

FE513 Practical aspects of database design

Final Exam

Please submit a SINGLE pdf report containing codes and outputs. Provide necessary explanations and comments to each question. If any plagiarism is found, you will NOT get any grade for the final exam.

1 SQL(50 pt)

1. Import given bank data into PostgreSQL database.
2. Create a primary key for the import table.
3. Find the highest asset observation for each bank.(i.e. There are 4 observations for each bank in a year. If there're 100 unique banks in total, then the result contains 100 observations. Each observation may belong to different quarters.) Sort the resulting table according to asset value. Report the first 10 observations of output table.
4. Show the query plan for question 1.3 using EXPLAIN tool.
5. Given the highest asset table from question 1.3, count how many observations are there for each quarter.
6. For the whole sample data, how many observations have asset value higher than 100,000 and liability value smaller than 100,000.
7. Each observation was given an 'idx' number. Find the average liability of observation with odd 'idx' number.
8. Find the average liability of observation with even 'idx' number. What's the difference between these two average number.
9. For each bank find all records with increased asset. The record with increase asset means one record's asset value is larger than the one of previous quarter. (For instance a bank (id: 123) has asset 30,000 in 3/31/02, asset 20,000 in 6/30/02 and asset 25,000 in 9/30/02. Then the record with bank id (123), asset value (25,000) and date (9/30/02) is recorded. Because its asset value is larger than asset value in 6/30/02.) Report the first 10 observation of output table.

2 User-defined Function in R(35 pt)

Write a function in R achieving following targets: (In the following, it describes the functionality that your function need to have. Do NOT write an individual function for each requirement. The desired answer is one single function who has the following methods.)

1. Download daily stock data using a given stock ticker for a given time period (set starting time, ending time and stock ticker as input variable).
2. Get the adjusted close price and consider this price data only for following tasks.
3. Perform a rolling window estimation on stock price vector to calculate the mean and standard deviation. For each time, you keep using a fixed size of data to calculate the mean and the standard deviation. (e.g. The whole sample contains 5 data points. You can perform a rolling window estimation with a window size of 2. In the first window (subset), you calculate the mean and standard deviation (std) using first two data points. Then you will roll the window one step ahead and calculate the mean and std using the second and the third data points. You repeat the same procedure using the fixed size of data until the end of whole sample.) The window size needs to be set as a input variable.
4. Store the statistical result of Q 2.3 into a dataframe. Plot this statistical dataframe using scatter plot. In the plot, x axis represents the index for each rolling window and y axis represents the statistical values. (Make sure you use different colors to differentiate between mean value and std value. Don't forget to include an legend.)
5. Return the statistical dataframe.
6. Test your function with suitable parameters.(1 or 2-year data and a rolling window size of 20). If you got a huge dataframe for the output, do NOT print the whole sample. Showing a part of it is enough.

To summarize, your function takes 4 arguments as inputs and return 1 object as output.

- Inputs: stock ticker, start time, end time, rolling window size
- Outputs: a dataframe

3 PostgreSQL API in R (15 pt)

Complete following tasks in R using PostgreSQL API:

- 1. Make a connection to your local PostgreSQL database.

- 2. Query the PostgreSQL database via API to get the original bank data. (The bank data that you import to PostgreSQL database in Q 1.1) Store the data into a dataframe.
- 3. Calculate asset growth rate for each quarter and each bank. (asset growth rate = (current quarter value - previous quarter value) / previous quarter value). The result start from second quarter, since we don't have all necessary data for first quarter calculation. Store the calculation result in a data frame.
- 4. Export the dataframe of Q 3.3 to the PostgreSQL database via API.