Based on the expected output, I'll need to create an SCD type 2 dimension table. First, I'll do a quick profile of the data. I'll check to see if there are any null values in the data and do a group by to check for duplicates. It turns out neither are issues so we can skip handling these.

The question defines a closed outlet as one that has no transactions for 30 consecutive days. This means that I'll need to think about creating records for the in-between periods when there are no recorded transactions in order to capture the closed periods in the SCD table.

To determine the number of days elapsed between transactions, we'll need to use the LAG function on the current transaction dates. We can then find the difference in dates. We'll add a status of 'open' to all of them since they are all transactions that occurred.

Now we'll need to add the records for closed periods and set their status as 'closed'. The 'open' records show the number of days between transactions over 30, so we'll use that to filter them. We add 1 to the previous transaction date to get the start of the 'closed' period, and subtract 1 from the current transaction date since it was closed up until the day before the new transaction was recorded.

We union the closed records with the original set of open records and use LAG on the status column to figure out when the status changes by comparing the current status to the previous status. If they're different, we mark the change, and use a rolling sum over the window to identify change streaks. We can later use this to group each streak change to get the valid_from and valid_to dates.

We now set the max streak to the same window of data (we'll come back to this), followed by getting the LEAD transaction date. This will be used for the valid_to date shortly.

Finally, we group the data and we use the MIN() of the current transaction date to set the valid_from date. We then compare the change streak against the group's max change streak value to determine the latest status period for each shop and set the valid_to date to far in the future to indicate that it is the current record. If the streak is less than the max streak for the group, we take the maximum of the LEAD transaction date and subtract by 1 to get the valid_to date.

To determine the number of stores closed in 2021 or 2022, we'll filter to get only the stores that were closed, and then sum them up. Each store appears multiple times, so we use distinct to get the number of unique stores that were closed.

To determine stores open at the start of each quarter, we'll create a CTE with the start dates for each quarter and join that to the SCD table. The join condition will implicitly filter the data to ensure that we only match against stores that were open at the start of each quarter.