

## Wishlist - Data수집 관련

### \*상품 category

(11번가 기준)

1. 4분류 : 대부분 4분류

대분류(브랜드/패션) - 중분류(브랜드 여성의류) - 소분류(원피스) - 세부분류(쉬폰원피스) - 상품

2. 3분류 : 소분류 이하 하나의 세부분류

대분류(가구/인테리어) - 중분류(수납가구) - 소분류(수납장) - 세부분류(수납장) - 상품

=> 4분류로 통일, 3분류의 경우에 소분류 이하를 소분류와 동일한 세부분류로 입력하면 됨.

### 1. 상품 category 고려한 Table Schema 짜기

#### <상품 Table>

상품코드 (KEY)	상품명	대분류	중분류	소분류	세부분류	가격	사용자 코드	판매자 코드??
INT64	STR	INT32	INT32	INT32	INT64	INT64	INT64	INT64

#### <사용자 Table>

사용자코드 (KEY)	사용자명	성별	나이	거주지	관심사1 (ML 결과, 혹은 통계 량)	관심사2 (ML결과, 통계량)
INT64	STR	STR	INT32	STR	INT32	INT32

#### <사용자가 저장한 상품>

저장코드(KEY)	사용자코드	상품코드	저장 날짜	우선순위
INT64	INT64	INT64	DATE	INT32

#### <판매자와 상품> ???

판매 코드(KEY)	판매자 코드	상품 코드
INT64	INT64	INT64

#### cf> 데이터 저장 예시

구분	데이터 항목	영문명	상세 설명
Demo.	ID	ID	고객의 고유 식별번호
	성별	GENDER	고객의 성별 (남성:1, 여성:2)

	연령대	AGE_PRD	고객의 연령대
	거주지	HOM_PST_NO	거주지역 신우편번호 앞 3자리 (서울 - 구 단위, 서울 외 지역 - 시/도 단위 변환)
쇼핑 업종 상품 구매 정보	ID	ID	고객의 고유 식별번호
	영수증번호	RCT_NO	구매 내역의 고유 식 별번호
	업종	BIZ_UNIT	쇼핑 5개 업종(A01/ A02/.../A05) (A01:백화점, A02: 대형마트, A03:슈퍼 마켓, A04:편의점, A05:드러그스토어)
	상품 소분류 코드	PD_S_C	제휴사 상품분류정보
	점포코드	BR_C	구매가 발생한 점포 코드성 정보
	구매일자	DE_DT	구매가 발생한 일자 (YYYY/MM/DD)
	구매시간	DE_HR	구매가 발생한 시각
	구매금액	BUY_AM	구매한 금액
	구매수량	BUY_CT	구매한 수량
쇼핑 외 업종 이용정보	ID	ID	고객의 고유 식별번호
	업종	BIZ_UNIT	9개 업종명 (B01:호텔, B02:여 행사, B03:면세점, C01:영화관, C02:테 마파크, C03:야구관 람, D01:패스트푸드, D02:패밀리레스토 랑, D03:카페)
	이용월	CRYM	이용이 발생한 월 (YYYY/MM)
	이용금액	U_AM	이용한 금액
	이용건수	U_CT	이용한 건수
쇼핑 업종 상품 분류 정보	업종	BIZ_UNIT	쇼핑 5개 업종(A01/ A02/.../A05)
	상품 소분류 코드	PD_S_C	상품 소분류 카테고리 코드성 정보

	소분류명	PD_S_NM	상품 소분류 카테고리 한글명
	중분류명	PD_M_NM	상품 중분류 카테고리 한글명
	대분류명	PD_H_NM	상품 대분류 카테고리 한글명

### 1. '상품명' 크롤링

- 대분류 13종
- 중분류 평균 15개
- 소분류 평균 10개
- 세부분류 평균 5개
- => 약 9750개 categories
- => 중분류 별로 크롤링 시작 -> Chrome Driver, Beautiful Soap
- => 약 200번의 (수동) 노가다가 필요
- => 코드 돌려봐야 알겠지만, 하루 20여개 정도는 가능할듯,, 혹은 중분류 클릭도 자동화해서 13번으로 노가다 횟수 줄임

### 2. 일정

#### 2.1 9/28(토)

- 크롤링 코드 짜고, test
- 전처리 코드 짜고, 명사 추출 (패키지 : KoNLP / 시도 : KoBert)
- 최소 1개의 소분류에 대해 crawling / 전처리 완료하여 상품명 추출

#### 2.2 10/5(토)

- 자동화하여 상품명 크롤링 완료
- 스키마 점검

### 3. ML/DL 시도 (CNN + RNN),,,

- 알고리즘 미리 짜 놓고,
- 데이터 쌓이기 시작하면 시도

#### —————핵심가치 관련—————

- 생필품인지 vs 기호품인지
- 위시리스트는, 기호품
- 소비자가 믿는 것 : (광고성이 아닌) 리뷰 수, 리뷰 내용

<소비자에게 줄 수 있는 핵심 가치는 무엇일까?>

내가 사려고 했던, 필요한 물건들에 대한 알짜 쇼핑정보 제공?  
알짜 쇼핑정보 -> 개인이 올리는 글들을 긁어서 정리 (맘카페,,)

내가 사고 싶었던, 필요한 상품을 기억/관리할 수 있는것?  
내가 과거에 좋아했던 것보다, 지금 트렌드인 상품이 더 중요.

상품 가격비교??

상품 매진임박 알림?? -> 판매자와의 connection이 필요,,