algoseek
THE MARKET DATA COMPANY

US Equity Industry Standard Trade Only Minute Bar Guide

version 1.1 (Oct 2023)

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 through 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 Industry Standard Trade Only Minute Bar Adjusted data is built from trades for all listed stocks, ETNs, ETFs, ADRs, and funds from 16+ US exchanges and marketplaces based on Industry Standard logic. This dataset contains OHLC prices for each minute during the full trading session including pre and post-market hours.

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 to view the current state of the market.

FINRA TRF AND ODD LOT 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. *Odd lot executions can create unrealistic high/low trade prices in an OHLC bar.*

DATA ORGANIZATION AND FILE FORMAT

algoseek provides Equity market data in plain text CSV files. The first row of 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 for a separate trading day are zip-compressed.

Table 1: Sample Trade Only Minute Bar Data

SecId	33449	33449	33449
Date	20200825	20200825	20200825
Ticker	AAPL	AAPL	AAPL
TimeBarStart	09:30	09:31	09:32
FirstTradePrice	498.76	499.58	499.35
HighTradePrice	500.75	500.75	499.38
LowTradePrice	498.57	498.55	496.96
LastTradePrice	499.63	499.2	497.3106
VolumeWeightPrice	499.11041	499.59889	497.78382
Volume	1059318	305868	434849
TotalTrades	8387	5379	8305

Table 2 below summarizes the name, description, and data type for each data field (column) in the Industry Standard Equity Trade Only Minute Bar CSV file.

Table 2: CSV File Fields Schema

Field	Type	Description

	(Format)	
SecId	integer	algoseek unique security identifier
Date	string (yyyymmdd)	Trading date in yyyymmdd format
Ticker	string	Symbol name
TimeBarStart	string (time)	Start Time of the Bar. For minute bars, the format is HH:MM. For the second bars, the 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 Industry Standard Trade Only Minute Bar dataset covers the entire trading day from the start of pre-market trading to post-hour 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: During holidays e.g. Christmas Eve or early market closure for other reasons, trading only happens half-day at 13:00:00. Download algoseek's early closing date 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). algoseek provides a reference list for US holidays when the stock market is closed at s3://us-equity-market-holidays/holidays.csv (direct download link: https://us-equity-market-holidays.s3.amazonaws.com/holidays.csv).

Timestamps in Excel. Excel fails when importing timestamp fields as Excel automatically tries to convert milliseconds and nanoseconds to Excel time format. Import timestamp fields as Text instead.

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.

```
sum(Trade_Shares x (Trade_Price - NBBOMidpoint)) /
                sum(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: Trade Only Minute Bar Sample Data

SecId	3753921	3753921	3753921
Date	20200924	20200924	20200924
Ticker	GAL	GAL	GAL
TimeBarStart	10:06	10:13	10:25
FirstTradePrice	38.157	38.16	38.1
HighTradePrice	38.157	38.16	38.1
LowTradePrice	38.157	38.16	38.1

LastTradePrice	38.157	38.16	38.1
VolumeWeightPrice	38.157	38.16	38.1
Volume	100	100	105
TotalTrades	1	1	1

This implies there were no trades during the bar period 10:07 through 10:12, and between 10:13 and 10:25.

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. INDUSTRY STANDARD BAR CALCULATIONS FROM TRADE EVENTS

This section describes Industry Standard 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.

Excluded data

- Exclude any trade with Price = 0
- Exclude any trade with Quantity = 0
- Exclude Trade Cancel events
- Exclude the original trade that was canceled later

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
1	tCash
2	tNextDay
9	tDerivativelyPriced
11	tSold
13	tExtendedHours
18	tStockOption
20	tAveragePrice
22	tPriceVariation
23	tRule155
24	tOfficialClose
25	tPriorReferencePrice
26	tOfficialOpen
27	tCapElection
31	tOddLot

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
5	tIntermarketSweep
6	tOpeningPrints
7	tClosingPrints
10	tFormT
14	tOutOfSequence
21	tCross
29	tTradeThroughExempt

Also, data are shifted by 1 second starting from 09:31. For example,

09:30: from 09:30:00.000 to 09:31:00.999;

09:31: from 09:31:01.000 to 09:32:00.999;

09:32: from 09:32:01.000 to 09:33:00.999, and so on.