

1. 데이터 셋

선물용 소품을 중심으로 다양한 소품류를 판매하는 영국의 온라인 쇼핑몰 업체로, 주요 고객은 도매업자(wholesalers)임

※ Reference: <https://www.kaggle.com/carrie1/ecommerce-data>

2. 데이터 명세

컬럼명	정의	데이터 타입	값 예시 및 분포	비고 (전처리 규칙)
InvoiceNo	거래 번호	String	572344, 548813, C581490...	C가 앞에 붙은 거래번호 존재 (거래 취소건)
StockCode	물건 번호	String	85123A, 71053, POST, B...	5글자 미만인 데이터 제거 (POST, D, C2, DOT, M, S, m ...)
Description	물건에 대한 설명	String	WHITE HANGING HEART T-LIGHT	물건의 유형을 패턴으로 확인할 수 있음 (BOX OF, WHITE, HAND WARMER ...) StockCode에 의존적
Quantity	구매 개수	Integer	1, 2, 3, 4 ... 최솟값: -80995, 평균: 9.55, 중앙값: 3.0, 최댓값: 80995	최대 개수: 80995 (이상치로 판단됨, 제거 필요) 0 이하인 개수 : 10624 (결측치로 판단, 제거 필요)
InvoiceDate	구매 시간	TimeStamp	2010-12-01 08:26:00 UTC 최솟값: 2010-12-1, 최댓값: 2011-12-9	
UnitPrice	개별 가격 (달러)	Float	2.55, 3.39 ... 최솟값: -11062, 평균: 4.61, 중앙값: 2.08, 최댓값: 38970	0 이하인 경우 데이터셋에서 삭제
CustomerID	구매자 번호	Integer	17850, 13047 ...	Null 값 존재함 (비로그인, 게스트 구매자로 추정됨) - 전체 거래 건의 25%
Country	구매한 국가	String	United Kingdom	