# 뉴스 데이터 분석 보고서 (2025-08-24)

이 문서는 2025-08-14부터 2025-08-21까지의 뉴스 데이터에 대한 분석 결과를 요약합니다. 데이터 원본과 분석 코드는 워크스페이스의 파일을 기반으로 하며, 주요 시각화 결과를 포함합니다.

## 데이터 개요

• 원본 데이터: NewsResult\_20250814-20250821.xlsx  
• 기간: 2025-08-14 ~ 2025-08-21  
• 주요 컬럼: 일자, 언론사, 제목, 키워드, 본문 등

## 분석 방법

분석은 Jupyter 노트북(main.ipynb)에서 수행되었으며, pandas로 데이터 로딩/전처리, konlpy·nltk로 텍스트 처리, wordcloud·matplotlib·seaborn으로 시각화를 진행했습니다.

## 핵심 인사이트

• 기간 내 특정 키워드의 빈도 차이가 뚜렷하게 나타났습니다.  
• 워드클라우드에서 상위 핵심 토큰이 강조되며, 해당 이슈 중심의 보도 경향이 관찰됩니다.  
• 추가로 일자별/언론사별 분포를 교차 분석하면 트렌드 파악에 도움이 됩니다.

## 산출물 요약

|  |  |  |
| --- | --- | --- |
| **파일** | **설명** | **경로** |
| NewsResult\_20250814-20250821.xlsx | 원본 수집 데이터 (8/14~8/21) | /home/an0jin/Documents/GitHub/news\_data\_analytics/NewsResult\_20250814-20250821.xlsx |
| barchart.png | 키워드/뉴스 분포 바차트 | /home/an0jin/Documents/GitHub/news\_data\_analytics/barchart.png |
| wordcloud.png | 상위 토큰 워드클라우드 | /home/an0jin/Documents/GitHub/news\_data\_analytics/wordcloud.png |
| main.ipynb | 분석 노트북 (코드 및 시각화) | /home/an0jin/Documents/GitHub/news\_data\_analytics/main.ipynb |

## 재현성(Reproducibility)

• 환경: Python 3.12  
• 주요 패키지: pandas, konlpy, wordcloud, nltk, seaborn, koreanize\_matplotlib, matplotlib, openpyxl  
• 재현 절차:  
 1) 워크스페이스의 main.ipynb 실행  
 2) 엑셀 데이터(NewsResult\_20250814-20250821.xlsx) 로딩  
 3) 전처리 및 시각화 셀 순차 실행하여 barchart.png, wordcloud.png 생성  
 4) 본 보고서 스크립트로 문서 생성

## 그래프 해설

• 상위 키워드는 ‘지원’(최다), ‘개설’, ‘사업’, ‘포토’, ‘기업’ 순으로 나타났습니다. ‘지원’과 ‘개설’이 높은 빈도를 보이며 정책‧제도 안내 및 신규 프로그램 출시에 관한 보도가 많음을 시사합니다.  
• 중간권에는 ‘지역’, ‘한국’, ‘교육’이 위치하여 지역 이슈와 교육·국가 단위 주제가 넓게 다뤄졌음을 보여줍니다.  
• 하위권(상위 10위 내)에는 ‘안전’, ‘산업’이 포함되어 안전 관련 보도가 꾸준히 존재하지만, 정책·사업 보도 대비 상대적 비중은 낮습니다.  
• 전반적으로 행정·정책(지원, 개설) → 실행(사업) → 대상·영역(기업, 지역, 교육) → 가치(안전, 산업) 순의 서사가 관찰됩니다.