

IE 421 High Frequency Trading Technology

Group 2

Final Report

Xiaoyue Chen

Chenglin Miao

Luwei Wang

Xi Teng

Project Scope

- Create a MySQL database to store tick and order book information for selected stocks on IEX.
- Analyze market data, stock correlation, spread, order flow imbalance, bid and ask depth, and VWAP.

Data Collection

The `download_iex_pcaps.py` file is designed for automatically downloading and managing market data files from the Investors Exchange (IEX) Trading platform, specifically targeting the DEEP and TOPS data feeds. It provides a way to fetch a list of available market data files within a specified date range and subsequently downloads these files to a designated local directory. The script includes functionality to parse command-line arguments for specifying the date range and download path, ensuring that only the relevant data files for the given period are retrieved and stored.

Additionally, it incorporates checks to prevent re-downloading files that already exist locally, thus optimizing the use of network and storage resources.

Data Parsing

The `parse_iex_pcap.py` file is structured to handle various types of market data messages such as trading status, operational halt, price level updates, and trade reports. It includes a `BasicPcapParser` class that manages the parsing process, including reading pcap files, extracting IEX messages, and processing different message types. Handling Multiple Message Types: The script can parse and process different IEX message types, each with specific data structures.

- **Output File Generation:** It has the capability to generate output files in both plain text and gzip formats. These files contain detailed information about trades, book updates, and timestamps extracted from the pcap data.
- **Symbol Tracking:** The script tracks specific stock symbols of interest, allowing for focused data extraction.
- **Data Extraction and Conversion:** It efficiently handles the extraction of relevant data from the messages and converts timestamps from raw formats to human-readable forms.

Database Construction

First use the `parse_iex_pcap.py` file to parse packets received from IEX change into order book and trading data of targeting symbols. Figure is a screenshot of the order book of 2022/8/1 order book of AAPL, AMZN, FB, NFLX, GOOGL, and MSFT. Figure b is a screenshot of the trading log of the above stocks on 2022/8/1.

COLLECTION_TIME	MESSAGE_ID	MESSAGE_TYPE	SYMBOL	BID_PRICE_1	BID_SIZE_1	BID_PRICE_2	BID_SIZE_2	BID_PRICE_3	BID_SIZE_3	ASK_PRICE_1	ASK_SIZE_1	ASK_PRICE_2	ASK_SIZE_2	ASK_PRICE_3	ASK_SIZE_3
30:00.1	317131	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135.07	100	140.55	100	143.43	100
30:00.1	317369	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135.01	100	135.07	100	140.55	100
30:00.1	317396	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135.01	100	140.55	100	143.43	100
30:00.1	317418	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135.01	300	140.55	100	143.43	100
30:00.1	318231	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135.01	100	140.55	100	143.43	100
30:00.1	318441	BID_UPDATE	AMZN	134.9	1	134.75	1	129.35	100	135.01	100	140.55	100	143.43	100
30:00.1	319515	ASK_UPDATE	AMZN	134.9	1	134.75	1	129.35	100	135	100	135.01	100	140.55	100
30:00.1	319525	ASK_UPDATE	AMZN	134.9	1	134.75	1	129.35	100	135	100	140.55	100	143.43	100
30:00.1	319544	ASK_UPDATE	AMZN	134.9	1	134.75	1	129.35	100	135	300	140.55	100	143.43	100
30:00.1	320186	ASK_UPDATE	AMZN	134.9	1	134.75	1	129.35	100	135	100	140.55	100	143.43	100
30:00.1	320222	BID_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	100	140.55	100	143.43	100
30:00.1	320425	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	100	135.01	100	140.55	100
30:00.1	321737	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	100	135.01	300	140.55	100
30:00.1	322312	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	200	135.01	300	140.55	100
30:00.1	322654	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	400	135.01	300	140.55	100
30:00.1	322707	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	400	135.01	100	140.55	100
30:00.1	322737	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	100	135	400	135.01	100
30:00.1	322751	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	100	135	300	135.01	100
30:00.1	322870	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	100	135	300	140.55	100
30:00.1	323231	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	300	140.55	100	143.43	100
30:00.2	324025	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	400	140.55	100	143.43	100
30:00.2	324186	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	135	300	140.55	100	143.43	100
30:00.2	324717	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	100	135	300	140.55	100
30:00.2	324721	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	300	135	300	140.55	100
30:00.2	324728	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	300	135	100	140.55	100
30:00.2	326888	ASK_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	300	140.55	100	143.43	100
30:00.2	327123	BID_UPDATE	AMZN	134.9	1	129.35	100	95	10	134.98	300	140.55	100	143.43	100
30:00.2	327209	BID_UPDATE	AMZN	134.9	1	134.87	244	129.35	100	134.98	300	140.55	100	143.43	100

Fig. a

COLLECTION_TIME	MESSAGE_ID	MESSAGE_TYPE	SYMBOL	PRICE	SIZE	TRADE_ID	TRADE_FLAGS
00:17.7	50525	T	AMZN	135.3	10	2805953	EXTENDED_HOURS ODD_LOT
01:27.7	53279	T	AMZN	135.27	10	2991128	EXTENDED_HOURS ODD_LOT
01:53.8	54165	T	AMZN	135.24	1	3040570	EXTENDED_HOURS ODD_LOT
03:24.0	58058	T	AAPL	162.08	50	3221106	EXTENDED_HOURS ODD_LOT
03:24.0	58059	T	AAPL	162.08	50	3221111	EXTENDED_HOURS ODD_LOT
03:24.0	58060	T	AAPL	162.09	10	3221116	EXTENDED_HOURS ODD_LOT
03:24.0	58061	T	AAPL	162.09	10	3221121	EXTENDED_HOURS ODD_LOT
03:24.0	58062	T	AAPL	162.09	10	3221127	EXTENDED_HOURS ODD_LOT
03:24.0	58063	T	AAPL	162.09	50	3221132	EXTENDED_HOURS ODD_LOT
05:09.2	62450	T	MSFT	279.65	25	3424325	EXTENDED_HOURS ODD_LOT
05:42.2	63938	T	AMZN	134.94	15	3501200	EXTENDED_HOURS ODD_LOT
05:42.2	63939	T	AMZN	134.94	50	3501212	INTERMARKET_SWEEP EXTENDED_HOURS ODD_LOT
06:59.0	66323	T	AAPL	161.97	10	3658304	EXTENDED_HOURS ODD_LOT
07:41.2	67613	T	AMZN	135	50	3742045	EXTENDED_HOURS ODD_LOT
09:56.1	72204	T	AAPL	162	50	3991818	EXTENDED_HOURS ODD_LOT
10:10.0	72936	T	AAPL	161.99	10	4016933	EXTENDED_HOURS ODD_LOT
10:10.0	72937	T	AAPL	161.98	50	4016940	EXTENDED_HOURS ODD_LOT
10:10.0	72938	T	AAPL	161.98	50	4016945	EXTENDED_HOURS ODD_LOT
10:10.0	72939	T	AAPL	161.98	50	4016950	EXTENDED_HOURS ODD_LOT
10:20.5	73240	T	AAPL	161.98	50	4034792	EXTENDED_HOURS ODD_LOT
10:20.5	73241	T	AAPL	161.98	50	4034797	EXTENDED_HOURS ODD_LOT
10:52.3	74609	T	AMZN	135.03	10	4104127	EXTENDED_HOURS ODD_LOT
15:00.9	82499	T	MSFT	279.35	9	4585877	EXTENDED_HOURS ODD_LOT
15:16.6	83066	T	GOOGL	116	34	4624411	EXTENDED_HOURS ODD_LOT
15:16.6	83067	T	GOOGL	116	34	4624416	EXTENDED_HOURS ODD_LOT
18:15.6	87401	T	AMZN	135.02	50	4936165	EXTENDED_HOURS ODD_LOT
18:19.3	87550	T	AAPL	161.89	10	4941082	EXTENDED_HOURS ODD_LOT
18:31.6	87910	T	AAPL	161.89	1	4970128	EXTENDED_HOURS ODD_LOT
19:23.1	89942	T	AMZN	134.9	67	5057767	EXTENDED_HOURS ODD_LOT
19:23.1	89943	T	AMZN	134.91	33	5057772	EXTENDED_HOURS ODD_LOT
20:37.8	92318	T	AMZN	135.01	10	5245052	EXTENDED_HOURS ODD_LOT
20:37.8	92319	T	AMZN	135	50	5245057	EXTENDED_HOURS ODD_LOT
20:37.8	92320	T	AMZN	134.99	10	5245062	EXTENDED_HOURS ODD_LOT

Fig. b

After parsing the data, we constructed a data table of the order book and the trading logs. Figure c is the data table of the order book, and figure b is the data table of the trading log.

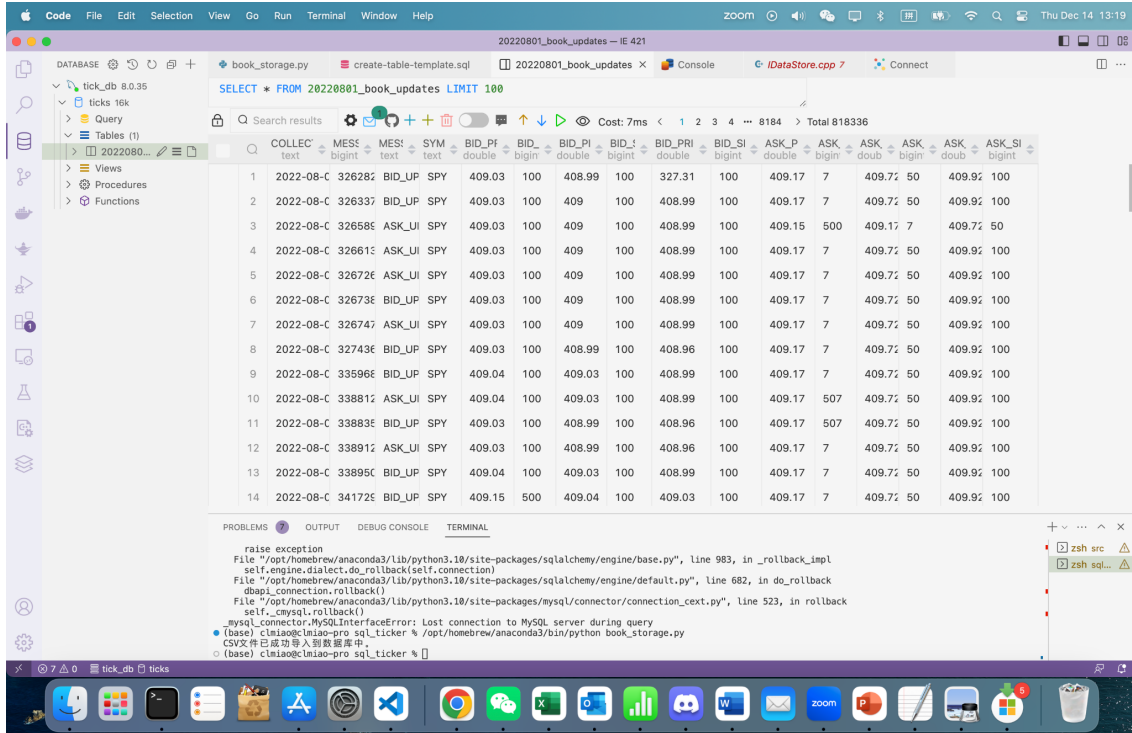


Fig. c

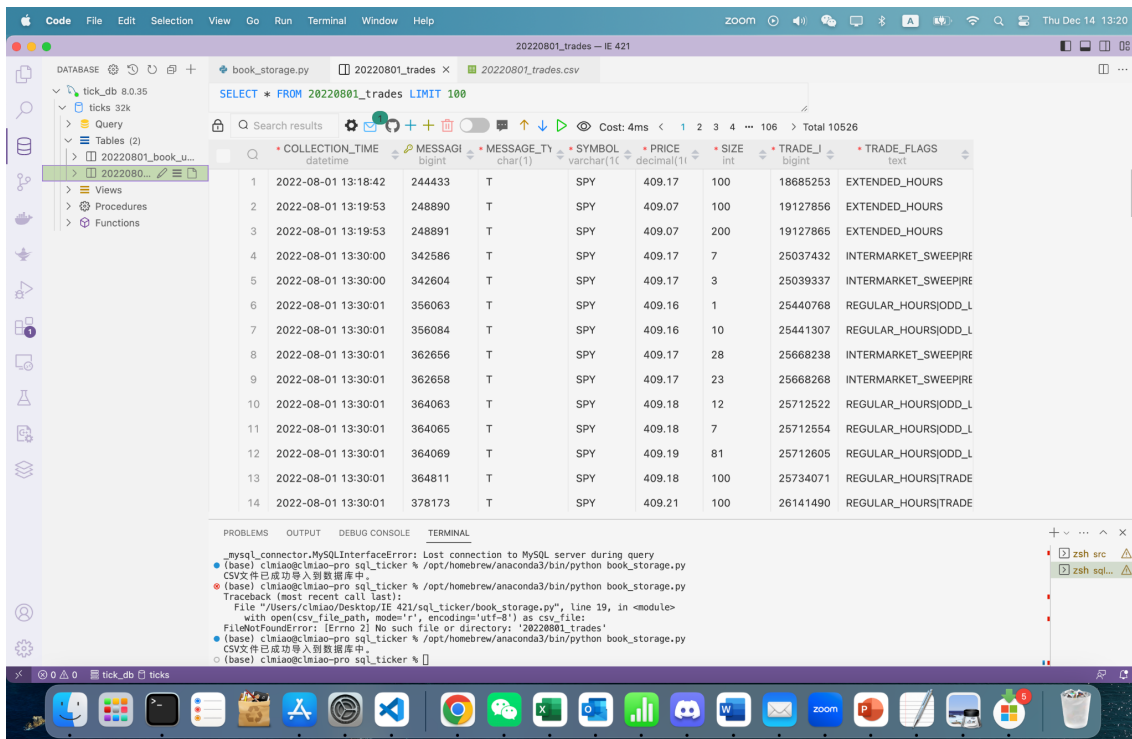


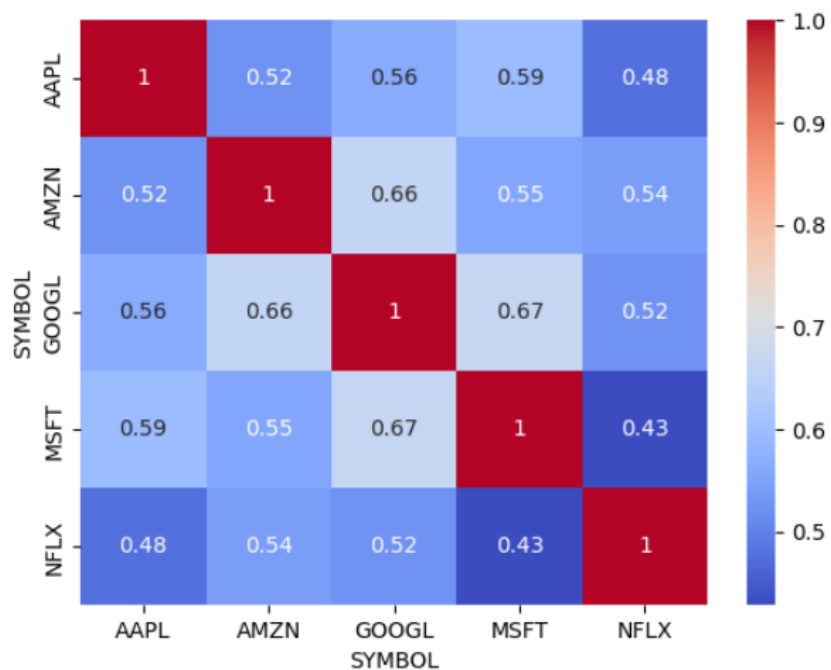
Fig. d

Data Analysis

According to the data we generated from the parser and loaded in the database, we chose five symbols to do the analysis. The five chosen symbols are NFLX, MSFT, GOOGL, AMZN and AAPL.

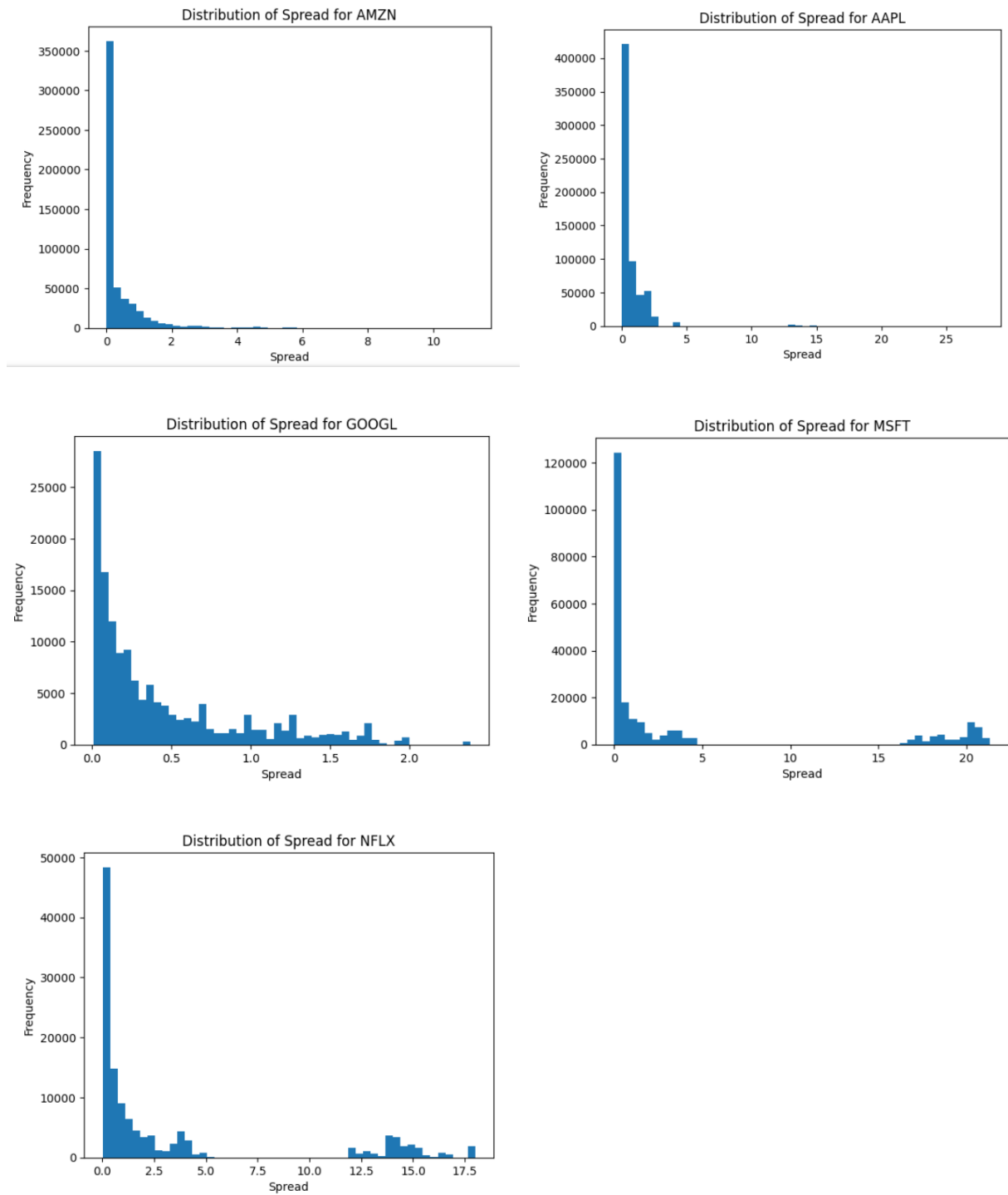
Data Correlation

Based on the trade information, we tried to find out the correlation between the trade data. We generated the message from the figure below:

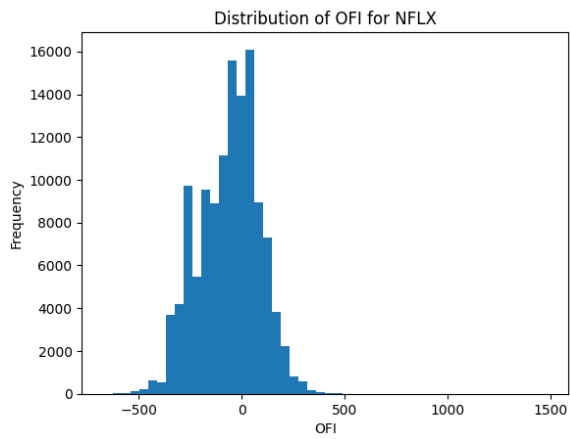
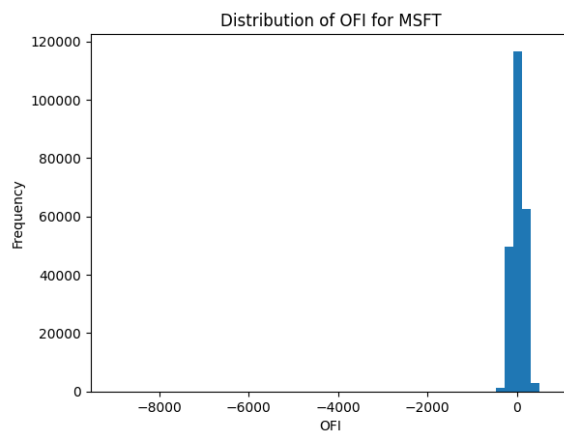
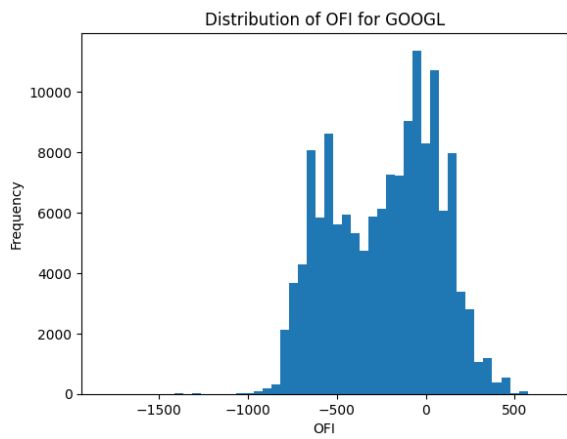
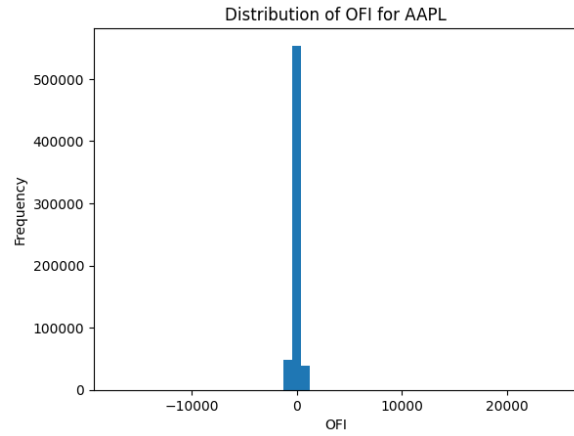
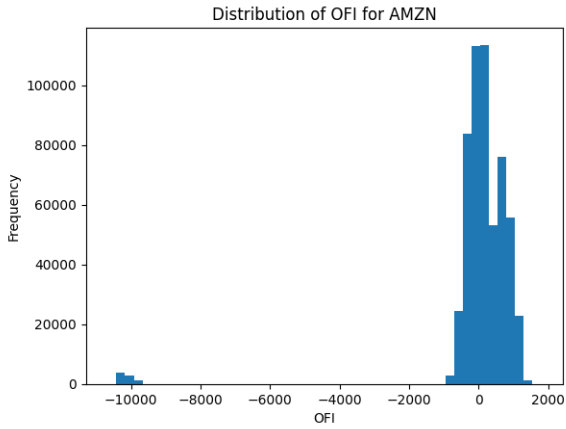


Market Depth Analytics

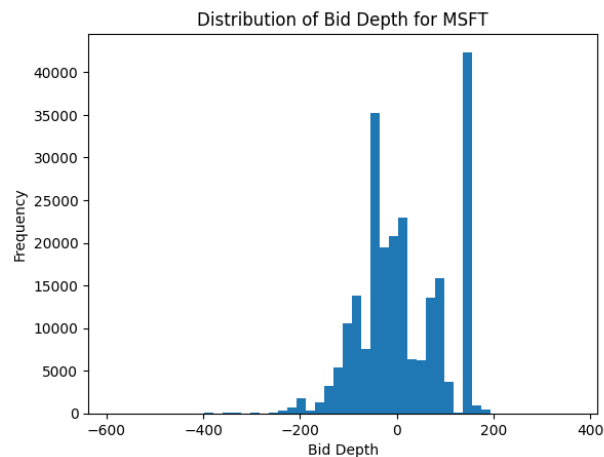
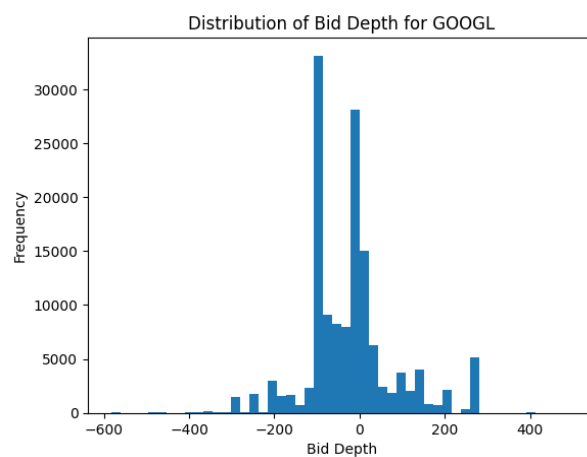
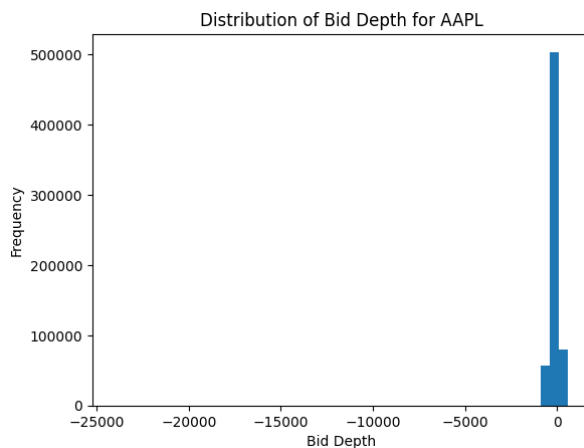
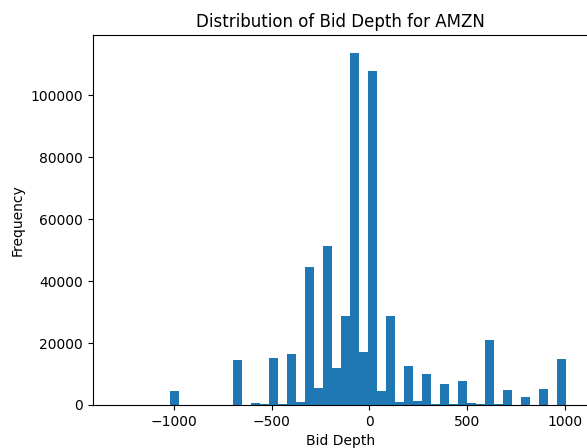
For each symbol, we do the statistics of the distribution of the spread. And the result is shown below in the histogram.



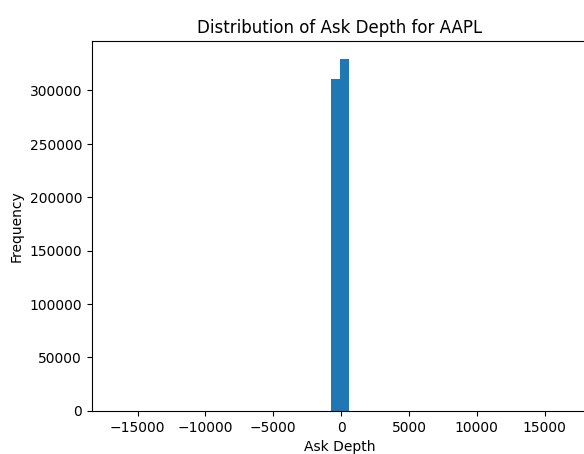
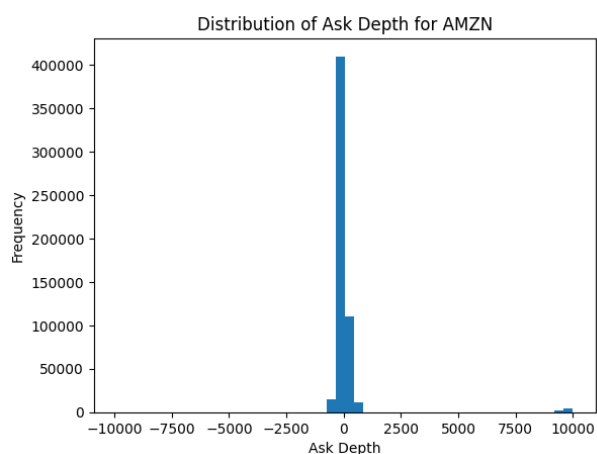
For each symbol, we calculated the distribution of OFI. Following is the result:

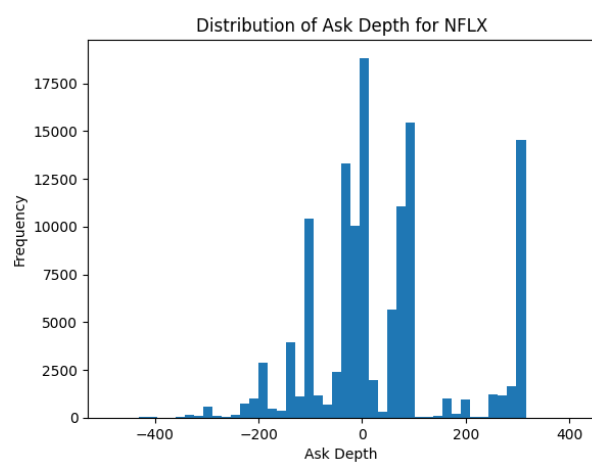
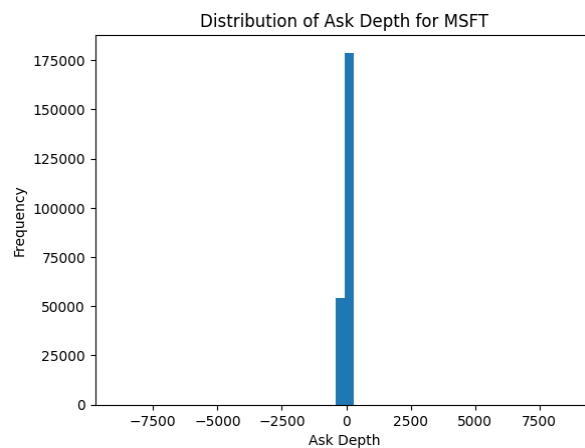
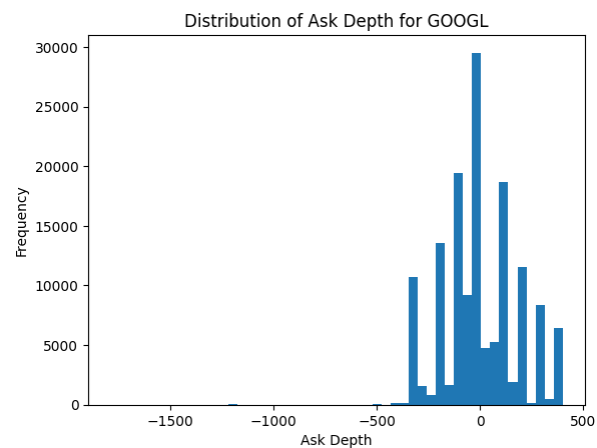


For each symbol, we calculated the distribution of the bid depth. Following is the result:



For each symbol, we calculated the distribution of ask depth. Following is the result:





For each symbol, we calculated the volume-weighted average price. Following is the result:

