

This is my assignment instruction:

Find (or ask an LLM to generate!) a dataset that includes time series for two or more separate items. For example, you could use end of day stock or cryptocurrency prices since Jan 1, 2022 for several instruments.

Use window functions (in SQL or dplyr) to calculate the year-to-date average and the six-day moving averages for each item.

You may work in a small team on this assignment.

I choose this dataset

[https://www.kaggle.com/datasets/ibrahimshahrukh/tesla-stock-price-historical-dataset-2010-2025?](https://www.kaggle.com/datasets/ibrahimshahrukh/tesla-stock-price-historical-dataset-2010-2025?select=tesla_stock_data_2010_2025.csv)
`select=tesla_stock_data_2010_2025.csv`

can you give me motivation why tesla data is compatible

Your chosen dataset — the Tesla historical stock price dataset from Kaggle — is a strong and fully compatible choice for this assignment. Here's a clear motivation you can use in your submission:

Motivation for Using Tesla Stock Data (2010–2025)

I selected the **Tesla Stock Price Historical Dataset (2010–2025)** from Kaggle because it satisfies all the assignment requirements and provides meaningful time-series financial data suitable for window function analysis.

1 It Contains Clear Time Series Data

The dataset includes daily stock prices for Tesla, Inc., with a date column and numerical variables such as:

- Open price
- Close price
- High and Low prices
- Volume

Because the data is recorded daily over many years, it forms a structured time series — ideal for calculating moving averages and year-to-date metrics.

2 It Allows Window Function Calculations

The dataset supports:

✓ Year-to-Date (YTD) Average

Using a window function partitioned by year and ordered by date, we can calculate cumulative averages from January 1 through each trading day of the year.

Example logic:

- Partition by `YEAR(date)`
- Order by `date`
- Use cumulative average window frame

This directly satisfies the assignment requirement to compute year-to-date averages.

✓ Six-Day Moving Average

Using a rolling window frame (e.g., `ROWS BETWEEN 5 PRECEDING AND CURRENT ROW`), we can calculate a 6-day moving average of the closing price.

This demonstrates practical use of window functions for smoothing time series data.

3 Financial Data is Ideal for Moving Averages

Stock prices are commonly analyzed using moving averages in finance. A 6-day moving average helps:

- Smooth short-term volatility
- Identify trends
- Reduce noise in daily price movements

Therefore, the dataset is not only technically suitable but also conceptually meaningful for time-series analysis.

4 Can Be Extended to Multiple Items

Although the Kaggle dataset focuses on Tesla alone, it can easily be extended by:

- Combining it with stock data from other companies (e.g., Apple, Microsoft)
- Or adding cryptocurrency price datasets

This would allow partitioning by stock ticker and calculating window functions for multiple instruments, fully matching the requirement of “two or more separate items.”

5 Large Enough for Robust Analysis

The dataset spans 2010–2025, providing:

- Multiple market cycles
- High volatility periods
- Thousands of observations

This ensures reliable moving average and YTD calculations.

Conclusion

The Tesla stock dataset is compatible with the assignment because:

- It contains structured daily time-series data
- It supports window function calculations
- It enables both year-to-date averages and rolling six-day averages
- It is financially meaningful and analytically rich
- It can be expanded to include multiple items

If you'd like, I can also help you write:

- The SQL window function code
- The dplyr solution in R
- Or a short paragraph formatted exactly for submission 😊



why tesla stock gets lots of attention

Tesla, Inc. gets a lot of attention for several interconnected reasons — financial, technological, and cultural:

1 Leadership & Public Persona

Tesla is strongly associated with its CEO, Elon Musk, one of the most high-profile business figures in the world.

- His social media activity frequently moves markets.
- His involvement in other major companies like SpaceX and X Corp. keeps him constantly in the news.
- Investor sentiment often reacts to his public statements.

Because Musk is highly visible, Tesla stock receives outsized media coverage.

2 Pioneer in Electric Vehicles (EVs)

Tesla is considered a leader in the modern electric vehicle industry.

- It helped push EVs into the mainstream.
- It competes with major automakers transitioning from gasoline vehicles.
- Governments worldwide are promoting EV adoption, increasing relevance.

Investors view Tesla not just as a car company, but as a technology and clean-energy company.

3 High Volatility

Tesla stock is known for large price swings.

- Rapid growth periods
- Sharp corrections
- Strong reactions to earnings announcements

High volatility attracts:

- Short-term traders
- Options traders
- Media attention

Volatile stocks naturally generate more headlines.

4 Massive Market Growth

Tesla experienced extraordinary stock appreciation in certain periods (especially 2020–2021).

At one point, its market capitalization surpassed many traditional automakers combined. This dramatic growth made it a major topic in financial markets.

5 Innovation Narrative


Tesla is involved in multiple advanced technologies:

- Autonomous driving (Full Self-Driving software)
- Battery innovation

- Energy storage and solar products
- AI development

This positions Tesla as more than an automotive company — investors often see it as a

ChatGPT   Free offer 

 Temporary Chat 

6 Retail Investor Popularity

Tesla is one of the most actively traded stocks among retail investors.

- Frequently discussed on online investing forums
- Strong fan and critic communities
- Heavy social media engagement

High retail participation amplifies attention.

7 Broader Economic Impact

Tesla is part of major stock indices like the S&P 500, meaning:

- It influences index funds
- It affects retirement portfolios
- Institutional investors track it closely

Short Summary (If You Need a Concise Answer for Assignment)

Tesla stock receives significant attention because of its high volatility, rapid growth history, strong association with Elon Musk, leadership in the electric vehicle industry, and its

positioning as both a technology and automotive company. Its market influence and heavy retail investor interest further amplify media coverage and public focus.

If you want, I can also help you turn this into a short academic-style paragraph for submission.



...

what is Anticipated Challenges in this dataset and task

Here are well-structured **anticipated challenges** you can include in your assignment when using the Tesla, Inc. stock dataset for time-series window analysis:

Anticipated Challenges in the Dataset and Task

1 Data Quality Issues

Stock datasets often contain:

- Missing trading days (weekends and holidays)
- Possible null values
- Stock splits or adjusted price changes
- Duplicate rows

If not handled properly, these can distort:

- Year-to-date (YTD) averages
- Six-day moving averages

You may need to clean the data before applying window functions.

2 Non-Trading Days

Financial markets are closed on weekends and holidays.

This creates gaps in dates, meaning:

- A "6-day moving average" refers to 6 trading days, not 6 calendar days.
- Calculations must be based on ordered rows, not calendar intervals.

In SQL or dplyr, this requires careful ordering by date.

3 Large Dataset Size

The dataset spans 2010–2025, which may include thousands of daily observations.

Potential issues:

- Slower query execution
 - Memory limitations (especially in R)
 - Performance considerations when using window functions
-

4 Year-to-Date Partitioning Complexity

To calculate YTD averages correctly, you must:

- Extract the year from the date
- Partition by year
- Reset the cumulative average at January 1

Mistakes in partitioning can cause:

- Averages to carry over across years
 - Incorrect cumulative calculations
-

5 Initial Period Moving Average Problem

For the first few rows of each year:

- There are fewer than six previous observations
- The 6-day moving average may return NULL (SQL)
- Or partial averages (depending on settings)

You must decide whether to:

- Allow partial windows
 - Filter out early rows
 - Or handle them explicitly
-

6 Only One Item in the Dataset

The assignment requires "two or more separate items."

The Tesla dataset contains only one company:

- You may need to merge it with another stock dataset
- Or create a simulated second instrument

Without multiple items, you cannot demonstrate partitioning by item.

7 High Volatility

Tesla stock is known for large price swings.

This can:

- Create extreme values
 - Influence moving averages heavily
 - Make interpretation more complex
-

8 Understanding Window Frame Definitions

In SQL, specifying the correct window frame is critical:

Example:

```
sql
```

 Copy code

```
ROWS BETWEEN 5 PRECEDING AND CURRENT ROW
```

If incorrectly defined (e.g., RANGE instead of ROWS), the moving average may be computed incorrectly.

Understanding the difference is essential.

Short Academic Version (If Needed)

The main anticipated challenges include data cleaning, handling non-trading days, correct year-based partitioning for YTD calculations, defining proper window frames for six-day moving averages, and addressing the limitation that the dataset contains only one instrument. Additionally, volatility and performance considerations may impact analysis accuracy and efficiency.

If you'd like, I can also help you write a brief "Limitations" section to strengthen your report.

