5 HW #1B: Basic Pandas

Repeat HW #1A, this time using Pandas. In order to receive full credit, please turn in a document which is python code containing what would be run to return the data asked The following provides a template that you may wish to use.

The best approach to learning from these problems is to complete them using pen and paper, working by yourself and then using your group to double check your results. The First Five problems provide a short overview of the core concepts in the assignment, so make sure that you understand them. The Main Problems section contains questions which range from easy to very difficult. Remember to don't get stuck! If a problem is taking a long time or is too difficult, *use your group!*

```
import pandas as pd
import numpy as np

df2010 = pd.read_csv( '/Users/ncross/git/sqlnotes/newserver/data/2010.tdf',
        sep='\t', engine='python', names=['symb', 'retdate', 'opn', 'high', 'low',
        'cls', 'vol', 'exch'])

df2011 = pd.read_csv( '/Users/ncross/git/sqlnotes/newserver/data/2011.tdf',
        sep='\t', engine='python', names=['symb', 'retdate',  'opn', 'high', 'low',
        'cls', 'vol', 'exch'])

dffnd = pd.read_csv( '/Users/ncross/git/sqlnotes/newserver/data/fnd.tdf',
        sep='\t', engine='python', names=['gvkey', 'datadate', 'fyear', 'indfmr',
        'consol', 'popsrc', 'datafmt' , 'tic', 'cusip', 'conm', 'fyr', 'cash', 'dp',
        'ebitda', 'emp', 'invt', 'netinc', 'ppent', 'rev' , 'ui', 'cik'])

## Question #1
ans = df2010.loc[(df2010.loc[:, 'symb']=='AAPL'), :]
print(ans.head())

## Question #2
df2010C = df2010.copy()
df2010C = df2010C.loc[(df2010.loc[:, 'retdate'] =='07-Jan-2010') &
        (df2010.loc[:, 'symb']=='AAPL'), :]
df2010C.loc[:, 'diff'] = df2010C.loc[:, 'opn'] - df2010C.loc[:, 'cls']
ans = df2010C
print(ans.head())
```

First Five

Using the 2010 stocks data, write a query that returns the following.

1. All columns relating to AAPL.
2. All columns from the table and a column with the difference between open and close (open - close) for AAPL on the 7th of January.
3. Write a query which returns the stock symbol, the date, the open and close price for the top five differences (open - close) in 2010 for only those stocks on the New York Stock Exchange (NYSE).
4. The days when AAPL has a volume more than 20 million and where the high is $3 or more dollars greater than the low. Write it twice, once to return a series and once as a dataFrame.
5. Write a query which returns 3 columns: the return date, SYMB and volume, but only for stocks that

have a volume larger than 200 million

Main Problems

1. Write a query which returns all information about about Google (GOOG), NetFlix (NFLX), Amazon (AMZN) and Microsoft (MSFT) in 2010.
2. Write a query which returns the date and symbol of the largest "one-day gainer", that is the stock which has the highest close - open on the NYSE.
3. Write a query which returns the date and symbol of the largest "one-day percentage gainer", that is the stock which has the highest (close - open) / open on the NYSE.
4. Consider stocks on the NYSE which had a volume of more than 1 million. Which stocks (symbol and date) had their open price the same as their low and their closing price the same as their high?
5. Consider stocks on the NYSE which had a volume of more than 1 million. Which stocks (symbol and date) had their closing price the same as their low and their opening price the same as their high?
6. Consider stocks on the NYSE which had a volume of more than 1 million. Of those days which a stock had either (a) open = low and close = high or (b) open = high and close = low, which symbol and date has the largest volume traded?
7. Which company (ticker symbol) had the highest net income over all the years that are in the FND table?
8. Which company (ticker symbol) had the highest net income in fiscal year 2011 (use the FND table)?
9. Which company (ticker symbol) had the lowest, non-zero, net income over all years (use the FND table)?
10. Which company (ticker symbol), which had a net-income per employee over $1,000, had the largest number of employees (over all years)? Keep units in mind (use the FND table)!

Even more problems

1. Which company (ticker symbol) had the lowest, non-zero, net income in fiscal year 2011 (use the FND table)?
2. Of the companies which had more than 1,000 employees in 2011 which had the highest net income per employee in 2011 (use the FND table)?