

데이터 전처리 관련 정보

- 아이디 마스킹
 - 개인 정보 노출 우려가 있는 게임 사용자 아이디, 파티 아이디, 아이템 아이디 등은 모두 해싱 처리 후 제공됩니다.
- 데이터 표준화
 - 모든 수치형 데이터는, 아래의 방식으로 표준화 처리되어 제공됩니다.

$$x_{new} = \frac{x - \mu}{\sigma}$$

- 모든 날짜 정보는 제공되는 데이터의 최초 날짜를 기준으로 인덱싱 처리하여 제공됩니다.
 - 주는 최초 시작 주부터 일주일씩 1~8의 번호 표기
 - 날짜는 요일 별로 1~7의 번호 표기
 - ◆ 각 주는 수요일을 시작일로 함
 - ◆ 1: 수요일, 2: 목요일, 3: 금요일, 4: 토요일, 5: 일요일, 6: 월요일, 7: 화요일
 - 시간 정보는 그대로 유지
 - 예를 들어, 2번째 주의 화요일 오후 1시의 경우, 다음과 같이 표기

Week	Day	Hour
2	7	13:00:00

파일 별 스키마 및 컬럼 설명

데이터는 학습용 데이터와 테스트용 데이터로 크게 구분되며 train_label.csv 를 제외하고 나머지 데이터는 학습용 데이터와 테스트용 데이터의 형식이 동일합니다.

전체 데이터는 아래와 같이 구성됩니다.

- train_label.csv: 학습 대상 계정들의 고유 아이디 및 이탈 여부 (Retained, week, month, 2month)를 나타내는 파일
- train_activity.csv, test_activity.csv: 유저의 인게임 활동량을 일주일 단위로 집계한 파일

- train_party.csv, test_party.csv: 유저간 파티 구성 관계를 집계한 파일
- train_guild.csv, test_guild.csv: 문파 구성 관계를 집계한 파일
- train_trade.csv, test_trade.csv: 유저간 거래이력을 집계한 파일
- train_payment.csv, test_payment.csv: 유저별 주간 결제 금액을 집계한 파일

각 파일별 상세 스키마는 다음과 같습니다.

- train_label.csv

acc_id	계정 아이디
label	이탈 여부 <ul style="list-style-type: none"> ● week : 1주내 이탈 ● month : 4주내 이탈 ● 2month : 8주내 이탈 ● retained : 비이탈

- train_activity.csv

- 고유 아이디(acc_id)로 유저 구분이 되며, 각 주별 일일 활동 정보의 합계를 나타냄

wk	활동 주 (1~8)
acc_id	계정 아이디
cnt_dt	해당 주에 한번 이상 접속한 일수
play.time	플레이 시간 (단위: 초)
npc.exp	NPC 사냥 일반 경험치
npc.hongmun	NPC 사냥 홍문 경험치
quest.exp	퀘스트 일반 경험치
quest.hongmun	퀘스트 홍문 경험치
item.hongmun	아이템 홍문 경험치
game.combat.time	전투 시간 (단위: 초)
get.money	재화 획득량
duel.cnt	결투 참여횟수
duel.win	결투 승리 횟수
partybattle.cnt	전장 참여 횟수
partybattle.win	전장 승리 횟수
cnt.enter.inzone.solo	솔로 인전 입장 횟수
cnt.enter.inzone.light	라이트 인전 입장 횟수

cnt.enter.inzone.skilled	숙련 인던 입장 횟수
cnt.enter.inzone.normal	라이트/숙련 인던 입장 횟수
cnt.enter.raid	레이드 참여 횟수
cnt.enter.raid.light	라이트 레이드 참여 횟수
cnt.enter.bam	바람평야 입장 횟수
cnt.clear.inzone.solo	솔로 인전 완료 횟수
cnt.clear.inzone.light	라이트 인전 완료 횟수
cnt.clear.inzone.skilled	숙련 인던 완료 횟수
cnt.clear.inzone.normal	라이트/숙련 인던 완료 횟수
cnt.clear.raid	레이드 완료 횟수
cnt.clear.raid.light	라이트 레이드 완료 횟수
cnt.clear.bam	바람평야 완료 횟수
normal.chat	일반 채팅 횟수
whisper.chat	귓속말 채팅 횟수
district.chat	지역 채팅 횟수
party.chat	파티 채팅 횟수
guild.chat	문파 채팅 횟수
factions.chat	세력 채팅 횟수
cnt.use.buffitem	버프 아이템 사용 횟수
gathering.cnt	채집 횟수
making.cnt	제작 횟수

■ train_party.csv

- 유저간 파티 구성 관계를 나타내는 파일

party_start_week	파티 생성 주 (1~8)
party_start_day	파티 생성 일 (1~7)
Party_start_time	파티 생성 시간 (00:00:00 ~ 23:59:59)
Party_end_week	파티 종료 주 (1~8)
party_end_day	파티 종료 일 (1~7)
party_end_hour	파티 종료 시간 (00:00:00 ~ 23:59:59)
party_members_acc_id	파티 구성원 아이디 리스트 파티 참여했던 모든 구성원들의 아이디 기록

■ train_guild.csv

- 문파 구성 관계를 나타내는 파일

- 데이터 제공 기간 마지막 시점을 기준으로 추출

guild_id	길드 고유 아이디
guild_member_acc_id	길드원 아이디 리스트

■ train_trade.csv

- 유저간 거래이력을 나타내는 파일

trade_week	거래 발생 주차 (1 ~ 8)
trade_day	거래 발생 일 (1 ~7)
trade_time	거래 발생 시간 (00:00:00 ~ 23:59:59)
source_acc_id	주는 계정 아이디
target_acc_id	받는 계정 아이디
item_type	아이템 종류 <ul style="list-style-type: none"> ● money : 금 ● grocery : 잡화 ● weapon : 무기 ● costume : 옷 ● gem : 보석 ● accessory : 액세서리
item_amount	거래된 아이템 수량

■ train_payment.csv

- 유저별 주간 결제 금액

payment_week	결제 주 (1~8)
acc_id	결제 유저 아이디
payment_amount	해당 주 총 결제액