**미래에셋 머신러닝 경진대회 - 주식거래내역으로 매수 상위종목 예측**

단계별 안내를 확인하고, 요구사항을 만족하는 프로젝트를 완성하세요.

* 박정환

* sj2059@pusan.ac.kr

* 10월 4일 23:59 마감

[STEP 1. 데이터 구조 설명 및 채점 기준 안내](https://programmers.co.kr/assignments/16206/challenges/635#collapseStepFirst)

**주식거래내역으로 매수 상위종목 예측**

**개요**

과거 거래 데이터를 이용해, 각 그룹에서 7월에 가장 많은 고객이 매수할 상위 3개 종목 예측

* 훈련 데이터 기간 : 2019년 7월 ~ 2020년 6월
* 예측 기간: 2020년 7월 1일~28일 (4주)
* 상위 3개 종목 : 2020년 7월에 그룹 내 매수 고객 수가 가장 높을 3개 종목

**데이터**

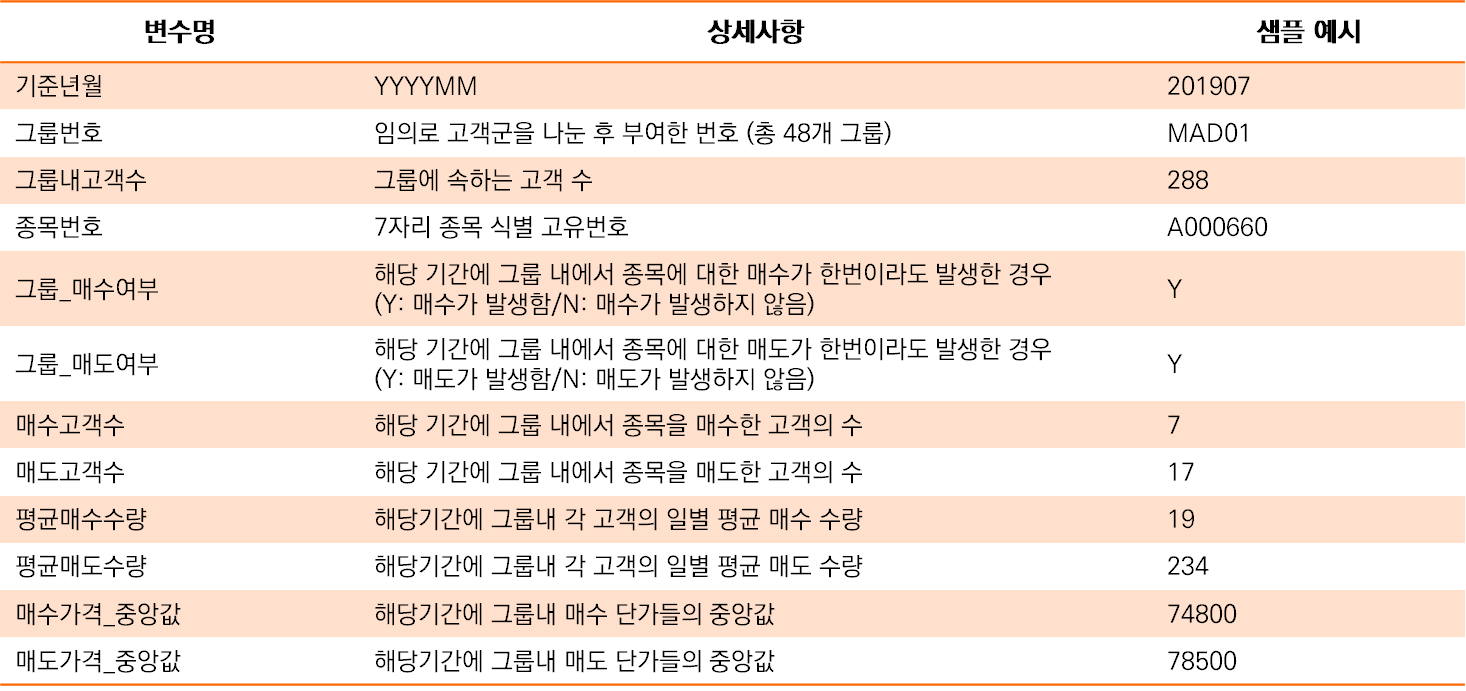
본 문제에서는 다음 두 데이터가 주어집니다.

**1. trade\_train.csv**

이 csv 파일은 2019년 7월 ~ 2020년 6월 월별 매수/매도 데이터를 담고 있습니다.

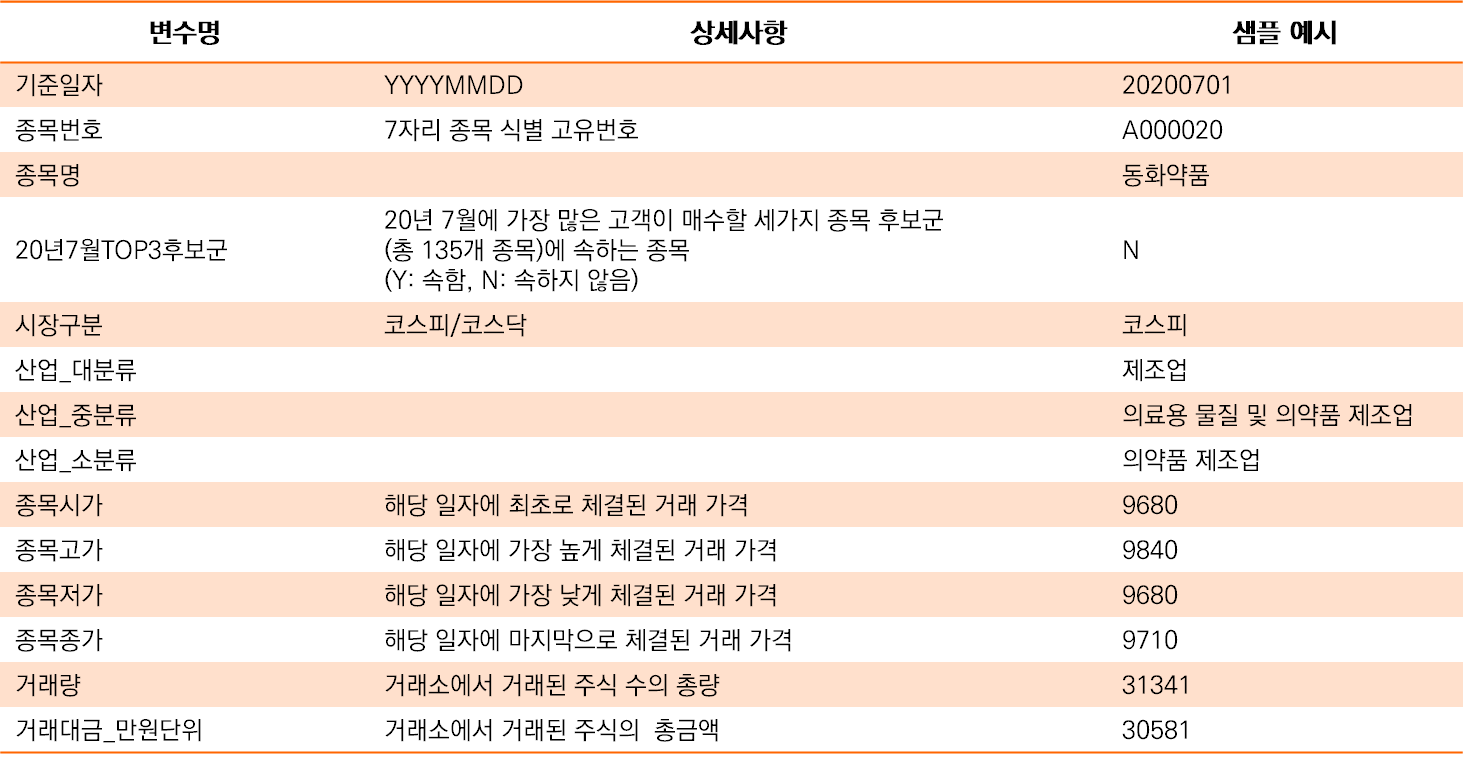
csv 파일에는 총 48개 그룹, 19118명의 고객, 1086개 종목이 있으며, 비식별화 조치에 따라 <매수고객수>와 <매도고객수> 필드는 0 또는 3 이상의 정수로 구성되어 있습니다.

csv 파일의 필드는 다음과 같습니다.



**2. stocks.csv**

이 csv 파일에는 일자별 종목 정보 데이터가 담겨 있습니다. csv 파일의 필드는 다음과 같습니다.



**과제**

주어진 두 데이터를 이용하여 각 그룹이 20년 7월에 **가장 많은 고객이 매수할** Top 3 종목을 맞춰주세요. 결과 파일은 다음과 같은 형식의 csv 파일이어야 합니다. 또한 csv 파일은 다음 양식을 만족해야 합니다.

1. row는 그룹 이름 순으로 정렬되어있어야 합니다. 즉, row는 MAD01, MAD02, ..., MAD48 순이어야 합니다.
   * 그룹명의 MAD 뒤에 붙는 번호는 두 자리 수여야 합니다. 한자리 수인 경우, 숫자 앞에 0을 붙여주세요.
   * 예를 들어 **MAD8**이 아니라 **MAD08**과 같이 작성해주세요.
2. Top 3 종목 또한 종목번호순으로 정렬되어있어야 합니다.
   * 예를 들어, MAD1 그룹이 매수할 Top 3 종목을 A001260, A000440, A000520으로 예측한 경우, csv 파일에는 세 종목을 사전 순으로 A000440, A000520, A001260 처럼 나열해야 합니다.

| **그룹명** | **종목번호1** | **종목번호2** | **종목번호3** |
| --- | --- | --- | --- |
| MAD01 | A014570 | A081150 | A238120 |
| MAD02 | A087260 | A147760 | A285130 |
| MAD03 | A002780 | A003410 | A011040 |
| MAD04 | A043340 | A045390 | A290550 |
| .. | .. | .. | .. |
| MAD32 | A042040 | A072130 | A148150 |
| MAD33 | A019170 | A066130 | A285130 |
| .. | .. | .. | .. |
| MAD48 | A004835 | A044480 | A241560 |

<제출 파일 예시>

※ 주의: 각 그룹이 7월에 **가장 많이 매수할** Top 3 종목을 맞춰야 합니다. 거래대금, 거래량 등의 기준이 아닌, 고객이 매수한 종목순(매수고객수(unique)가 높은 순) 입니다.  
※ 주의: 하단의 데이터가 다운로드지 않는다면 [대회 페이지](https://programmers.co.kr/competitions/252/2020-miraeasset)부터 다시 입장해주세요. 데이터는 화면을 로드한 후 일정 시간이 지나면 다운로드가 되지 않습니다. [대회 페이지](https://programmers.co.kr/competitions/252/2020-miraeasset)에서 [테스트 입장] 버튼을 눌러 재입장하고 과제를 다시 열어 데이터를 다운로드하여 보세요.

[STEP 2. 데이터 다운로드](https://programmers.co.kr/assignments/16206/challenges/635#collapseStepSecond)

테스트를 시작하기 전에 개발 환경 설정을 다운로드 받으세요.

[훈련 데이터 파일](https://grepp-ds-test-data.s3-ap-northeast-2.amazonaws.com/production/trade_train_set.zip?X-Amz-Expires=600&X-Amz-Date=20200924T064850Z&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIGGYA7NALKSFWYJA%2F20200924%2Fap-northeast-2%2Fs3%2Faws4_request&X-Amz-SignedHeaders=host&X-Amz-Signature=1c5ca35b60d92690da463e15f17f3c9da2a7d4dde2e73091a70388a7d8d2de15)

[ZIP file format](https://grepp-ds-test-data.s3-ap-northeast-2.amazonaws.com/production/trade_train_set.zip?X-Amz-Expires=600&X-Amz-Date=20200924T064850Z&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIGGYA7NALKSFWYJA%2F20200924%2Fap-northeast-2%2Fs3%2Faws4_request&X-Amz-SignedHeaders=host&X-Amz-Signature=1c5ca35b60d92690da463e15f17f3c9da2a7d4dde2e73091a70388a7d8d2de15)

[예시 출력 데이터 파일](https://grepp-ds-test-data.s3-ap-northeast-2.amazonaws.com/production/answer_sheet.csv?X-Amz-Expires=600&X-Amz-Date=20200924T064850Z&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIGGYA7NALKSFWYJA%2F20200924%2Fap-northeast-2%2Fs3%2Faws4_request&X-Amz-SignedHeaders=host&X-Amz-Signature=2a77d08dbffe8c89e18a5867fc340bb1b0788dcf5ac751bccb93e6a5f89d67d1)

[ZIP file format](https://grepp-ds-test-data.s3-ap-northeast-2.amazonaws.com/production/answer_sheet.csv?X-Amz-Expires=600&X-Amz-Date=20200924T064850Z&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIGGYA7NALKSFWYJA%2F20200924%2Fap-northeast-2%2Fs3%2Faws4_request&X-Amz-SignedHeaders=host&X-Amz-Signature=2a77d08dbffe8c89e18a5867fc340bb1b0788dcf5ac751bccb93e6a5f89d67d1)

[STEP 3. 결과 파일 업로드하기](https://programmers.co.kr/assignments/16206/challenges/635#collapseStepThird)

데이터 파일 제출하기

* 파일 포맷은 CSV 이어야합니다.
* 최종 점수 계산에 사용될 제출물을 선택하지 않으면 공개 리더 보드의 제출 점수에 따라 자동으로 선택됩니다.
* 공개 점수는 제출된 데이터의 일부만 가지고 채점하기 때문에 실제 최종 점수와 다를 수 있습니다.
* 예측 결과가 가장 좋은 데이터를 선별해서 올려주세요.

 결과 파일 선택

첨부된 파일이 없습니다.

파일 업로드

[STEP 4. 최종 점수로 반영할 데이터 선택하기 & 코드 제출](https://programmers.co.kr/assignments/16206/challenges/635#collapseStepFifth)

| **제출 목록** | **공개 점수** | **최종 점수 사용 데이터** |
| --- | --- | --- |
| **;^)**  STEP 3에서 파일을 업로드 하면 이곳에 표시되고, 최종 점수에 사용할 데이터를 선택할 수 있습니다. | | |
|  |  |  |

[리더보드 보기](https://programmers.co.kr/competitions/252/2020-miraeasset?tab=leaderboards)

코드 제출하기

* 제출한 코드에 따라 추가 점수를 받을 수 있습니다.
* [코드 파일]에는 상단의 지문에서 요구한 파일을 제출해주세요. 지문에 파일에 대한 언급이 없다면 Jupyter Notebook, python, R 파일 등을 올려주세요.
* [HTML 파일]도 업로드하면 평가에 도움이 됩니다. 코드에 Jupyter Notebook을 제출했다면, 해당 파일의 HTML 버전도 올려주세요.  
  Jupyter Notebook을 사용하지 않았다면 제출하지 않아도 무방합니다.
* [HTML 파일]은 Jupyter Notebook에서 file-download as-HTML 메뉴를 통해 받을 수 있습니다.

양식의 맨 위

 코드 파일 선택

첨부된 파일이 없습니다.

양식의 맨 아래

양식의 맨 위

 HTML 파일 선택

첨부된 파일이 없습니다.

양식의 맨 아래