



US Equity Trade Only Minute Bar Adjusted Guide

version 1.2 (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 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	4
DATA SOURCE	4
STANDARD AND NO-FINRA/TRF VERSIONS	5
FINRA TRF	5
Round Lot and Odd Lot	5
The No FINRA and Odd Lots Minute Bars	5
DATA ORGANIZATION AND FILE FORMAT	5
Time Range	7
Bar Notes	8
Skipped Bars	8
Corporate Events	9
DAILY UPDATES	10
APPENDIX A. FREQUENTLY ASKED QUESTIONS	12



INTRODUCTION

algoseek Trade Only Adjusted Minute Bar data is built from trades for all listed stocks, ETNs, ETFs, ADRs, and funds from 16+ US exchanges and marketplaces. This dataset contains OHLC prices and additional fields for pricing adjusted by corporate events during the full trading session, including pre- and post-market hours.

For all information about corporate events applied to adjust the data and implemented logic for adjusting, see the 'Corporate Events' section.

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.

STANDARD AND NO-FINRA/TRF VERSIONS

Two versions of Equity Trade Only Adjusted 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. *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

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.

There are two data aggregation options for these datasets:

tradedate: one CSV file with data per symbol per trading day

SecId: one CSV file with data for all trading days per Security ID - a unique security identifier used by algoseek that remains unchanged when the ticker changes

Both aggregation options provide the same data fields.



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).

Data files are organized with one file per symbol per trading day or by SecId. For example, all trade bars for IBM or AAPL tickers on March 3, 2020, are each stored in a separate CSV file under a tradedate aggregation. In the case of SecId-based aggregation, all data for the security with an ID 33449 (AAPL) for a single year is stored in a single CSV file.

Due to the large dataset size, each CSV file is gzip-compressed, so the uncompressed data is on average seven times larger than the compressed.

Table 1: Sample Trade Only Minute Bar Adjusted 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
FirstTradePriceAdjusted	124.69	124.895	124.8375
HighTradePriceAdjusted	125.1875	125.1875	124.845
LowTradePriceAdjusted	124.6425	124.6375	124.24
LastTradePriceAdjusted	124.9075	124.8	124.3276
VolumeWeightPriceAdjusted	124.7776	124.8997	124.446
VolumeAdjusted	4237272	1223472	1739396

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



Table 2: CSV File Fields Schema

Field	Type (Format)	Description
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
FirstTradePriceAdjusted	decimal	Backward adjusted price of the first trade
HighTradePriceAdjusted	decimal	Backward adjusted trade with the highest price
LowTradePriceAdjusted	decimal	Backward adjusted trade with the lowest price
LastTradePriceAdjusted	decimal	Backward adjusted price of the last trade
VolumeWeightPriceAdjusted	decimal	Backward adjusted volume weighted average price
VolumeAdjusted	integer	Backward adjusted total number of shares traded

Time Range

The Trade Only Adjusted Minute Bar dataset covers the entire trading day from the start of pre-market trading to post-hour 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: 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



US Equity Trade Only Adjusted Minute Bar Guide

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.

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

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: Trade Only Adjusted 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
FirstTradePriceAdjusted	38.157	38.16	38.1
HighTradePriceAdjusted	38.157	38.16	38.1
LowTradePriceAdjusted	38.157	38.16	38.1
LastTradePriceAdjusted	38.157	38.16	38.1
VolumeWeightPriceAdjusted	38.157	38.16	38.1
VolumeAdjusted	100	100	105

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.



Corporate Events

Table 4: Corporate Event's effect on Price and/or Volume

Corporate Event	Affects Price	Affects Volume
Bonus issue in the same class	Yes	Yes
Bonus issue in a different class	Yes	No
Capital Reduction	Yes	Yes
Consolidation	Yes	Yes
Distribution	Yes	No
Cash Dividend	Yes	No
Script dividend in the same class	Yes	Yes
Script dividend in a different class	Yes	No
De-merger	Yes	No
Entitlement in the same class	Yes	No
Entitlement in a different class	Yes	No
Capital Return	Yes	No
Rights in the same class	Yes	No
Rights in a different class	Yes	No
Security Swap	Yes	Yes
Reclassification	Yes	Yes
Any subdivision (by any stock split, stock dividend, reclassification, recapitalization or otherwise) or combination (by the reverse stock split, reclassification, recapitalization, or otherwise) of the Class A Common Stock.	Yes	Yes

DAILY UPDATES

algoseek's Equity Trade Only Minute Bar Adjusted dataset contains backward adjusted prices and volume. Whenever a new corporate event (e.g., dividend, split, etc) is published for a SecId, all data for this SecId is rebuilt. In this case, you need to re-download the data for this security.

To access the list of updated SecIds for a specific trading day, please follow this s3 file path:
`s3://bucket_name/daily-changes/yyyymmdd.csv`



`bucket_name` can one of the following:

`us-equity-1min-trades-adjusted-secid-all`

`us-equity-1min-trades-adjusted-secid-yyyy`

where `yyyy` - year, `mm` - month, and `dd` - day.

There are a couple of ways to keep the data updated:

1. If you work with the data aggregated by SecId and would like to have the whole history for a specific security in one file, you can use the AWS CLI SYNC command on the `us-equity-1min-trades-adjusted-secid-all` bucket. This setup will download only files from the S3 bucket that are different from the local copy. To sync your local directory and S3 bucket, please make sure to point the SYNC command to your local copy of the dataset. In the meantime, this approach may be expensive (\$80 per month for the universe of 12000 tickers) if you regularly update the full universe of securities. To overcome it, please check the approach below.
2. You can work with the data aggregated by SecId and by year. In this case, you can use AWS SYNC on `us-equity-1min-trades-adjusted-secid-yyyy` buckets that you need, where `yyyy` is the year. If you still need to keep the whole history for each security in a single file, you can merge them on your side locally. It will be cheaper than updating data using the `us-equity-1min-trades-adjusted-secid-all` bucket (\$10 versus \$80 per month for the universe of 12000 tickers).
3. If you are working with the data aggregated by trade date, you can run the SYNC command for each `us-equity-1min-trades-adjusted-yyyy` bucket separately.
4. Also, algoseek provides a Python [script](#) that you can use to keep data updated for both aggregation types: by SecId and by ticker. Using this script, you can easily update data for a list of tickers or SecIds, and merge all data for each SecId.

Daily Update Time

Data aggregated by trade date is built by 3 am T+1 ET.

Data aggregated by SecId is built by 5 am T+1 ET.



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.