Directions: Please submit your Python code file, formatted as in previous assignments.

The grader should be able to run your code when it is placed in the same directory as the input data file. Be sure that your code loads any libraries you are using.

Background: The folder `Stocks.zip` contains nearly 300 files, each including daily data for a specific stock, with stock ticker symbol given in the file name. Each observation includes the following:

```
Date = date information recorded
Open = opening stock price
High = high stock price
Low = low stock price
Close = closing stock price
Volume = number of shares traded
Adj Close = closing price adjusted for stock splits (ignored for this assignment)
```

The time interval covered varies from stock to stock. There are some missing records, so the data is incomplete. Note that some dates are not present because the exchange is closed on weekends and holidays. Those are not missing records. Dates outside the range reported for a given stock are also not missing records, these are just considered to be unavailable. Answer the questions below based on the data available in the files.

1. Use the collective data to determine when the market was open from January 2, 2005 to December 31, 2014. (Do not use external data for this question.) Report the number of days the market was open for each year 2005-2014. (Include the year and the number of days in table form.)

2. Determine the total number of missing records for all stocks for the period 2005-2014.

3. For the period 2005-2014, find the 10 stocks (plus ties) that had the most missing records, and the 10 stocks (plus ties) with the fewest missing records. (For the latter, don't include stocks that have no records for 2005-2014.) Report the stocks and the number of missing records for each.

4. Repeat the previous question, this time for the stocks with the greatest and least proportion of missing records during 2005-2014. Report the stocks and the number of missing records for each.

5. Identify the top-10 dates (plus ties) in 2005-2014 that were missing from the most stocks.

6. For each stock, impute (fill in) the missing records using linear interpolation. For instance, if $d_1 < d_2 < d_3$ are dates, and P_1 and P_3 are opening prices on dates d_1 and d_3, respectively, then we estimate P_2 (the opening price on date d_2) with

$$P_2 = \frac{(d_3 - d_2)P_1 + (d_2 - d_1)P_3}{d_3 - d_1}$$

The same formula is used for the other missing values of High, Low, Close, and Volume.

Use the imputed values to recalculate the Python Index (see Assignment 10 for the formula) for the open dates in 2007-2013. (Remember that weekends and holidays are not open dates, so don't impute those.)

   (a) Find the Open, High, Low, and Close for the imputed Python Index for each day the market was open in October 2010. Give a table the includes the Date, Open, High, Low, and Close, with one date per row.

   (b) Determine the mean Open, High, Low, and Close imputed Python index for each month in 2008-2012, and report that in a table that includes the month and year together with the corresponding Open, High, Low, and Close.