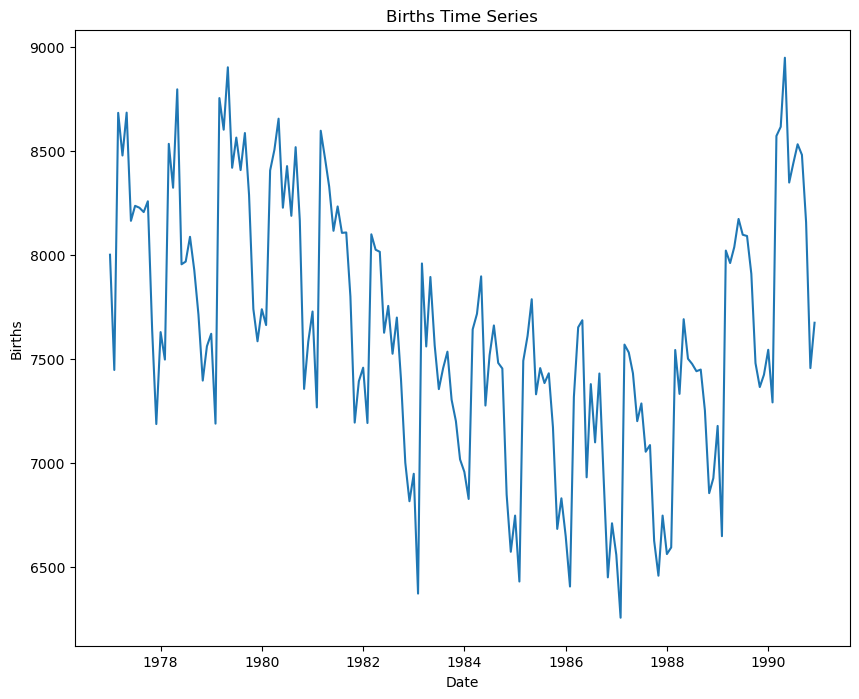
Assignment 2: Preparing and Exploring Time Series Data

1. To answer this question, you will need to download the DAILY\_BIRTHS data set.  This data set contains the number of daily births in Quebec, Canada from January 1, 1977 to Dec 31, 1990.  You will use the techniques that we have learned in class to practice accumulating, indexing and plotting a time series.
   1. Create a monthly time series for the total number of births occurring each month. Produce a time plot of the series and copy and paste the plot below. (time index = month, aggregation method = total)



* 1. Create a yearly time series for the average number of births occurring each year. Produce a time plot of the series and copy and paste the plot below. (time index = year, aggregation method = average)

A graph showing the growth of the stock market

Description automatically generated

1. In this question, you will be exploring a real data set in preparation for modeling. The data set includes hourly air pollution information from the nationally controlled Changping air-quality monitoring site in China. The data were provided by the Beijing Municipal Environmental Monitoring Center. The meteorological data are sourced from the weather station nearest the monitoring site and were provided by the China Meteorological Administration. The data are from the time period March 1, 2013 to February 28, 2017. Missing data are denoted as NA.

Data Dictionary:

|  |  |
| --- | --- |
| Variable Name | Variable Description |
| No | Row number |
| Year | Year |
| Month | Month |
| Day | Day |
| Hour | Hour |
| PM2.5 | PM2.5 concentration (ug/m^3) |
| PM10 | PM10 concentration (ug/m^3) |
| SO2 | SO2 concentration (ug/m^3) |
| NO2 | NO2 concentration (ug/m^3) |
| CO | CO concentration (ug/m^3) |
| O3 | O3 concentration (ug/m^3) |
| TEMP | Temperature (degrees Celsius) |
| PRES | Barometric pressure (hPa) |
| DEWP | Dew point temperature (degrees Celsius) |
| RAIN | Precipitation (mm) |
| Wd | Wind direction |
| WSPM | Wind speed (m/s) |
| Station | Name of the air quality monitoring site |

You will need to apply the data preparation and exploratory analysis techniques that you have learned up to this point in order to understand the patterns that are present in the data.

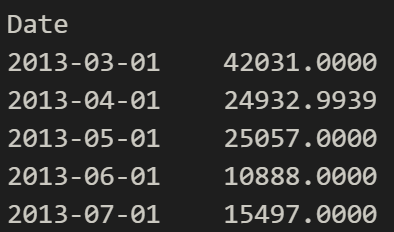
NOTE: You will be analyzing the NO2 concentrations, so your Python series will be based on that column from the data.

Please provide your answers below.

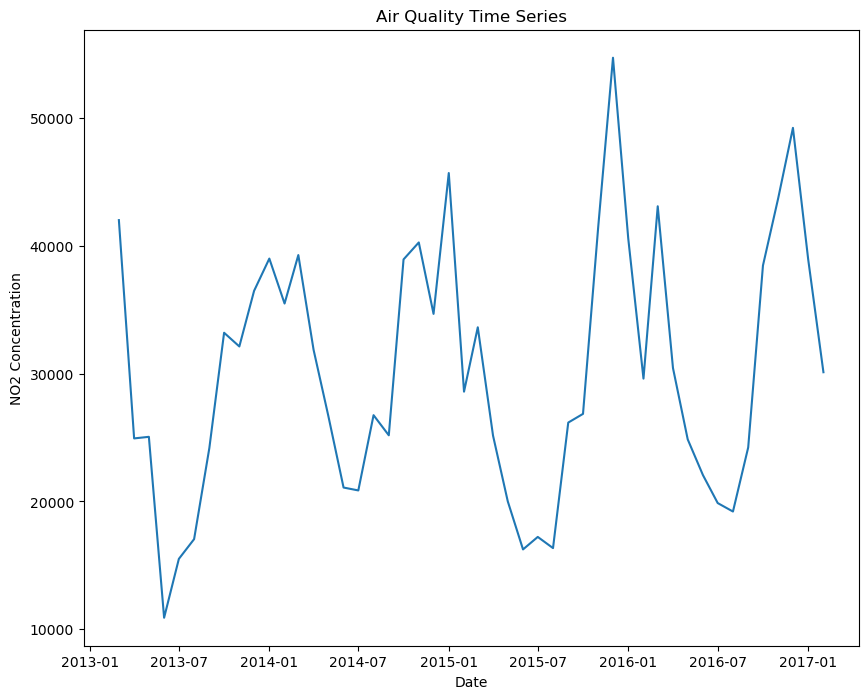
1. Read the PRSA\_Changping.csv file into Python. Are there any missing values present in the data?

Yes, there are missing values in the dataset.

1. Your ultimate goal with these data is to produce a monthly forecast of the total NO2 concentration at the Changping air quality monitoring station. With this in mind, you will need to accumulate the data using a monthly time index before continuing with your analysis. Print out the first few rows of the monthly data series and copy and paste them below.



1. Using the monthly series that you created in part (b), generate a time plot of data. Copy and paste the plot below. How would you describe this series (e.g., are there any outstanding features of the series – outliers, trend, seasonality)?



This graph has seasonality since it rises and falls the same way every period.

1. Is this a white noise series? Paste the relevant output from Python below. Be sure to state your hypotheses along with your conclusion.

Ho: The series is white noise

Ha: The series is not white noise

A screenshot of a black and white screen

Description automatically generated

Lb\_pvalue is < 0.05 so we are going to reject the null hypothesis and determine that the series is not white noise.

1. Generate plots of the autocorrelation function and the partial autocorrelation function for this series. Copy and paste the plots below. What do you learn about the series by looking at these plots?

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

1. Is this series stationary? Paste the relevant output from Python below. Be sure to state your hypotheses along with your conclusion.

Ho: Series is not stationary

Ha: Series is stationary

A black background with numbers and percentages

Description automatically generated



The p-value determined using the ADF test is < 0.05. Therefore, we reject the null hypothesis and determine that the series is stationary.