ASSIGNMENT
Part - 1

Disclaimer: For the purpose of this assignment, you are free to use as much Google as you want for research purposes.

High Frequency Trading firms derive a major part of the alpha from speed. A part of winning this speed game comes from analysing extremely granular data. Currently, the most granular-level of data available is Tick-by-tick data, popularly known as TBT data. It's an exchange-side dissemination of all the order and trade related activity that is happening in all its instruments in the form of the real time order book of that instrument up to various levels (here level indicates price level). This information is communicated to the market participants (who have subscribed to this service) in real time. Each instrument level TBT file can be categorized into the following four types based upon the type of tick info -

1. N - denotes a New Order in the orderbook of that instrument. A unique order ID called ExchID('ExchId1' in the dataset given) is assigned to the order and stays so till the order remains in the orderbook.
2. M - denotes a Modification of an existing order in the orderbook of that instrument
3. X - denotes a cancellation of an order previously existing in the orderbook of that instrument
4. T - denotes a trade that happened in that instrument, resulting in the removal of the order from the orderbook.

TBT files for two or more instruments can be compared with each other using the Exchange timestamp, which is a unified clock and is available in nanoseconds. In the dataset given, it corresponds to 'ExchTstamp' and is multiplied by 10^9, therefore reported in nanoseconds. The 'ExchTstamp' is an epoch timestamp.

For all future references, the word 'Side' refers to Buy(B) or Sell(S). The side published for the 'N' ,'M' and 'X' tick is the side for which the order was sent into the system - a buy or a sell order. The side published for the 'T' tick is the side that got hit in the trade, also called the passive side. It means that the order was already present in the orderbook.

Regarding the trade tick 'T', one has to note that based upon the type of trade, the exchange id of the other orders are published. The passive side order's exchange id is always published in the 'T' tick line. Sometimes, the exchange id against which the trade happened is also published in the column 'ExchId2' ('ExchId2' in the dataset given). It is not necessary that ExchId2 will always be populated whenever a trade happens. However, of the two IDs, one ID will always be published whenever a 'T' tick is published. Value traded in a single trade can be defined as - Price at which the trade happened times the volume traded.

Participants trade in the markets using a variety of orders. Some popular order types are -

1. Limit Order - An order that allows the price to be specified while entering the order into the system.
2. Market Order - An order to buy or sell securities at the best price obtainable at the time of entering the order.
3. IOC - An Immediate or Cancel (IOC) order allows a Trading Member to buy or sell a security as soon as the order is released into the market, failing which the order will be removed from the market. Partial match is possible for the order, and the unmatched portion of the order is cancelled immediately.

You have been given the datasets of two instruments - one being a European Call option and the other is a European put option. The data is for the same strike 'X', same expiry 'E', same underlying 'U', same day 'D' and for the same time period 'P' of the day and is in sync with each other using the concept of 'Exchange Timestamp' earlier mentioned. The column 'qty' denotes - Quantity. The column 'price' denotes - Price. All prices in the dataset given are in paise. 'SecurityId' is the unique instrument identifier. 'SeqNum' stands for sequence number and is a counter used by Exchange. (Not relevant to this assignment).

The first task of your assignment is to identify all the IOC trades in the given data sets using the following condition -

1. In case of an IOC trade, there is no new 'N' tick available for the IOC order.
2. Straightaway, a trade tick 'T' is published and only the exchange ID of the side that got hit in the trade is published.
3. That means either ExchId1 or ExchId2 is missing.
4. The order that got hit in the trade has to be present in the orderbook before, ergo; its exchId should be there in the orderbook before the trade in the form of 'N' (New) or 'M' (Modify) ticks.

You have to identify all the IOC trades done in both the instruments (individually) and have to report the following -

1. Total Number of IOC trades.
2. Total Value traded in IOC trades.

Part - 2
A POPULAR OPTION TRADING STRATEGY ON VOLATILITY - STRADDLE!

A straddle is a trading strategy that involves options. To use a straddle, a trader buys/sells a Call option and a Put option simultaneously for the same underlying asset at a certain point of time provided both options have the same expiry date and same strike price. A trader enters such a neutral combination of trades when the price movement is not clear. In an ideal situation, the two opposite trades can offset losses if either of the options fails. In this strategy, one can go 'either' long (buy) on both options i.e. Call & Put, 'or' short (sell) both. The eventual outcome of the strategy depends entirely on the quantum of price movement on the security in question. In other words, the degree of price movement (Volatility), rather than the direction of price movement, affects the outcome.

Your second task is to identify the straddles traded. For this the following assumption is used -

"All straddles are traded using IOC orders and both the legs of the trade are done simultaneously in real time."

Therefore in order to identify the straddles bought or sold in both the instruments you have to check the following conditions -

1. Identify all the IOC trades.
2. Identify all the IOC trades that were done simultaneously in both the instruments within a +/- 250 nanoseconds window.
3. The side of both the trades has to be the same. If an IOC buy trade in instrument1 was done at time t, then an IOC buy trade also has to happen in instrument2 in time t +/- 250 nanoseconds for it to be categorized under straddles. In case of an IOC buy trade, a 'S' sell side trade will be published in the 'T' tick (as explained in Part 1 of the assignment)
4. The minimum of the two traded quantities shall be counted as the quantity of the straddle trade.

Once identified, you have to report the following -

1. Total straddles bought at strike X. (Count) and the value traded.
2. Total straddles sold at strike X. (Count) and the value traded.
3. Net straddles traded - Bought - Sold at strike X. (Count). And value traded.

BONUS QUESTION -

How will your answers change if we alter the window for scouting straddles from 250 nanoseconds to:

1. 100 nanoseconds
2. 500 nanoseconds
3. 1 microsecond
4. 10 microseconds
5. 500 microseconds
6. 1 millisecond
7. 100 milliseconds
8. 1 second

SUBMISSION-

1. Entire python notebook or script which should be properly commented.
2. Readme document to run your script.
3. A summary containing the following -
    a. Answers of the questions and values asked for above.
    b. What will happen if both the legs of the straddles are not traded simultaneously and have a time difference between them?
    c. Do you suggest any other order type for trading straddles?
    d. What do you think SeqNum might be used for?
    e. Did you enjoy this assignment and what did you learn from this assignment?

Tip - *In case you are unable to come up with a solution or understand the assignment, you are free to arrange for a call through HR within 24 hours of the assignment being given to you.*

ALL THE BEST