

# DS School 교육용 ZIGZAG 데이터 명세서

## 1. user\_event\_logs.csv

- a. 지그재그 앱 이용자 로그 (105,815건)
- b. 기간 : 2018년 6월 11일
- c. 앱 버전 : v5.5 이용자
- d. 컬럼 별 명세
  - i. timestamp : 이벤트 발생 시간 (한국 시간 기준)
  - ii. user\_id : 이용자 고유 식별자
  - iii. event\_origin : 이벤트가 발생한 앱 위치
    - 1. event\_origin 값 별 의미
      - a. goods\_search\_result : 특정 검색어의 상품 검색 결과  
(Ex: goods\_search\_result/반팔티)
      - b. shops\_ranking : '쇼핑몰 랭킹' 영역
      - c. shops\_bookmark : '즐거찾기' 영역
      - d. category\_search\_result : 카테고리 검색 결과 (Ex: category\_search\_result/상의)
      - e. my\_goods : '내 상품' 영역
  - iv. event\_name : 발생한 이벤트 명
    - 1. event\_name 값 별 의미
      - a. app\_page\_view : 앱 내 화면 이동
      - b. enter\_browser : 앱 내 클릭을 통해, 특정 웹페이지로 진입
      - c. add\_bookmark : 특정 쇼핑몰을 즐겨찾기 추가
      - d. remove\_bookmark : 특정 쇼핑몰을 즐겨찾기 제거
      - e. add\_my\_goods : 특정 상품을 내 상품 추가
      - f. remove\_my\_goods : 특정 상품을 내 상품 제거
  - v. event\_goods\_id : 이벤트가 발생한 상품 고유 식별자
    - 1. 상품 관련 이벤트가 아닌 경우, 공백
  - vi. event\_shop\_id : 이벤트가 발생한 쇼핑몰 고유 식별자
    - 1. 쇼핑몰 관련 이벤트가 아닌 경우, 공백

## 2. order\_info.csv

- a. 지그재그 앱 이용자의 상품 주문 정보 (867건)

b. 컬럼 별 명세

- i. timestamp : 상품 주문 발생 시간 (한국 시간 기준)
- ii. user\_id : 이용자 고유 식별자
- iii. goods\_id : 주문한 상품의 고유 식별자
- iv. shop\_id : 주문한 쇼핑몰의 고유 식별자
- v. price : 상품 주문 금액 (단위: ₩)

### 3. user\_info.csv

a. 지그재그의 이용자 정보 (이용자 10,000명을 연령 기준 무작위 균등 추출)

b. 컬럼 별 명세

- i. user\_id : 이용자 고유 식별자 (비식별화)
- ii. os : 이용자의 기기 OS
- iii. age : 이용자가 입력한 연령 (한국 나이 기준)
  - 1. 범위 : 15 ~ 45
  - 2. 값이 45인 경우, 45세 이상을 의미
  - 3. 값이 -1 인 경우, 결측을 의미

### 4. shop\_info.csv

a. 지그재그에 입점한 쇼핑몰 정보 (200개, 2018년 6월 11일 자 지그재그 랭킹 기준 상위 1,000개 쇼핑몰 중 200개 무작위 추출)

b. 컬럼 별 명세

- i. shop\_id : 쇼핑몰 고유 식별자 (입점 순)
- ii. name : 쇼핑몰 명 (실제 쇼핑몰 명과 무관한 가상의 명칭)
- iii. category : 쇼핑몰이 등록한 카테고리
  - 1. '의류', '가방', '슈즈', '란제리&파자마', '액세서리', '비치웨어', '빅사이즈', '커플룩', '피트니스', '임부복', '패션소품'
- iv. age : 쇼핑몰의 등록한 연령대
  - 1. category 값이 '의류'인 경우, age값 존재 (그 외의 경우, 공백)
  - 2. '10대', '20대 초반', '20대 중반', '20대 후반', '30대 초반', '30대 중반', '30대 후반' 값 중 최대 3가지 값을 연속으로 가질 수 있음
  - 3. 연속 되는 값의 구분자 : '/'
- v. style : 쇼핑몰이 등록한 스타일
  - 1. category 값이 '의류', '빅사이즈', '커플룩', '액세서리'인 경우, style 값 존재 (그 외의 경우, 공백)
  - 2. category 별 가질 수 있는 style 값

- a. 의류: '페미닌', '모던시크', '심플베이직', '러블리', '유니크', '미시스타일', '캠퍼스룩', '빈티지', '섹시글램', '스쿨룩', '로맨틱', '오피스룩', '럭셔리', '헐리웃스타일' 중 최대 2가지 값을 가질 수 있음
- b. 연속 되는 값의 구분자 : '/'
- c. 빅사이즈, 커플룩: 의류 영역 style 중 1가지 값을 가질 수 있음
- d. 액세서리: '페미닌', '심플시크', '키치', '펑키', '큐티', '볼드&에스닉' 중 1가지 값을 가질 수 있음

## 5. goods\_info.csv

- a. 지그재그에 등록 된 상품 정보 (7,188개, 제공되는 로그에서 이용자가 이벤트를 발생시킨 모든 상품)
- b. 컬럼 별 명세
  - i. goods\_id : 상품 고유 식별자 (최초 등록 순)
  - ii. timestamp : 마지막 상품 정보 업데이트 시간 (한국 시간 기준)
  - iii. shop\_id : 쇼핑몰 고유 식별자
  - iv. category : 상품의 카테고리
    - 1. '상의', '바지', '스커트', '원피스', '비치웨어', '아우터', '악세사리', '슈즈', '가방', '패션소품', '피트니스', '란제리&파자마' 중 1가지 값을 가질 수 있음
  - v. price : 상품의 가격 (단위: ₩)
    - 1. 상품에서 제공하는 옵션에 따라 실제 이용자에게 적용되는 가격은 다를 수 있음
  - vi. image\_type : 상품 대표 이미지의 확장자
    - 1. 'jpg', 'gif', 'png' 값을 가질 수 있음
  - vii. image\_width : 상품 대표 이미지의 너비 (단위: pixel)
  - viii. image\_height : 상품 대표 이미지의 높이 (단위: pixel)