

네이버 클릭수 읽어온 과정

1. 네이버 쇼핑에서 모든 카테고리 이름과 카테고리의 id를 크롤링으로 읽어온다. —> '날씨관련 코드'폴더 안의 파일들 참고....ㅈㅈ....
 1. 각 카테고리마다 id와 parent id가 있는데 parent id는 상위 카테고리의 id이다.
 2. 그냥 id만 긁어오면 상위 카테고리 정보를 알 수 없기 때문에 parent_id도 긁어와야한다 —> 'parent 크롤링' 파일 참고
 3. 하지만, 예외경우들이 있다. (예) 내가 중분류를 긁어왔지만 parent id가 대분류가 아닌, 중분류인 경우들이 있다. 즉, 소분류 카테고리가 중분류에서 긁혀옴
 4. 그런 것들은 소분류로 가야하는 것들이기때문에 이런것들 소분류로 분류해둔다 —> 'parent 정제'파일 참고
2. 긁어온 (카테고리 이름, id)를 활용하여 네이버 api에 전체연령, 전체기기, 전체 날짜를 기준으로 클릭수를 얻어온다. —> 'api' 파일 참고
3. 네이버 카테고리명 정제
 1. 네이버 카테고리는 카테고리명이 '점퍼'로 쓰여있고 중분류 이름을 봐야 여성 점퍼인지 아동점퍼인지 등 판단 가능하게 돼있다.
 2. 따라서, 중분류가 여성의류면, 여성을 따서 소분류명에 다 '여성'을 붙여준다.—> '네이버카테고리정제중'파일 참고
4. 대부분 수동으로 lpoint와 naver 카테고리 명을 붙였지만, 몇몇은 처리하기 어려운 경우들이 있어서 자동화함
 1. 수동으로 붙인것은 —>data_ctg2.xlsx 파일 참고
 2. 네이버 검색창에 그 lpoint소분류 이름 자체를 검색했을 때 나오는 첫번째 상품의 카테고리 이름을 네이버 카테고리이름과 매칭시킴
 3. 예를들어, lpoint 소분류에 '여성일반스포츠바지'가 있다면 네이버쇼핑에 '여성일반스포츠바지'를 검색어로 입력한 후 나오는 상품의 네이버 소분류 카테고리를 크롤링으로 읽어왔다. —> '네이버 읽기'파일 참고