Options Data Ingestion and Analysis Exercise

Task

Your firm has decided to subscribe to a new options data vendor. While the deal is not done yet, they have provided a sample CSV file containing two weeks of data for SPX. Your firm wants you to use this sample to build out an ingestion and analysis process for the data.

Note: The sample data provided differs from the production format. While this sample contains two weeks of options data for a single ticker (SPX) in one file, in production we will receive:

- Daily files (one per day)
- Each file containing options data for all tickers
- Files arriving each business day for processing

Please design your solution with this production scenario in mind, even though you're working with the sample format.

Requirements

1. **Data Ingestion**: Load the raw data into a database for safekeeping (PostgreSQL or SQLite)
2. **Analysis**: Implement at least 2 of the suggested analyses below, saving results in new database table(s)
3. **Testing:** Include unit tests for your analysis logic
4. **Documentation**: Include a README explaining your approach and how to run your solution

Deliverables

- Script(s) to ingest the data
- SQL or code to perform your chosen analyses
- Brief documentation of your design decisions

Please use Java or Python for any code written.

Note: We expect this to take approximately 4-6 hours. Focus on code quality and correctness rather than implementing every possible analysis.

Input File Specification

Column Name	Description
Ticker	Unique identifier for the underlying asset (e.g., SPX)
Date	The observation date (i.e., when this row of data was captured)
Expiration	The expiration date of the option contract
T	Time to expiration in days
Strike	Strike price of the option
CallPut	Type of option: "C" for Call, "P" for Put
symbol	Ticker-like string identifying the specific option
BestBid	Current best bid price (highest price a buyer is willing to pay)
BestOffer	Current best ask/offer price (lowest price a seller will accept)
Midpoint	Average of bid and ask prices: (BestBid + BestOffer) / 2
Volume	Number of contracts traded during the observation date
OpenInterest	Number of open contracts that haven't been closed or exercised
ImpliedVolatility	Model derived Volatility
Delta	Sensitivity of option price to changes in underlying price
Gamma	Sensitivity of delta to changes in the underlying price
Vega	Sensitivity of option price to changes in implied volatility
Theta	Sensitivity of option price to time decay (loss of value as expiration nears)
OptionID	Unique identifier for the option contract

Suggested Analyses

1. Put-Call Ratio

- **Formula**: Put Volume / Call Volume
- **Use**: Sentiment indicator; put volume > call volume indicates bearish sentiment
- **Implementation ideas**:
    - Calculate daily ratios
    - Break down by expiration date ranges
    - Identify significant changes day-over-day

2. Volume and Open Interest Analysis

- **Outlier Detection**: Identify options with statistically significant changes in volume or open interest compared to the previous day

- **Implementation ideas**:
  - Flag options where volume > 2x average volume
  - Track large changes in open interest (e.g., >50% increase/decrease)
  - Group by strike ranges or expiration buckets

3. Implied Volatility Analysis

- **Volatility Skew Calculation**:

  Background: In theory, options with the same expiration should have similar implied volatility regardless of strike price. In practice, they don't - this difference is called "skew." Traders pay attention to skew because it reveals market sentiment about potential price movements.

  How to calculate & implementation ideas:

  1. Group options by expiration date or time buckets
     a. Option A – By exact expiration
        i. Pro: Most granular view
        ii. Con: May have too few options per group for reliable averages
     b. Option B – By days to expiration (DTE) buckets:
        i. 0-7 days (weekly options)
        ii. 8-30 days (short-term)
        iii. 31-60 days (medium-term)
        iv. 61-180 days (long-term)
        v. 180+ days (very long-term)
  2. Within each group, separate calls and puts
  3. Calculate average implied volatility for different delta ranges:
     a. For calls: Find options with delta between 0.7-0.8 (these are in-the-money) and options with delta around 0.5 (at-the-money)
     b. For puts: Find options with delta between -0.3 to -0.2 (these are in-the-money) and options with delta around -0.5 (at-the-money)
  4. Calculate the skew:
     a. Call Skew = avg(IV of 70-80 delta calls) - avg(IV of 50 delta calls)
     b. Put Skew = avg(IV of -30 to -20 delta puts) - avg(IV of -50 delta puts)

4. Custom Analysis

Feel free to implement any additional analysis you find interesting or valuable.

Evaluation Criteria

We will evaluate your submission based on:

- **Code quality**: Clean, readable, maintainable code
- **Analysis implementation**: Correctness and efficiency of calculations
- **Error handling**: Graceful handling of edge cases and data issues
- **Documentation**: Clear explanation of your approach and assumptions

Submission Instructions

Please submit your solution as a single ZIP file containing:

- All code and SQL files
- README.md with setup and run instructions