

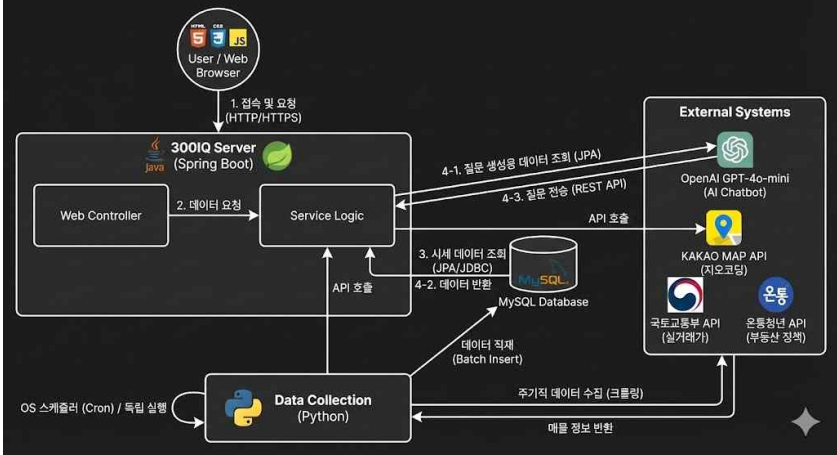
2025년 오픈소스 개발프로젝트 결과보고서

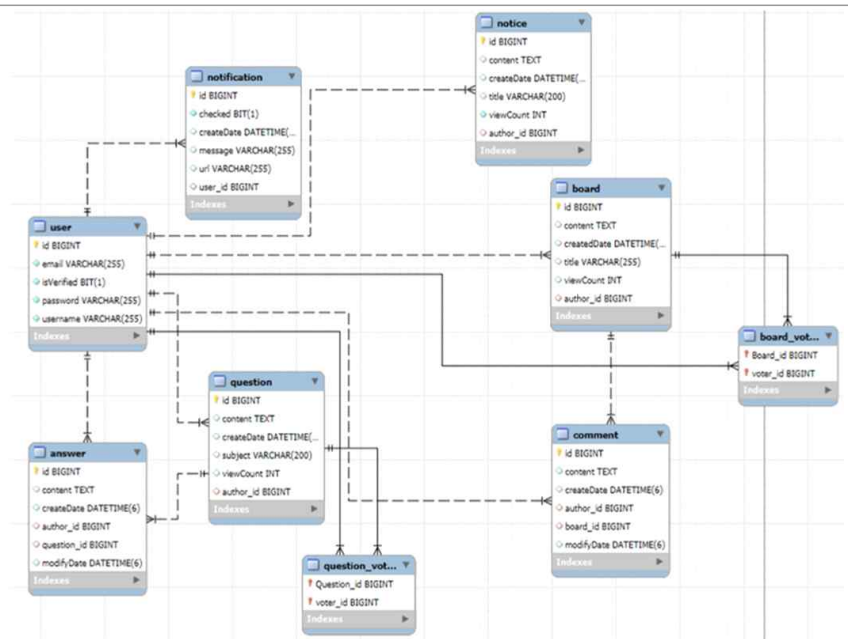
작성일자 : 2025. 12. 11

구 분	세 부 내 용		
팀 번호	7	팀 장(학번)	박성웅(2022041033)
팀 명	300IQ	팀원1(학번)	박정환(2022041038)
		팀원2(학번)	하희찬(2022041041)
		팀원3(학번)	

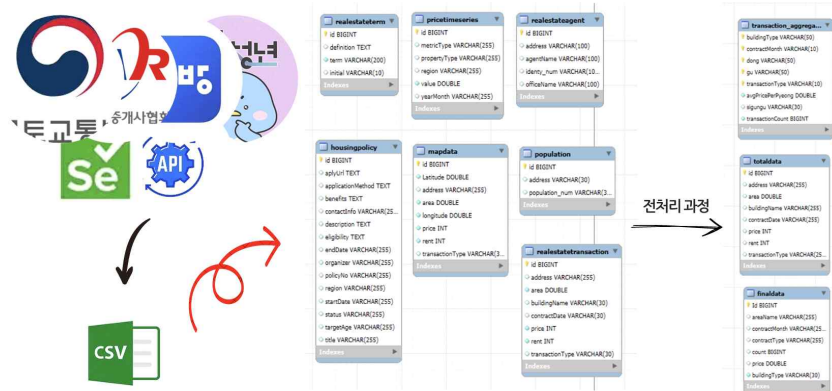
□ 결과보고서

프로젝트 개요	
프로젝트명 (한글)/(영문)	청주 부동산 실거래가 기반 분석 및 AI 어드바이저 플랫폼
GitHub URL	https://github.com/Open-300IQ
시연영상	https://youtu.be/2DmqeSTgyTY?si=H8S4qy/Y91_eRPACK
프로젝트 소개	300IQ의 프로젝트는 청주시(상당구, 서원구, 흥덕구, 청원구)의 최근 3년간의 부동산 실거래가를 시각화하여 제공하며, 사용자가 부동산 정책 및 매물에 대해 빠른 정보를 접하고 자유롭게 소통할 수 있는 커뮤니티 플랫폼으로 설계되었으며, OpenAI의 Gpt 모델을 활용하여 부동산 관련 질의응답을 제공합니다.
프로젝트 세부 내용	
개발 배경 및 목적	최근 '투자'에 관한 관심이 많이 쏠리기 시작하면서, 자연스럽게 부동산도 사람들의 주목의 대상이 되어가고 있습니다. 하지만 데이터포털에서 구할 수 있는 데이터로는 한눈에 비교할 수 있는 시각화된 데이터를 구할 수 없습니다. 이에 저희는 이번 프로젝트를 통해 지난 3년간의 데이터를 수집하고 이를 시각화하여 사람들이 직관적으로 시장의 흐름을 파악할 수 있도록 돕고자 하였습니다. 또한 '청주시'라는 국지적인 지역을 집중 분석함으로써 해당 지역의 보다 정밀한 데이터 분석을 통해 신뢰성 있는 정보를 제공하고자 하였으며, 단순한 정보 제공을 넘어 커뮤니티를 통한 정보 공유와 OpenAI의 GPT모델을 활용한 AI상담 기능을 도입하여 사용자가 복잡한 부동산 정책과 매물 정보를 쉽고 빠르게 습득하고 합리적인 의사결정을 내릴 수 있도록 도와주는 플랫폼을 구축하고자 하였습니다.
관련 논문 및 유사 프로그램	관련 논문 - 빅데이터 기반 생성형 AI를 활용한 부동산 정보제공 시스템 UX 디자인 개

	<p>발 연구 : 부동산 정보 제공 서비스 '오지라파' 사례를 중심으로, https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE12016780</p> <ul style="list-style-type: none"> - 공공데이터를 활용한 전세가격 결정요인 : 서울시 강남지역 아파트를 중심으로, https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE06075852 <p>유사 프로그램</p> <ul style="list-style-type: none"> - 호갱노노, https://hogangnono.com - 아실, https://asil.kr - 부동산지인, https://aptgin.com/root_main - 다방, https://dabangapp.com
<p>개발환경</p>	<p>컴퓨터 언어 : JAVA, HTML, Python 에디터 : Eclipse(JAVA, HTML), Vscodc(Python) 프로젝트 관리 도구 : Discord, GitHub-PullRequest, Trello 데이터베이스 : MySQL API : KAKAO MAP API, OpenAI GPT API 데이터 수집 : Python selenium (다방 부동산 매물)</p>
<p>시스템 구성 및 아키텍처</p>	 <p>The diagram illustrates the system architecture. It starts with a 'User / Web Browser' sending a request (1. 접속 및 요청 (HTTP/HTTPS)) to the '300IQ Server (Spring Boot)'. Inside the server, a 'Web Controller' sends a request (2. 데이터 요청) to 'Service Logic'. The 'Service Logic' interacts with a 'MySQL Database' through 'API 호출' (API calls). It also connects to 'External Systems' including 'OpenAI GPT-4o-mini (AI Chatbot)' and 'KAKAO MAP API (지오코딩)'. The 'Service Logic' sends data (3. 시세 데이터 조회 (JPA/JDBC)) to the database and receives data (4-1. 질문 생성용 데이터 조회 (JPA), 4-2. 데이터 반환) back. The 'Data Collection (Python)' component, running on 'OS 스케줄러 (Cron) / 독립 실행', performs '주기적 데이터 수집 (크롤링)' (periodic data collection/crawling) and '매물 정보 변환' (listing information conversion) to update the 'MySQL Database' via '데이터 직재 (Batch Insert)' (data direct insert/batch insert). The 'External Systems' also provide data to the 'MySQL Database' via 'API 호출' (API calls).</p> <p>- 전체 시스템 흐름 아키텍처</p>



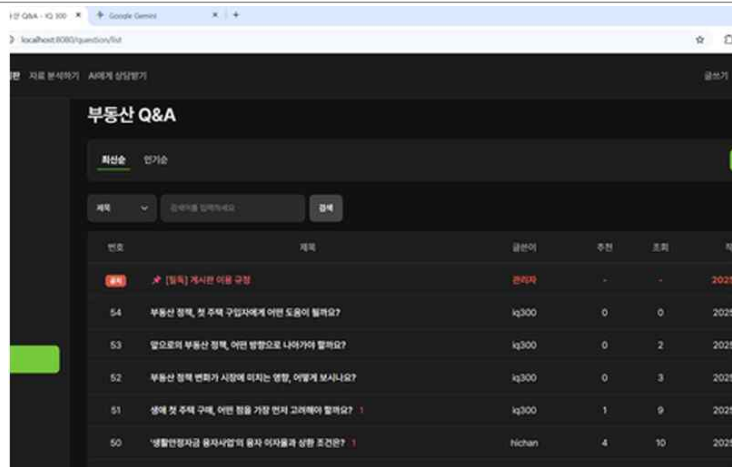
- 게시판 데이터베이스 아키텍처



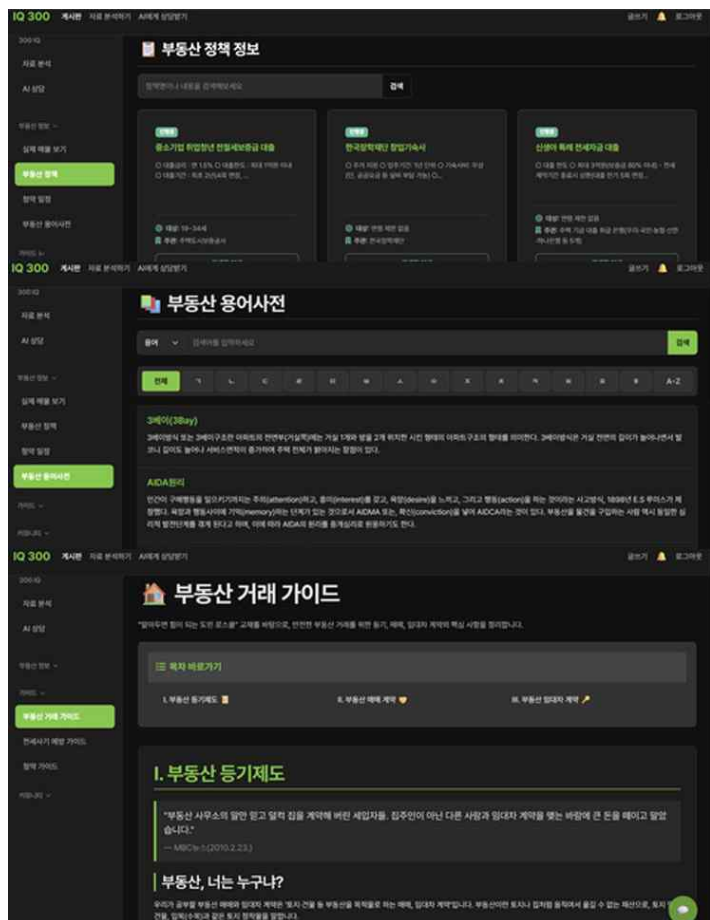
- 그 외 기능 데이터베이스 아키텍처

프로젝트 주요 기능
및 구조도

프로젝트 주요 기능
1. 프론트엔드



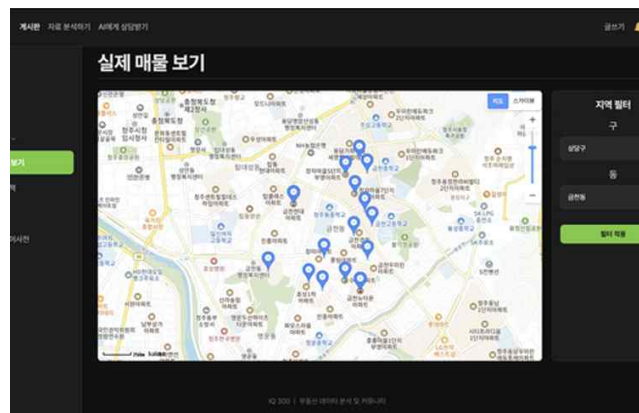
- **메인 화면(게시판 기능)** : 일반적으로 정보를 공유하는 커뮤니티의 성격을 띄고있는 게시판과 질문을 통해 새로운 정보를 얻을 수 있는 Q&A게시판이 있다. (AI 상담은 정해진 프롬프트가 있기에 AI 상담에서 제공할 수 없는 프롬프트 답변일 경우, Q&A가 필수적으로 필요하다고 판단)



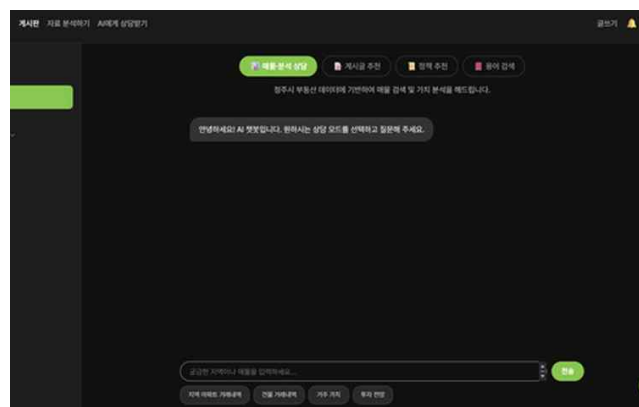
- **각 정보 제공 기능** : 놓치기 쉬운 각 혜택을 제공하는 정책들과 이해하기 힘들 용어들, 거래 시에 조심해야할 핵심 가이드 사항들(예 : 전세 사기 방지 가이드)을 정리하여 제공한다.



- **자료 분석 기능** : 청주시에 국한되는 데이터인 만큼 정밀한 분석을 통해 지역별, 건물 유형별, 거래 유형별로 평당 가격을 나누어 사용자가 직관적으로 해당 데이터를 쉽게 인식할 수 있도록 시각 정보를 제공한다.



- **매물 확인 기능** : 실거래가 데이터가 아닌 현재 시장의 매물을 지도에 마크로 표시하여 사용자가 매물을 한눈에 확인할 수 있도록 시각적 정보를 제공한다.



- **AI 상담 기능** : 4개의 탭에서 각 주제에 맞는 질문을 하고 질문에 대한 답변을 받을 수 있다. 총 7개의 프롬프트로 이루어져 있다.

2. 백엔드 서버

- **메인 화면(게시판 기능)** : 게시판 목록과 각 게시글에 대한 경로를 명시하고 html코드에 있는 get/post 요청을 인식하여 각 경로로의 연결, post 요청에 따른 데이터베이스의 수정 및 추가 등등의 작업이 발생하며, 데이터베이스의 내용을 읽어와 프론트엔드로 전달한다.
- **각 정보 제공 기능** : 사용자가 csv 분석 또는 스크래핑을 통해 받아온 정책 및 용어들을 데이터베이스에 저장한다. 각 정보 페이지에 해당하는 경로를 명시하고 get요청을 통한 각 경로로의 연결이 발생하며, 데이터베이스의 내용을 읽어와 프론트엔드로 전달한다.
- **자료 분석 기능** : 실거래가 데이터 csv파일을 파싱하여 데이터베이스에 저장하며, 저장한 데이터들의 2차 가공을 통해 사용자가 원하는 데이터로 가공한다. 여기서 가공된 데이터들을 프론트엔드로 전달하여 그래프의 각 요소로 출력되도록 한다.
- **매물 확인 기능** : 스크래핑을 통해 받아온 매물 데이터들을 테이블에 맞게 정돈 후 데이터베이스에 저장한다. 카카오맵 restAPI의 지오코딩 요청을 통해 각 주소의 좌표를 저장한 후 카카오맵API로 표현되는 지도 위에 해당 좌표에 해당하는 곳에 마크를 표시하고 세부 정보를 데이터베이스에서 받아와 프론트엔드로 전달한다.
- **AI 상담 기능** : 각 탭마다 post요청을 받는 4개의 경로를 명시하며 사용자의 버튼 입력에 따라 해당 4개의 경로에 따른 프롬프트를 실행한다. 또한 한 경로 내의 키워드 인식을 통해 다른 프롬프트를 실행시키기도 한다. 질문 추천을 제외한 모든 프롬프트들은 RAG형식을 통해 사용자 입력에 맞추어 데이터베이스의 정보를 기반으로 한 답변을 하도록 한다.

개발 과정

1. 주요 개발 단계

※ 피드백을 통한 주제 수정이 많이 이루어졌기 때문에 마지막 주제 선정 이후의 개발 과정만을 명시하였습니다.

① 요구사항 분석 및 기획

- **서비스 범위 정의** : 청주시 4개 구 한정
- **핵심 기능 확정** : 시세 분석, 게시판 기능, AI상담 챗봇 기능 등
- **테마 UI 확정** : 이전 주제의 프론트엔드 UI 디자인 계승 (검정 초록 기반의 UI디자인), 기초 메인화면 구축 완료

② 시스템 설계 및 환경 구축

- **아키텍처 설계** : 이전 주제 개발 상황의 아키텍처 계승 (Spring-Boot, MySQL, Gemini Flash 2.5 → OpenAI Gpt-4o-mini)
- **DB스키마 설계** : cheongju_db 스키마 아래 18개의 테이블 설계

- 개발 환경 구축 : 이전 주제 개발 환경 계승 (Eclipse, Github)

③ 데이터 수집 및 전처리

- 데이터 수집 : 국토부 등의 신뢰성 있는 사이트에서 3년치 실거래가 데이터(csv파일) 수집, 각 정책(csv파일) 및 용어 데이터(csv파일) 수집
- 수집된 데이터의 정규화 및 DB적재 : 각 csv파일의 데이터 -> MySQL에 저장. 저장된 테이블의 데이터를 기반으로 다른 테이블에 전처리를 통한 정규화된 데이터를 적재(자료 분석 및 AI상담에 사용할 데이터)

④ 각 기능 구현 및 통합

- 게시판 및 각 정보 제공 기능 : 기초 메인화면의 확장. 프론트엔드, 백엔드의 연결을 통한 게시판 기능 및 각 페이지로의 연결 구현
- 차트 자료 분석 기능 : 백엔드에서 DB와의 연결을 통해 받아온 데이터를 프론트엔드로 전달. 프론트엔드에서 시각화 차트를 구현하여 전달받은 데이터 출력
- AI분석 기능 : 백엔드에서 Gpt모델 호출 및 경로 입력을 통한 프롬프트 확정(프롬프트 엔지니어링을 통해 구축) -> 정규화된 데이터 및 사용자 입력 메시지를 전달하여 해당하는 답변 반환 -> 프론트엔드로 전달하여 대화창에 답변 출력.

⑤ 디버깅 & 피드백 반영

- 테스트 진행 및 오류 수정 : 각 개발 진행 단계마다 실행. 제대로 구현되지 않았거나 오류 발생 시 디버깅을 통해 코드 수정하여 기능 재구축(LLM모델 변경 Gemini Flash-2.5 → OpenAI Gpt-4o-mini)
- 사용자 시나리오 기반 피드백 반영 : 사용자의 입장에서 적절하지 않은 UI나 자료 디자인, 각 페이지의 접근성, 가독성이 떨어지는 답변 등등의 문제점을 발견 시 피드백을 통해 기능 수정. (차트 디자인 수정 및 사이드바 디자인 수정)
- 사용자 시나리오 기반 기능 추가 : 사용자의 입장에서 존재 시 매우 유리하면서 프로젝트에서 감당 가능한 기능 의논을 통해 추가(매물 정보 보기)
- 팀 티칭 기반 피드백 반영 : 팀 티칭 시 제공받은 피드백을 기반으로 전적 기능 및 디자인 추가/수정. (2번의 주제 수정 발생, 데이터 신뢰성 검사 및 프롬프트 추가)

2. 개발 방법론

- 본 프로젝트는 애자일 방법론을 기반으로 매주 Discord 상에서 개발 상황을 공유하는 스프린트 리뷰와 비슷한 형태의 화상 회의를 진행하여(매일 회의를 진행하지 않아 스크럼 방법론을 사용했다고 하기 어려움) 각 사용자 시나리오 기반 테스트를 통한 피드백을 통해 기능을 추가 및 수정하고 소멸 차트의 개념에 해당하는 부분을 각 기능별 개발상황으로 정리하여 어떤 개발이 남았는지 각 팀원이 인식한 후, 그 정보를 기반으로 개발에 다시 착수하는 방식으로 개발을 진행하였습니다.

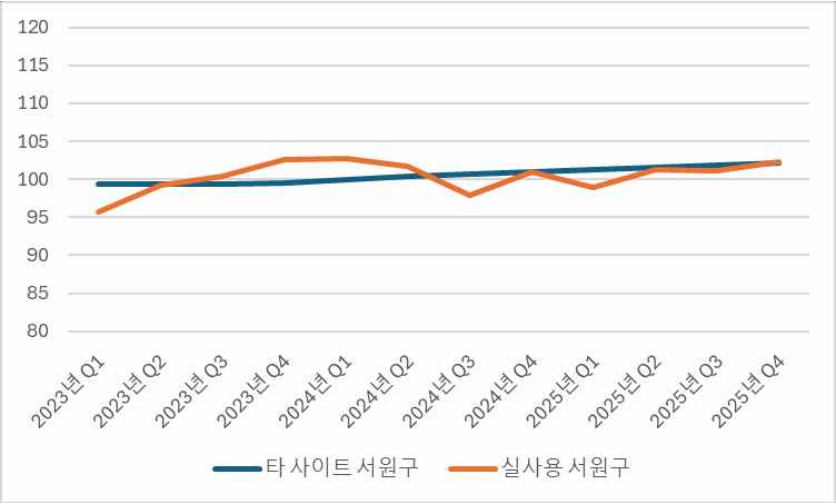
3. 협업 및 버전 관리

- 협업 도구(Github) : 각 팀원이 git을 통해 github의 상의 프로젝트 리포지토리에 각자 브랜치를 만들어 push한 후, pull request 기능을 통해 merge를 신청하고 해당 타임라인에서 소통을 통해 수정이 필요한 경우 수정하여 commit을 승인하는 방식으로 진행
- 의사소통(카카오톡, Discord) : 카카오톡을 통해 간단한 사항 및 요청 사항에 대한 소통을 이루고, Discord를 통해 데이터 파일의 전달 및 문제가 있는 코드 공유, 화상 회의 등의 소통을 진행.

	<p>4. 역할 분담</p> <p>※ 기본적으로 백엔드 및 프론트엔드의 구별은 존재하지 않으며(단, 전체적인 프론트의 디자인은 팀원 박정환이 담당해서 기초 디자인을 구현함.) 각 기능을 맡은 팀원이 백엔드와 프론트엔드를 같이 개발하는 방향으로 진행 (백엔드와 프론트엔드의 연결이 중요하기 때문에 각자 본인 방식의 코드로 연결하는 것이 개발에 용이하다 판단. 또한 위와 같은 맥락으로 해당하는 데이터의 전처리 역시 기본적으로 해당 기능 담당자가 처리)</p> <ul style="list-style-type: none"> - 박성웅(스크럼 마스터, 스크럼 팀원) : AI상담 7개 프롬프트 기능 개발 및 매물 정보 크롤링 / 전체적인 문서 및 ppt, 테스트 시나리오 구축 등 버전 관리 등 스크럼 마스터의 역할을 담당 - 박정환(제품 책임자, 스크럼 팀원) : 게시판, 캘린더, 각종 정보페이지 구현, 전체 layout디자인 및 구현 예정 기능 구상, 영상 보조 / 프로젝트 기능과 최종 목표 및 계획 등등 제품 책임자의 역할을 담당 - 하희찬(스크럼 팀원) : 자료 분석 기능 및 매물 정보 화면 개발, 실거래 데이터 수집 및 영상 제작 / 전체적인 데이터 수집과 데이터 신뢰성 분석을 담당 <p>프로젝트 실행 및 테스트 방법</p> <ul style="list-style-type: none"> - 사용자 시나리오 기반 테스트 진행 : 사용자가 이 프로젝트를 사용할 때의 시나리오 모델을 기반으로 테스트를 진행하여 실제로 사용자가 마주할만한 문제점을 찾는데 중점을 두어 실질적인 서비스에 더욱 가까이 갈 수 있도록 하였습니다.
기대효과 및 활용 분야	<p>기대 효과</p> <ul style="list-style-type: none"> - 부동산 정보의 비대칭성 해소 : 특정 지역의 정밀한 분석을 통한 데이터를 시각화하여 제공함으로써, 일반인과 전문가 사이의 정보 격차를 줄이며, 사용자는 객관적인 시세 파악에 한발짝 더욱 다가가게 되어 거래나 투자에 대한 우려를 줄일 수 있습니다. 또한 이 서비스에서 제공하는 부동산 정책과 용어 설명을 통해 일반인이 이해하기 힘든 복잡한 정보들을 더욱 이해하기 쉽도록 도와주며 시각화된 데이터와 AI상담 서비스, 커뮤니티 내용을 통한 교차 검증으로 매물 탐색에 소요되는 시간과 비용을 획기적으로 단축 시킵니다. <p>활용 분야</p> <ul style="list-style-type: none"> - 일반 실수요자 : 청주시 거주 대학생, 신혼부부, 사회초년생 등등 혼자서 매물을 판단하고자 하는 일반인들이 시세 파악, 적정 가격 판단을 위해 사용할 수 있습니다. - 공인중개사 및 프롭테크 서비스 연계 : 공인중개사에서 고객 상담시에 해당 서비스의 시각화 데이터를 이용할 수 있습니다. - 지역 커뮤니티 및 지자체 정책 모니터링 : 청주시 지자체에서 주거 정보나 커뮤니티의 관련 내용을 확인함으로써 개선 방향을 세울 수 있습니다.
기타 (프로젝트 추가 설명 등)	<p>프로젝트의 차별성</p> <ul style="list-style-type: none"> - 전국적인 데이터가 아닌 국지적인 데이터를 다루는 만큼 정보량으로는 따라잡지 못하지만 그만큼 정밀한 분석과 세부적인 상담 및 제안이 가능합니다. - 기존의 서비스들은 정보 전달에 많은 비중이 들어있다면 30IQ의 프로젝트는 정보를 실시간으로 보며 다른 사람과 해당 정보에 대해 공유하고 토론할 수 있는 플랫폼이라는 점과 최신 AI기술의 RAG형식 증강 검색을 통해 검증된 정규화 데이터를 기반으로 사용자에게 도움이 되는 조언을 주는

것이 가능합니다.

데이터 검증



t-test

지역	T-통계량 (T-Statistic)	P-값 (P-value)	평균 차이
상당구	-0.0995	0.9225	-0.11
서원구	1.5983	0.1383	0.8027
청원구	1.6885	0.1194	1.25
홍덕구	1.3898	0.1921	2.475

회귀분석

요인	계수 (Coefficient)	P-값	95% 신뢰구간 하한	95% 신뢰구간 상한
타 사이트 vs 실사용	1.104434	0.050882	-0.004341	2.21321
결정계수	0.446858	-	-	-
수정된 결정계수	0.343144	-	-	-

- t-test결과 p-값이 모두 유의수준 0.05보다 크기 때문에 타 사이트의 지수와 실제 사용한 데이터 지수 간의 평균 차이가 없다고 할 수 있습니다.
- 또한 회귀분석 결과도 p-값이 유의수준 0.05보다 크므로 타 사이트의 지수와 실제 사용한 지수 간의 평균 차이가 없다고 할 수 있습니다.
- 발생하는 오차는 타 사이트에서는 평당 실거래 데이터 이외의 외부요인과 지수들을 채용하여 나타난 오차라고 판단됩니다.

프로젝트 진행 소감 및 후기

- 박성웅 : 이번 프로젝트는 낯선 개발 환경과 새로운 언어에 도전하며, 단순한 코딩을 넘어 문제 해결 능력을 기르는 계기가 되었습니다. 초기에는 익숙하지 않은 툴과 문법으로 시행착오를 겪기도 했으나, 이를 극복하는 과정에서 새로운 기술 스택에 대한 빠른 적응력을 키울 수 있었습니다. 특히 본 강의의 핵심 목표인 게시판 CRUD 로직을 완벽하게 체화하였고, 나아가

Java 환경에서 RAG(검색 증강 생성) 기반의 프롬프트 엔지니어링을 적용해 AI 챗봇을 성공적으로 구현함으로써 기술적 깊이를 더했습니다. 또한 팀장으로서 프로젝트 일정 관리와 보고서 작성을 주도하며, 수정 사항을 체계적으로 반영하는 과정에서 버전 관리와 협업의 중요성을 몸소 체험했습니다. 이는 향후 전문 개발자로 성장하는 데 있어 기술과 소통 능력을 겸비하는데 매우 중요한 경험으로 작용할 것입니다.

- 박정환 : 사용자가 서비스를 이용하며 느끼는 경험을 최우선으로 고려하여 UI/UX를 설계했습니다. 특히 웹 애플리케이션 전반에 걸쳐 일관된 디자인과 사용성을 제공하기 위해 Thymeleaf의 Fragment 기능을 활용해 레이아웃 시스템을 구축했습니다. 이는 페이지별로 파편화될 수 있는 UI 요소를 중앙에서 통합 관리하게 해주었으며, 디자인 수정 사항을 전체 페이지에 즉각 반영할 수 있는 유연한 구조를 완성했습니다.
기능 구현 측면에서는 캘린더 등 이미 완성도 높은 오픈소스 라이브러리를 적재적소에 도입하는 전략을 취했습니다. 라이브러리 연동 과정에서 발생한 호환성 문제나 스크립트 충돌 오류를 해결하며 외부 코드를 우리 서비스에 맞게 최적화하는 방법을 익혔습니다. 이를 통해 사용자에게 친숙하고 직관적인 인터페이스를 제공하는 완성도 높은 서비스를 구현할 수 있었습니다.
- 하희찬 : 저는 이번 프로젝트에서 국토교통부에서 제공하는 실거래가를 가공 및 분석하였고 이를 Vega-Lite와 Vega-Embed 라이브러리를 사용하여 사용자의 선택에 따라 동적으로 차트를 만들어 시각화해주는 자료 분석 기능을 구현하였습니다. 또한, Kakao Maps API를 이용하여 사용자가 원하는 지역에 대한 매물을 보여주는 실매물 확인 기능을 구현하였습니다. 이번 프로젝트는 자바 언어를 사용하거나 스프링부트를 사용하는 등 제게 여러 모로 새로운 시도였습니다. 처음에는 새로운 환경에 적응하는게 힘들기도 하였으나 프로젝트를 진행하며 새로운 환경에서 적응하는 방법을 체득하며 개발자로서의 능력을 키울 수 있었습니다. 또한, 중간에 주제가 바뀌기도 하는 등의 문제도 있었지만 팀원들과 함께 차근차근 문제를 해결하며 팀워크를 키울 수 있어 좋은 경험이 된 것 같습니다.

향후 추진 계획

1. 데이터 분석 심화

- 비가격 데이터 연동 및 분석 : 현재 실거래 데이터 외에 유동인구, 인구 성향, 지역 호재 등등의 비가격적인 데이터들을 추가로 수집한 후 분석을 통해 시각 정보를 제공하고 조연 시스템에 추가 로직을 추가하는 계획입니다.
- 간이 예측 모델 도입 : 데이터의 범위를 더욱 늘려 학습하기에 충분한 양의 데이터를 수집한 후, 각 군집화한 데이터들을 시계열 분석 모델 등에 적용시켜 3~6개월 이후의 간이 시세 예측값을 도출한 후 해당 정보를 시각적 데이터로 제공하는 계획입니다.

2. AI기능 고도화

- 다양한 프롬프트 추가 : 현재 6개의 프롬프트가 제공되고 7개의 프롬프트가 실질적으로 작동하고 있지만, 모든 경우의 프롬프트를 처리하지 못하므로 이를 확장하는 계획입니다.(Langchain을 이용한 프롬프트 오케스트레이션)
→ **프롬프트 확장 예시)** 정책 기반 자금 설계 시뮬레이션 : 사용자의 자산·소득 정보와 복잡한 부동산 정책(취득세율, 생애최초/청년 혜택 등)을 실시간으로 연동하여, 단순 매물 추천을 넘어 '실제 가용 자금' 내에서 구매 가능한 최적의 매수 시나리오를 도출하는 계획입니다.
- 자연어 기반 데이터 탐색 : 시각화 필터링 기능을 AI기능과 연동시켜 질문에 대해 텍스트뿐 아니라 R언어와 연동하여 사용자가 요청한 데이터를 시각 정보로 만든 후, 텍스트와 함께 출력하게 하는 계획입니다.

3. 서비스 확장성

	<ul style="list-style-type: none">- 지역 확장 : 현재 청주시에 국한된 서비스를 전국으로 확장하여, 지역마다 정밀한 데이터 분석과 제안 시스템, 한눈에 보기 쉬운 시각적 이미지를 제공하는 계획입니다.- 데이터 자동화 파이프라인 : 현재 데이터 수집을 코드를 작성했다고 하지만, 데이터 수집에 걸리는 시간을 고려하여 미리 데이터 수집을 진행하고 서버를 구동하는 식으로 작동하고 있지만, 서버 구동 이후에도 자동으로 데이터를 수집하고 최신화하는 파이프라인을 구축하는 계획입니다.
--	--