**데이터 수집 문서**

1. 정보 콘텐츠 데이터

| 번호 | 1 |
| --- | --- |
| 데이터 명 | 유투브 견인사이 행동 영상 자막 |
| 저장 파일 명 | 2\_youtube\_doghumanbond\_data.csv |
| 데이터 출처 | <https://www.youtube.com/@%EA%B2%AC%EC%9D%B8%EC%82%AC%EC%9D%B4> |
| 데이터 수량 | 14 |
| 데이터 크기 | 29KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(영상 자막), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |

| 번호 | 2 |
| --- | --- |
| 데이터 명 | 견종 데이터 |
| 저장 파일 명 | 3\_dog\_breeds.csv |
| 데이터 출처 | <https://www.royalcanin.com/kr/dogs/breeds> |
| 데이터 수량 | 335 |
| 데이터 크기 | 537KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | index(순서), breed\_name(품종 이름), image\_url(견종별 사진), short\_description(간단 설명), full\_description(전체 설명) |
| 데이터 포맷 | csv |
| 비고 |  |

| 번호 | 3 |
| --- | --- |
| 데이터 명 | 블로그 포스팅 데이터 |
| 저장 파일 명 | 4\_blog\_posts.csv |
| 데이터 출처 | <https://m.blog.naver.com/vet6390/221813865545> |
| 데이터 수량 | 440 |
| 데이터 크기 | 1,346KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(글 내용), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |

| 번호 | 4 |
| --- | --- |
| 데이터 명 | 논문 데이터 |
| 저장 파일 명 | 7\_dog\_paper.csv |
| 데이터 출처 | <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002767552> |
| 데이터 수량 | 15 |
| 데이터 크기 | 33KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | id(페이지 번호), text(본문 내용) |
| 데이터 포맷 | csv |
| 비고 | 추천 콘텐츠에 쓰이지 않음 |

| 번호 | 5 |
| --- | --- |
| 데이터 명 | 하이닥 반려견 질환 칼럼 데이터 |
| 저장 파일 명 | 8\_hidoc\_articles.csv |
| 데이터 출처 | <https://news.hidoc.co.kr/news/articleList.html?sc_area=A&view_type=sm&sc_word=%EB%B0%98%EB%A0%A4%EA%B2%AC> |
| 데이터 수량 | 133 |
| 데이터 크기 | 542KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(글내용), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |

| 번호 | 6 |
| --- | --- |
| 데이터 명 | 핏펫 블로그의 반려견 관련 칼럼 데이터 |
| 저장 파일 명 | 10\_fitpet\_data.csv |
| 데이터 출처 | <https://www.fitpetmall.com/blog/category/dogs> |
| 데이터 수량 | 195 |
| 데이터 크기 | 706KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(내용), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |

| 번호 | 7 |
| --- | --- |
| 데이터 명 | 비마이펫 콘텐츠 데이터 |
| 저장 파일 명 | 11\_doglab\_all\_articles\_fix.csv |
| 데이터 출처 | <https://bemypet.kr/content/doglab?p=1> |
| 데이터 수량 | 1270 |
| 데이터 크기 | 8,376KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(내용), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |

1. QnA 데이터

| 번호 | 8 |
| --- | --- |
| 데이터 명 | 와요 질의응답 데이터 |
| 저장 파일 명 | 6\_wayo\_qna\_data.json |
| 데이터 출처 | <https://wayopet.com/trainer/qna> |
| 데이터 수량 | 2378 |
| 데이터 크기 | 9,364KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | url, 질문 제목, 증상과 행동, 시작된 시점, 보호자님 반응, 원인 분석, 솔루션 제안 |
| 데이터 포맷 | json |
| 비고 | 목적 : 파인튜닝용 데이터로 사용 |

| 번호 | 9 |
| --- | --- |
| 데이터 명 | qna데이터 |
| 저장 파일 명 | 9\_pet\_qna.csv |
| 데이터 출처 | <https://ghomeshop.com/board/%EB%B0%98%EB%A0%A4%EA%B2%AC-%EA%B5%90%EC%9C%A1%EC%83%81%EB%8B%B4/3001/> |
| 데이터 수량 | 78 |
| 데이터 크기 | 345KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | url(주소), question(질문), answer(답변) |
| 데이터 포맷 | csv |
| 비고 |  |

1. 대화 데이터

| 번호 | 10 |
| --- | --- |
| 데이터 명 | 강형욱의 퍼피 교육 유튜브 영상 자막 데이터 |
| 저장 파일 명 | 1\_youtube\_puppy\_dictionary.csv |
| 데이터 출처 | <https://www.youtube.com/playlist?list=PLVh3TM0B0Wtl67fiuSL5xNEc0uZYvEB18>  <https://www.youtube.com/playlist?list=PLVh3TM0B0WtksY4ZQVNdD0aTG1Qm1mQNM> |
| 데이터 수량 | 110 |
| 데이터 크기 | 824KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(영상 자막), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |

| 번호 | 11 |
| --- | --- |
| 데이터 명 | 댕쪽이 상담소 유튜브 영상 자막 데이터 |
| 저장 파일 명 | 5\_youtube\_dogconseling\_data.csv |
| 데이터 출처 | <https://www.youtube.com/@%EB%8C%95%EC%AA%BD%EC%9D%B4%EC%83%81%EB%8B%B4%EC%86%8C> |
| 데이터 수량 | 55 |
| 데이터 크기 | 805KB |
| 데이터 업데이트 주기 | 비정기 (최초 한 번 수집) |
| 컬럼 | title(제목), content(영상 자막), URL(주소), image(대표 이미지) |
| 데이터 포맷 | csv |
| 비고 |  |