

빅데이터 프로젝트 보고서

1. 데이터 수집 / 크롤링

- 국가동물보호정보시스템 Open API를 이용해 전국 유기·입양 동물 데이터를 월·지역 단위로 자동 수집
- 전국동물보호센터 정보표준데이터(Open API)를 통해 보호소 주소·연락처·위치(위도·경도) 정보 수집
- 농림축산검역본부 동물등록 현황 Open API를 이용해 연도·지역·출생연도별 등록두수 데이터 수집
- API 오류 대비 재시도 로직, 페이징 처리, 월 단위 청크 분할 등 안정적인 대용량 수집 구조 구현

initCollections.js

2. 데이터 저장 / 추출

- 수집한 데이터를 MongoDB에 abandonments / shelters / registrations 컬렉션으로 구분 저장
- Open API 응답을 종·품종·성별·중성화·지역·좌표 등이 표준화된 스키마로 정규화하여 저장
- unique 인덱스와 2dsphere 공간 인덱스를 기반으로 중복 없는 빠른 조회 구조 구성

upsert_abandonments_korea.js

upsert_registrations_korea.js

upsert_shelters_korea.js

3. 데이터 가공 / 정제

- Pandas를 활용해 날짜형식 변환, 종·품종 분리, 성별·중성화 매핑, 지역 정보(sido/sigungu) 정규화 수행
- 몸무게, 나이, 출생연도 등 수치형 변수 파싱
- 계절(season), 요일(weekday), 월(month) 등 파생 변수 생성
- 보호소 정보와 유기동물 데이터를 보호소명 기준으로 조인하여 공간 데이터 확장

clean_animals.py

4. 데이터 분석

- 연도·월·계절별 유기동물 발생 추이 분석(Time Series)
- 시도·시군구별 유기·등록 패턴 비교(Spatial Analysis)
- 성별·중성화 여부·종·품종·나이 등 다변수 상관 분석 수행
- RandomForest 기반 '처리결과(processState)' 예측 모델 구축 및 성능 평가

analysis_model.py

5. 데이터 시각화

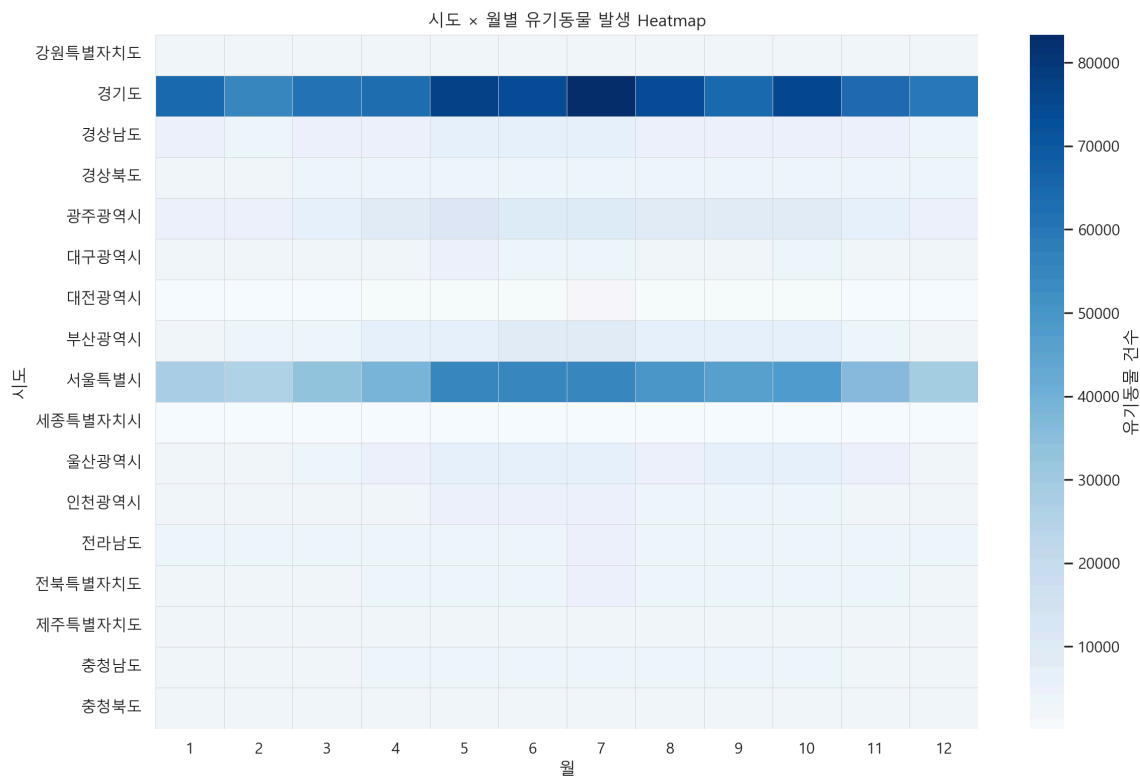
- Matplotlib-Seaborn 기반 시각화(선그래프, 막대그래프, 파이차트, Heatmap 등)
- folium 기반 전국 보호소 지도(HTML) 생성
- 등록 대비 유기 비율(1000마리당 유기) 비교 그래프 제작
- Tableau용 집계 데이터셋 생성(시·도×월×처리결과, 출생연도별 등록현황 등)

[visualization.py](#)

6. 분석 결과

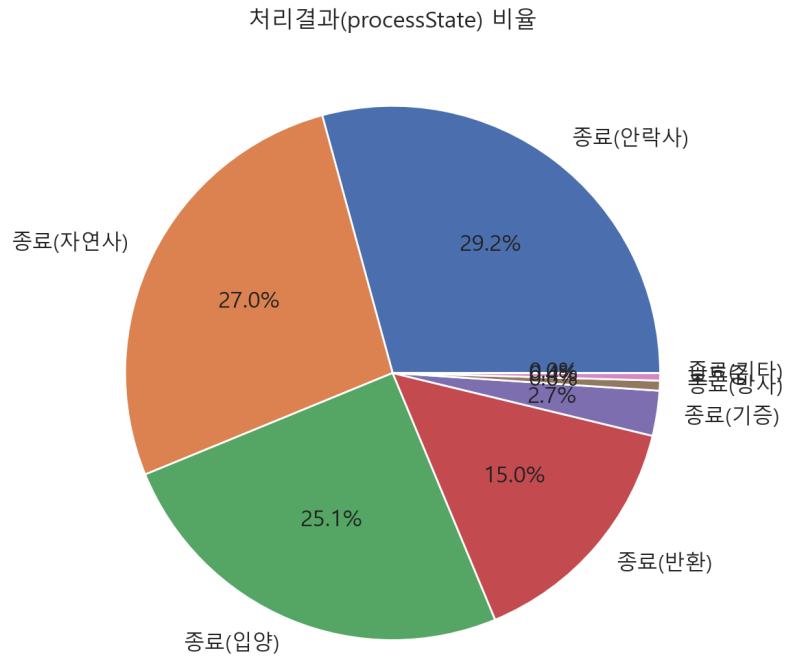
본 프로젝트에서는 전국 단위 유기동물 데이터, 동물등록 현황, 보호소 데이터를 통합하여 시계열·지역·종별·처리결과·등록 대비 유기 비율·보호소 분포 등 다양한 분석을 수행하였다. 아래는 모든 시각화(13개 이미지 + 1개 HTML 지도)를 반영한 종합 분석 결과이다.

1) 시도 × 월별 유기동물 발생 Heatmap



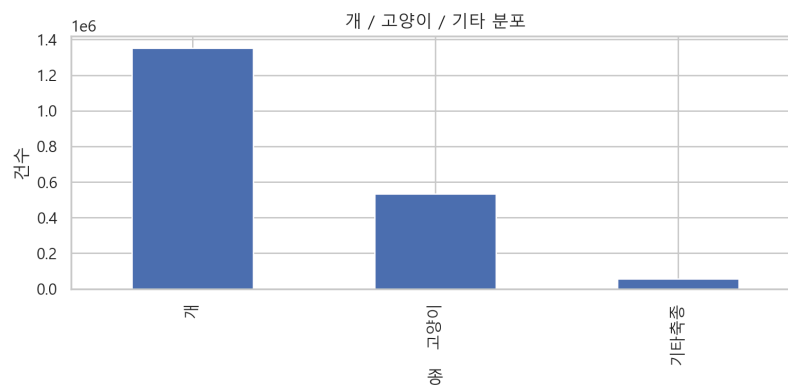
- 경기도와 서울특별시가 유기 건수에서 압도적인 비중을 보인다.
- 5~8월 사이 급증하는 계절 패턴이 뚜렷하게 나타난다.
- 광역시·도 전체가 유사한 월별 패턴을 보이나 규모 차이는 매우 크다.
- 여름철 유기 증가 원인: 이동량 증가, 여행, 외부 노출 증가, 번식기 영향 등.

2) 처리결과(processState) 비율 분석



- 안락사(29.2%) + 자연사(27.0%) = 전체의 절반 이상이 생존하지 못하는 구조
- 입양은 25.1% 수준으로 매우 낮다.
- 반환(15%)은 등록제 실효성 부족을 시사한다.
- "보호 → 입양" 흐름이 약하고 "보호 → 자연사" 비중이 높은 것은 보호 기간 장기화 문제를 반영한다.

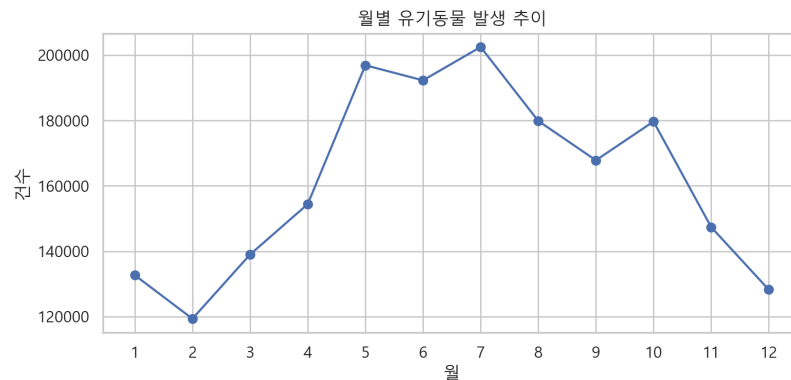
3) 개 / 고양이 / 기타 종 분포



- 개가 압도적으로 많아 약 130만 건 이상

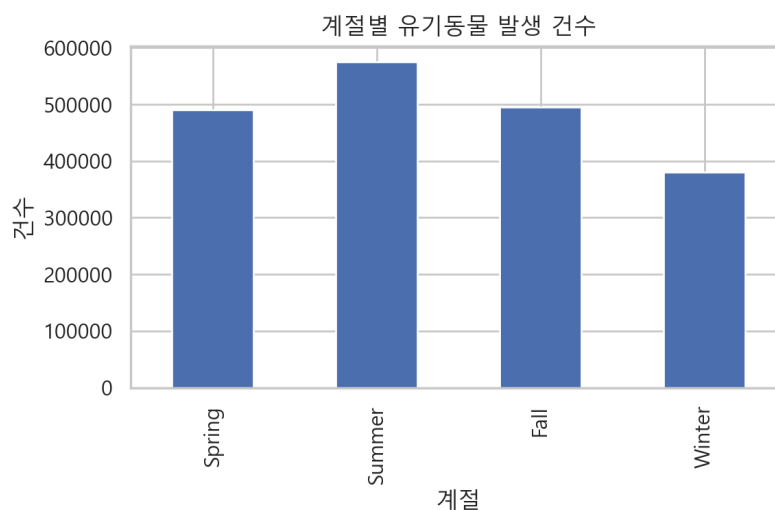
- 고양이는 약 50만 건 수준
- 기타종은 극히 적음
- 유기동물 문제는 **개·고양이에 집중된 구조임**을 확인.

4) 월별 유기동물 발생 추이



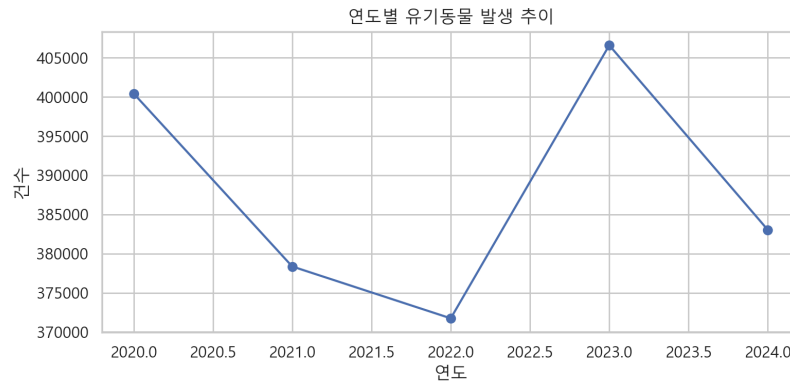
- 1~2월 낮음 → 5월부터 증가 → **7~8월 최고치** → 겨울철 재감소
- 전체적으로 여름철 유기 증가 패턴이 매우 강함
- 이는 Heatmap 패턴과 동일하게 **전국 공통 현상**임을 보여줌.

5) 계절별 유기동물 발생



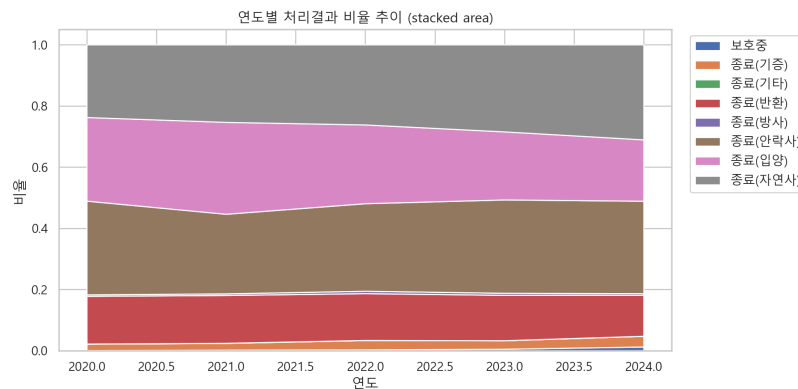
- **여름 > 봄 ≈ 가을 > 겨울** 순
- 여름철 유기 발생은 다른 계절 대비 뚜렷하게 많다.
- 계절성 요인은 향후 예측 모델 설계에서도 중요한 특징임.

6) 연도별 유기동물 발생 추이



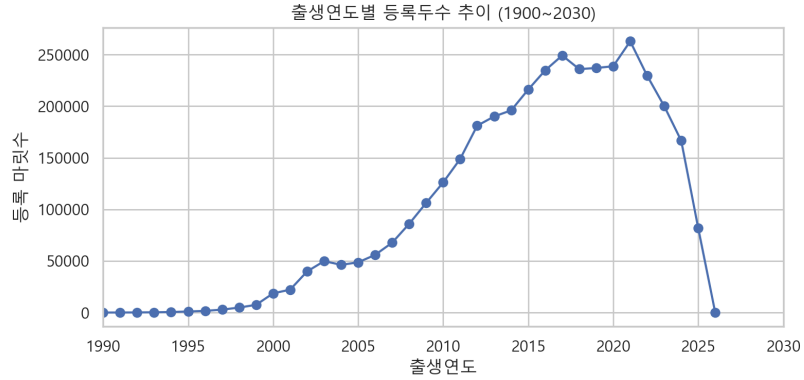
- 2020년 → 2021년 감소 → 2022년 최저점 → **2023년 다시 증가**
- 장기적으로 유기동물 감소세가 확실하다고 보기 어렵다.
- 2024년은 아직 연도 데이터 수집 중이라 소폭 감소한 것처럼 보임.

7) 연도별 처리결과 변화



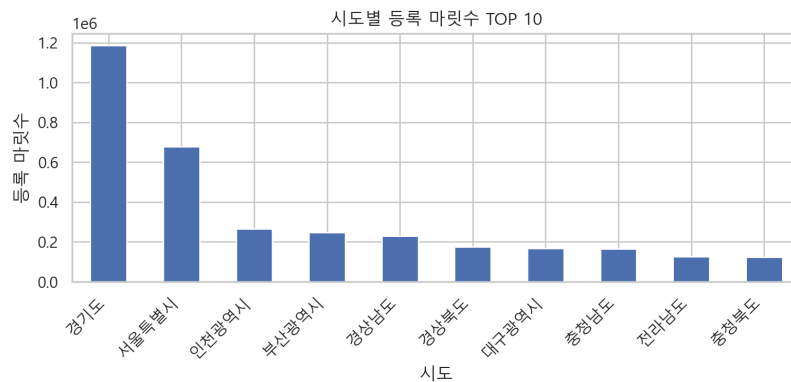
- 입양 비율은 시간이 지나도 큰 증가 없음
- 안락사 비율은 오히려 유지 또는 증가
- 자연사 비율도 높은 수준에서 유지
- 반환율 역시 정체
- 즉, 구조 이후 결과가 개선되지 않은 구조적 문제가 존재한다.

8) 출생연도별 등록두수 추이



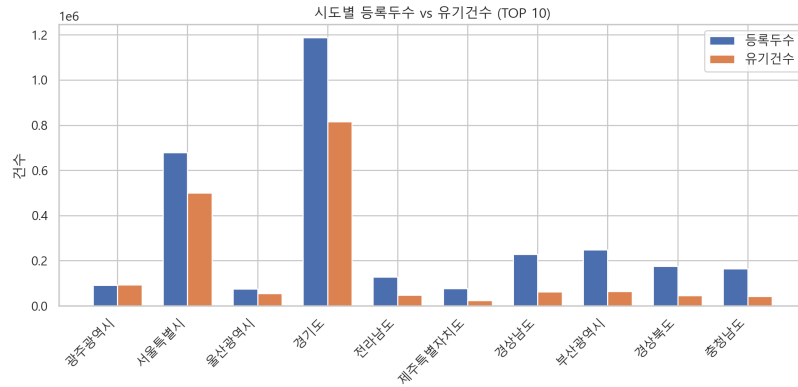
- 1990년대 이후 꾸준히 증가하다 **2010년대 후반~2020년대 초반에 정점**
- 등록제 정착 효과가 매우 크다는 것을 보여주는 지표
- 2024~2026년 감소는 데이터 누락 또는 실제 등록 감소 가능성 모두 존재
- 등록두수의 급증은 반려동물 양육 증가와 직접적으로 연결된다.

9) 시도별 등록 마릿수 TOP 10



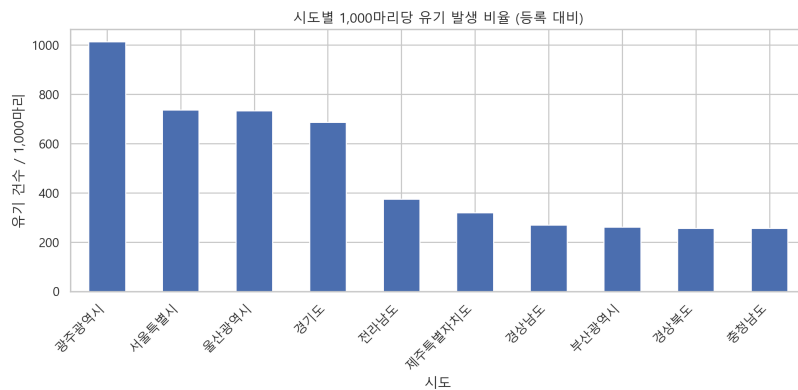
- 경기도 > 서울특별시 > 인천광역시 순으로 등록 개체 수가 많다.
- 인구 및 반려동물 양육 비율과 유사한 패턴을 보임
- 등록은 수도권 중심으로 압도적으로 많다.

10) 시도별 등록두수 vs 유기견수 비교 (TOP 10)



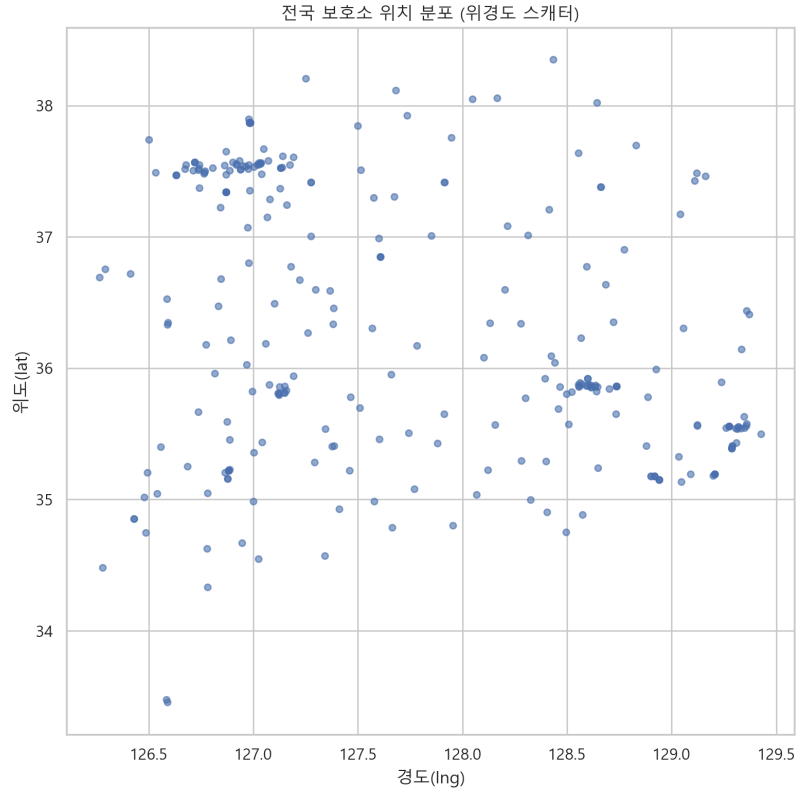
- 등록이 많은 지역이 유기도 많지만, 이는 전체 모수 차이로 인한 현상
- 상대적 유기율을 보려면 등록 대비 유기 비율을 함께 봐야 함
- 서울·경기의 유기 건수 자체는 높지만, **등록 대비 유기율은 중간 수준**
- 반면 전남·충북 등은 등록이 적음에도 유기율은 높게 나타남 → **관리 격차 존재**

11) 시도별 1,000마리당 유기 발생 비율



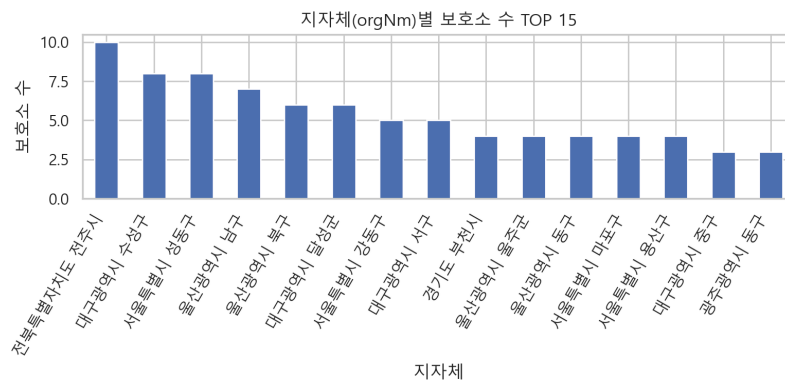
- 광주광역시가 전국 1위로 매우 높은 유기율 기록
- 서울·울산·경기도가 뒤를 잇는다
- 이 지표는 단순 유기 건수가 아닌 **정규화된 유기율이기 때문에 정책 반영 가치가 높음**
- 즉, 등록두수가 적더라도 유기율이 높은 지역은 관리 사각지대임을 의미한다.

12) 보호소 위치 분포 (Scatter Plot)



- 보호소는 전국적으로 존재하지만 **수도권·부산·경남에 밀집**
- 강원·충청·전북 일부는 보호소 밀도가 낮아
 - 구조 요청 대응 속도 저하
 - 보호소 간 이동 거리 증가
 - 지역 격차 문제 발생
- 전국 보호 체계가 균등하지 않다는 중요한 시사점을 제공한다.

13) 지자체별 보호소 수 TOP 15



- 전주시, 대구 수성구·달성군, 수원시, 성동구 등 대도시 중심으로 보호소 수가 많음
- 그러나 보호소 수가 많다고 유기 감소로 이어지지는 않음

- 보호소의 규모·수용능력·운영 효율성이 실제 문제 해결의 핵심임을 시사함.

14) 전국 보호소 지도 (인터랙티브 folium 지도)

[sh_map_shelters.html](#)

- 전국 보호소 위치를 위경도 기반으로 시각화한 HTML 지도
- 사용자는 실제 지도를 확대·축소하며 지역별 분포 확인 가능
- 수도권과 대도시에서 밀집된 보호소 구조가 더욱 명확하게 드러난다
- 반면 강원도·충북·전북 일부 지역은 보호소 접근성이 낮아
 - 구조 지연
 - 보호 능력 부족
 등의 문제가 발생할 수 있다.
- 이 지도는 지역별 신규 보호소 설치, 권역별 구조센터 설계, 자원 배분에 활용 가능한 자료이다.

종합 결론

본 프로젝트 분석 결과 다음과 같은 핵심 인사이트를 확인하였다.

1. 유기는 여름철(5~8월)에 집중되며 전국 공통 계절 패턴이 존재한다.
2. 입양률은 낮고 안락사·자연사 비율이 절반 이상으로 매우 높은 구조적 문제를 보인다.
3. 등록두수 자체보다 '등록 대비 유기율'이 지역 문제를 훨씬 정확하게 반영한다.
4. 광주·서울·울산 등은 유기율이 높아 집중 관리 대상 지역으로 확인된다.
5. 보호소는 수도권·부산·경남에 편중되어 지방의 보호 인프라가 취약하다.
6. 종·성별·중성화 여부·나이 등 다양한 요인이 처리결과 및 유기 패턴에 영향을 준다.
7. 분석 결과는 보호소 배치 재설계, 계절별 집중 단속·관리 정책, 등록제 강화, 입양 캠페인 확대 등 정책적 시사점을 제공한다.