Data 607 - Extra Credit - Window Functions

William Jasmine

2022-09-14

The chunk below downloads the time series data from the "OpenIntro" package. More specifically, the data imported below is from the `sp500_1950_2018` dataset and includes daily financial metrics for the S&P 500 market for all trading days from 1950-2018.

```
data(sp500_1950_2018)
```

The query below demonstrates how to calculate both the the year-to-date average and the six-day moving averages for the opening and high daily price of the S&P 500 market. For simplicity's sake, the only data used for this analysis was those observations between 2018-01-01 and 2018-12-31.

```
query <- "
WITH prep_data AS (
  SELECT
    *,
    CEIL((row_num - 1) / 6) + 1 as week
  FROM (
    SELECT
      *,
      ROW_NUMBER() OVER(ORDER BY date) as row_num
    FROM sp500_1950_2018
    WHERE
      date BETWEEN '2018-01-01' AND '2018-12-31'
    ORDER BY date
  )
)

SELECT
  date,
  open,
  high,
  AVG(open) OVER(ORDER BY date) as open_YTD_avg,
  AVG(high) OVER(ORDER BY date) as high_YTD_avg,
  AVG(open) OVER(PARTITION BY week ORDER BY week) as open_6_day_rolling_avg,
  AVG(high) OVER(PARTITION BY week ORDER BY week) as high_6_day_rolling_avg
FROM prep_data
WHERE
  date BETWEEN '2018-01-01' AND '2018-12-31'
ORDER BY 1
"
df <- sqldf(query)
head(df)
```

```
##         Date    Open    High open_YTD_avg high_YTD_avg open_6_day_rolling_avg
## 1 2018-01-02 2683.73 2695.89     2683.730     2695.890               2721.007
```

```
## 2 2018-01-03 2697.85 2714.37     2690.790     2705.130                2721.007
## 3 2018-01-04 2719.31 2729.29     2700.297     2713.183                2721.007
## 4 2018-01-05 2731.33 2743.45     2708.055     2720.750                2721.007
## 5 2018-01-08 2742.67 2748.51     2714.978     2726.302                2721.007
## 6 2018-01-09 2751.15 2759.14     2721.007     2731.775                2721.007
##   high_6_day_rolling_avg
## 1               2731.775
## 2               2731.775
## 3               2731.775
## 4               2731.775
## 5               2731.775
## 6               2731.775
```

The `prep_data` CTE above shows how to use the `ROW_NUMBER` function to determine which 6-day "week" the observation falls into, and then the following query uses the CTE and window functions to determine the YTD and 6-day rolling average of both the opening and high daily price of of the S&P 500 stock index.