

卓识基金数据测评项目 成果报告

令狐冠杰

https://github.com/lah-dee-dah/zsquant_assessment.git



大纲

- 数据检查
- 分钟bar合成

数据检查

- 统计LocalTime的单调性
- 分交易所统计 (UpdateTime, UpdateMillisec) 单调性
- 探索每个交易所的tick的推送频率
- 探索ActionDay, TradingDay和实际交易时间的关系
- 对现有数据进行数据检查

统计LocalTime的单调性

- 结果：LocalTime字段的19位时间戳严格单调递增

分交易所统计UpdateTime, UpdateMullisec单调性

- 方法：
 - 将UpdateTime, UpdateMillisec合并为新字段UpdateTime_full
 - 将UpdateTime_full 分为0点之前和0点之后两组分别检查，避免跨日导致的误判
 - 在以上基础上根据交易所分组进行检查
 - 在以上基础上根据交易所和合约名称分组进行检查

分交易所统计UpdateTime, UpdateMullisec单调性

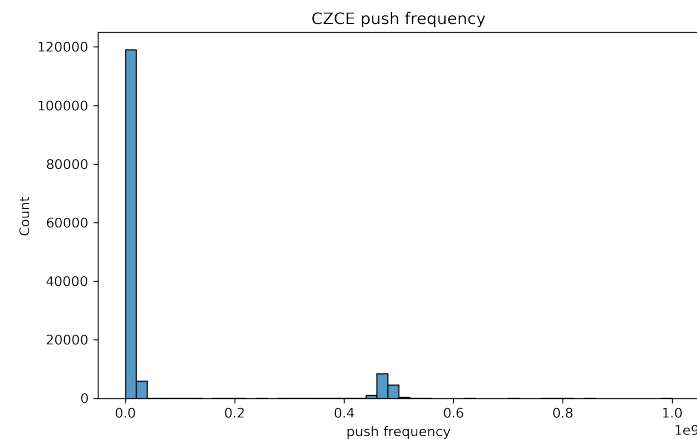
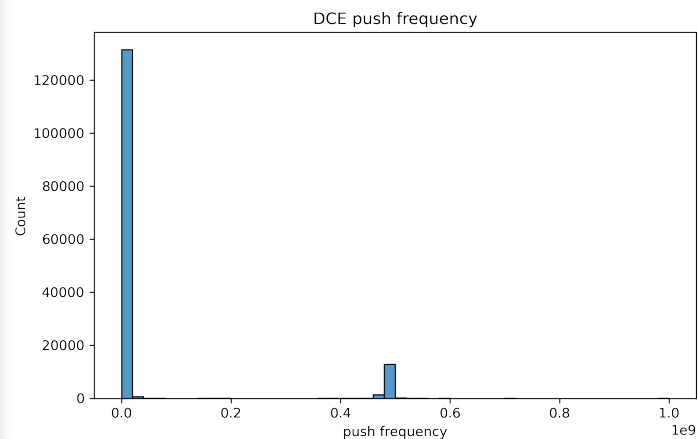
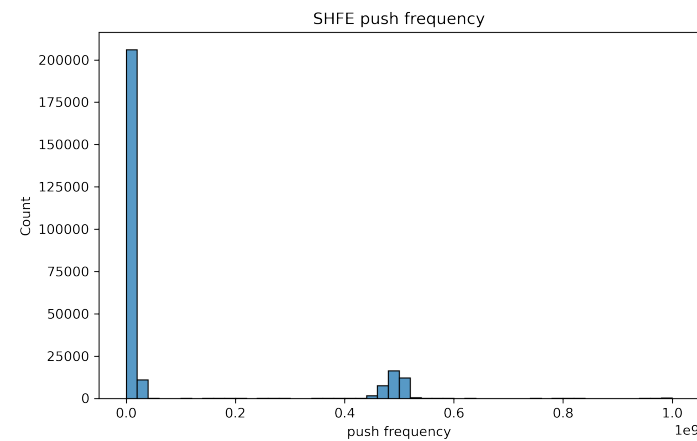
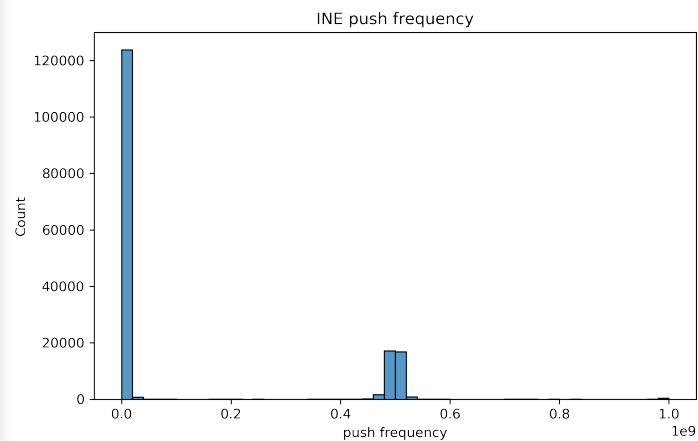
- 结果：
 - 仅对交易所分组：INE和SHFE的UpdateTime, UpdateMillisec具有单调性，CZCE和DCE的UpdateTime, UpdateMillisec不具有单调性
 - 对交易所，合约名称分组：所有合约对应的UpdateTime, UpdateMillisec都具有单调性

每个交易所的tick的推送频率

- 方法：
 - 根据交易所分组后，通过求LocalTime字段的一阶差分获得推送间隔，用每组推送间隔画直方图观察推送频率的规律
 - 在交易所分组的基础上再对合约进行分组，求得推送间隔后再根据交易所汇总并画直方图
 - 统计UpdateMillisec的分布辅助判断

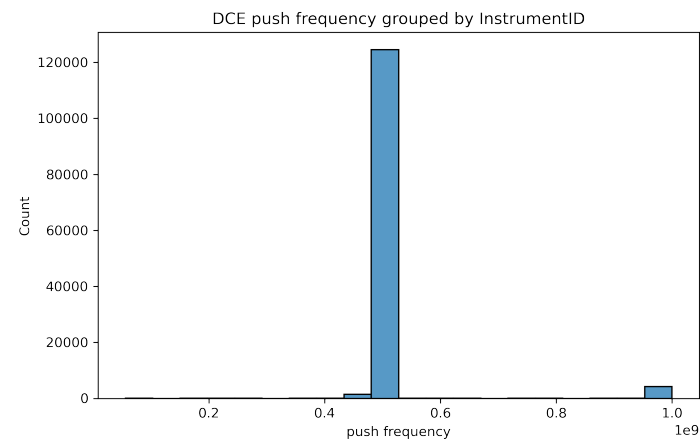
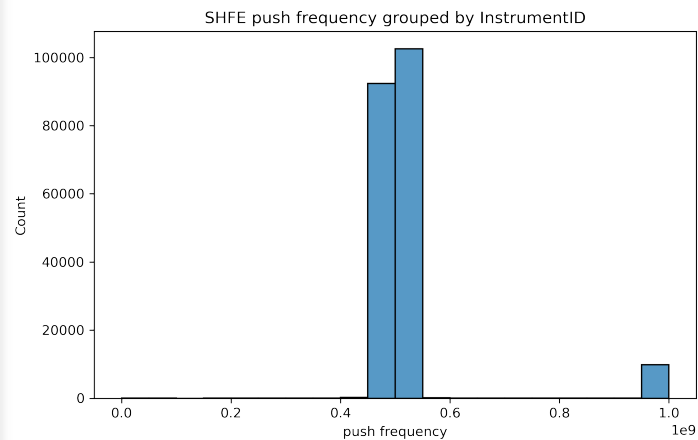
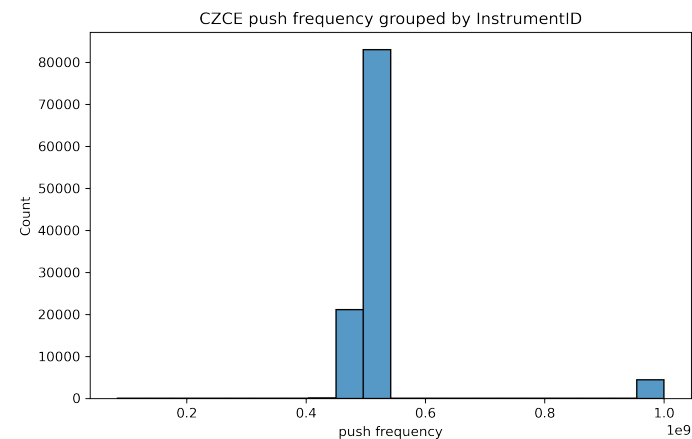
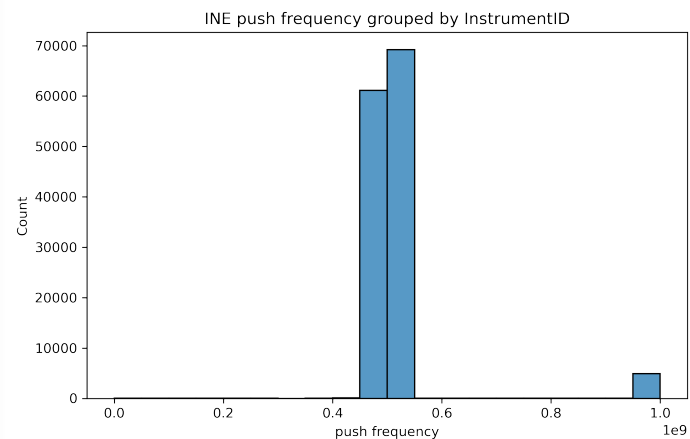
只根据交易所分组

- 推送间隔可聚成两组，分别是微秒级和百毫秒级，以微秒级为主。百毫秒级主要聚集在0.5秒左右。
- 推测，每秒钟传两次，微秒级的间隔是将整包tick分成每个合约的tick所用的时间。



根据交易所和合约名称分组

- 根据合约名进行差分后，微秒级的推送间隔被去除了。说明微秒级的时间间隔是对同一批推送数据进行处理所消耗的时间。



ActionDay, TradingDay和实际交易时间的关系

- 方法：
 - 将LocalTime转为datetime格式以便比较。
 - 观察LocalTime和UpdateTime的极值，确认所有数据都是夜盘数据。
 - 根据交易所分组分别统计ActionDay, TradingDay和LocalTime的取值分布，LocalTime的取值精确到day。

DCE交易所统计结果

- 从统计结果可以看出，数据接收时间为12月21日，但ActionDay和TradingDay都记录为12月22日。
- 结论：DCE交易所夜盘数据的ActionDay和TradingDay都是交易发生的自然日的下一个交易日。

DCE

ActionDay distribution:

20211222 146635

Name: ActionDay, dtype: int64

TradingDay distribution:

20211222 146635

Name: TradingDay, dtype: int64

LocalTime distribution:

20211221 146635

Name: LocalTime_datetime, dtype: int64

SHFE交易所统计结果

- 从统计结果可以看出，LocalTime和ActionDay基本可以对齐，但TradingDay全部归为12月22日。
- 结论：SHFE交易所夜盘数据的ActionDay在0点以前是交易发生的自然日，在0点之后是交易发生的下一个交易日。TradingDay是交易发生的下一个交易日。

SHFE

ActionDay distribution:

20211221 176367

20211222 79531

Name: ActionDay, dtype: int64

TradingDay distribution:

20211222 255898

Name: TradingDay, dtype: int64

LocalTime distribution:

20211221 176358

20211222 79540

Name: LocalTime_datetime, dtype: int64

CZCE交易所统计结果

- 从统计结果可以看出, LocalTime, ActionDay和TradingDay都记录为12月21日。
- 结论：DCE交易所夜盘数据的ActionDay和TradingDay都是交易发生的自然日。

CZCE

ActionDay distribution:

20211221 139737

Name: ActionDay, dtype: int64

TradingDay distribution:

20211221 139737

Name: TradingDay, dtype: int64

LocalTime distribution:

20211221 139737

Name: LocalTime_datetime, dtype: int64

SHFE交易所统计结果

- 从统计结果可以看出，LocalTime和ActionDay基本可以对齐，但TradingDay全部归为12月22日。
- 结论：INE交易所夜盘数据的ActionDay在0点以前是交易发生的自然日，在0点之后是交易发生的下一个交易日。TradingDay是交易发生的下一个交易日。

INE

ActionDay distribution:

20211221 117435

20211222 44964

Name: ActionDay, dtype: int64

TradingDay distribution:

20211222 162399

Name: TradingDay, dtype: int64

LocalTime distribution:

20211221 117431

20211222 44968

Name: LocalTime_datetime, dtype: int64

对现有数据进行数据检查

- 检查缺失值：无缺失值
- 检查价格是否为负或0：价格全部为正
- 检查LastPrice是否超过涨停跌停板：无LastPrice超过涨停跌停板
- 检查InstrumentID是否对应唯一的ExchangeID：InstrumentID有唯一对应ExchangeID（每个合约只在一个交易所交易）

对现有数据进行数据检查

- 检查Volume的单调性：Volume单调递增
- 检查volume异常值：存在volume=0的情况
- 检查Turnover的单调性：Turnover单调递增
- 检查Turnover异常值：存在Turnover=0的情况，与Volume相符

分钟bar合成

- 代码详见 “卓识基金project_code.ipynb”
- 结果转为pandas dataframe后存储为csv文件 “minbar_result.csv”
- 过滤掉Volume为0的tick

分钟bar合成 成果展示

- 合约名称：TA201

TA201 candlestick plot



分钟bar合成 成果展示

- 合约名称：ag2201

ag2201 candlestick plot



分钟bar合成 成果展示

- 合约名称：sc2204

