

## 8 HW #3B: Subqueries in Pandas

Repeat HW #3A, 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 same requirements as in HW #1B apply.

Three terms that are defined in this assignment:

- **Profit Margin:** Net Income divided by Revenue.
- **Turnover:** Revenue divided by Inventory.
- **Dollar-volume:** This is the dollar value of stocks traded based on the closing price, so equal to the closing price of the shares traded multiplied by the volume.

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!*

```
## Initial Information
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'])
```

### First Five

1. Using the daily stock data from 2010, return an array of the unique trading days in 2010.
2. Return the symbol, return date and the dollar volume traded for the highest dollar volume traded stocks in 2010 on the NYSE.
3. Using the 2010 data, return the stock (symbol only) with the largest volume on Jan 11th that also appears on Dec 1st.
4. Using the 2010 data, return the stock symbol and a column called “HFlag” which is equal to 1 if the high - low is greater than 1 and zero otherwise. Only return those companies whose stock symbol begins or ends with “A”.
5. Write a query which returns (a) the date, (b) closing price and (c) a flag (“gt30”) which is equal to “1” when the closing price is greater than \$30.00 and “0” otherwise for “AAPL” in 2010.

### Main Problems

1. Return the list of symbols that exist in 2011, but not 2010.

2. Using the fnd data, return company name, year and the a column called “HFlag” which is equal to 1 if the company has a net income larger than \$1 Billion dollars and 0 otherwise. Only include those companies whose name begins with “B”.
3. Using the fnd data, which ticker symbols have a net income to employee ratio greater than \$1,000 in fiscal year 2010 and also have a net income between 20 and 30 million dollars in 2011?
4. Using the fnd data, which companies (company name), in fiscal year 2010 had a profit margin greater than 20%, turnover more than 2 and more than 10,000 employees?
5. The lowest five symbols by volume from Janaury 11th, 2010 that have a volume between 1 million and 10 million on December 1st, 2011. In other words, of those stocks which had between 1 and 10 million shares traded on December 1st, 2011, which five have the lowest volume traded on January 11th, 2010.
6. Of the stocks (symbols) that existed in 2011, but not in 2010, which had the highest closing price in 2011?
7. Which symbols were in the top 500 of dollar volume on the 2nd, 3rd and 4th days of February 2011 (The stock needs to be in the top 500 for all days)?
8. Of the symbols that had volume between 100,000 and 1,000,000 on the 2nd and 3rd of February 2011, which had volume greater than 5,000,000 on the 4th on February?
9. Generate the following dataset:
  - company name, ticker symbol, revenue for all companies whose name begin with “A” or “a”
  - A column, revflag which is 1 if revenue is greater than \$25,000,000 and 0 otherwise.
10. Generate the following dataset:
  - company name, ticker symbol, revenue, inventory and employee information from fiscal year 2010
  - A column called turnflag which is 1 for companies with turnover greater than 2, 0 otherwise
  - For a company to be included it must have revenue, inventory and employee all greater than zero for both 2010 and 2011

### Additional Problems

1. Of the stocks (ticker symbols) that have a net income to revenue ratio (called a profit margin) greater than 20%, which have more than 25,000 employees in fiscal year 2011?
2. We define revenue divided by inventory as the turnover. It expresses how many times the inventory has turned-over during the year in the form of sales. For companies (ticker symbols) with revenue between 1 and 2 million dollars in 2010, what company has the highest turnover in 2011?
3. Of the stocks (ticker symbol) that have profit margin greater than 20% in 2010, which had a profit margin greater than 30% in fiscal year 2011?
4. Of the stocks (ticker symbols) that have a net-income to employee ratio greater than \$1,000 in fiscal year 2010 and more than 1,000 employees in 2011, what is the highest profit margin in fiscal year 2011 and what is the ticker symbol?
5. Of the stocks (ticker symbols) that have a net-income to employee ratio greater than \$1,000 in fiscal year 2010 and more than 1,000 employees in 2011, what is the lowest profit margin in fiscal year 2011?

6. Of the stocks (ticker symbols) that have a net-income to employee ratio greater than \$1,000 in fiscal year 2010 and between 1,000 and 2,000 employees in 2011, what is the highest profit margin in fiscal year 2011 and what is the ticker symbol?
7. Of the companies (ticker symbols) with turnover between 1 and 2 in 2010, which companies also had a net income to employee ratio greater than \$1,000 in 2010?
8. Of the companies (ticker symbols) with turnover between 1 and 2 in 2010, which companies also had a net income to employee ratio greater than \$1,000 in 2011?
9. Write a select statement to generate the following datasets:
  - company name, ticker symbol, revenue, inventory and employee information from both 2010 and 2011 fiscal years.
  - A column called `invtfld` which is equal to 1 for companies with turnover between 2 and 3, 2 for turnover between 3 and 4 and 5 for turnover greater than 4 and zero otherwise.
  - A column called `invProfit` which is equal to 1 for companies with less than 20% profit margin and turnover greater than 2, 2 for companies with profit margin greater than 40% and turnover greater than 2 and 0 otherwise.
  - A column called `EmployeeProfit` which is equal to 0 for companies that have profit margins between 20% and 40% and have more than 10,000 employees, is equal to a company's profit margin if the margin is less than 20%, is equal to twice the number of employees (if it exists) if the profit margin is greater than 40% and is -1 otherwise.
10. What are the symbols and dollar volume traded for the companies with the top 5 dollar volume traded (based on closing price) on February 3rd 2010 (NYSE only)?