CS 795/895 –Introduction to Data Science, HW3

You have just been hired as an analyst for an investment firm. Your first assignment is to analyze data for stocks in the S&P 500. The S&P 500 is a stock index that contains the 500 largest publicly traded companies.

You have been given two sources of data to work with. The first is an XML file that contains the Symbol (ticker), company name, sector, and industry for every stock in the S&P 500, as of summer 2016.
The second is a CSV file that contains pricing information for stocks in the S&P 500 between August 2009 and August 2010. There is one row in the CSV file for every stock, on every date that the market was open. Each row contains the date as a string, the stock's ticker, the day's opening price, the day's high price, the day's low price, the day's closing price, and the volume traded that day.

Provided Files SP500_ind.csv and SP500_symbols.xml

Write a Python module in **Lastname-hw3.py** that includes the functions from the following activities.

- Read the .csv file into a DataFrame called "csv_data" and .xml file to a dictionary called "xml_dict" in your python module.
- Generate a list of unique symbol values from the csv_data and name the list "ticker" using unique() method.
- Complete the following functions in your python module

```
def ticker_find(xml_dict, ticker):
        """This function takes in the xml_dict and the list that contains a
        Symbol (ticker).  Return the name of the ticker
        Ex: for ticker "A", the function returns Agilent Technologies Inc
        """
def calc_avg_open(csv_data, ticker):
        """This function takes in the csv_data and a ticker.
         Return the average opening price for the stock as a float.
        """
def vwap(csv_data, ticker):
        """This function takes in the csv_data and a ticker.  Return the volume weighted average price (VWAP)
        of the stock.  In order to do this, first find the average price of the stock on each day.  Then, multiply
        that price with the volume on that day.  Take the sum of these values.  Finally, divide that value by the
        sum of all the volumes.
        (hint: average price for each day = (high + low + close)/3)
        """
```

Using the functions you have created:

1. Read the XML file and CSV file and,
2. Find the VWAP and average opening price associated with each of the tickers found.
   Use ticker_find with the xml_dict and ticker to find the associated name of the stock. Display the data as in the example below. (Hint: Write a loop that takes the "tickers" from the list and call the functions to display the name, average open and vwap values).
   American Express Co 39.92 39.72
   No data in SP500 23.04 22.85
   AutoZone Inc 169.52 167.22
3. To perform these calculations, you should call above functions in a logical order, with the appropriate parameters.
4. Note that stocks move in and out of the S&P 500. Some stocks may be represented in the CSV file, but not in the XML file (and vice-versa). Display "No data in SP500" for the names of these tickers.
5. You **cannot use** any regular python loops inside calc_avg_open and vwap functions. Instead use the DataFrame techniques learned in Pandas.

What to turn in:

Lastname-hw3.py should contain the following information at the top:

```
CS795
HW3
@author:
```