Lab 08 (Module 09)

Time Series

We will be using the Alcohol_Sales.csv data set for this part of the lab. The data set contains the monthly number of bottles sold from 1992 - 2019.

1. The date information is in a non-standard format. You will need to convert it to a datetime. The function `strptime()` is one way to do this as shown in this link.
2. Make the date value an index using `set_index()`
3. Plot the number of units sold vs time. Answer the following:
    a. Are data independent?
    b. Is the data stationery?
4. Make the data stationery using differencing. This link shows several methods for differencing. Use `shift()`.
5. The link above provides a function for determining if the differencing made the data stationery. It is reproduced here, removing the hard coded variable name. Use this function to test whether the series is stationary.

```
def get_stationarity(timeseries, varname):
    """Title: get_stationerity
    Function for testing whether time series data is stationery.

    Args:
        timeseries - Dateframe containing time series
        varname - Name of the column containing the series values
    """
    # rolling statistics
    rolling_mean = timeseries.rolling(window=12).mean()
    rolling_std = timeseries.rolling(window=12).std()

    # rolling statistics plot
    original = plt.plot(timeseries, color='blue', label='Original')
    mean = plt.plot(rolling_mean, color='red', label='Rolling Mean')
    std = plt.plot(rolling_std, color='black', label='Rolling Std')
    plt.legend(loc='best')
    plt.title('Rolling Mean & Standard Deviation')
    plt.show(block=False)

    # Dickey-Fuller test:
    result = adfuller(timeseries[varname])
    print('ADF Statistic: {}'.format(result[0]))
    print('p-value: {}'.format(result[1]))
    print('Critical Values:')
    for key, value in result[4].items():
```

```
        print('\t{}: {}'.format(key, value))
```

   a. Did differencing make the data stationery?
   b. Do you see another issue with the data? Hint: Think about the other regression assumptions, apart from independence.

6. Generate ACF and PACF plots of the differenced data. For how to create the plots, see this link. The function `plot_pacf` is also available in `statsmodels`.
   a. Would a time series model be appropriate?
   b. Will this require an ARIMA or a SARIMA model? Explain.