**Data Science Practice Problem**

The following data sets consist of trade executions/trades from an equity trading platform, the corresponding market data, and instrument reference data. The goal of the exercise is to perform the corresponding (below tasks) transformations and calculations to show your understanding of data manipulation/engineering in python. The final output should be one file, in any format you feel relevant, with the goal of it being used in a data analytics platform / BI platform / database. As well as the corresponding code of how you got to your output, which can be rerun on another machine.

**What are we looking for**

* Please provide all your code and the final output file.
* Please use python.
* You are free to choose your project structure/style e.g. Jupyter notebook, python project, docker etc..

**Data Sets**

* Executions.parquet
* Marketdata.parquet
* Refdata.parquet

**Task**

1. Start:
   1. Count the number of executions within the executions.parquet file, determine the unique number of [‘Venue’]s and the date of executions. Log output this information.
2. Data Cleaning:
   1. Filter executions.paraquet for only CONTINUOUS\_TRADING trades.
   2. Log output the # of executions.
3. Data Transformation:
   1. Add column [‘side’], if quantity is negative, side = 2, if quantity is positive side = 1.
   2. Complement the data with refdata.parquet
      1. Add the primary ticker [‘primary\_ticker’]
      2. Add the primary mic [‘primary\_mic’]
4. Calculations:
   1. Best bid price and best ask (bbo) – bbo data in marketdata.parquet
      1. Find bbo price at execution, 1 second before execution and 1 second after execution from the marketdata.parquet file and add this data into the final output file – respective column table names [‘best\_bid’,’best\_ask’, ‘best\_bid\_min\_1s’,’best\_ask\_min\_1s’, ‘best\_bid\_1s’,’best\_ask\_1s’]
   2. Mid-Price – bbo data in marketdata.parquet
      1. Find the Mid-Price at execution, 1s before the execution and 1s after the execution – respective column table names [‘mid\_price’, ‘mid\_price\_min\_1s’ ‘mid\_price\_1s’]
   3. Calculate Slippage [‘slippage’] at execution price
      1. For SELL: (execution\_price – best\_bid) / (best\_ask – best\_bid)
      2. For BUY : (best\_ask – execution\_price) / (best\_ask – best\_bid)
5. Please provide performance metrics on your program

**Thank you and Good Luck.**