필요 산출물 정리(3rd 1st)

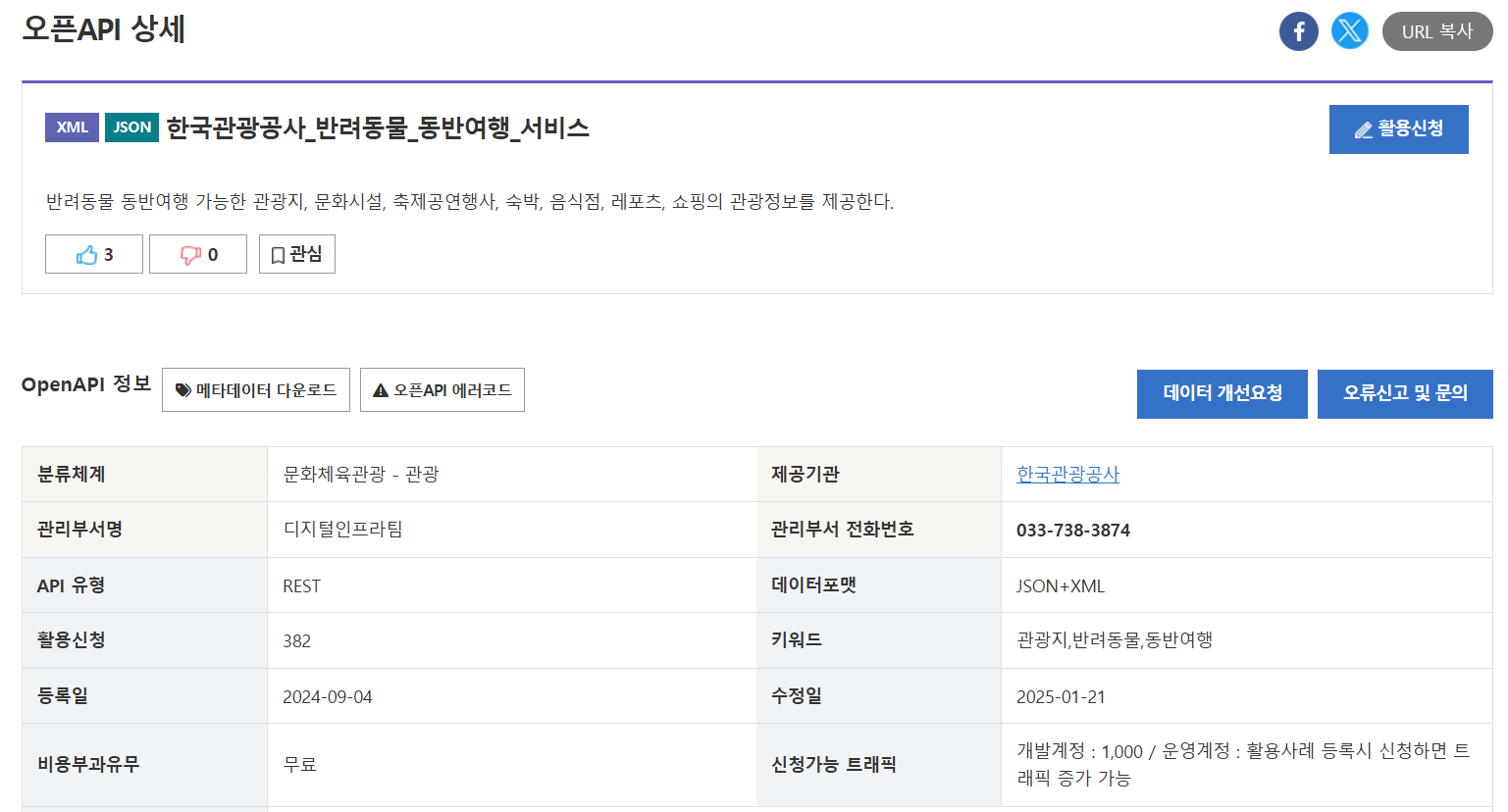
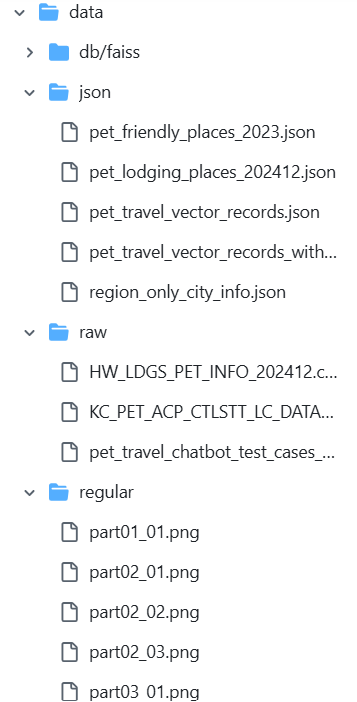
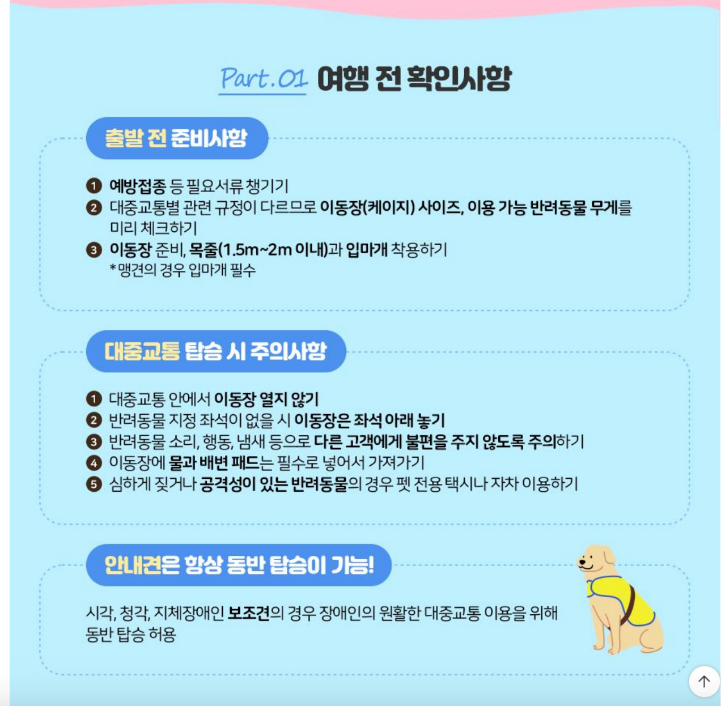
1. 수집된 데이터 및 데이터 전처리 문서

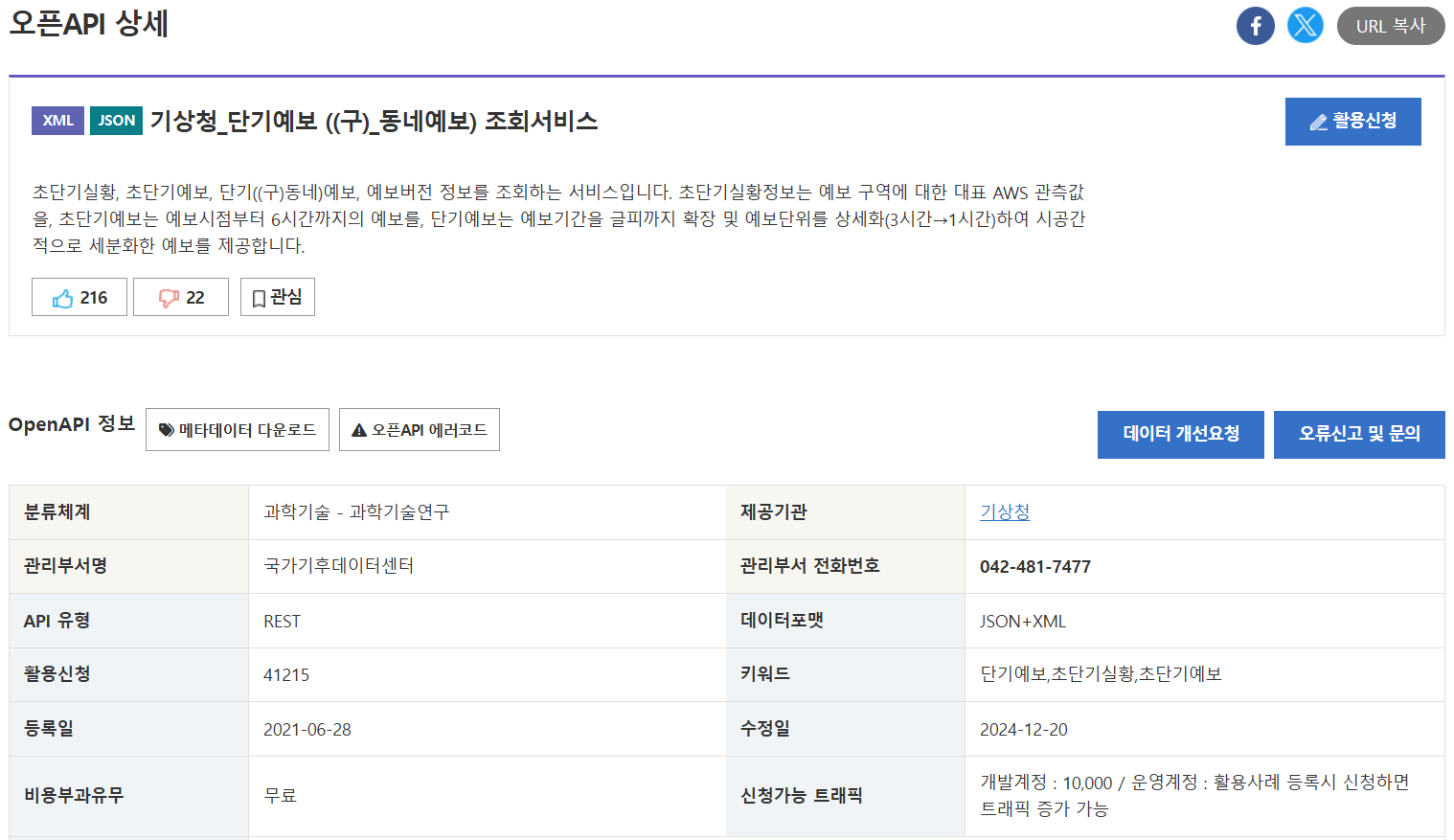
1) 데이터 및 데이터 전처리 문서

- 공공 데이터 포털 반려동물 동반가능 문화시설 위치 데이터

- 국가 배포물 (반려동물 동반 이동시 유의사항) 텍스트 추출

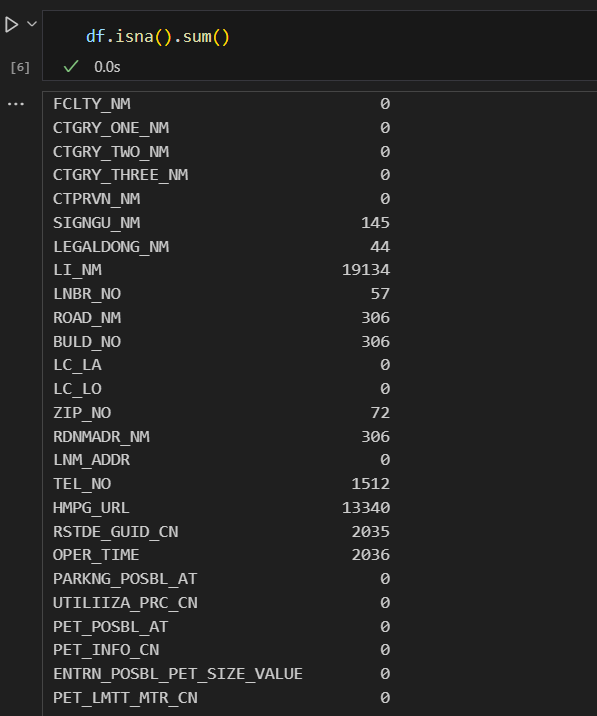
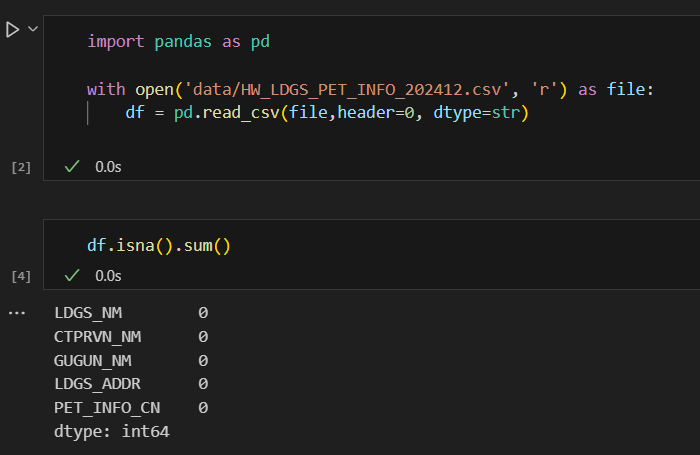
- 공공 API 및 날씨 API 데이터 활용



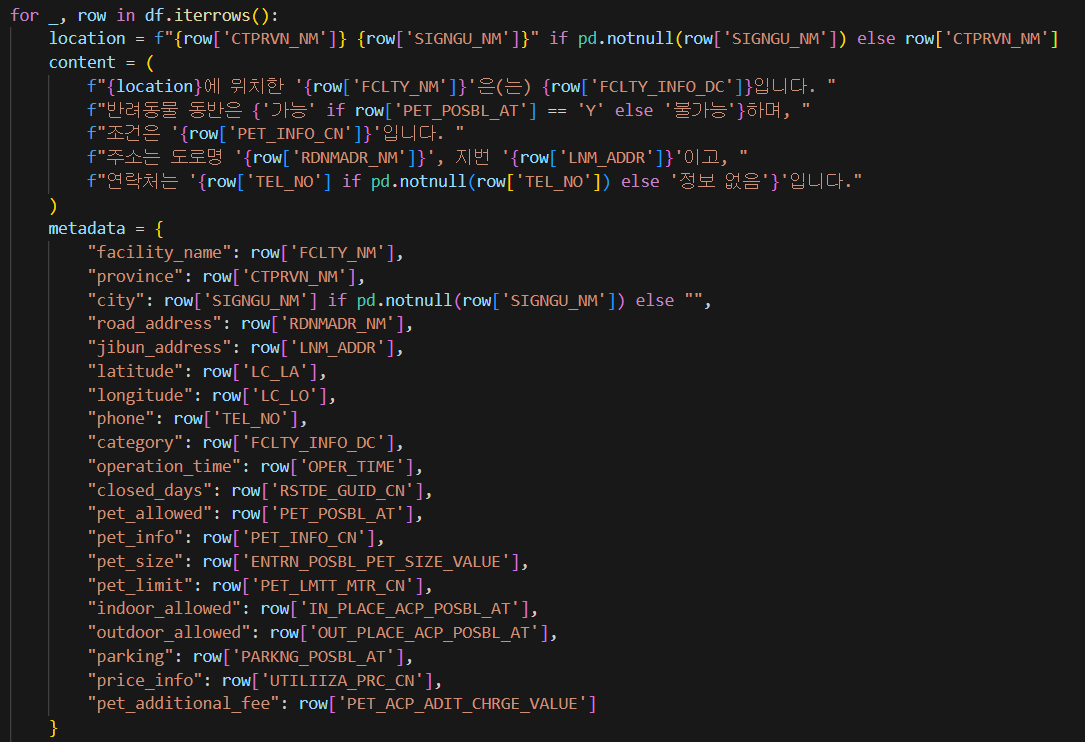


2) 데이터 품질(결측치, 이상치) (완료)

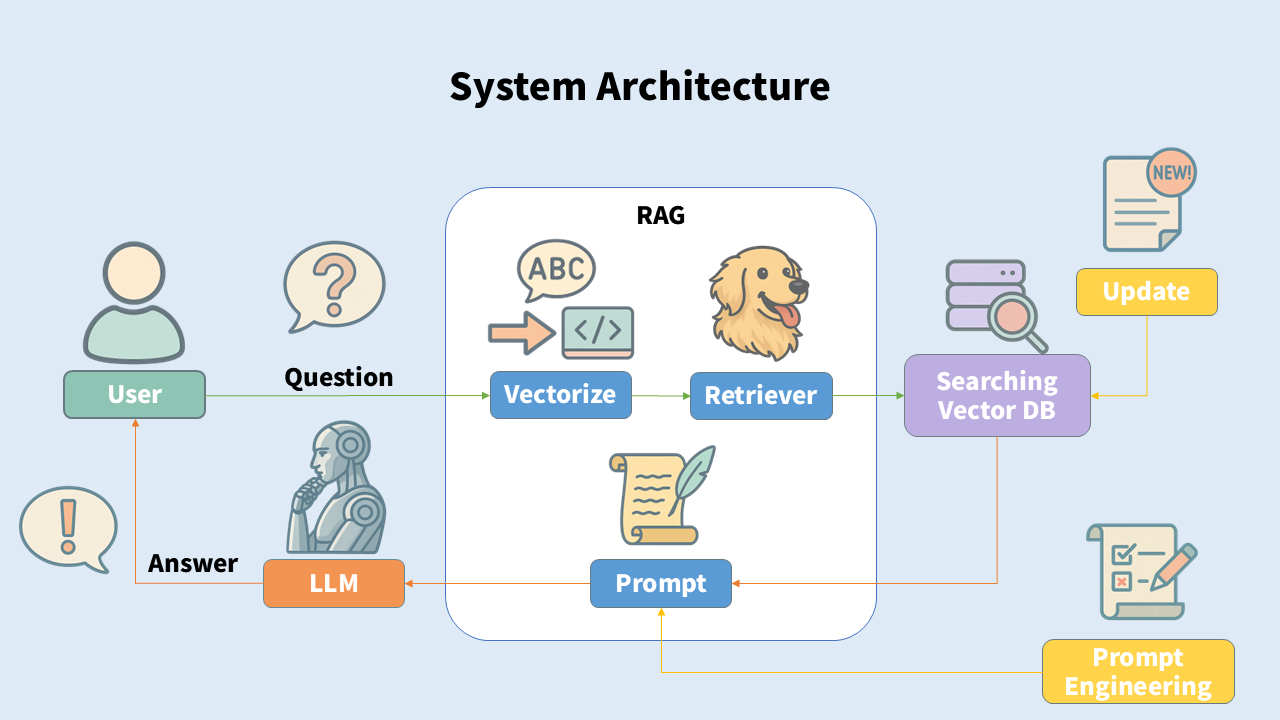
- 공공 데이터 활용

- 결측치 처리 : 주요 정보가 아닌 컬럼은 메타데이터 제외

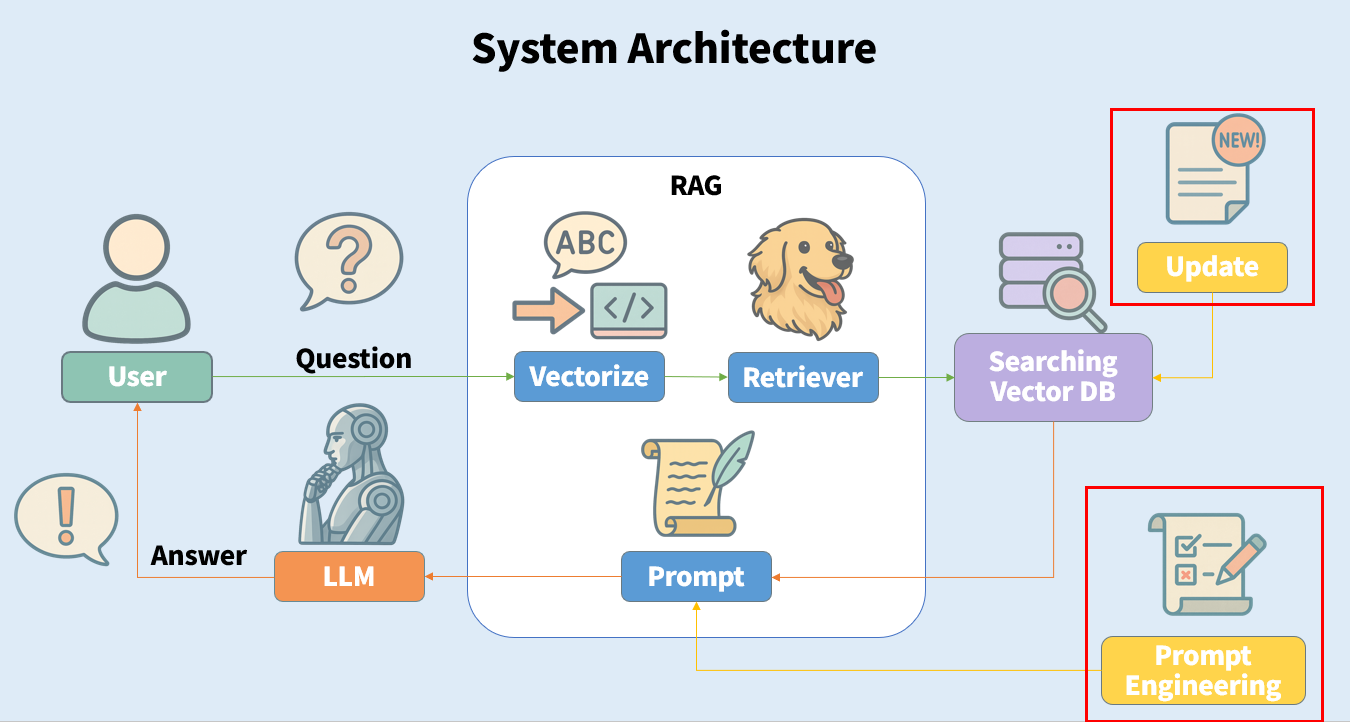
3) 메타데이터 활용 여부

- json 변환 코드 자료 첨부 (1번과 동일)

2. 시스템 아키텍처

 1) RAG와 벡터DB, LLM 연동 구조

2) 확장성 및 유지보수 관점 고려 여부

- 벡터 DB 업데이트 및 프롬프트 엔지니어링으로 패치 / 업데이트 가능

3. 개발된 소프트 웨어

1) 코드 가독성 및 주석 처리 (코드 참조)

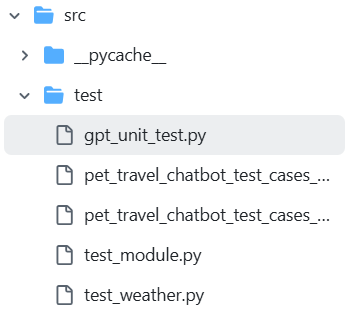
- 개발 후 한번에 정리

- 일률화하여 각 함수 및 코드마다 주석처리 완료



2) 벡터 검색과 LLM 연동 정확도

- 벡터DB 및 API 처리 테스트 코드



3) 에러 처리 및 예외 대응

- 각 함수마다 try, except 구문활용

- 에러 및 데이터 처리 오류 시 오류구문 프린트

 - 즉각 해당 함수를 디버깅 대응 가능하도록 처리함

4. 테스트 계획 및 결과 보고서

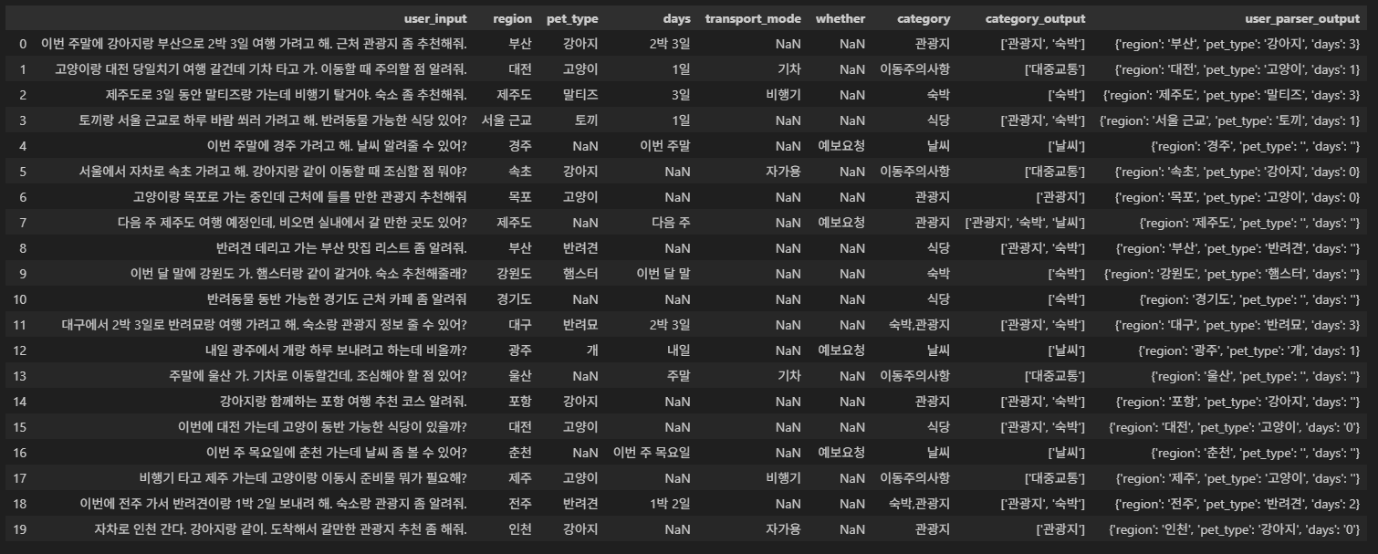
1) 테스트 케이스의 적절성 및 다양성

- 테스트 케이스 20개 제작



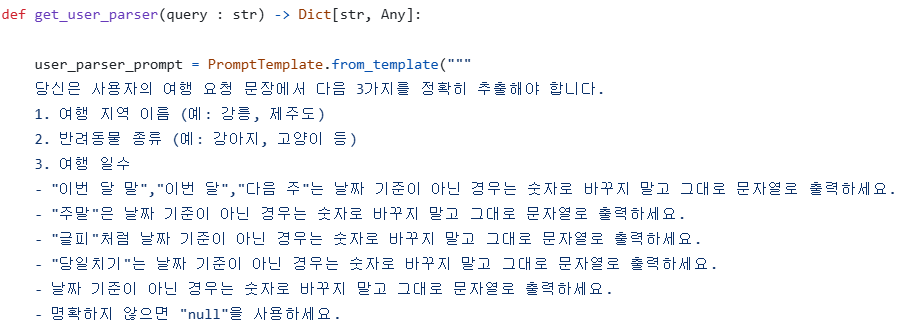
2) 테스트 결과 분석 및 문제점 도출 능력

- 테스트 케이스 실행 후 나온 결과

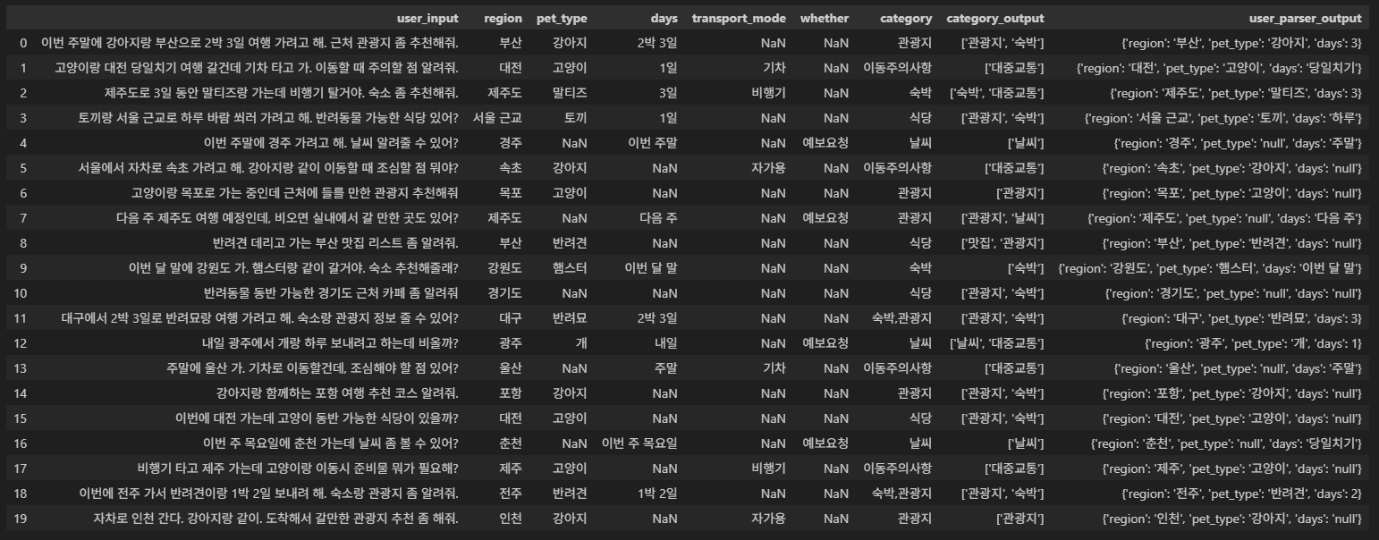


- 특정 언어(당일치기 = 1일)등에 대한 LLM 분석이 부족함을 확인

- 프롬프트 엔지니어링 시도



- 프롬프트 엔지니어링(2차) 후 변환한 결과

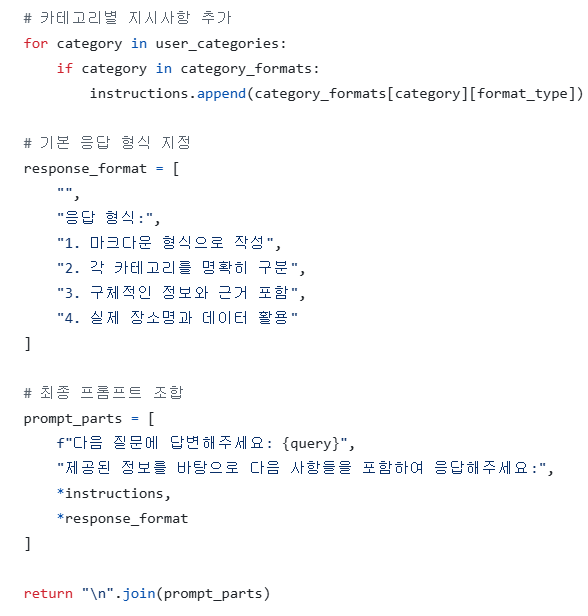


* 카테고리가 없는 경우 NULL로 반환하도록 변경
* 날짜와 일정과 관련된 뭉뚱그려진 언어는 확실한 시기로 반환
* LLM에서 현재 시간과 비교하여 판단하도록 전달

3) 환각 방지 및 성능 개선 시도 포함 여부

- 벡터 DB 자료 부족 시 API 활용

 - 실제 공공데이터 및 현재 날씨 검색 및 반환

 - 프롬프트에 실제 사례 검색 유도 추가

* 실존하지 않는 업체가 결과로 표시되는 문제 발견
* 네이버 API 연결, 결과 중 현존하지 않는 업체는 결과 삭제

