**데이터 전처리 과정 – {사용언어: 파이썬, 사용라이브러리: Pandas}**

1. 데이터프레임 포맷 정비
   1. CSV를 불러오는 과정에서 ‘Product Name’ 컬럼의 길이가 긴 경우, 뒷 부분이 잘리면서 다음 열로 한 칸 씩 밀리는 문제 발생 (인코딩 과정에서 발생한 문제로 추측됨)
   2. 반복문을 활용하여 잘린 문자열을 동일 행 ‘Product Name’ 컬럼 뒤쪽에 합쳐주고, 밀린 데이터들을 원상 복구 시키는 작업을 통하여 해결
2. 컬럼명에 공백이나 ‘-‘ 이 포함된 문제
   1. 컬럼명에 공백이나 ‘-‘ 이 포함되어 MySQL에서 작업 시 어려움이 있을 것으로 예상
   2. 공백과 ‘-‘를 ‘\_’로 치환하여 해결
3. ‘Product Name’ 컬럼 문제
   1. ‘Product Name’ 컬럼의 데이터들에 인코딩 과정에서 깨진 문자열이 발생하는 문제를 비롯하여 여러 문제 발생
   2. 파이썬 정규표현식 라이브러리 (RE)를 사용하여 알파벳과 숫자를 제외하고 모두 빼는 작업 시도 -> 문제 지속해서 발생
   3. ‘Product Name’으로 Grouping 진행 시 항목당 수가 지나치게 적어 분석 시 사용하지 않을 것 같다는 결론을 내림. -> 컬럼 Drop
4. Date Format 문제
   1. Date Format이 ‘%d-%m-%Y’ 형태로 이루어져 있어 분석 진행 시 어려움 발생
   2. Date Format을 ‘%Y-%m-%d’ 형태로 변환하여 문제 해결