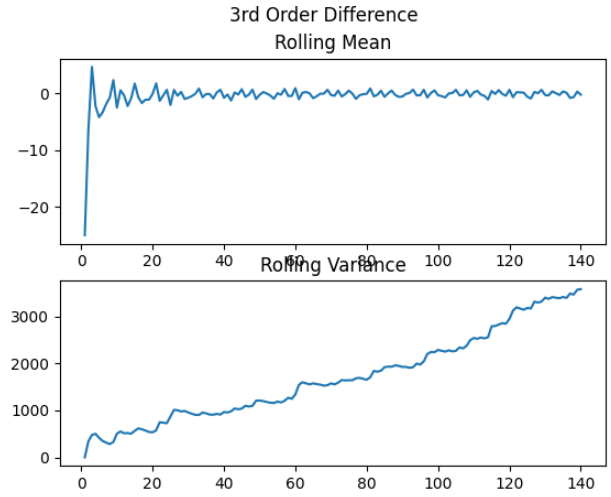
7. Perform a 2nd order non-seasonal differencing and repeat step 3 and 4. Is the dataset stationary? Explain why.




```
ADF Statistic: -16.384232
p-value: 0.000000
Critical Values:
    1%: -3.482
    5%: -2.884
    10%: -2.579
```

ADF-test: p-value <0.05, fail to reject null hypothesis; assume stationary.
The plots seem to present non-stationary, but this test is subjective.
2nd order dataset is stationary per the ADF-test results.

8. Perform a 3rd order non-seasonal differencing and repeat step 3 and 4. Is the dataset stationary? Explain why.




```
ADF Statistic: -9.434675
p-value: 0.000000
Critical Values:
    1%: -3.483
    5%: -2.885
    10%: -2.579
```

ADF-test: p-value <0.05, fail to reject null hypothesis; assume stationary.
The plots seem to present non-stationary, but this test is subjective.
3rd order dataset is stationary per the ADF-test results.

9. If the procedures in steps 5, 6 & 7 does not make the dataset stationary then perform a log transformation of the original raw dataset followed by a 1st order differencing. Repeat step 3 and 4. This step should make the dataset stationary which means the rolling mean and variance is stabilize and the ADF-test confirms stationarity.

A log transformation is not needed since both the 2nd and 3rd order datasets can be assumed to be stationary per the ADF-test results.