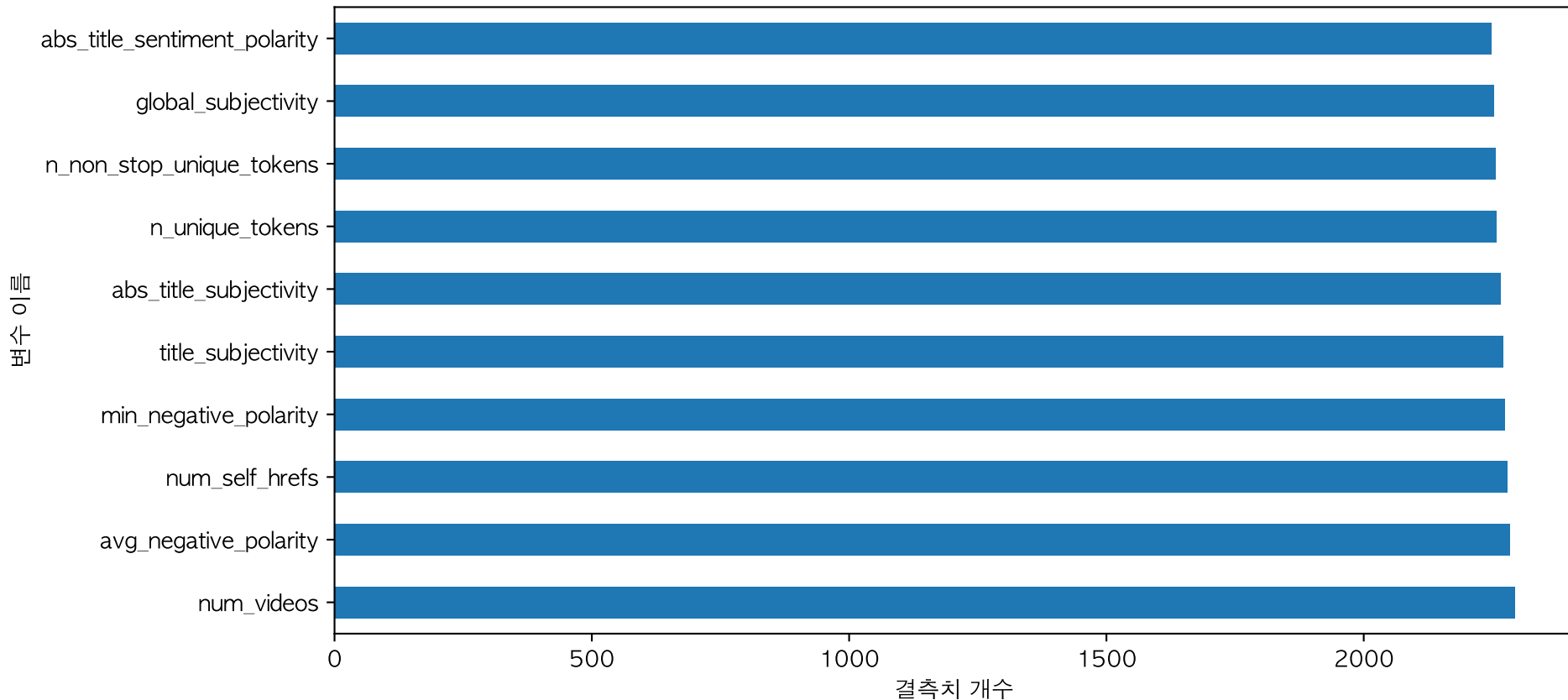
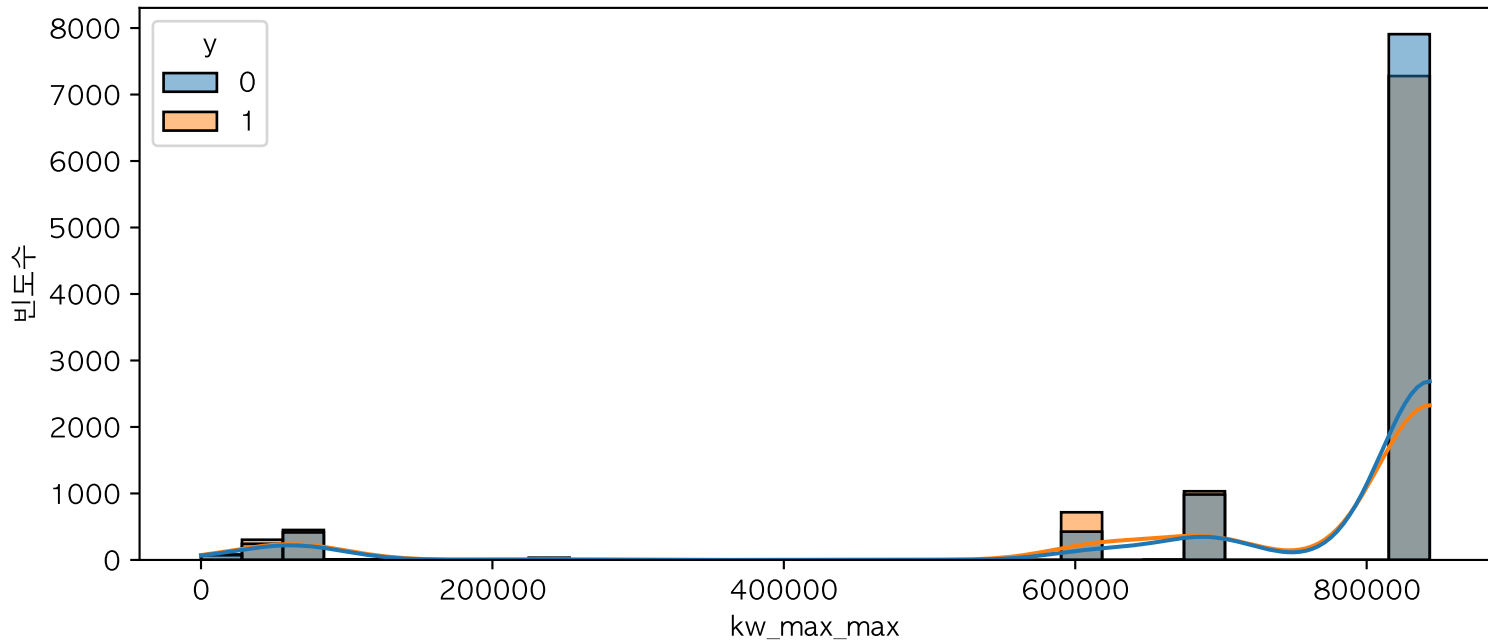


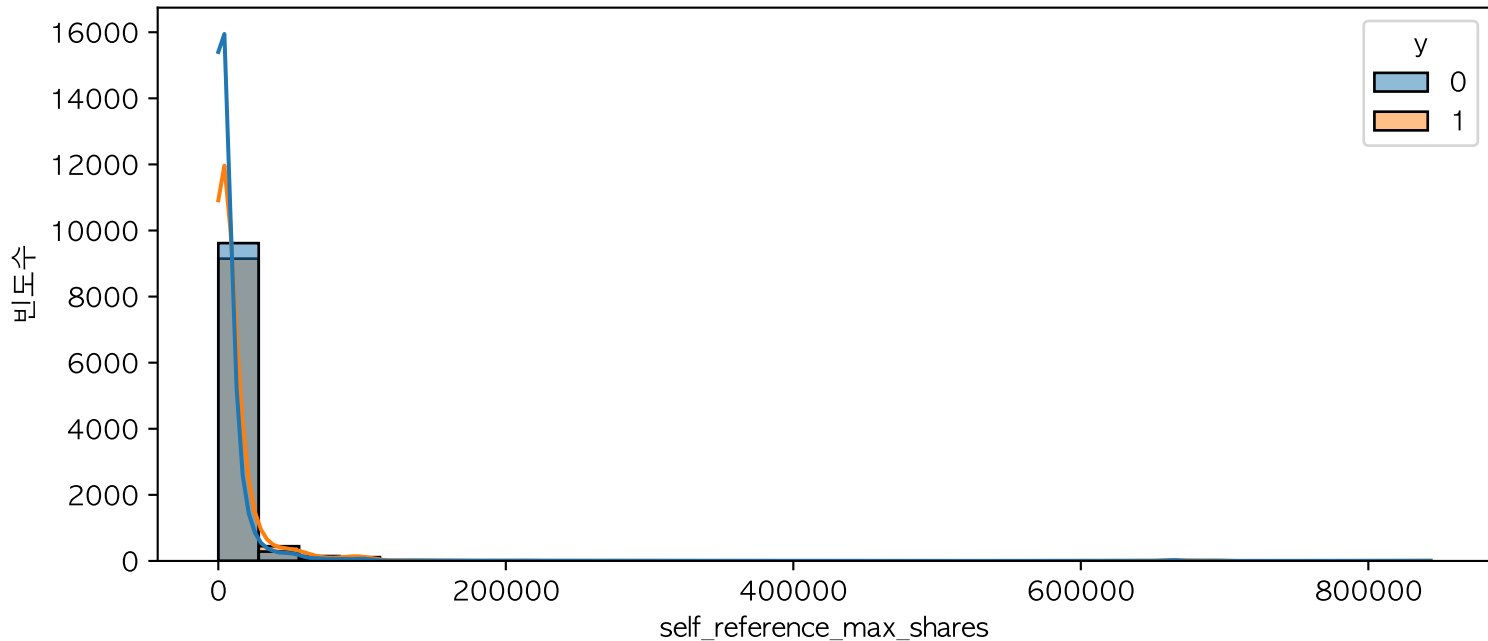
결측치 많은 변수 TOP 10
→ 결측치를 우선적으로 처리해야 할 주요 변수 파악



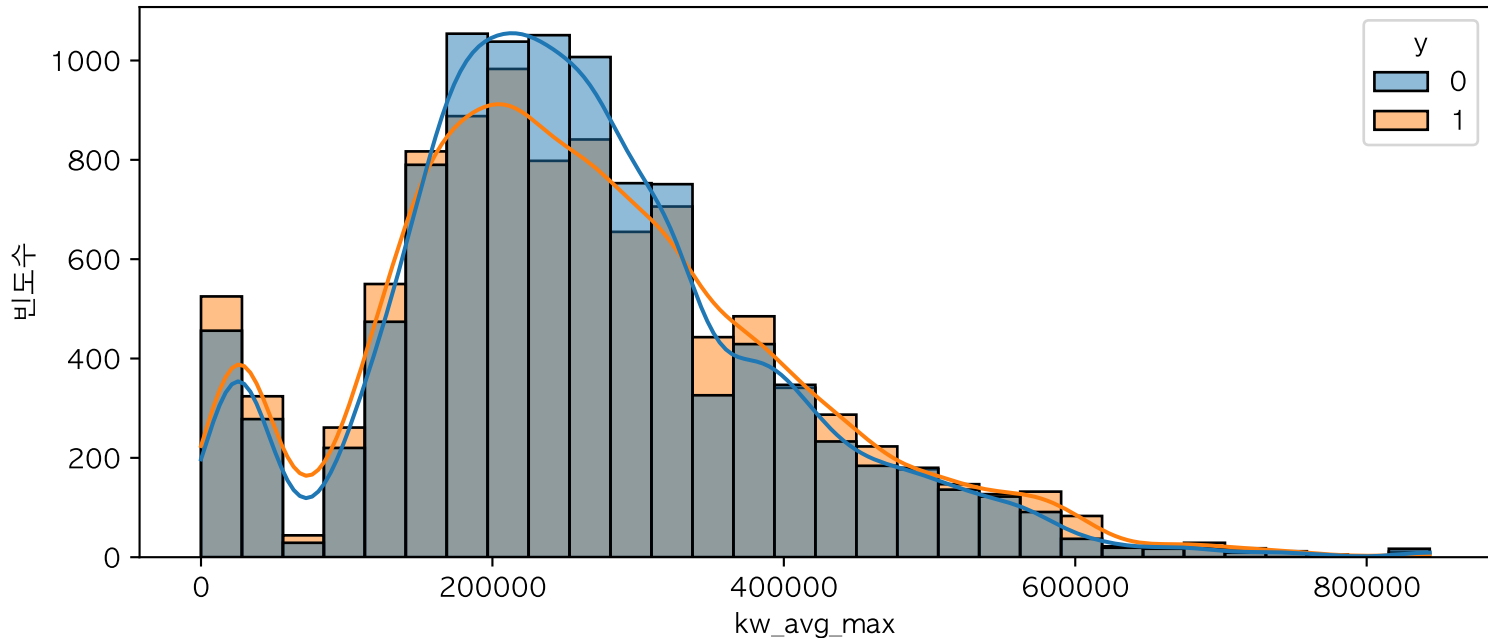
kw_max_max 분포 (y=0 vs y=1)
→ y=0일 때 키워드 최대값이 더 큼



self_reference_max_shares 분포 (y=0 vs y=1)
→ 인기 기사일수록 자기 참조 수치가 높게 나타남



kw_avg_max 분포 (y=0 vs y=1)
→ 평균 최대 키워드 값은 y=1에서 살짝 높음



수치형 변수 간 상관관계 Heatmap
→ 변수 간 강한 상관성을 시각적으로 파악

