

미래에셋증권 X NAVER Cloud 공동주관

2023 빅데이터 페스티벌

데이터 설명자료

미래에셋증권 DataScience팀

작성일자: 2023.06.30

본 자료는 미래에셋증권이 제작한 것이며, 투자권유를 위한 광고물로 활용될 수 없고, 투자자에게 배포될 수 없습니다. 본 자료에 수록된 내용은 신뢰할만한 자료 및 정보로부터 얻어진 것이나 당사는 그 정확성이나 안정성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 투자 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.

MIRAE ASSET
미래에셋증권

1. 고객데이터
2. 시장데이터
3. 뉴스데이터
4. 유의사항

1. 고객데이터 - 데이터개요

- 미래에셋증권 고객의 거래이력 및 인구통계학적 정보와 4개월간의 자산/거래/채널접속 내역을 제공
- 파일명 : cs_data.csv

스냅샷 정보

고객정보

연령대
 고객등급
 관리사원존재여부

거래이력

최고자산액/시기
 국내주식 최초/최종 매수연월
 국내주식 거래월수
 공모주 청약 종목수
 매도패턴별 거래액 비중
 ...

시계열 정보

자산내역

주식/현금성 자산액
 주식 종목수
 보유 상위 종목명
 ...

거래내역

매수/매도 체결횟수
 매수/매도액
 매수 1위 종목명
 매도 1위 종목명
 입금/출금액
 ...

접속내역

HTS/MTS 접속일수

- ✓ 범주화, 상하위 1% 데이터 대체, 결측치 처리 등의 방법을 통해 익명 처리된 데이터를 제공

1. 고객데이터 - 데이터개요

- 시계열 정보 : 4개월간의 자산 및 거래/접속내역 데이터이며, 후위 표시자(M1, M1_1, ..)로 시기 구분

데이터 기간	2022.12(M1) ~ 2023.3(M4) 4개월
자산내역	월 말일 기준 데이터 제공 : M1, M2, M3, M4
거래내역 접속내역	월별 초순,중순,하순 동안의 데이터 집계내역 제공 : M1_1, M1_2, M1_3, M2_1, M2_2,, M4_1, M4_2, M4_3 : 초순(1~10일), 중순(11일~20일), 하순(21~말일)을 각각 _1, _2, _3으로 구분

➤ Example

칼럼명	상세사항
DMST_AST_EVAL_M1	2022년 12월 말일 기준 국내주식 자산 평가액
DMST_AST_EVAL_M2	2023년 1월 말일 기준 국내주식 자산 평가액
DMST_AST_EVAL_M3	2023년 2월 말일 기준 국내주식 자산 평가액
DMST_AST_EVAL_M4	2023년 3월 말일 기준 국내주식 자산 평가액
BUY_CNT_M1_1	2022년 12월 1일~10일간의 매수체결횟수
BUY_CNT_M1_2	2022년 12월 11일~20일간의 매수체결횟수
BUY_CNT_M1_3	2022년 12월 21일~31일간의 매수체결횟수
BUY_CNT_M3_3	2023년 2월 21일~28일간의 매수체결횟수

1. 고객데이터 – 데이터 컬럼 및 명세

- 자산 내역 : 4개월간의 월 말일 기준 자산내역. NULL값은 데이터가 존재하지 않거나 값이 0인 경우

데이터 상세

컬럼명 (후위표시자 제외)	상세사항	샘플 예시
CASH_AST	월 말일 기준 예수금 등 현금성 자산액	150000
DMST_AST_EVAL	월 말일 기준 국내주식 자산 평가액	10000000
DMST_AST_PCHS	월 말일 기준 국내주식 자산 매입액	3600000
OVST_AST_EVAL	월 말일 기준 해외주식 자산 평가액	9700000
OVST_AST_PCHS	월 말일 기준 해외주식 자산 매입액	9200000
DMST_ITM_CNT	월 말일 기준 국내주식 보유종목수 (해외주식 : OVST)	2
DMST_AST1(2,3)_ITM	월 말일 기준 평가액 1(2,3)위 국내주식 종목번호 (해외주식 : OVST)	218420
DMST_AST1(2,3)_MKT	월 말일 기준 평가액 1(2,3)위 국내주식 거래소 (해외주식 : OVST)	코스피
DMST_AST1(2,3)_EVAL	월 말일 기준 평가액 1(2,3)위 국내주식 평가액 (해외주식 : OVST)	9100000
DMST_AST1(2,3)_PCHS	월 말일 기준 평가액 1(2,3)위 국내주식 매입액 (해외주식 : OVST)	8300000

1. 고객데이터 - 데이터 컬럼 및 명세

- 거래/접속내역 : 4개월간의 초순, 중순, 하순 별 집계내역. NULL값은 데이터가 존재하지 않거나 값이 0인 경우

데이터 상세

컬럼명 (후위표시자 제외)	상세사항	샘플 예시
BUY_CNT	특정 기간 동안의 주식 매수체결횟수	45
SEL_CNT	특정 기간 동안의 주식 매도체결횟수	42
BUY_ITM_CNT	특정 기간 동안의 주식 매수종목수	31
SEL_ITM_CNT	특정 기간 동안의 주식 매도종목수	31
BUY_AMT	특정 기간 동안의 주식 매수액 합계	24000000
SEL_AMT	특정 기간 동안의 주식 매도액 합계	23000000
BUY1(2,3)_ITM	특정 기간 동안의 주식 매수액 1(2,3)위 종목번호	196700
BUY1(2,3)_MKT	특정 기간 동안의 주식 매수액 1(2,3)위 종목 거래소	코스닥
BUY1(2,3)_AMT	특정 기간 동안의 주식 매수액 1(2,3)위 종목 매수액 합계	3200000
SEL1(2,3)_ITM	특정 기간 동안의 주식 매도액 1(2,3)위 종목번호	170030
SEL1(2,3)_MKT	특정 기간 동안의 주식 매도액 1(2,3)위 종목 거래소	코스닥
SEL1(2,3)_AMT	특정 기간 동안의 주식 매도액 1(2,3)위 종목 매도액 합계	2700000
STK_IN(OUT)	특정 기간 동안의 주식 입고(출고)액 합계	14000000
MNY_IN(OUT)	특정 기간 동안의 입금(출금)액 합계	800000

1. 고객데이터 – 데이터 컬럼 및 명세

- 거래/접속내역 : 4개월간의 초순, 중순, 하순 별 집계내역. NULL값은 데이터가 존재하지 않거나 값이 0인 경우

데이터 상세

컬럼명 (후위표시자 제외)	상세사항	샘플 예시
CONN_DYS	특정 기간 동안의 HTS 또는 MTS 접속일수	10
MTS_DYS	특정 기간 동안의 MTS 접속일수	3
HTS_DYS	특정 기간 동안의 HTS 접속일수	5

- ✓ MTS, HTS 모두 동시접속한 날이 있는 고객의 경우 $CONN_DYS < MTS_DYS + HTS_DYS$ 가능

1. 고객데이터 - 데이터 컬럼 및 명세

- 스냅샷 정보 : 2022년 12월 기준. NULL값은 데이터가 없거나 값이 0인 경우
(e.g. 3년간 국내ETF 매수내역 없음)

데이터 상세

컬럼명	상세사항	샘플 예시
AGE_TCD	10세 기준 연령대	60
MT_EP_EXIST_YN	관리사원 존재여부	Y
LST_BEST_EA	과거 최고자산액	23000
LST_BEST_YM	과거 최고자산액 기록 연월 (YYYYMM)	202206
DMST_FST_BUY_YM	이전 3년(19.12~22.11)간 국내주식(ETF제외) 최초 매수연월	202007
DMETF_FST_BUY_YM	이전 3년(19.12~22.11)간 국내ETF 최초 매수연월	202111
OVST_FST_BUY_YM	이전 3년(19.12~22.11)간 해외주식(ETF제외) 최초 매수연월	202111
OVETF_FST_BUY_YM	이전 3년(19.12~22.11)간 해외ETF 최초 매수연월	202205
DMST_FIN_BUY_YM	이전 3년(19.12~22.11)간 국내주식 최종 매수연월 (DMETF, OVST, OVETF 컬럼도 있으며, 상품구분은 상동)	202208
DMST_TR_MONTHS_CNT	이전 3년(19.12~22.11)간 국내주식 매수 또는 매도 발생 연월수 (DMETF, OVST, OVETF 컬럼도 있으며, 상품구분은 상동)	18
APY_ITM_CNT	이전 3년(19.12~22.11)간 공모주 청약에 참여한 종목 개수	4
APY_FIN_YM	이전 3년(19.12~22.11)간 마지막으로 참여한 공모주 청약연월	202204

1. 고객데이터 - 데이터 컬럼 및 명세

- 스냅샷 정보 : 2022년 12월 기준. NULL값은 데이터가 없거나 값이 0인 경우
(e.g. 3년간 국내ETF 매수내역 없음)

데이터 상세

컬럼명	상세사항	샘플 예시
DAY_TR_RATIO	이전 6개월(22.6~22.11) 매도액 중 당일매수매도 발생 매도액 비중	0.0
SWING_TR_RATIO	이전 6개월(22.6~22.11) 매도액 중 매수 후 5일 이내 매도액 비중	1.0
MONTHS_TR_RATIO	이전 6개월(22.6~22.11) 매도액 중 매수 후 30일 이내 매도액 비중	0.4
MID_TR_RATIO	이전 6개월(22.6~22.11) 매도액 중 매수 후 180일 이내 매도액 비중	0.2
HLD_TR_RATIO	이전 6개월(22.6~22.11) 매도액 중 매수 후 1년 이내 매도액 비중	1.0
YEARS_TR_RATIO	이전 6개월(22.6~22.11) 매도액 중 매수 N()=2)년 이후 매도액 비중	0.4

- ✓ 반올림 처리 등으로 고객별 TR_RATIO의 합계는 1이 아닐 수 있음

2. 시장데이터

- 4개월간의 일자 별 미국지수/국내지수와 이전 3년간 미래에셋증권에서 진행된 공모주 청약일정 제공

시장 지수

국내
KOSPI
KOSDAQ

미국
DOW
NASDAQ
S&P500

공모주 청약 일정

공모주
월별 청약 종목번호

데이터 상세

파일명	칼럼명	상세사항	샘플 예시
mkt_idx.csv	BAS_DT	기준일자	20221226
	SP500	S&P500지수	3839.50
	DOW	다우지수	33147.25
	NASDAQ	나스닥지수	10466.48
	KOSPI	코스피지수	2317.14
	KOSDAQ	코스닥지수	694.68
apy_itm_hist.csv	BAS_YM	기준연월	201912
	APY_ITM	공모주 청약 진행된 종목 번호	062970

✓ mkt_idx.csv : NULL값은 국내/미국 영업일이 아닌 경우를 의미

3. 뉴스 데이터

- 데이터 정보 : 22.12.01 ~ 23.06.30 미래에셋증권 MTS/HTS에서 제공됐던 뉴스 데이터

데이터 상세

칼럼명	상세사항	샘플 예시
DATE_TIME	뉴스 수신시각 (YYMMDD_HHmmSS)	20230610_080025
WRITER	뉴스 제공자	로이터(AI뉴스)
SUBCATEGORY	뉴스 카테고리 (정치/경제/사회/기술 中 1, 당사 주제분류모델 활용)	경제
TITLE	뉴스 제목	ChatGPT가 아시아 부동산 산업을 뒤흔들다..
IMPORTANCE	뉴스 중요도 (0 ~ 100 사이, 당자 중요도평가 모델 활용)	38.3
ITEM_NAME	뉴스 연관 종목명	SK네트웍스
TAG_LIST	뉴스 관련 태그 (' ' 기호로 구분)	아시아 부동산 AI
CONTENT	뉴스 내용	HTML

- ✓ AI뉴스의 경우, 당사 알고리즘을 통해 번역/요약되어 서비스되는 영문 뉴스

4. 유의사항

- csv 파일 특성상 데이터타입 보존이 되지 않을 수 있습니다. 종목코드 등 숫자 문자열과 같은 경우를 고려하여 타입 캐스팅에 유의해 데이터를 사용하시기 바랍니다.
- 제공되는 데이터는 빅데이터 페스티벌 공모전을 위해 공유되는 데이터로, 임의적인 외부 활용에서 발생 가능한 민/형사상 책임은 외부 활용 당사자에게 있습니다.
- 제공되는 데이터를 모두 활용할 필요는 없으며, 외부 데이터만 사용하는 등 데이터 출처는 자유롭게 사용 가능합니다. 외부 데이터 활용 과정에서 발생 가능한 민/형사상 책임은 당사자에게 있습니다.

미래에셋증권 X NAVER Cloud 공동주관

2023
빅데이터
페스티벌

감사합니다