

데이터 수집

수집 데이터: 이미지, 작품명, 작가명, 제작시기, 가격(판매 희망가격)

우리 모델에서 사용할 설명변수 외에 가격에 영향을 미칠만한 다른 변수들의 영향력을 줄이기 위하여 수집 대상 데이터의 범위를 한정함

- 고정 변수: 회화, 고유 작품, 중간 크기(40x100cm)
- 가격(y값) 범위: \$100 – \$2.5만
- 고려사항: 균등한 y값 분포로 수집하기 위하여, 가격대별 크롤링
- 제약사항: 웹사이트 검색 필터 1개 당 최대 3천 건 크롤링 한계