

data exploration

September 26, 2019

```
[1]: ! aws s3 ls s3://deutsche-boerse-xetra-pds/2018-01-01/
```

```
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO0.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO1.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO2.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO3.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO4.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO5.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO6.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO7.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO8.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETRO9.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR10.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR11.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR12.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR13.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR14.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR15.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR16.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR17.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR18.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR19.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR20.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR21.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR22.csv
2018-04-04 20:03:45      136 2018-01-01_BINS_XETR23.csv
```

```
[6]: ! aws s3 ls s3://deutsche-boerse-xetra-pds/2018-01-02/
```

```
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO0.csv
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO1.csv
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO2.csv
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO3.csv
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO4.csv
2018-04-04 20:03:45     1106 2018-01-02_BINS_XETRO5.csv
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO6.csv
2018-04-04 20:03:45      136 2018-01-02_BINS_XETRO7.csv
2018-04-04 20:03:45    981913 2018-01-02_BINS_XETRO8.csv
```

```

2018-04-04 20:03:45      973324 2018-01-02_BINS_XETR09.csv
2018-04-04 20:03:45      909783 2018-01-02_BINS_XETR10.csv
2018-04-04 20:03:45      780858 2018-01-02_BINS_XETR11.csv
2018-04-04 20:03:45      725804 2018-01-02_BINS_XETR12.csv
2018-04-04 20:03:45      774361 2018-01-02_BINS_XETR13.csv
2018-04-04 20:03:45     1039513 2018-01-02_BINS_XETR14.csv
2018-04-04 20:03:45     1152927 2018-01-02_BINS_XETR15.csv
2018-04-04 20:03:45      744488 2018-01-02_BINS_XETR16.csv
2018-04-04 20:03:45         136 2018-01-02_BINS_XETR17.csv
2018-04-04 20:03:46         136 2018-01-02_BINS_XETR18.csv
2018-04-04 20:03:46         136 2018-01-02_BINS_XETR19.csv
2018-04-04 20:03:46         136 2018-01-02_BINS_XETR20.csv
2018-04-04 20:03:46         136 2018-01-02_BINS_XETR21.csv
2018-04-04 20:03:46         136 2018-01-02_BINS_XETR22.csv
2018-04-04 20:03:46         136 2018-01-02_BINS_XETR23.csv

```

```

[2]: ! aws s3 cp s3://deutsche-boerse-xetra-pds/2018-01-02/ ./data/2018-01-02/
      ↪--recursive

```

```

download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR06.csv
to data/2018-01-02/2018-01-02_BINS_XETR06.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR02.csv
to data/2018-01-02/2018-01-02_BINS_XETR02.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR01.csv
to data/2018-01-02/2018-01-02_BINS_XETR01.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR05.csv
to data/2018-01-02/2018-01-02_BINS_XETR05.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR00.csv
to data/2018-01-02/2018-01-02_BINS_XETR00.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR04.csv
to data/2018-01-02/2018-01-02_BINS_XETR04.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR03.csv
to data/2018-01-02/2018-01-02_BINS_XETR03.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR07.csv
to data/2018-01-02/2018-01-02_BINS_XETR07.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR17.csv
to data/2018-01-02/2018-01-02_BINS_XETR17.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR19.csv
to data/2018-01-02/2018-01-02_BINS_XETR19.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR18.csv
to data/2018-01-02/2018-01-02_BINS_XETR18.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR20.csv
to data/2018-01-02/2018-01-02_BINS_XETR20.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR22.csv
to data/2018-01-02/2018-01-02_BINS_XETR22.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR21.csv
to data/2018-01-02/2018-01-02_BINS_XETR21.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR23.csv

```

```

to data/2018-01-02/2018-01-02_BINS_XETR23.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR09.csv
to data/2018-01-02/2018-01-02_BINS_XETR09.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR11.csv
to data/2018-01-02/2018-01-02_BINS_XETR11.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR13.csv
to data/2018-01-02/2018-01-02_BINS_XETR13.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR12.csv
to data/2018-01-02/2018-01-02_BINS_XETR12.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR10.csv
to data/2018-01-02/2018-01-02_BINS_XETR10.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR08.csv
to data/2018-01-02/2018-01-02_BINS_XETR08.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR14.csv
to data/2018-01-02/2018-01-02_BINS_XETR14.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR15.csv
to data/2018-01-02/2018-01-02_BINS_XETR15.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-02/2018-01-02_BINS_XETR16.csv
to data/2018-01-02/2018-01-02_BINS_XETR16.csv

```

```

[8]: # TODO can be omitted, because Kafka nor S3 are directly used
import os
os.environ['PYSPARK_SUBMIT_ARGS'] = '--packages org.apache.spark:
↳spark-sql-kafka-0-10_2.11:2.4.4,org.apache.hadoop:hadoop-aws:2.7.3_
↳pyspark-shell'

```

```

[1]: import findspark
findspark.init()

import pyspark

```

```

[2]: from pyspark.sql import SparkSession

spark = SparkSession.builder.appName("SimpleApp").getOrCreate()

```

```

[3]: spark

```

```

[3]: <pyspark.sql.session.SparkSession at 0x7fd04bfd11d0>

```

```

[14]: data = spark.read.format("csv").option("header", "true").load("./data/
↳2018-01-02/")
data.show()

```

```

+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
|      ISIN|Mnemonic|      SecurityDesc|SecurityType|Currency|SecurityID|
Date| Time|StartPrice|MaxPrice|MinPrice|EndPrice|TradedVolume|NumberOfTrades|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

```

-----+-----+-----+-----+-----+-----+-----+
|AT0000AOE9W5|    SANT|S+T AG (Z.REG.MK...|Common stock|    EUR|
2504159|2018-01-02|15:00|    19.19|    19.19|    19.19|    19.19|    100|
1|
|AT0000743059|    OMV|    OMV AG|Common stock|    EUR|
2504175|2018-01-02|15:00|    53.28|    53.28|    53.28|    53.28|    18|
1|
|AT0000969985|    AUS|AT+S AUSTR.T.+SYS...|Common stock|    EUR|
2504191|2018-01-02|15:00|    24.15|    24.2|    24.15|    24.15|    401|
3|
|CA32076V1031|    FMV|FIRST MAJESTIC SI...|Common stock|    EUR|
2504197|2018-01-02|15:00|    5.84|    5.84|    5.84|    5.84|    203|
1|
|DE0005933931|    EXS1|ISHS CORE DAX UCI...|    ETF|    EUR|
2504265|2018-01-02|15:00|    111.44|    111.44|    111.42|    111.42|    527|
2|
|LU0274211480|    DBXD|DB X-TRACK.DAX ET...|    ETF|    EUR|
2504269|2018-01-02|15:00|    125.26|    125.32|    125.26|    125.32|    2465|
2|
|DE000A0B9N37|    A8A|    JDC GROUP AG O.N.|Common stock|    EUR|
2504284|2018-01-02|15:00|    9|    9|    9|    9|    1050|
1|
|DE000AOCAYB2|    WIN|DIEBOLD NIXDORF ...|Common stock|    EUR|
2504286|2018-01-02|15:00|    73.7|    73.7|    73.7|    73.7|    63|
1|
|DE000AOD6554|    NDX1|    NORDEX SE O.N.|Common stock|    EUR|
2504290|2018-01-02|15:00|    9.206|    9.22|    9.2|    9.2|    1426|
5|
|DE000AOD8QZ7|    EXSE|ISH.S.EUR.SMALL 2...|    ETF|    EUR|
2504291|2018-01-02|15:00|    28.63|    28.63|    28.63|    28.63|    750|
1|
|DE000AOD8Q07|    EXSI|ISHARES EURO STOXX...|    ETF|    EUR|
2504292|2018-01-02|15:00|    37.96|    37.96|    37.96|    37.96|    784|
1|
|DE000A0HGQF5|    MF6|    MAGFORCE AG|Common stock|    EUR|
2504308|2018-01-02|15:00|    6.5|    6.51|    6.5|    6.51|    650|
2|
|DE000A0HHJR3|    CLIQ|CLIQ DIGITAL AG ...|Common stock|    EUR|
2504310|2018-01-02|15:00|    8.46|    8.46|    8.46|    8.46|    9|
1|
|DE000A0HN5C6|    DWNI|DEUTSCHE WOHNEN S...|Common stock|    EUR|
2504314|2018-01-02|15:00|    35.72|    35.72|    35.72|    35.72|    258|
2|
|DE000A0H08M3|    EXH1|IS.S.E.600 OIL+G...|    ETF|    EUR|
2504330|2018-01-02|15:00|    31.575|    31.58|    31.57|    31.57|    3383|
6|
|DE000A0JKHC9|    N7G|NANOGATE SE INH...|Common stock|    EUR|
2504339|2018-01-02|15:00|    46.65|    46.65|    46.65|    46.65|    8|

```

```

1|
|DE000A0JK2A8|      AR4|AURELIUS EQ.OPP. ...|Common stock|      EUR|
2504340|2018-01-02|15:00|      57.45|      57.45|      57.45|      57.45|      289|
2|
|DE000A0JL9W6|      VBK|VERBIO VER.BIOENE...|Common stock|      EUR|
2504343|2018-01-02|15:00|      8.175|      8.175|      8.175|      8.175|      276|
1|
|DE000A0JQ5U3|      M5Z|      MANZ AG|Common stock|      EUR|
2504345|2018-01-02|15:00|      34.15|      34.2|      34.15|      34.2|      305|
3|
|DE000A0KEXC7|      V3S|VECTRON SYSTEMS ...|Common stock|      EUR|
2504348|2018-01-02|15:00|      21|      21|      21|      21|      200|
1|
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

```

```
[18]: data.createOrReplaceTempView("data")
```

```
[25]: spark.sql("SELECT ISIN, Date, StartPrice, EndPrice, TradedVolume FROM data_
↪LIMIT 10").show()
```

```

+-----+-----+-----+-----+-----+-----+
|      ISIN|      Date|StartPrice|EndPrice|TradedVolume|
+-----+-----+-----+-----+-----+
|AT0000A0E9W5|2018-01-02|      19.19|      19.19|      100|
|AT0000743059|2018-01-02|      53.28|      53.28|      18|
|AT0000969985|2018-01-02|      24.15|      24.15|      401|
|CA32076V1031|2018-01-02|       5.84|       5.84|      203|
|DE0005933931|2018-01-02|     111.44|     111.42|      527|
|LU0274211480|2018-01-02|     125.26|     125.32|     2465|
|DE000A0B9N37|2018-01-02|       9|       9|     1050|
|DE000A0CAYB2|2018-01-02|      73.7|      73.7|       63|
|DE000A0D6554|2018-01-02|      9.206|       9.2|     1426|
|DE000A0D8QZ7|2018-01-02|      28.63|      28.63|      750|
+-----+-----+-----+-----+-----+-----+

```

```
[22]: data.printSchema()
```

```

root
|-- ISIN: string (nullable = true)
|-- Mnemonic: string (nullable = true)
|-- SecurityDesc: string (nullable = true)
|-- SecurityType: string (nullable = true)
|-- Currency: string (nullable = true)
|-- SecurityID: string (nullable = true)

```

```

|-- Date: string (nullable = true)
|-- Time: string (nullable = true)
|-- StartPrice: string (nullable = true)
|-- MaxPrice: string (nullable = true)
|-- MinPrice: string (nullable = true)
|-- EndPrice: string (nullable = true)
|-- TradedVolume: string (nullable = true)
|-- NumberOfTrades: string (nullable = true)

```

```
[28]: spark.sql("SELECT count(ISIN), count(distinct(ISIN)) FROM data").show()
```

```

+-----+-----+
|count(ISIN)|count(DISTINCT ISIN)|
+-----+-----+
|      68154|              1568|
+-----+-----+

```

```
[31]: spark.sql("SELECT ISIN, Date, Time, StartPrice, EndPrice, TradedVolume FROM_
→data WHERE ISIN = 'DE000A0H08M3']").show()
```

```

+-----+-----+-----+-----+-----+-----+
|      ISIN|      Date| Time|StartPrice|EndPrice|TradedVolume|
+-----+-----+-----+-----+-----+-----+
|DE000A0H08M3|2018-01-02|15:00|    31.575|    31.57|      3383|
|DE000A0H08M3|2018-01-02|15:01|    31.585|    31.585|      1091|
|DE000A0H08M3|2018-01-02|15:02|    31.595|    31.595|      2783|
|DE000A0H08M3|2018-01-02|15:03|    31.59|    31.59|       300|
|DE000A0H08M3|2018-01-02|15:04|    31.59|    31.605|      3623|
|DE000A0H08M3|2018-01-02|15:06|    31.6|    31.6|      2183|
|DE000A0H08M3|2018-01-02|15:08|    31.595|    31.595|      2182|
|DE000A0H08M3|2018-01-02|15:09|    31.61|    31.615|      2183|
|DE000A0H08M3|2018-01-02|15:11|    31.595|    31.59|      2940|
|DE000A0H08M3|2018-01-02|15:12|    31.605|    31.6|      1200|
|DE000A0H08M3|2018-01-02|15:13|    31.605|    31.615|      4147|
|DE000A0H08M3|2018-01-02|15:14|    31.615|    31.615|       186|
|DE000A0H08M3|2018-01-02|15:16|    31.61|    31.605|      3766|
|DE000A0H08M3|2018-01-02|15:18|    31.64|    31.635|      2210|
|DE000A0H08M3|2018-01-02|15:19|    31.64|    31.64|       747|
|DE000A0H08M3|2018-01-02|15:20|    31.65|    31.645|      2498|
|DE000A0H08M3|2018-01-02|15:21|    31.66|    31.66|       764|
|DE000A0H08M3|2018-01-02|15:23|    31.655|    31.67|      2510|
|DE000A0H08M3|2018-01-02|15:24|    31.64|    31.64|      2182|
|DE000A0H08M3|2018-01-02|15:26|    31.63|    31.625|      2815|
+-----+-----+-----+-----+-----+-----+

```

only showing top 20 rows

```
[3]: ! aws s3 ls s3://deutsche-boerse-xetra-pds/2018 --summarize --human-readable
      ↪--recursive | grep "Total Size"
```

Total Size: 1.9 GiB

```
[4]: ! aws s3 ls s3://deutsche-boerse-xetra-pds/2019 --summarize --human-readable
      ↪--recursive | grep "Total Size"
```

Total Size: 1.5 GiB

```
[5]: ! aws s3 ls s3://deutsche-boerse-xetra-pds/2018-01-03/
```

```
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO0.csv
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO1.csv
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO2.csv
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO3.csv
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO4.csv
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO5.csv
2018-04-04 20:03:46      136 2018-01-03_BINS_XETRO6.csv
2018-04-04 20:03:47      136 2018-01-03_BINS_XETRO7.csv
2018-04-04 20:03:46    961262 2018-01-03_BINS_XETRO8.csv
2018-04-04 20:03:47    987516 2018-01-03_BINS_XETRO9.csv
2018-04-04 20:03:48    917254 2018-01-03_BINS_XETR10.csv
2018-04-04 20:03:48    865492 2018-01-03_BINS_XETR11.csv
2018-04-04 20:03:48    786016 2018-01-03_BINS_XETR12.csv
2018-04-04 20:03:48    857821 2018-01-03_BINS_XETR13.csv
2018-04-04 20:03:49   1045421 2018-01-03_BINS_XETR14.csv
2018-04-04 20:03:49   1193576 2018-01-03_BINS_XETR15.csv
2018-04-04 20:03:49    758825 2018-01-03_BINS_XETR16.csv
2018-04-04 20:03:50      136 2018-01-03_BINS_XETR17.csv
2018-04-04 20:03:50      136 2018-01-03_BINS_XETR18.csv
2018-04-04 20:03:50      136 2018-01-03_BINS_XETR19.csv
2018-04-04 20:03:50    1650 2018-01-03_BINS_XETR20.csv
2018-04-04 20:03:50      136 2018-01-03_BINS_XETR21.csv
2018-04-04 20:03:50      136 2018-01-03_BINS_XETR22.csv
2018-04-04 20:03:50      136 2018-01-03_BINS_XETR23.csv
```

```
[6]: ! aws s3 cp s3://deutsche-boerse-xetra-pds/2018-01-03/ ./data/2018-01-03/
      ↪--recursive
```

```
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETRO0.csv
to data/2018-01-03/2018-01-03_BINS_XETRO0.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETRO2.csv
to data/2018-01-03/2018-01-03_BINS_XETRO2.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETRO1.csv
to data/2018-01-03/2018-01-03_BINS_XETRO1.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETRO6.csv
to data/2018-01-03/2018-01-03_BINS_XETRO6.csv
```

```

download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR07.csv
to data/2018-01-03/2018-01-03_BINS_XETR07.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR05.csv
to data/2018-01-03/2018-01-03_BINS_XETR05.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR04.csv
to data/2018-01-03/2018-01-03_BINS_XETR04.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR18.csv
to data/2018-01-03/2018-01-03_BINS_XETR18.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR03.csv
to data/2018-01-03/2018-01-03_BINS_XETR03.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR17.csv
to data/2018-01-03/2018-01-03_BINS_XETR17.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR20.csv
to data/2018-01-03/2018-01-03_BINS_XETR20.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR21.csv
to data/2018-01-03/2018-01-03_BINS_XETR21.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR22.csv
to data/2018-01-03/2018-01-03_BINS_XETR22.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR23.csv
to data/2018-01-03/2018-01-03_BINS_XETR23.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR19.csv
to data/2018-01-03/2018-01-03_BINS_XETR19.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR08.csv
to data/2018-01-03/2018-01-03_BINS_XETR08.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR16.csv
to data/2018-01-03/2018-01-03_BINS_XETR16.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR13.csv
to data/2018-01-03/2018-01-03_BINS_XETR13.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR12.csv
to data/2018-01-03/2018-01-03_BINS_XETR12.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR11.csv
to data/2018-01-03/2018-01-03_BINS_XETR11.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR10.csv
to data/2018-01-03/2018-01-03_BINS_XETR10.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR09.csv
to data/2018-01-03/2018-01-03_BINS_XETR09.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR14.csv
to data/2018-01-03/2018-01-03_BINS_XETR14.csv
download: s3://deutsche-boerse-xetra-pds/2018-01-03/2018-01-03_BINS_XETR15.csv
to data/2018-01-03/2018-01-03_BINS_XETR15.csv

```

[26]: *# TODO can be omitted, because Spark infers schema automatically*

```

# saving and loading schemas
# source: https://stackoverflow.com/questions/56707401/
→ save-schema-of-dataframe-in-s3-location

```



```

# funcs
from pyspark.sql.functions import *
from pyspark.sql.types import *

schema = data.schema.json()

import json

with open('schema/schema.json', 'w+') as F:
    json.dump(schema, F)

with open('schema/schema.json', 'r') as F:
    saved_schema = json.load(F)

saved_schema

new_schema = StructType.fromJson(json.loads(saved_schema))

new_schema

```

[26]: StructType(List(StructField(ISIN,StringType,true),StructField(Mnemonic,StringType,true),StructField(SecurityDesc,StringType,true),StructField(SecurityType,StringType,true),StructField(Currency,StringType,true),StructField(SecurityID,StringType,true),StructField(Date,StringType,true),StructField(Time,StringType,true),StructField(StartPrice,StringType,true),StructField(MaxPrice,StringType,true),StructField(MinPrice,StringType,true),StructField(EndPrice,StringType,true),StructField(TradedVolume,StringType,true),StructField(NumberOfTrades,StringType,true)))

```

[41]: full_data = data = spark.read.format("csv").option("header", "true").
      ↪option("inferSchema", "true").load("./data/*")
      full_data.printSchema()

```

```

root
|-- ISIN: string (nullable = true)
|-- Mnemonic: string (nullable = true)
|-- SecurityDesc: string (nullable = true)
|-- SecurityType: string (nullable = true)
|-- Currency: string (nullable = true)
|-- SecurityID: integer (nullable = true)
|-- Date: timestamp (nullable = true)
|-- Time: string (nullable = true)
|-- StartPrice: double (nullable = true)
|-- MaxPrice: double (nullable = true)
|-- MinPrice: double (nullable = true)
|-- EndPrice: double (nullable = true)
|-- TradedVolume: integer (nullable = true)
|-- NumberOfTrades: integer (nullable = true)

```

```
[42]: data.createOrReplaceTempView("full_data")
```

```
[52]: spark.sql("SELECT ISIN, Date, COUNT(1) FROM full_data WHERE ISIN =  
        ↳ 'DE000A0H08M3' GROUP BY 1,2 ORDER BY 2").show()
```

ISIN	Date	count(1)
DE000A0H08M3	2018-01-02 00:00:00	167
DE000A0H08M3	2018-01-03 00:00:00	87

```
[71]: spark.sql("SELECT ISIN, Date, Time, StartPrice, EndPrice FROM full_data WHERE  
        ↳ ISIN = 'DE000A0H08M3' ORDER BY Date, Time").show(1000)
```

ISIN	Date	Time	StartPrice	EndPrice
DE000A0H08M3	2018-01-02 00:00:00	08:09	31.61	31.61
DE000A0H08M3	2018-01-02 00:00:00	08:12	31.66	31.665
DE000A0H08M3	2018-01-02 00:00:00	08:19	31.705	31.705
DE000A0H08M3	2018-01-02 00:00:00	08:41	31.6	31.6
DE000A0H08M3	2018-01-02 00:00:00	08:52	31.575	31.555
DE000A0H08M3	2018-01-02 00:00:00	08:54	31.56	31.56
DE000A0H08M3	2018-01-02 00:00:00	09:06	31.605	31.605
DE000A0H08M3	2018-01-02 00:00:00	09:07	31.62	31.62
DE000A0H08M3	2018-01-02 00:00:00	09:10	31.635	31.635
DE000A0H08M3	2018-01-02 00:00:00	09:17	31.64	31.64
DE000A0H08M3	2018-01-02 00:00:00	09:39	31.63	31.63
DE000A0H08M3	2018-01-02 00:00:00	09:53	31.595	31.595
DE000A0H08M3	2018-01-02 00:00:00	09:57	31.585	31.58
DE000A0H08M3	2018-01-02 00:00:00	10:36	31.55	31.55
DE000A0H08M3	2018-01-02 00:00:00	10:52	31.55	31.55
DE000A0H08M3	2018-01-02 00:00:00	11:00	31.55	31.55
DE000A0H08M3	2018-01-02 00:00:00	11:04	31.565	31.565
DE000A0H08M3	2018-01-02 00:00:00	11:08	31.565	31.565
DE000A0H08M3	2018-01-02 00:00:00	11:13	31.565	31.56
DE000A0H08M3	2018-01-02 00:00:00	11:23	31.555	31.555
DE000A0H08M3	2018-01-02 00:00:00	11:30	31.56	31.56
DE000A0H08M3	2018-01-02 00:00:00	11:49	31.575	31.575
DE000A0H08M3	2018-01-02 00:00:00	11:54	31.58	31.58
DE000A0H08M3	2018-01-02 00:00:00	12:01	31.6	31.6
DE000A0H08M3	2018-01-02 00:00:00	12:15	31.62	31.62
DE000A0H08M3	2018-01-02 00:00:00	12:49	31.6	31.6
DE000A0H08M3	2018-01-02 00:00:00	13:11	31.59	31.59

DE000A0H08M3 2018-01-02 00:00:00 13:13	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 13:15	31.58	31.58
DE000A0H08M3 2018-01-02 00:00:00 13:16	31.58	31.58
DE000A0H08M3 2018-01-02 00:00:00 13:18	31.59	31.585
DE000A0H08M3 2018-01-02 00:00:00 13:19	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 13:20	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 13:21	31.585	31.585
DE000A0H08M3 2018-01-02 00:00:00 13:22	31.58	31.58
DE000A0H08M3 2018-01-02 00:00:00 13:24	31.58	31.58
DE000A0H08M3 2018-01-02 00:00:00 13:26	31.58	31.58
DE000A0H08M3 2018-01-02 00:00:00 13:27	31.58	31.58
DE000A0H08M3 2018-01-02 00:00:00 13:28	31.595	31.595
DE000A0H08M3 2018-01-02 00:00:00 13:31	31.6	31.6
DE000A0H08M3 2018-01-02 00:00:00 13:32	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 13:34	31.59	31.585
DE000A0H08M3 2018-01-02 00:00:00 13:35	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 13:36	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 13:37	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 13:38	31.565	31.565
DE000A0H08M3 2018-01-02 00:00:00 13:40	31.575	31.575
DE000A0H08M3 2018-01-02 00:00:00 13:42	31.565	31.565
DE000A0H08M3 2018-01-02 00:00:00 13:43	31.565	31.565
DE000A0H08M3 2018-01-02 00:00:00 13:44	31.56	31.56
DE000A0H08M3 2018-01-02 00:00:00 13:46	31.555	31.55
DE000A0H08M3 2018-01-02 00:00:00 13:47	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 13:48	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 13:50	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 13:51	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 13:52	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 13:53	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 13:54	31.56	31.56
DE000A0H08M3 2018-01-02 00:00:00 13:56	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 13:57	31.56	31.56
DE000A0H08M3 2018-01-02 00:00:00 13:59	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 14:00	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 14:02	31.56	31.56
DE000A0H08M3 2018-01-02 00:00:00 14:04	31.56	31.56
DE000A0H08M3 2018-01-02 00:00:00 14:05	31.565	31.58
DE000A0H08M3 2018-01-02 00:00:00 14:08	31.58	31.595
DE000A0H08M3 2018-01-02 00:00:00 14:09	31.6	31.6
DE000A0H08M3 2018-01-02 00:00:00 14:11	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:12	31.595	31.595
DE000A0H08M3 2018-01-02 00:00:00 14:13	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:14	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:15	31.605	31.605
DE000A0H08M3 2018-01-02 00:00:00 14:16	31.605	31.605
DE000A0H08M3 2018-01-02 00:00:00 14:17	31.6	31.6
DE000A0H08M3 2018-01-02 00:00:00 14:18	31.595	31.595

DE000A0H08M3 2018-01-02 00:00:00 14:20	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:21	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:23	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:24	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 14:26	31.6	31.605
DE000A0H08M3 2018-01-02 00:00:00 14:29	31.595	31.595
DE000A0H08M3 2018-01-02 00:00:00 14:30	31.59	31.57
DE000A0H08M3 2018-01-02 00:00:00 14:32	31.565	31.58
DE000A0H08M3 2018-01-02 00:00:00 14:33	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 14:36	31.585	31.585
DE000A0H08M3 2018-01-02 00:00:00 14:37	31.57	31.545
DE000A0H08M3 2018-01-02 00:00:00 14:38	31.535	31.535
DE000A0H08M3 2018-01-02 00:00:00 14:40	31.565	31.535
DE000A0H08M3 2018-01-02 00:00:00 14:41	31.535	31.535
DE000A0H08M3 2018-01-02 00:00:00 14:42	31.54	31.54
DE000A0H08M3 2018-01-02 00:00:00 14:43	31.535	31.535
DE000A0H08M3 2018-01-02 00:00:00 14:45	31.505	31.505
DE000A0H08M3 2018-01-02 00:00:00 14:47	31.5	31.5
DE000A0H08M3 2018-01-02 00:00:00 14:48	31.52	31.52
DE000A0H08M3 2018-01-02 00:00:00 14:50	31.525	31.525
DE000A0H08M3 2018-01-02 00:00:00 14:51	31.53	31.53
DE000A0H08M3 2018-01-02 00:00:00 14:53	31.545	31.545
DE000A0H08M3 2018-01-02 00:00:00 14:54	31.56	31.56
DE000A0H08M3 2018-01-02 00:00:00 14:55	31.56	31.555
DE000A0H08M3 2018-01-02 00:00:00 14:56	31.555	31.555
DE000A0H08M3 2018-01-02 00:00:00 14:57	31.57	31.57
DE000A0H08M3 2018-01-02 00:00:00 14:58	31.575	31.57
DE000A0H08M3 2018-01-02 00:00:00 14:59	31.565	31.565
DE000A0H08M3 2018-01-02 00:00:00 15:00	31.575	31.57
DE000A0H08M3 2018-01-02 00:00:00 15:01	31.585	31.585
DE000A0H08M3 2018-01-02 00:00:00 15:02	31.595	31.595
DE000A0H08M3 2018-01-02 00:00:00 15:03	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 15:04	31.59	31.605
DE000A0H08M3 2018-01-02 00:00:00 15:06	31.6	31.6
DE000A0H08M3 2018-01-02 00:00:00 15:08	31.595	31.595
DE000A0H08M3 2018-01-02 00:00:00 15:09	31.61	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:11	31.595	31.59
DE000A0H08M3 2018-01-02 00:00:00 15:12	31.605	31.6
DE000A0H08M3 2018-01-02 00:00:00 15:13	31.605	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:14	31.615	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:16	31.61	31.605
DE000A0H08M3 2018-01-02 00:00:00 15:18	31.64	31.635
DE000A0H08M3 2018-01-02 00:00:00 15:19	31.64	31.64
DE000A0H08M3 2018-01-02 00:00:00 15:20	31.65	31.645
DE000A0H08M3 2018-01-02 00:00:00 15:21	31.66	31.66
DE000A0H08M3 2018-01-02 00:00:00 15:23	31.655	31.67
DE000A0H08M3 2018-01-02 00:00:00 15:24	31.64	31.64
DE000A0H08M3 2018-01-02 00:00:00 15:26	31.63	31.625

DE000A0H08M3 2018-01-02 00:00:00 15:28	31.62	31.635
DE000A0H08M3 2018-01-02 00:00:00 15:29	31.64	31.64
DE000A0H08M3 2018-01-02 00:00:00 15:30	31.645	31.645
DE000A0H08M3 2018-01-02 00:00:00 15:31	31.65	31.65
DE000A0H08M3 2018-01-02 00:00:00 15:32	31.64	31.635
DE000A0H08M3 2018-01-02 00:00:00 15:35	31.63	31.63
DE000A0H08M3 2018-01-02 00:00:00 15:36	31.65	31.655
DE000A0H08M3 2018-01-02 00:00:00 15:38	31.64	31.64
DE000A0H08M3 2018-01-02 00:00:00 15:39	31.63	31.63
DE000A0H08M3 2018-01-02 00:00:00 15:41	31.625	31.625
DE000A0H08M3 2018-01-02 00:00:00 15:42	31.615	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:43	31.61	31.61
DE000A0H08M3 2018-01-02 00:00:00 15:45	31.625	31.625
DE000A0H08M3 2018-01-02 00:00:00 15:46	31.62	31.62
DE000A0H08M3 2018-01-02 00:00:00 15:47	31.615	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:48	31.615	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:49	31.625	31.625
DE000A0H08M3 2018-01-02 00:00:00 15:50	31.63	31.635
DE000A0H08M3 2018-01-02 00:00:00 15:52	31.62	31.615
DE000A0H08M3 2018-01-02 00:00:00 15:54	31.635	31.65
DE000A0H08M3 2018-01-02 00:00:00 15:55	31.655	31.66
DE000A0H08M3 2018-01-02 00:00:00 15:56	31.665	31.665
DE000A0H08M3 2018-01-02 00:00:00 15:59	31.645	31.645
DE000A0H08M3 2018-01-02 00:00:00 16:00	31.63	31.625
DE000A0H08M3 2018-01-02 00:00:00 16:02	31.64	31.635
DE000A0H08M3 2018-01-02 00:00:00 16:04	31.63	31.63
DE000A0H08M3 2018-01-02 00:00:00 16:05	31.615	31.615
DE000A0H08M3 2018-01-02 00:00:00 16:06	31.62	31.62
DE000A0H08M3 2018-01-02 00:00:00 16:07	31.61	31.605
DE000A0H08M3 2018-01-02 00:00:00 16:08	31.61	31.61
DE000A0H08M3 2018-01-02 00:00:00 16:09	31.62	31.615
DE000A0H08M3 2018-01-02 00:00:00 16:10	31.63	31.635
DE000A0H08M3 2018-01-02 00:00:00 16:12	31.63	31.63
DE000A0H08M3 2018-01-02 00:00:00 16:13	31.645	31.655
DE000A0H08M3 2018-01-02 00:00:00 16:14	31.66	31.665
DE000A0H08M3 2018-01-02 00:00:00 16:16	31.66	31.67
DE000A0H08M3 2018-01-02 00:00:00 16:19	31.65	31.65
DE000A0H08M3 2018-01-02 00:00:00 16:20	31.615	31.615
DE000A0H08M3 2018-01-02 00:00:00 16:22	31.61	31.59
DE000A0H08M3 2018-01-02 00:00:00 16:23	31.585	31.585
DE000A0H08M3 2018-01-02 00:00:00 16:24	31.585	31.585
DE000A0H08M3 2018-01-02 00:00:00 16:26	31.59	31.59
DE000A0H08M3 2018-01-02 00:00:00 16:28	31.605	31.605
DE000A0H08M3 2018-01-02 00:00:00 16:29	31.6	31.6
DE000A0H08M3 2018-01-03 00:00:00 08:35	31.855	31.855
DE000A0H08M3 2018-01-03 00:00:00 08:40	31.815	31.815
DE000A0H08M3 2018-01-03 00:00:00 08:47	31.83	31.83
DE000A0H08M3 2018-01-03 00:00:00 08:48	31.83	31.83

DE000A0H08M3 2018-01-03 00:00:00 09:00	31.85	31.85
DE000A0H08M3 2018-01-03 00:00:00 09:03	31.855	31.855
DE000A0H08M3 2018-01-03 00:00:00 09:11	31.88	31.88
DE000A0H08M3 2018-01-03 00:00:00 09:12	31.88	31.88
DE000A0H08M3 2018-01-03 00:00:00 09:20	31.905	31.905
DE000A0H08M3 2018-01-03 00:00:00 09:24	31.89	31.89
DE000A0H08M3 2018-01-03 00:00:00 09:32	31.875	31.875
DE000A0H08M3 2018-01-03 00:00:00 09:34	31.89	31.89
DE000A0H08M3 2018-01-03 00:00:00 09:35	31.885	31.885
DE000A0H08M3 2018-01-03 00:00:00 09:36	31.88	31.875
DE000A0H08M3 2018-01-03 00:00:00 09:40	31.905	31.905
DE000A0H08M3 2018-01-03 00:00:00 09:41	31.91	31.91
DE000A0H08M3 2018-01-03 00:00:00 09:44	31.93	31.93
DE000A0H08M3 2018-01-03 00:00:00 09:47	31.935	31.935
DE000A0H08M3 2018-01-03 00:00:00 09:50	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 09:54	31.945	31.945
DE000A0H08M3 2018-01-03 00:00:00 09:56	31.945	31.945
DE000A0H08M3 2018-01-03 00:00:00 09:57	31.935	31.935
DE000A0H08M3 2018-01-03 00:00:00 10:00	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 10:12	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 10:19	31.97	31.97
DE000A0H08M3 2018-01-03 00:00:00 10:21	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 10:28	31.98	31.98
DE000A0H08M3 2018-01-03 00:00:00 10:29	31.965	31.965
DE000A0H08M3 2018-01-03 00:00:00 10:32	31.975	31.975
DE000A0H08M3 2018-01-03 00:00:00 10:34	31.97	31.97
DE000A0H08M3 2018-01-03 00:00:00 10:40	31.98	31.98
DE000A0H08M3 2018-01-03 00:00:00 10:44	31.965	31.965
DE000A0H08M3 2018-01-03 00:00:00 10:47	31.975	31.975
DE000A0H08M3 2018-01-03 00:00:00 10:49	31.985	31.985
DE000A0H08M3 2018-01-03 00:00:00 10:50	31.98	31.98
DE000A0H08M3 2018-01-03 00:00:00 10:55	31.985	31.985
DE000A0H08M3 2018-01-03 00:00:00 10:57	31.97	31.97
DE000A0H08M3 2018-01-03 00:00:00 11:06	31.965	31.965
DE000A0H08M3 2018-01-03 00:00:00 11:16	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 11:17	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 11:30	31.965	31.965
DE000A0H08M3 2018-01-03 00:00:00 12:15	31.935	31.935
DE000A0H08M3 2018-01-03 00:00:00 12:26	31.985	31.99
DE000A0H08M3 2018-01-03 00:00:00 12:27	31.99	31.99
DE000A0H08M3 2018-01-03 00:00:00 13:12	31.98	31.99
DE000A0H08M3 2018-01-03 00:00:00 13:14	31.98	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:15	31.975	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:17	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 13:18	31.95	31.945
DE000A0H08M3 2018-01-03 00:00:00 13:19	31.945	31.945
DE000A0H08M3 2018-01-03 00:00:00 13:35	31.945	31.945
DE000A0H08M3 2018-01-03 00:00:00 13:36	31.945	31.945

DE000A0H08M3 2018-01-03 00:00:00 13:38	31.945	31.945
DE000A0H08M3 2018-01-03 00:00:00 13:39	31.945	31.945
DE000A0H08M3 2018-01-03 00:00:00 13:40	31.94	31.94
DE000A0H08M3 2018-01-03 00:00:00 13:42	31.935	31.935
DE000A0H08M3 2018-01-03 00:00:00 13:43	31.93	31.93
DE000A0H08M3 2018-01-03 00:00:00 13:44	31.95	31.95
DE000A0H08M3 2018-01-03 00:00:00 13:45	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 13:49	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:50	31.95	31.95
DE000A0H08M3 2018-01-03 00:00:00 13:51	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:52	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:53	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:55	31.955	31.955
DE000A0H08M3 2018-01-03 00:00:00 13:56	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:57	31.96	31.96
DE000A0H08M3 2018-01-03 00:00:00 13:59	31.95	31.95
DE000A0H08M3 2018-01-03 00:00:00 14:20	31.97	31.975
DE000A0H08M3 2018-01-03 00:00:00 14:23	31.965	31.965
DE000A0H08M3 2018-01-03 00:00:00 14:25	31.965	31.965
DE000A0H08M3 2018-01-03 00:00:00 14:34	31.97	31.97
DE000A0H08M3 2018-01-03 00:00:00 14:36	31.975	31.975
DE000A0H08M3 2018-01-03 00:00:00 14:37	31.985	31.985
DE000A0H08M3 2018-01-03 00:00:00 14:44	31.995	31.995
DE000A0H08M3 2018-01-03 00:00:00 14:50	32.0	32.0
DE000A0H08M3 2018-01-03 00:00:00 14:51	31.995	31.995
DE000A0H08M3 2018-01-03 00:00:00 14:52	31.995	31.995
DE000A0H08M3 2018-01-03 00:00:00 14:54	32.0	32.0
DE000A0H08M3 2018-01-03 00:00:00 14:56	32.02	32.02
DE000A0H08M3 2018-01-03 00:00:00 15:04	32.02	32.02
DE000A0H08M3 2018-01-03 00:00:00 15:25	32.025	32.03
DE000A0H08M3 2018-01-03 00:00:00 15:33	32.05	32.05
DE000A0H08M3 2018-01-03 00:00:00 15:34	32.05	32.05
DE000A0H08M3 2018-01-03 00:00:00 15:42	32.09	32.09
DE000A0H08M3 2018-01-03 00:00:00 16:20	32.06	32.06
DE000A0H08M3 2018-01-03 00:00:00 16:28	32.075	32.075

+-----+-----+-----+-----+

```
[73]: spark.sql("SELECT first(ISIN), Date, first(StartPrice), last(EndPrice),  

  ↳sum(TradedVolume) FROM full_data WHERE ISIN = 'DE000A0H08M3' GROUP BY Date").  

  ↳show()
```

+-----+-----+-----+-----+	
-----+-----+	
first(ISIN, false)	Date first(StartPrice, false) last(EndPrice,
false) sum(TradedVolume)	
+-----+-----+-----+-----+	
-----+-----+	

31.6	6641			
DE000A0H08M3	2018-01-02 00:00:00	08:52	31.575	31.61 31.555
31.555	53702			
DE000A0H08M3	2018-01-02 00:00:00	08:54	31.56	31.61 31.56
31.56	25000			

```

+-----+-----+-----+-----+-----+-----+-----+-----+
-----+

```

```

[107]: spark.sql("""
SELECT
    ISIN,
    Date,
    Time,
    StartPrice,
    First_value(StartPrice) OVER ( partition BY Date ORDER BY Time) AS
↪FirstStartPrice,
    EndPrice,
    Last_value(EndPrice) OVER ( partition BY Date ORDER BY Time) AS
↪LastEndPrice,
    TradedVolume,
    ROW_NUMBER() OVER ( partition BY DATE ORDER BY Time DESC ) AS RK
FROM full_data
WHERE ISIN = 'DE000A0H08M3'
-- AND Date = '2018-01-03'
AND TIME BETWEEN '08:00' AND '09:00'""").createOrReplaceTempView("tmp")

```

```

[123]: spark.sql("""
SELECT
    ISIN,
    Date,
    FirstStartPrice AS OpeningPrice,
    last(LastEndPrice) AS ClosingPrice,
    sum(TradedVolume) AS DailyTradedVolume,
    format_number(last(LastEndPrice) / Lag(last(LastEndPrice)) OVER (partition
↪BY ISIN ORDER BY Date) *100 -100,2) AS PctChgPrvCls
FROM tmp
GROUP BY 1, 2, 3""").show()

```

	ISIN			
Date	OpeningPrice	ClosingPrice	DailyTradedVolume	PctChgPrvCls

```

+-----+-----+-----+-----+-----+-----+-----+-----+
-----+

```

DE000A0H08M3	2018-01-02 00:00:00	31.61	31.56	89019
null				

|DE000A0H08M3|2018-01-03 00:00:00| 31.855| 31.85| 3733|
0.92|

+-----+-----+-----+-----+-----+-----+
-----+

[]: