# IEX Intraday Data
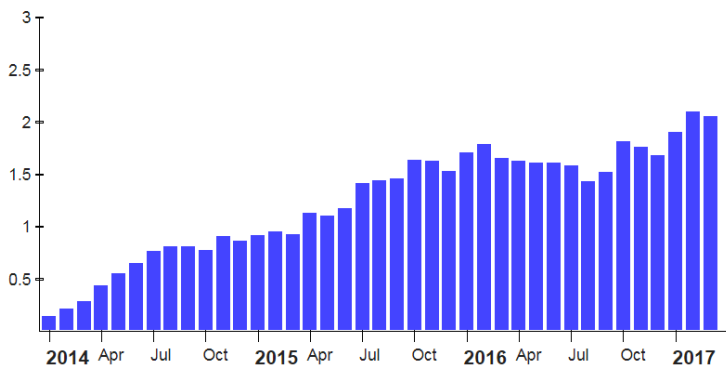
## Overview

The IEX Investors Exchange provides a wealth of great data.  Even better, they currently provide all of this data for free!  According to Wikipedia, the IEX makes up about 2% of total trading volume in the United States with 175 million shares traded daily as of September 2018.

*% Market Share*



While only 1 company is actually listed on this exchange (Interactive Brokers, listed on October 1, 2018), over 8,000 stocks are traded on this exchange.  IEX provides APIs with data for all these stocks.  This data includes historical stock prices, company info, financial statements, dividends, and split information.  They also provide detailed intraday data available in either real time or as a download at the end of the day.

This paper will examine the detailed intraday data.  We will provide a quick overview of the files offered and some sample data from one file.  This data offers an interesting glimpse into the activity occurring on the IEX exchange.  However, there are several potential issues and idiosyncrasies of this data that should be understood if you are attempting to use it.  It remains to be seen whether this data will end up being useful or not.

## File Format and Parsers

IEX provides two main files for intraday data:

- DEEP – Detailed information including every trade, every update to the order book, and a variety of other information.
- TOPS – Nearly identical to DEEP except that instead of getting updates to the entire order book, only changes to the best bid and best offer (the top of the order book) are included.

2 years of daily files are available on their website at:

**IEX Intraday Data**

https://iextrading.com/trading/market-data/

Both of these formats are binary PCAP (packet capture) files.  It looks like they simply record the outgoing traffic from their servers as they run them throughout the day.  The network data is saved in a format that can reproduce every detail of that daily feed.  The data requires a PCAP file parser to get past the packet-level data and get to higher-level message segments.  These binary messages then need to be decoded using the IEX binary message specifications.  I did this using the Java libraries: iextrading4j-hist and Pcap4j.

I downloaded TOPS data from 2018-12-28.  The file was 924 MB compressed and 3.8 GB uncompressed.  The rest of this paper will examine data contained in that file.

**Message Types**

A quick summary of message types for the day revealed the following counts:

| Message Type | Count | Description |
|---|---:|---|
| QUOTE_UPDATE | 27,120,370 | Change to the best bid or offer for a stock |
| TRADE_REPORT | 1,189,305 | Notification that a trade occurred |
| SHORT_SALE_PRICE_TEST_STATUS | 9,104 | |
| TRADING_STATUS | 8,786 | |
| OPERATIONAL_HALT_STATUS | 8,720 | |
| AUCTION_INFORMATION | 2,872 | Auction information for IEX-listed stocks |
| OFFICIAL_PRICE_MESSAGE | 8 | Official open and close prices for IEX-listed stocks |
| SYSTEM_EVENT | 6 | |
| SECURITY_DIRECTORY | 4 | |
| Total | 28,339,175 | |

Most of the information in this feed involves QUOTE_UPDATEs.  Each of these is a change to the top of the order book.  There were also 1.1 million trade reports sent this day.  At first I was excited to see that auction information and official open and close prices might be included in this data feed.  However, these ended up only being for IEX-listed stocks.  As mentioned, there is only 1 stock currently listed on the exchange.  However, there are 3 symbols that appear to be test stocks.  The 8 OFFICIAL_PRICE_MESSAGES provide the opening and closing prices for these 4 stocks.

**Quotes and Trade Messages**

The quote and trade messages are the most interesting to examine.  Each of these refers to a specific stock so we can count the number of each messages by stock symbol.  Below are tables showing the top 20 quoted and traded stocks on this day:

**IEX Intraday Data**

*Top 20 Traded Stocks*

| Symbol | Trades | Quote Updates | Trades/min | Quotes/min |
|--------|--------|---------------|------------|------------|
| FB | 14,347 | 174,133 | 36.8 | 446.5 |
| EFA | 9,506 | 170,299 | 24.4 | 436.7 |
| SPY | 9,434 | 294,360 | 24.2 | 754.8 |
| GE | 9,386 | 16,660 | 24.1 | 42.7 |
| BAC | 7,344 | 18,180 | 18.8 | 46.6 |
| T | 6,208 | 21,585 | 15.9 | 55.3 |
| MSFT | 5,431 | 10,358 | 13.9 | 26.6 |
| ORCL | 5,052 | 16,804 | 13.0 | 43.1 |
| XLF | 4,979 | 43,208 | 12.8 | 110.8 |
| PFE | 4,617 | 14,807 | 11.8 | 38.0 |
| TWTR | 4,447 | 13,597 | 11.4 | 34.9 |
| SQQQ | 4,394 | 23,252 | 11.3 | 59.6 |
| EEM | 4,163 | 169,279 | 10.7 | 434.0 |
| F | 3,979 | 11,612 | 10.2 | 29.8 |
| FDC | 3,898 | 9,488 | 10.0 | 24.3 |
| C | 3,604 | 11,307 | 9.2 | 29.0 |
| FCX | 3,313 | 14,749 | 8.5 | 37.8 |
| QQQ | 3,213 | 104,985 | 8.2 | 269.2 |
| XOM | 3,200 | 17,937 | 8.2 | 46.0 |
| XOP | 3,130 | 172,172 | 8.0 | 441.5 |

*Top 20 Quoted Stocks*

| Symbol | Trades | Quote Updates | Trades/min | Quotes/min |
|--------|--------|---------------|------------|------------|
| SPY | 9,434 | 294,360 | 24.2 | 754.8 |
| AAPL | 2,937 | 281,356 | 7.5 | 721.4 |
| VOO | 673 | 210,558 | 1.7 | 539.9 |
| GOOG | 459 | 204,517 | 1.2 | 524.4 |
| GOOGL | 874 | 192,104 | 2.2 | 492.6 |
| FB | 14,347 | 174,133 | 36.8 | 446.5 |
| IWM | 2,915 | 172,477 | 7.5 | 442.2 |
| XOP | 3,130 | 172,172 | 8.0 | 441.5 |
| EFA | 9,506 | 170,299 | 24.4 | 436.7 |
| EEM | 4,163 | 169,279 | 10.7 | 434.0 |
| V | 2,655 | 156,923 | 6.8 | 402.4 |
| AMZN | 2,110 | 140,648 | 5.4 | 360.6 |
| GUSH | 202 | 122,216 | 0.5 | 313.4 |
| IWF | 492 | 121,855 | 1.3 | 312.4 |
| VT | 828 | 115,168 | 2.1 | 295.3 |
| NFLX | 2,838 | 107,007 | 7.3 | 274.4 |
| OEF | 244 | 106,076 | 0.6 | 272.0 |
| ACWI | 697 | 106,039 | 1.8 | 271.9 |
| QQQ | 3,213 | 104,985 | 8.2 | 269.2 |
| ACWV | 177 | 102,965 | 0.5 | 264.0 |

As shown, the top-traded stocks several times per minute and have several hundred quote updates each minute.  There are 787 stocks that trade at least once per minute on average.

**IEX Intraday Data**

However, the majority of stocks (6,491) that trade 100 or fewer times per day. This means that we should have pretty good, minute-level data for the equivalent of the S&P 500 or even the Russell 1000 (the top 500 and 1000 stocks in the U.S.). However, we may run into sparse data problems as we get into the more lightly traded stocks.

**Sample Data: Kroger Stock**

To get a better feel for this data we'll dig into one stock in particular: Kroger (symbol KR). This was the 286th most-traded stock on the exchange this day and ranked 683rd in quote updates. It had 837 trades (about 2.1 trades per minute) and 9,237 quote updates (23.7/minute). A simple program that reads the Pcap data and outputs trade summaries produced the following:

| Timestamp | DateTime | Price | Size |
|---|---|---|---|
| 1546007418156053643 | 2018-12-28T09:30:18.156053643 | 27.5100 | 100 |
| 1546007419060025458 | 2018-12-28T09:30:19.060025458 | 27.5050 | 97 |
| 1546007422300034994 | 2018-12-28T09:30:22.300034994 | 27.5050 | 100 |
| 1546007422629714545 | 2018-12-28T09:30:22.629714545 | 27.4850 | 100 |
| 1546007545136970943 | 2018-12-28T09:32:25.136970943 | 27.5700 | 46 |
| 1546007711336604987 | 2018-12-28T09:35:11.336604987 | 27.7000 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 300 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 100 |
| 1546007749588518985 | 2018-12-28T09:35:49.588518985 | 27.7100 | 200 |
| 1546007751034668030 | 2018-12-28T09:35:51.034668030 | 27.7100 | 100 |
| 1546007760716598841 | 2018-12-28T09:36:00.716598841 | 27.7200 | 100 |
| 1546007769333217834 | 2018-12-28T09:36:09.333217834 | 27.6500 | 100 |
| 1546007784358131944 | 2018-12-28T09:36:24.358131944 | 27.6500 | 100 |
| 1546007947964705977 | 2018-12-28T09:39:07.964705977 | 27.6800 | 5 |
| 1546007957886342783 | 2018-12-28T09:39:17.886342783 | 27.6850 | 100 |
| 1546007957886342783 | 2018-12-28T09:39:17.886342783 | 27.6850 | 7 |
| 1546008003660215969 | 2018-12-28T09:40:03.660215969 | 27.6750 | 50 |
| 1546008010897524750 | 2018-12-28T09:40:10.897524750 | 27.6650 | 100 |

The first column shows the timestamp as it appears in the raw IEX messages. This is a nanosecond epoch that provides the number of nanoseconds since 1970. The second column is a translated value of this timestamp into New York time. We then see the price and size values we would expect to see with trade information. There are some additional fields in the IEX messages including a unique trade ID and a "first message timestamp" (used to correlate messages) and some sale condition flags to provide more information about the trade.

If we plot trades throughout the day we can compare them to data from Yahoo! Finance to see how much of the price activity we capture. Here's what we see on IEX:

**IEX Intraday Data**
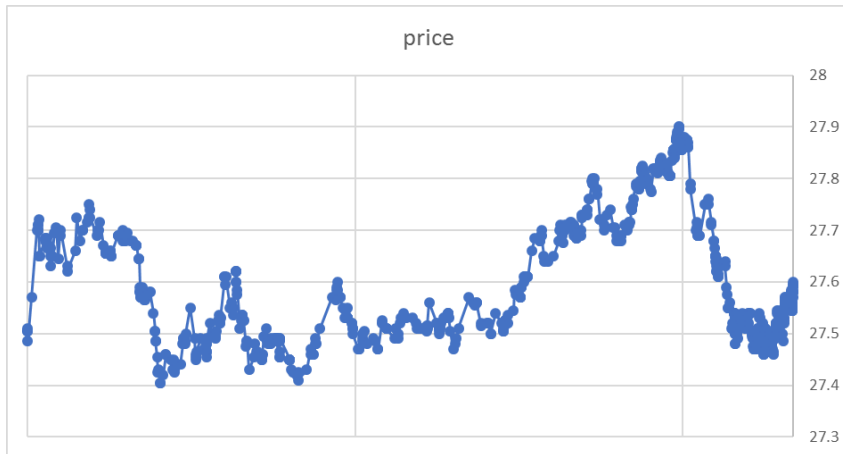


And here's what we see on Yahoo!:



If we overlay these it actually looks like we capture all of the stock price movement pretty well:

**IEX Intraday Data**

Now, we might wonder if the quote data could give us an even better view into the price movement. We have 10 times as many quote updates as we do trade reports, so this would seem to make sense at first. However, when we print out the quote update messages we see the following:

| Timestamp | DateTime | Bid | Ask | Bid Size | Ask Size |
|---|---|---|---|---|---|
| 1545999485575239125 | 2018-12-28T07:18:05.575239125 | 0.0000 | 0.0000 | 0 | 0 |
| 1546007400001389113 | 2018-12-28T09:30:00.001389113 | 0.0000 | 0.0000 | 0 | 0 |
| 1546007401009195258 | 2018-12-28T09:30:01.009195258 | 0.0000 | 27.5700 | 0 | 200 |
| 1546007401019510266 | 2018-12-28T09:30:01.019510266 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007401162768907 | 2018-12-28T09:30:01.162768907 | 0.0000 | 27.5500 | 0 | 200 |
| 1546007401694757156 | 2018-12-28T09:30:01.694757156 | 27.4000 | 27.5500 | 100 | 200 |
| 1546007402527516716 | 2018-12-28T09:30:02.527516716 | 27.4000 | 27.5500 | 100 | 100 |
| 1546007414162395879 | 2018-12-28T09:30:14.162395879 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007422630632804 | 2018-12-28T09:30:22.630632804 | 0.0000 | 27.5300 | 0 | 100 |
| 1546007422978753857 | 2018-12-28T09:30:22.978753857 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007424599825921 | 2018-12-28T09:30:24.599825921 | 0.0000 | 30.0000 | 0 | 100 |
| 1546007516088475261 | 2018-12-28T09:31:56.088475261 | 27.5400 | 30.0000 | 100 | 100 |
| 1546007534159838341 | 2018-12-28T09:32:14.159838341 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007544303189435 | 2018-12-28T09:32:24.303189435 | 27.5700 | 30.0000 | 100 | 100 |
| 1546007545136970943 | 2018-12-28T09:32:25.136970943 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007546800474054 | 2018-12-28T09:32:26.800474054 | 26.3500 | 27.6200 | 200 | 100 |
| 1546007548471284850 | 2018-12-28T09:32:28.471284850 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007590934068918 | 2018-12-28T09:33:10.934068918 | 27.6600 | 30.0000 | 100 | 100 |
| 1546007616023139501 | 2018-12-28T09:33:36.023139501 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007637192350417 | 2018-12-28T09:33:57.192350417 | 27.6500 | 30.0000 | 100 | 100 |

Two things are rather concerning:

- There are multiple times where we have no bid or ask
- The spread between the bid and the ask is huge (as much as $3.00-4.00)

So how does this affect trading? The following data is the trade report data we viewed earlier but this time combined with the best bid and offer at the time:
Plotting the spread throughout the day shows that this is not just something that happened early in the day either:
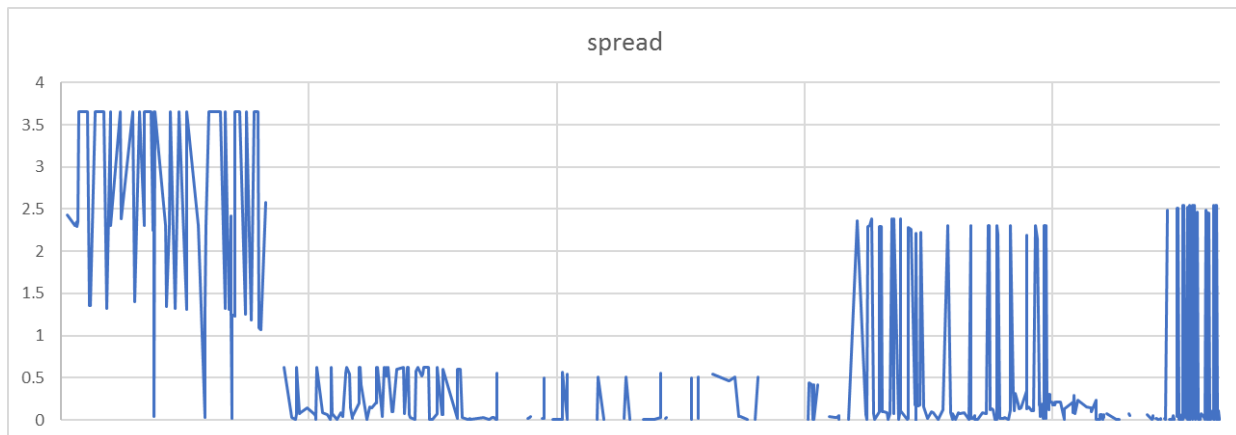
**IEX Intraday Data**

This last chart was actually obtained by combining trade data with the best bid and offer when the trade occurred.  The spread is the spread at the time the trade occurred.  This data set is shown below:

| Timestamp | DateTime | Price | Size | Bid | Ask | Bid Size | Ask Size |
|---|---|---|---|---|---|---|---|
| 1546007418156053643 | 2018-12-28T09:30:18.156053643 | 27.5100 | 100 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007419060025458 | 2018-12-28T09:30:19.060025458 | 27.5050 | 97 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007422300034994 | 2018-12-28T09:30:22.300034994 | 27.5050 | 100 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007422629714545 | 2018-12-28T09:30:22.629714545 | 27.4850 | 100 | 0.0000 | 27.5500 | 0 | 100 |
| 1546007545136970943 | 2018-12-28T09:32:25.136970943 | 27.5700 | 46 | 27.5700 | 30.0000 | 100 | 100 |
| 1546007711336604987 | 2018-12-28T09:35:11.336604987 | 27.7000 | 100 | 27.7000 | 30.0000 | 100 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 100 | 27.6500 | 30.0000 | 100 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 300 | 27.6500 | 30.0000 | 100 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 100 | 27.6500 | 30.0000 | 100 | 100 |
| 1546007749461754260 | 2018-12-28T09:35:49.461754260 | 27.7000 | 100 | 27.6500 | 30.0000 | 100 | 100 |
| 1546007749588518985 | 2018-12-28T09:35:49.588518985 | 27.7100 | 200 | 27.7100 | 30.0000 | 300 | 100 |
| 1546007751034668030 | 2018-12-28T09:35:51.034668030 | 27.7100 | 100 | 27.7100 | 30.0000 | 100 | 100 |
| 1546007760716598841 | 2018-12-28T09:36:00.716598841 | 27.7200 | 100 | 27.6500 | 30.0000 | 100 | 100 |
| 1546007769333217834 | 2018-12-28T09:36:09.333217834 | 27.6500 | 100 | 27.6500 | 30.0000 | 100 | 100 |
| 1546007784358131944 | 2018-12-28T09:36:24.358131944 | 27.6500 | 100 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007947964705977 | 2018-12-28T09:39:07.964705977 | 27.6800 | 5 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007957886342783 | 2018-12-28T09:39:17.886342783 | 27.6850 | 100 | 26.3500 | 30.0000 | 200 | 100 |
| 1546007957886342783 | 2018-12-28T09:39:17.886342783 | 27.6850 | 7 | 26.3500 | 30.0000 | 200 | 100 |
| 1546008003660215969 | 2018-12-28T09:40:03.660215969 | 27.6750 | 50 | 26.3500 | 27.7100 | 200 | 103 |
| 1546008010897524750 | 2018-12-28T09:40:10.897524750 | 27.6650 | 100 | 26.3500 | 27.7100 | 200 | 103 |

We see that trades occur even when there are no bids.  We also see that when the spreads are large (such as 27.65 and 30.00) that trades still occur in the middle of the spread at 27.70.

There are times where the spreads shrink to just a penny or two (as we'd expect on a modern exchange) and then trades do appear to be executing against the order book:

Daniel Rogers                                                                                   2018-12-31

## IEX Intraday Data

| Timestamp | DateTime | Price | Size | Bid | Ask | Bid Size | Ask Size | Spread |
|---|---|---|---|---|---|---|---|---|
| 1546015517869473679 | 2018-12-28T11:45:17.869473679 | 27.4250 | 100 | 27.4100 | 27.4400 | 100 | 100 | 0.03 |
| 1546015635374450689 | 2018-12-28T11:47:15.374450689 | 27.4200 | 100 | 27.4200 | 27.4300 | 100 | 100 | 0.01 |
| 1546015671895883648 | 2018-12-28T11:47:51.895883648 | 27.4250 | 100 | 27.4100 | 27.4300 | 100 | 100 | 0.02 |
| 1546015675062538794 | 2018-12-28T11:47:55.062538794 | 27.4100 | 100 | 27.4100 | 27.4200 | 100 | 100 | 0.01 |
| 1546015942271088954 | 2018-12-28T11:52:22.271088954 | 27.4300 | 100 | 27.4000 | 27.4300 | 100 | 100 | 0.03 |
| 1546016053148277062 | 2018-12-28T11:54:13.148277062 | 27.4700 | 100 | 27.4700 | 27.4800 | 100 | 200 | 0.01 |
| 1546016055157941572 | 2018-12-28T11:54:15.157941572 | 27.4600 | 100 | 27.4600 | 27.4700 | 100 | 200 | 0.01 |
| 1546016117720650761 | 2018-12-28T11:55:17.720650761 | 27.4600 | 52 | 27.4300 | 27.4600 | 100 | 200 | 0.03 |
| 1546016157620832161 | 2018-12-28T11:55:57.620832161 | 27.4600 | 100 | 27.4300 | 27.4600 | 100 | 200 | 0.03 |
| 1546016193027235817 | 2018-12-28T11:56:33.027235817 | 27.4800 | 100 | 27.4700 | 27.4800 | 100 | 200 | 0.01 |
| 1546016214995572426 | 2018-12-28T11:56:54.995572426 | 27.4900 | 100 | 27.4800 | 27.4900 | 100 | 100 | 0.01 |
| 1546016217730988880 | 2018-12-28T11:56:57.730988880 | 27.4800 | 100 | 27.4800 | 28.0300 | 100 | 100 | 0.55 |
| 1546016340673226997 | 2018-12-28T11:59:00.673226997 | 27.5100 | 10 | 0.0000 | 28.0300 | 0 | 100 | 28.03 |
| 1546016716405329250 | 2018-12-28T12:05:16.405329250 | 27.5700 | 100 | 0.0000 | 27.5700 | 0 | 100 | 27.57 |
| 1546016839254135666 | 2018-12-28T12:07:19.254135666 | 27.5650 | 100 | 27.5500 | 27.5700 | 100 | 200 | 0.02 |
| 1546016867985948468 | 2018-12-28T12:07:47.985948468 | 27.5850 | 100 | 27.5600 | 27.5900 | 100 | 100 | 0.03 |
| 1546016867995677693 | 2018-12-28T12:07:47.995677693 | 27.5900 | 100 | 27.5600 | 27.5900 | 100 | 100 | 0.03 |
| 1546016882977866533 | 2018-12-28T12:08:02.977866533 | 27.6000 | 100 | 27.5600 | 27.6000 | 100 | 100 | 0.04 |
| 1546016887239675914 | 2018-12-28T12:08:07.239675914 | 27.5850 | 100 | 0.0000 | 27.6000 | 0 | 200 | 27.6 |
| 1546016916946082959 | 2018-12-28T12:08:36.946082959 | 27.5700 | 100 | 27.5700 | 27.5800 | 100 | 200 | 0.01 |
| 1546016953460481286 | 2018-12-28T12:09:13.460481286 | 27.5700 | 600 | 0.0000 | 28.0300 | 0 | 100 | 28.03 |
| 1546016954052217750 | 2018-12-28T12:09:14.052217750 | 27.5700 | 100 | 0.0000 | 28.0300 | 0 | 100 | 28.03 |
| 1546017043550384027 | 2018-12-28T12:10:43.550384027 | 27.5500 | 100 | 27.5500 | 27.5600 | 100 | 200 | 0.01 |
| 1546017092499344351 | 2018-12-28T12:11:32.499344351 | 27.5300 | 100 | 0.0000 | 27.5300 | 0 | 100 | 27.53 |

Notice that in this data the bids disappear at one point but trades still occur at reasonable prices.

Investigation into this issue found the following article:

https://mechanicalmarkets.wordpress.com/2017/08/14/liquidity-disincentives-iex/

This sheds some light on the subject by explaining that most of the liquidity on the IEX exchange is "dark liquidity."  This makes more sense if you understand the types of orders that the exchange allows:

- Market
- Limit
- Primary Peg (pegged to the national best bid/offer)
- Midpoint Peg
- Discretionary Peg

Market and limit orders are well known.  The peg orders are more interesting.  These basically place orders based on the NBBO (national best bid/offer) which is a combination of data from all the U.S. exchanges (the SIP feed, as Michael Lewis explains in his book Flash Boys).  People who place pegged orders are basically asking to trade at the NBBO, whatever it might be. These orders do not show up in the order books, and we are not notified of their presence.  However, it appears that most of the trades are executing at these NBBO prices without hitting the order book.

Out of 837 trades this day only 407 appear to have hit the order book and traded at the bid or ask prices on the IEX exchange.  So this raises the question: is the order book information really that useful?  When the spreads are wide it doesn't seem to be very helpful.  The trade report data is

much more valuable since it provides insight to the NBBO prices regardless of what's happening on IEX's order book.

Philosophically, we might also wonder why anyone would bother posting trades to IEX's order book. Unlike other exchanges that might pay for these types of orders (most exchanges pay liquidity makers and charge liquidity takers), IEX charges every trade a flat fee of $0.0009 per share. This is regardless of whether you place a market order, peg order, or limit order. So why would anyone place a limit order that would hit the books? I'm not sure. The article above mentions that IEX still has some work to do to show that it can assist in "price discovery" via its order book, or is it just executing trades at prices based on other exchanges order books? It seems like the latter is predominantly the case at this point – at least for as much as 50% of the day in this rather highly-traded stock.

**Summary**

Overall, the detailed intraday data feeds from IEX are pretty interesting for the visibility they provide into activity on the exchange. The trade data may even be somewhat useful as it provides a decent view into minute-level data on 500-1000 stocks. The order book data is less useful as the order book does not do a good job of representing the NBBO or orders placed on other exchanges. The spreads on the IEX order book are unrealistic at many times throughout the day. Sometimes the bids and asks disappear completely.

It would be preferable to get minute-level data from a source that more closely follows the NBBO for all stocks or shows trades across all exchanges. Such a feed is produced by the CTA (Consolidated Tape Association) and is used as a principal component of many stock exchanges and trade execution engines. However, they would like several thousand dollars per month for this data. If you prefer data that's free, the IEX data is probably the best that you're going to be able to find.

Also remember: if you don't need intraday data; if you are instead interested in daily OHLC data and higher-level info such as corporate info, dividends, and splits, IEX provides a series of high-quality APIs to get this data. The data in these feeds appears to take into account OHLC prices from all exchanges (not just IEX) and could be very useful. They currently don't charge anything for access to these APIs (which is crazy!). They do plan to offer subscription access in the future (through a service called IEX cloud). Their site says that while they plan to keep IEX data always free, they plan to charge $9.00/month to access more comprehensive data that includes other sources. $9.00/month for this data is still a steal. I suspect we'll find much more value in this data than in the intraday feeds.