

크롤링 데이터 형식

1. storeInfo.csv (총 14개 열)

- 컬럼 내용

store_id: pk (int)

region: 지역 (string)

store_name: 이름 (string)

store_x: 위도 (float)

store_y: 경도 (float)

store_addr: 지번주소 (string)

store_addr_new: 도로명주소 (string)

store_tel: 전화번호 (string)

open_hours: 영업시간 (string)

website: 웹사이트 (가게 자체 웹사이트) (string)

s_link: 식신 사이트 가게 고유 ID (string)

g_link: 구글맵 가게 링크 (string)

d_link: 다이닝코드 가게 고유 ID (string)

n_link: 네이버 링크 가게 고유 ID (string)

- 데이터 예시

store_id: 18

region: 서울

store_name: 한신포차 강서구청점

store_x: 37.552961

store_y: 126.852461

store_addr: 서울특별시 강서구 화곡6동 1115-26번지

store_addr_new: 서울특별시 강서구 화곡로58길 25

store_tel: 02-2690-1356

open_hours: ('매일', '오후 5시 - 오전 7시') (*형식은 상관없음)

website: <http://www.hanshinpocha.com>

s_link: 315163

g_link:

<https://www.google.com/maps/place/%ED%95%9C%EC%8B%A0%ED%8F%AC%EC%B0%A8+%EA%B0%95%EC%84%9C%EA%B5%AC%EC%B2%AD%EC%A0%90/data=!4m2!3m1!1s0x357c9c115c07f6dd:0xd9f3bfabfc7dadee!10m1!1e1>

d_link: oaJ1TZCQe9qm

n_link: 33684829

2. review.csv

- 컬럼 내용

store_id: storeInfo의 가게 store_id (fk) (int)

portal_id: 포털사이트별 아이디 (int)

(식신: 1001, 구글: 1002, 다이닝코드: 1003, 네이버:1004)

date: 리뷰 작성일자 (string) (0000-00-00)

score: 리뷰 점수 (float)

review: 리뷰 텍스트 (string)

preprocessed_review: 전처리 된 리뷰 텍스트 (string)

+ 예측 점수, 긍/부정 예측 (int)

- 데이터 예시

store_id: 18

portal_id: 1002

date: 0000-00-00

score: 5

review: 역시 백종원계열 맛있음

preprocessed_review: 역시 백종원 계열 맛있음