12 HW #5B: Aggregate Functions and Dates

Repeat HW #5A, 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 queries below rely on information from the stock return data. To load the data use the following commands. **Note: these are different than the previous commands because they load retdate as a date, rather than a string**

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

df2010D = pd.read_csv('/Users/ncross/git/sqlnotes/newserver/data/2010.tdf',
        sep='\t', engine='python', names=['symb', 'retdate', 'opn', 'high', 'low',
        'cls', 'vol', 'exch'], parse_dates=['retdate'])

df2011D = pd.read_csv('/Users/ncross/git/sqlnotes/newserver/data/2011.tdf',
        sep='\t', engine='python', names=['symb', 'retdate', 'opn', 'high', 'low',
        'cls', 'vol', 'exch'], parse_dates=['retdate'])

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. Which day of the week (0,1,2,...) had the largest number of shares traded?

2. Which day of the week (0,1,2,..) has the highest average shares traded?

3. Which day of the week-month (January-Monday, January-Tuesday, etc.) combination had the highest average return (close - open)? Note that both day of the week and month can be kept as integers.

4. Write a query which returns 3 columns and 5 rows with each row should represent a day of the week. One column should be the english day of the week ("Monday,", "Tuesday," etc.) while the next column should be equal to the average number of shares traded on that day from stocks that have a volume traded between 1 million and 2 million shares on that day ("C2"). The final column ("C3") should be the average number of shares traded on that day from stocks that had a volume traded outside of 1 million to 2 million.

5. Write a query which returns the maximum closing price for each symbol in 2010, sorting the results the final table from from high-to-low closing price.

Main Questions

1. Which quarter in 2010 has the most trading days?[7]

---

[7]Define Q1 as Jan-Mar, Q2 as Apr-Jun, etc.

2. Write a query which returns symbol and a column "DFlag", which is equal to 1 if the max closing price in 2010 is larger than 100, 2 if the max closing price in 2010 is between 50 and 100 and 3 if the max closing price is less than 50. There should be one row per symbol.

3. Write a query which returns the number of distinct symbols of each type of Dflag (from the previous problem). This should be 3 rows and 2 columns (one of the columns should indicate what each row means).

4. Write a query which returns the number of distinct symbols of each type of Dflag (from the previous problems), this should be 3 columns and a single row.

5. Calculate the number of distinct trading days per month in 2010. This should return 12 rows with 2 columns.

6. For each symbol, calculate the difference between the maximum and minimum closing price for December, 2010. Only include those stocks with 22 observations (there are 22 trading days in December, 2010).

7. Calculate the average difference between the maximum and minimum closing price for Tuesdays in January, 2010 for stocks on NYSE. The max and min should be calculated per-stock and then averaged. Only include those stocks with 4 observations which fulfill the criteria.[8]

8. Calculate the average closing price for Tuesday in January 2010 from the NYSE. Only include those stocks with 4 observations which fulfill the criteria. In other words, calculate the average price for each stock and then take the average of that number.

9. Calculate the average closing price for all stocks on the NYSE, by month, in 2010. Only include those stocks which have a closing price greater than \$100 in *2011*.

10. Calculate the average closing price in 2010 for all stocks (NYSE only) which are "not extreme". We define a stock as not extreme if the closing price is less than .1% of the max closing price (for all stocks) for the entire year. In other words, identify those stocks which are not extreme and then calculate their average price.

---

[8]There are 4 Tuesday trading days in January, 2010.