algoseek
THE MARKET DATA COMPANY

US Equity Trade Only Guide

version 2.7 (Jul 2021)

CONTACT US

We are here to help you do great things with our market and reference data. For questions, feedback, and other concerns, you may reach our team of experts using the following contact information:

algoseek customer support

support@algoseek.com

(+1) 646 583 1832

algoseek sales

sales@algoseek.com

(+1) 646 583 1832

TABLE OF CONTENTS

INTRODUCTION

algoseek Trade Only data contain all trades for all listed stocks, ETNs, ETFs, ADRs, and funds from 15+ US exchanges and marketplaces. Trade Only data files are organized into a single format feed where events are ordered by the time received with nanosecond timestamps starting from 2016 and millisecond timestamps before. The entire trading session includes early and late hours from 04:00 to 20:00 EST.

SIP DATA FEED

algoseek collects live data from the Securities Information Processor (also known as the "Consolidated Feed") via its co-located ticker plant servers in Equinix NY2 and NY4 data center with 10Gb fiber connection for ultra-low latency.

The Securities Information Processor (SIP) includes Tape A and Tape B covered by the Consolidated Tape Association (CTA) plan and Tape C covered by the Unlisted Trading Privileges (UTP) plan. The SIP links the US markets by processing and consolidating all protected bid/ask quotes and trades from every trading venue into a single and easily consumable data feed.

The idea behind the creation of SIP was to form a national market system where investors and professionals can access real-time price information. Law prohibits exchanges from sending their quotes and trades to direct feeds before sending them to the SIP. In the highly fragmented world of US equities, the SIP is an easy way for people to get a view of the current state of the market. The SIP acts as the benchmark for regulators and others to determine the National Best Bid and Offer (NBBO). It also publishes other important information, such as short sale restrictions and regulatory halts.

DATA DISTRIBUTION AND COLLECTION

There are four major national listing Exchanges in the US: NYSE, NYSE Market (old AMEX), NASDAQ, and BATS - and 14 other national and regional exchanges and market centers. Please refer to Table 4 for a full list of market centers. When a company goes "public" via Initial Public Offering (IPO), Direct Listing, or a merger with SPAC, it is officially listed on a listing exchange and becomes available for trading on multiple other exchanges.

Direct feeds from the exchanges are different and you'll need to connect to all the exchanges to get a full picture of the market liquidity, so it is used by high-frequency / ultra-low latency trading firms with co-located computers.

SIP data provides an alternative way of collecting market data. When subscribing to CTA and UTP plans directly or via a vendor, you are receiving all the trades and top-of-book quotes from all exchanges in a consolidated feed.

CTA Plan

Trades and quotes data for stocks listed on NYSE and AMEX were distributed by the Consolidated Tape Association (CTA). The CTA has two services:

> **Consolidated Tape System (CTS):** an electronic system that collates real-time exchange-listed trade data, such as price and volume
>
> **Consolidated Quotation System (CQS):** an electronic service that provides quotation information and also includes issues traded by FINRA member firms in the third market

Any market center (exchange, dark pool, etc.) must report any trades for stocks that are officially listed on the NYSE or AMEX to CTA.

To learn more, please visit https://www.ctaplan.com/index.

UTP Plan

NASDAQ listed stocks have their trades and quotes reported through Unlisted Trading Privileges (UTP). Under the UTP Plan, all US exchanges that quote and trade NASDAQ listed securities must provide their data to a centralized SIP for data consolidation and dissemination.

NASDAQ, in its current role as the SIP for the UTP plan, supports the following data feed products:

> **UTP Quotation Data Feed (UQDF)** provides the best bid and offer (BBO) quotes from the UTP participants as well as the consolidated national best bid and offer (NBBO) quotes for securities listed on the NASDAQ Stock Market.
>
> **UTP Trade Data Feed (UTDF)** provides trade data from the UTP participants for securities listed on the NASDAQ Stock Market.

To learn more, please visit http://www.utpplan.com/overview.

Data Normalization

algoseek collects data from multicast CTS/CQS/UTDF/UQDF feeds, then normalizes and consolidates the data to deliver it in CSV format.

“As-is” Data

Live data from exchanges is not perfect. It is susceptible to issues such as crossed NBBO, bad prices, and bad trades, as well as exchange publishing mistakes such as out-of-sequence packets, but it is the data clients receive in real trading.

In order to provide the most realistic scenarios for clients to work with historical data, by default, we do not modify the data received live from exchanges.

If you need cleaned data, please contact us and specify your needs.

DATA ORGANIZATION AND FILE FORMAT

algoseek provides Equity market data in plain-text CSV files. The first row of the CSV file is a fixed header, and then rows of data corresponding to individual events (see Table 1). By default, data is organized into one file per symbol per trading day. For example, all events for ticker AAPL on Mar 3, 2020, are stored in one CSV file.

Due to the large data size, CSV files are gzip-compressed (having a csv.gz extension) with a compression ratio of about 8:1.

Table 1: Sample Trade Only Data

Date	Timestamp	Event Type	Ticker	Price	Quantity	Exchange	Conditions
20200128	09:30:00.004925261	TRADE	AAPL	312.38	58	ARCA	80000401
20200128	09:30:00.015300106	TRADE	AAPL	312.38	37	NASDAQ	a0000421
20200128	09:30:00.015303789	TRADE	AAPL	312.38	43	NASDAQ	a0000421
20200128	09:30:00.015362118	TRADE	AAPL	312.38	4	NASDAQ	a0000421
20200128	09:30:00.017242046	TRADE	AAPL	312.38	53	NASDAQ	80000401
20200128	09:30:00.017245998	TRADE	AAPL	312.38	47	NASDAQ	80000401
20200128	09:30:00.017394309	TRADE	AAPL	312.33	91	NASDAQ PSX	80000001
20200128	09:30:00.032299881	TRADE	AAPL	312.38	53	NASDAQ	80000401
20200128	09:30:00.032303999	TRADE	AAPL	312.38	84	NASDAQ	80000401
20200128	09:30:00.057034599	TRADE	AAPL	312.38	4	EDGA	a0000021

Table 2 below provides the name, description, and data type for each data field.

Table 2: CSV File Fields Schema

Field	Type (Format)	Description
Date	string (yyyymmdd)	Trading date in yyyymmdd format
Timestamp	string (HH:MM:SS.mmm or HH:MM:SS.mmmuuunnn)	Event timestamp in nanoseconds (milliseconds before 2016)
EventType	string	The type of event
Ticker	string	Symbol name
Price	decimal	The price of Trade. It can be up to 4 decimal places for sub-penny prices.
Quantity	integer	The number of shares. For some event types, this can be 0.
Exchange	string	The exchange or reporting venue
Conditions	hexadecimal (base 16)	Conditions applicable to the trade

Time Range

The Trade Only dataset covers the entire trading day from the start of pre-market trading to the end of after-hours trading (EST time):

Pre-Market Hours: 04:00:00 to 09:29:59

Market Hours: 09:30:00 to 16:00:00

Post-Market Hours: 16:00:01 to 20:00:00

Note: Occasionally trade and quote events are recorded several minutes after 20:00.

Market Holidays and Early Closes

The stock market is closed for trading on most US holidays. For reference, algoseek publishes a list of historical holidays which is available at `s3://us-equity-market-holidays/holidays.csv` (direct download link: https://us-equity-market-holidays.s3.amazonaws.com/holidays.csv).

Markets sometimes close early at 13:00:00 on the day before holidays such as Independence Day and Thanksgiving. You can download algoseek's early close date and time list from AWS S3 storage at

`s3://us-equity-market-holidays/earlycloses.csv` (or use a direct link us-equity-market-holidays.s3.amazonaws.com/earlycloses.csv).

Timestamp

The event timestamp has a nanosecond resolution, and the time zone is EST. Timestamp field takes the format of HH:MM:SS.mmmuuunnn, for example, 09:31:01.723317846, where

HH: Hour

MM: Minute

SS: Seconds

mmm: Milliseconds

uuu: Microseconds

nnn: Nanoseconds

Before 2016 events were published with millisecond timestamps (HH:MM:SS.mmm format). For example, 09:32:00.321.

Timestamps in Excel. Excel fails when importing timestamp fields as Excel automatically tries to convert milliseconds and nanoseconds to Excel time format. When importing timestamp, you can import as Text fields instead.

Event Types

Table 3 contains names and descriptions of event types present in data files.

Table 3: Event Types in Equity Trade Only Dataset

Event Type	Description
TRADE	Trade at the exchange or reported to the market center
TRADE NB	Trade at national best price
TRADE CANCELLED	Previous trade cancellation (reduce total traded volume by this amount). Trade corrections are constructed as a cancellation of the previous trade with the corrected trade following the cancellation

Note: usually a cancelation event does not immediately follow the trade it cancels and can be published up to a few hours after the trade event occurred.

National Best Bid and Offer Events

TRADE NB refers to a trade at the National Best Bid Offer (NBBO). Trades are not always executed at NBBO prices. Examples of trades that are not at NBBO prices include:

Some FINRA trades: FINRA trades are TRF reports of trades that are not done on the public Exchanges, and are not subject to the regulation of Rule 611.

Late report from the exchange: due to late reporting from a market center.

Extended hours trades: no NBBO in pre/post-market hours.

ISO orders: An Intermarket Sweep Order (ISO) can be executed outside of the NBBO by an exchange.

Other trade-through exemptions

For backtesting with trades, the best practice is to use the Trade NB reports.

Market Center ID

Table 4 contains all the 19 SIP Market Center IDs (including FINRA).

Table 4: algoseek Value for SIP Market Center IDs

	Exchange Name	SIP Market Center ID	algoseek Value
NYSE Group of Exchanges			
1	New York Stock Exchange LLC	N	NYSE
2	NYSE Arca, Inc.	P	ARCA
3	NYSE American, LLC	A	AMEX
4	NYSE National, Inc.	C	NSE
5	NYSE Chicago, Inc.	M	CSE
Nasdaq Group of Exchanges			
6	Nasdaq, Inc.	T (CTS) / Q (UTP)	Nasdaq
7	Nasdaq BX, Inc.	B	Nasdaq BX
8	Nasdaq PHLX LLC	X	NASDAQ PSX
9	Nasdaq ISE, LLC	I	ISE

CBOE Group of Exchanges			
10	Cboe Exchange, Inc.	W	CBOE
11	Cboe BYX Exchange, Inc.	Y	BATS Y
12	Cboe BZX Exchange, Inc.	Z	BATS
13	Cboe EDGA Exchange, Inc.	J	EDGA
14	Cboe EDGX Exchange, Inc.	K	EDGX
Independent Venues			
15	Investors' Exchange LLC (IEX)	V	IEX
16	Long-Term Stock Exchange (LTSE)	L	LTSE
17	MIAX Pearl, LLC (MIAX)	H	MIAX
18	MEMX LLC (MEMX) (Member's Exchange)	U	MEMX
19	Financial Industry Regulatory Authority	D	FINRA
Other Values			
20	Unknown		UNKNOWN
21	Invalid		INVALID

Normalized Condition Flags

algoseek Trade Only dataset normalizes the Sales Condition Modifiers from UTP/CTS plans into a 32-bit bitmask. Each flag shows a condition applicable to the trade. The flag value indicates a bit position of the flag value inside the unsigned 32-bit integer value in the base-16 (hex) format.

Note: the bit position 0 corresponds to the rightmost bit and 31 is the leftmost bit.

Download algoseek's sample python code for converting the Conditions flag value into its text format from https://github.com/aquanyc/algoseek_public/tree/master/utils/flag_decoder

Table 5: Trade Condition Flags

Bit Mask Position	Flags	Description
Settlement Type		
0	tRegular	A trade made without stated conditions is deemed as a regular way for settlement on the third business day following the transaction date.
1	tCash	A transaction that requires delivery of securities and payment on the same day the trade takes place.
2	tNextDay	A transaction that requires the delivery of securities on the first business day following the trade date.
3	tSeller	A Seller's Option transaction gives the seller the right to deliver the security at any time within a specific period, ranging from not less than two calendar days to not more than sixty calendar days. A security offered "Seller's Option" may command a lesser price than if offered "Regular Way".
4	tYellowFlag	Market centers will have the ability to identify regular trades being reported during specific events as out of the ordinary by appending a new sale condition code Yellow Flag ("Y") on each transaction reported to the UTP SIP. The new sale condition "Y" will be eligible to update all market centers and consolidated statistics.
Reason for Trade-Through Exemption		
5	tIntermarketSweep	The transaction that constituted the trade-through was the execution of an order identified as an Intermarket Sweep Order.
6	tOpeningPrints	The trade that constituted the trade-through was a single priced opening transaction by the market center.
7	tClosingPrints	The transaction that constituted the trade-through was a single priced closing transaction by the market center.
8	tReOpeningPrints	The trade that constituted the trade-through was a single priced reopening transaction by the market center.
9	tDerivativelyPriced	The transaction that constituted the trade-through was the execution of an order at a price that was not based, directly or indirectly, on the quoted price of the security at the time of execution, and for which the material terms were not reasonably determinable at the time the commitment to execute the order was made.
Extended Hours / Sequence Type		
10	tFormT	Trading in extended hours enables investors to react quickly to events that typically occur outside regular market hours such as

		earnings reports. However, liquidity may be constrained during such FormT trading, resulting in wide bid-ask spreads.
11	tSold	Sold Last is used when a trade prints in sequence but is reported late or printed in conformance to the One- or Two-Point Rule.
12	tStopped	The transaction that constituted the trade-through was the execution by a trading center of an order for which, at the time of receipt of the order, the trading center had guaranteed execution at no worse than a specified price (a “stopped order”), where: (i) the stopped order was for the account of a customer; (ii) the customer agreed to the specified price on an order-by-order basis; and (iii) the price of the trade-through transaction was for a stopped buy order lower than the National Best Bid in the security at the time of execution, or for a topped sell order higher than the National Best Offer in the security at the time of execution.
13	tExtendedHours	Identifies a trade that was executed outside of regular primary market hours and is reported as an extended-hours trade.
14	tOutOfSequence	Identifies a trade that takes place outside of regular market hours.
Other Types		
15	tSplit	An execution in two markets when the specialist or market maker in the market first receiving the order agrees to execute a portion of it at whatever price is realized in another market to which the balance of the order is forwarded for execution.
16	tAcquisition	A transaction made on the exchange as a result of an exchange acquisition.
17	tBunched	A trade representing an aggregate of two or more regular trades in a security occurring at the same price either simultaneously or within the same 60-second period, with no individual trade exceeding 10,000 shares.
18	tStockOption	Stock-Option Trade is used to identify cash equity transactions that are related to options transactions and therefore potentially subject to cancellation if market conditions of the options leg(s) prevent the execution of the stock-option order at the price agreed upon.
19	tDistribution	Sale of a large block of stock in such a manner that the price is not adversely affected.
20	tAveragePrice	A trade where the price reported is based upon an average of the prices for transactions in a security during all or any portion of the trading day.
21	tCross	Indicates that the trade resulted from a market center’s crossing session.

22	tPriceVariation	Indicates a regular market session trade transaction that carries a price that is significantly away from the prevailing consolidated or primary market value at the time of the transaction.
23	tRule155	To qualify as an NYSE AMEX Rule 155, from time to time, a specialist may arrange for the sale or purchase of a block of security or other large numbers of shares of securities at a single “clean-up” price. Generally, such a sale or purchase is outside of the current market. Such sale or trade is designated as a Rule 155 trade.
24	tOfficialClose	Indicates the ‘Official’ closing value as determined by a market center. This transaction report will contain the market center generated closing price.
25	tPriorReferencePrice	A sale condition that identifies a trade based on a price at a prior point in time, i.e. more than 90 seconds prior to the time of the trade report. The execution time of the trade will be the time of the prior reference price.
26	tOfficialOpen	Indicates the ‘Official’ opening value as determined by a Market Center. This transaction report will contain the market center generated opening price.
27	tCapElection	The CAP Election Trade highlights sales as a result of a sweep execution on the NYSE, whereby CAP orders have been elected and executed outside the best price bid or offer and the orders appear as "repeat" trades at subsequent execution prices. This indicator provides additional information to market participants that an automatic sweep transaction has occurred with repeat trades as one continuous electronic transaction.
28	tAutoExecution	A sale condition code that identifies an2 NYSE trade that has been automatically executed without the potential benefit of price improvement.
29	tTradeThroughExempt	Denotes whether a trade is exempt (Rule 611) and when used jointly with certain Sale Conditions, will more fully describe the characteristics of a trade.
30		Not in use
31	tOddLot	Denotes the trade is an odd lot less than 100 shares.

APPENDIX A. TRADE CONDITION FLAGS EXAMPLES

Flags are eight-character hexadecimal representations of an unsigned 32-bit integer. The following examples focus on Opening and Closing prices.

Example 1

```
09:30:00.948, TRADE NB, AIZ, 32.42, 700, NASDAQ, 20200040
```

Flag: `20200040`

Position: `10987654321098765432109876543210`

Binary: `00100000001000000000000001000000`

```
tOpeningPrints 6

tCross 21

tTradeThroughExempt 29
```

Example 2

```
09:30:00.948, TRADE, AIZ, 32.42, 700, NASDAQ, 04000000
```

Flag: `04000000`

Position: `10987654321098765432109876543210`

Binary: `00000100000000000000000000000000`

```
tOfficialOpen 26
```

Example 3

```
09:30:00.555, TRADE NB, INTC, 36.62, 761347, NASDAQ, 20200041
```

Flag: `20200041`

Position: `10987654321098765432109876543210`

Binary: `00100000001000000000000001000001`

```
tRegular 0

tOpeningPrints 6

tCross 21

tTradeThroughExempt 29
```

APPENDIX B. FREQUENTLY ASKED QUESTIONS

Why are time-based columns not recognized properly when I try importing data to Excel?

Older versions of Excel will automatically try to convert the “Timestamp” field into an Excel format timestamp but this fails when “Timestamp” is in HH:MM:SS.mmm (millisecond) or HH:MM:SS.mmmuuunnn (nanosecond) format. For timestamp with the nanosecond (millisecond) format, import the data using Excel “From Text” option and set the data type for column “Timestamp” to “Text,” so Excel does not automatically try to convert it.