



US Equity Trade Only Minute Bar Guide

version 1.5 (Jul 2021)



algoseek | the market data company

We provide research market data for machine learning and quantitative trading




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	4
DATA SOURCE	4
STANDARD AND NO-FINRA/TRF VERSIONS	4
DATA ORGANIZATION AND FILE FORMAT	5
APPENDIX A. FREQUENTLY ASKED QUESTIONS	9
APPENDIX B. BAR CALCULATIONS FROM TRADE EVENTS	10



INTRODUCTION

algoseek Trade Only Minute Bar data is built from trades for all listed stocks, ETNs, ETFs, ADRs, and funds from 16+ US exchanges and marketplaces.

algoseek provides two versions of US Equity Trade Only Minute Bar datasets: a standard version and a version without FINRA/TRF trades. In this document, they are collectively referred to as Trade Only Minute Bar datasets.

The algoseek Trade Only Minute Bar datasets are designed for quantitative trading, backtesting, machine learning and other applications.

Data files are in CSV (Comma-Separated Values) format. An individual CSV file is created for each active ticker on each trading day, and these data files are arranged in a flat-file database by date and then by ticker. If there are no trades during the entire day for a ticker, an empty CSV file with no bars will be created.

Note: All features and behavior of the dataset will be described in terms of a 1-minute resolution bar. In the meantime, all information applies to other resolutions (for example, 1-second) as well.

DATA SOURCE

algoseek Trade Only Minute Bar datasets are built from “as-is” tick data collected from live SIP feed algoseek’s co-located ticker plant servers in Equinix NY2 and NY4 data centers, connected with 10Gb fiber for 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 SIP disseminates and calculates critical regulatory information, including the National Best Bid and Offer (NBBO) and Limit Up Limit Down (LULD) price bands, among other important regulatory information such as short sale restrictions and regulatory halts. In the highly fragmented world of US equities, the SIP is an easy way for people view the current state of the market.

STANDARD AND NO-FINRA/TRF VERSIONS

Two versions of Trade Only Minute Bar datasets are available for algoseek clients:



Standard Minute Bar: includes all trades from the SIP feed

No-FINRA/TRF and Odd Lots: includes only trades executed on public exchanges and excludes all FINRA/TRF reports and odd lot (less than 100 shares) trades

FINRA TRF

Equity trades are executed on Public Exchanges (e.g. NASDAQ, BATS, NYSE, ARCA, etc.) and off the public exchanges in Dark Pools, Broker-Dealer internal crossing, and Block Trades. Regulation National Market System (NMS) requires all trades to be reported. There are currently three FINRA Trade Reporting Facilities (TRF) affiliated with registered national securities exchanges and provide FINRA members with a mechanism for reporting transactions affected otherwise than on an exchange.

Regulation NMS allows up to 10 seconds after the Trade execution time for the trade report to be sent to an exchange's TRF for publication. The delay can result in TRF Trade reports printed on the market data feed being out of the current NBBO.

Round Lot and Odd Lot

A round lot (or board lot) is a normal unit of trading of a security, which currently is 100 shares of stock in the US. Any quantity less than 100 shares is referred to as an odd lot. Odd lots are not subject to the Regulation NMS rules requiring execution to be within the current NBBO. Broker-dealers send odd lots to the exchange paying the most rebate per share and not the best execution price.

Note: Odd lot executions can create unrealistic high/low trade prices in an OHLC bar.

The No FINRA and Odd Lots Minute Bars

By excluding TRF trades and odd-lot trades, the no-FINRA/TRF and Odd Lots Trade Only Minute Bar dataset provides "clean" and easy-to-use data for hedge funds and market makers to trade and backtest trading algorithms based on the trades from Public Exchanges.

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 bars (Table 1). By default, data is organized into one file per symbol per trading day. For example, all trade bars 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 Minute Bar Data

Date	Ticker	Time Bar Start	First Trade Price	High Trade Price	Low Trade Price	Last Trade Price	Volume Weight Price	Volume	Total Trades
20200128	AAPL	09:30	312.38	312.93	312.33	312.85	312.5619	785594	3029
20200128	AAPL	09:31	312.87	312.92	312.1806	312.39	312.52706	209850	2466
20200128	AAPL	09:32	312.3650	313.24	312.30	313.2136	312.82281	220340	2492
20200128	AAPL	09:33	313.21	314.00	312.50	313.97	313.65868	308824	2956
20200128	AAPL	09:34	313.9836	314.03	312.50	313.32	313.73172	289758	2835

Table 2 (below) provides the name, base event, default value, brief description, and data type for each data field (column) in the Equity Trade Only Minute Bar CSV file.

Table 2: CSV File Fields Schema

Field	Type (Format)	Description
Date	string (yyyymmdd)	Trading date in yyyymmdd format
Ticker	string	Symbol name
TimeBarStart	string (time)	Start Time of the Bar. For minute bar format is HH:MM. For second bar format is HH:MM:SS
FirstTradePrice	decimal	Price of the first trade
HighTradePrice	decimal	Trade with the highest price
LowTradePrice	decimal	Trade with the lowest price
LastTradePrice	decimal	Price of the last trade
VolumeWeightPrice	decimal	Volume weighted average price
Volume	integer	Total number of shares traded
TotalTrades	integer	Total number of trades

Time Range

The Trade Only Minute Bar datasets cover the entire trading day from the start of pre-market trading to the end of after-hours trading (ET time):

Pre-Market Hours: 04:00:00 to 09:30:00 (excluding)

Market Hours: 09:30:00 to 16:00:00 (excluding)



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

Note: Occasionally, minute bars are extended several minutes past 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 holiday,s 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).

Bar Notes

Time Bar Start Format: One-second bar 13:03:01 is from time greater than 13:03:01 to less than 13:03:02.

One-minute bar 11:04 is from time greater than 11:04 to less than 11:05.

Single Event: For bars with only one trade,

`FirstTradePrice = HighTradePrice = LowTradePrice = LastTradePrice`

Price Field Example: If two trades occur with the first trade price higher than the second trade price, then,

`FirstTradePrice = Price of the first trade`

`HighTradePrice = Price of the first trade`

`LowTradePrice = Price of the second trade`

`LastTradePrice = Price of the second trade`

VolumeWeightPrice: the volume-weighted price is calculated as a dollar volume a sum of all trades divided by the total number of shares traded

$$\frac{\text{sum}(\text{Trade_Shares} \times (\text{Trade_Price} - \text{NBBOMidpoint}))}{\text{sum}(\text{Trade_Shares})}$$



Skipped Bars

The bars are trade-based, so there is no quote-related data, and a bar is only created when there is at least one trade during the bar period. When there are no trades during certain minutes, the timestamps are skipped as exhibited in Table 3.

Table 3: Sample Trade Only Minute Bar Data

Date	Ticker	Time Bar Start	First Trade Price	High Trade Price	Low Trade Price	Last Trade Price	Volume Weight Price	Volume	Total Trades
20200128	GDP	10:42	7.53	7.53	7.5201	7.5201	7.52017	756	9
20200128	GDP	10:43	7.53	7.53	7.53	7.53	7.53	11	1
20200128	GDP	10:45	7.53	7.53	7.53	7.53	7.53	100	1
20200128	GDP	10:54	7.53	7.53	7.53	7.53	7.53	10	2
20200128	GDP	11:05	7.56	7.58	7.56	7.58	7.56813	585	7

This implies there were no trades during the bar period 10:44, 10:46 through 10:53, and between 10:55 and 11:05.

Fewer bars will be displayed for thinly traded stocks or outside regular market hours due to a lack of activities. When only the header row is present, the security was not traded at all during the day.



APPENDIX A. FREQUENTLY ASKED QUESTIONS

Why do I see empty files with just a header line for some tickers?

An empty file is created for some tickers with low liquidity with no trades during the trading day but Bid/Ask quotes were published.



APPENDIX B. BAR CALCULATIONS FROM TRADE EVENTS

This section describes logic for minute bar calculations based on events from the Trade Only dataset. Please also refer to the Equity Trade Only Guide for more details on data fields and condition flags used.

There is a separate logic for the Standard Bars dataset and Bars with FINRA/TRF and Odd Lots Excluded.

Standard Trade Only Minute Bar

Excluded data

- Exclude any trade with Price = 0 (not applied to Trade Cancel)
- Exclude any trade with Quantity = 0

You should also exclude any event with one or more flags listed in Table 4.

Table 4: Flags for Trade Events to be Excluded During Bar Calculations

Bit Mask Position	Flags
14	tOutOfSequence
20	tAveragePrice
22	tPriceVariation
23	tRule155
24	tOfficialClose
25	tPriorReferencePrice
26	tOfficialOpen

Included data

- Include both NBBO and exchange trades

You should only include events with one or more flags listed in Table 5. If the event has any of the exclude flags enabled, it is not included. If the event does not contain any flags from the include list, it is not included in bar calculations.

**Table 5: Flags for Trade Events to be Included During Bar Calculations**

Bit Mask Position	Flags
0	tRegular
1	tCash
2	tNextDay
5	tIntermarketSweep
6	tOpeningPrints
7	tClosingPrints
10	tFormT
13	tExtendedHours
21	tCross
29	tTradeThroughExempt
31	tOddLot

No-FINRA/TRF and Odd Lots Trade and Quote Minute Bar

Excluded data

- Exclude any trade with Price = 0 (not applied to Trade Cancel)
- Exclude any trade with Quantity = 0
- Exclude any trade with Exchange = FINRA

Also, you should exclude any event with one or more flags listed in Table 6.

Table 6: Flags for Trade and Quote Events to be Excluded During Bar Calculations (No-FINRA/TRF Dataset)

Bit Mask Position	Flags
14	tOutOfSequence
20	tAveragePrice
22	tPriceVariation
23	tRule155



24	tOfficialClose
25	tPriorReferencePrice
26	tOfficialOpen
31	tOddLot

Included data

- Include both NBBO and exchange trades

You should only include events with one or more flags listed in Table 7. If the event has any of the exclude flags enabled, it is not included. If the event does not contain any flags from the include list, it is not included in bar calculations.

Table 7: Flags for Trade and Quote Events to be Included During Bar Calculations (No-FINRA/TRF Dataset)

Bit Mask Position	Flags
0	tRegular
1	tCash
2	tNextDay
5	tIntermarketSweep
6	tOpeningPrints
7	tClosingPrints
10	tFormT
13	tExtendedHours
21	tCross
29	tTradeThroughExempt