Learning plan2_gakyeong_bae

Detail

- Study 1 hour on M, W, F

week 7-8 : Understand Temporal/ Spatial analysis

Topic :

- Lazy evaluation
- exploding data
- window function
- sliding-window
- forward chaining validation
- sedona

Resource :

- https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.Window.partitionBy.html -https://sedona.apache.org/latest-snapshot/ -https://sparkbyexamples.com/pyspark-tutorial/#google_vignette

what I will learn:

1. How Spark delays execution until an action (show(), collect(), count()) is called. How lazy evaluation helps optimize queries by minimizing unnecessary computations.
2. How to flatten arrays, structs, and nested JSON in PySpark. How explode() and posexplode() work for expanding data. Performance considerations when working with large datasets.
3. How to use pyspark.sql.Window for partitioning data. How to apply moving averages, cumulative sums, and trend analysis.
4. How to load spatial data (shapefiles, GeoJSON, WKT) in Sedona. How to perform spatial joins between census block groups and ZIP codes.