# Analyst Report

## Pre Processing

When only looking for NaN and XT condition codes did the stocks fall to 99 values when the email specified 100?

Given Boomberg example at the top, hence why it went down to 99?

Everything was on a day to day basis i.e. when it closed the day was restarted and the mean and median calculations were continued. Only problem was this is do you need to tick change times over night or does that not count?

To handle this I would look at slicing the index as soon as you hit the negative value – indicating a new day – calculating the tick difference and then the to midnight from the closing and from midnight for the opening time. Add the two together and then you get the overnight time between the tick change, add this to a list and then perform the same action, then add the list into the mean calculating at the end.

## Generating the Time Data

From the specification

“This data is over several days and so when no trading occurs there are large time gaps to take into account so as not to skew the figures.”

What defines no trading?

1. When trade volume is 0 – searched the dataframe and there was no point where this occurred
2. Trading hours - Between hours of 9.30 and 4.30? What about extended hours?
3. When all the values stay the same?
4. Is it only when the update type was 1 – so asking to change price etc does not constitute a trade

It says there were data over 4 days – according to python – the data only had one date and it looked like the stocks never reset a seconds at midnight i.e there are 86000 seconds in a day, seconds typically started at I looked at several stocks and the date never changed.

I checked to see if the time reset itself, basically looked for a negative difference, if it does this 4 times then that means it is 4 days as suggested.

Check to see when the stock market opened etc etc, state that it was done on a day by day basis i.e. longest time between trades could have been the absolute number between the first trade on one day and the last trade on the previous day.

The start and end times for the data looked comparable to the opening and closing times of the Swedish stock exchange.

## Generating Tick Info

With tick data – can you just reset time prices to zero, as they can still change price over night before trading opens again? That way I could just follow the same logic, look for when price changes and add up the differences between he changes.

This one I may have to actually, did you want to include over night tick changes values?

Remove the values where there is no tick change – it doesn’t matter if there are loads of zeros in-between the price has not changed.

## Generating Bid Ask Spread

Using the same logic for the time data, I just took the difference between the bid and the ask price for each stock and then found the mean and median.

## Generating Examples of the round number effect

Didn’t specify that for price you wanted whole price or just end with a 0. Doesn’t matter for trade volumes as they are ill int64. Whereas trade prices are floats

## Future Work & Code improvements

If this was a one of piece of work, this code works well – it is not massively demanding, runs in couple of seconds and will output the data as requested.

If this was a recurring process and this was needed to be done repeatedly, I would do the following:

1. Structure the incoming csv to a common format i.e. look at the csv Itarle UK typically receives and format it so that the program could be as universal as possible.
2. Functions – already done this, no point in doing this for the time difference or the bid ask spread as it is 1-3 lines of code – would be no computational advantage
3. Threading – you could get the three functions to run as separate threads and therefore reduce the time. This would be more advantageous for larger data sets.
4. Conversion to Java.

If I continue to use python I would also:

1. UI – upload csvs and they will do it all for you – so anyone could do it.