

17 HW #8AO: SQL Window Functions: [TBD]

THIS ONE IS REPLACED AND NEEDS TO BE FIXED

Using only the functions and syntax that we have learned in class, please provide a query to answer the following questions. If a dataset is not specified, please use the 2010 dataset. **Do not create any tables or views.**

Before beginning the assignment, *please read the data dictionary to better understand the data.* When doing so, keep an eye on data types for different columns as well as table organization.

- If no year information is provided for a financial question, assume 2010.
- If the query returns a significant number of rows, please only copy a few rows in your response.

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!*

First Five

1. What is the median daily closing price on January 4th, 2010?
2. For stocks in 2010, write a query which creates a dataset containing closing price, symbol, retdate and the nominal change between yesterday's closing price and today's opening price. Ignore holes in the data, so that if the stock misses a day the change in price is from the last time listed.
3. Write a query which returns five columns: symbol, return date, closing price, the moving average of the price (covering the last two days it was traded, but not including the current day) and the difference between that moving average and the current price.
4. Using the FND data and an analytic function, return the number of stocks alphabetically before each stock (e.g. if "A" was the first company would be 0, AA would be 1, etc.) in 2010. Make sure to only include each name once. Feel free to include or exclude company names that begin with a number.¹⁴
5. Without using an analytic function answer the same question as above.

Main Problems

1. For each stock in 2010, return the largest average daily return $((\text{close} - \text{open})/\text{open})$ by the first letter of the symbol. In other words, there should be one row for each first letter of the ticker symbol. The dataset should return (a) the symbol which has the largest, (b) the total number of symbols which begin with that character and (c) the average daily return for the symbol.
2. Repeat the above, this time without using an analytic function. Make sure that you *aren't* joining on a float as joining on a floating variable will lead to uneven results.
3. For stocks in 2010, write a query which creates a dataset containing closing price, symbol, retdate and the nominal change between yesterday's stock (symbol) price and today. If there is a missing day then the nominal change should be missing (which is different from the above question).

¹⁴Comparisons of the form $\text{string} \leq \text{string}$ do alphabetical comparison. Also keep in mind that you can join using any conditional expression.

4. Return a dataset which contains symbol and the number of trading days that the price is within 10% of the max price for that year for each stock. For example, if the max price of a symbol is \$100, then return the number of times that price of that stock is ≥ 90 .
5. For each symbol return the number of days it took to reach its maximum closing price for that year. If a stock is not traded on a day, then that should *not* count toward the total days. Note that there should be one row per symbol in the final dataset.
6. Repeat the previous question without using any analytic functions.
7. For each stock symbol return the number of days it took to reach its maximum closing price for that year. If a stock is not traded on a day, then it should count toward the total days.

Extra Problems

1. In the Transaction data, what percentage of users, who start by purchasing a Unit end up Subscribing?
2. What percentage of users, in the transaction data, purchase both a Unit and a subscription?
3. What is the average amount of time between Unit Purchases?
4. Calculate the 25, 50 and 75 percentile of the amount of revenue generated in the first six months (per user). Only include those users who made their first purchase more than six months ago. Make sure that this query moves with time: if I run this query next month it should return updated data.