

데이터분석프로그래밍 최종 보고서

실거래가 분석을 통한 강원대 인근 부동산 시장 인사이트 도출

202112878 장영웅

1. 데이터셋 개요

1.1 선택한 데이터셋 소개

국토교통부에서 제공하는 60제곱미터 이하 단독/다가구 주택의 실거래가 데이터, 네이버 부동산에서 제공하는 원룸 및 투룸 매물 정보 데이터 - (효자동, 후평동, 석사동, 퇴계동)

1.2 데이터셋의 주요 특징

국토교통부에서 제공하는 데이터는 특정 기간의 실거래가 및 기타 정보(지역별 해당 연도 혹은 해당 월에 계약된 부동산의 면적과 보증금, 전/월세, 계약기간 등)를 확인할 수 있다. 이 데이터를 활용하면 부동산 거래의 성수기/비수기 같은 계절적 요인도 확인할 수 있다.

네이버 부동산 사이트에서 제공하는 데이터는 현재 매물로 나와 있는 부동산의 데이터(보증금, 전/월세, 관리비, 면적, 입주 가능일, 정확한 주소, 주변 대중교통, 건물의 준공 연도 등)를 수집할 수 있다. 네이버 부동산 데이터는 지도상의 정확한 위치 정보도 포함되어 있어, 공간적 분석을 통해 강원대와의 거리와 가격의 상관관계를 분석할 수 있다.

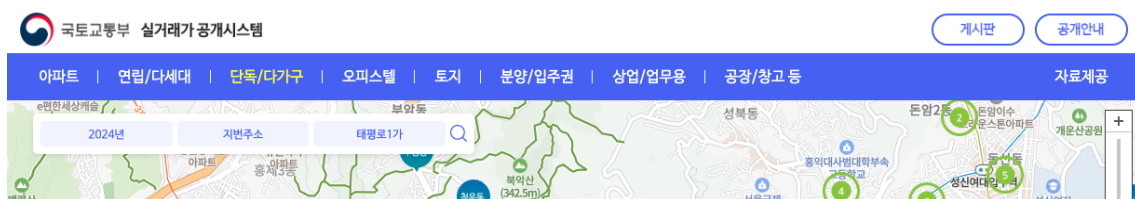
국토교통부 데이터는 과거 실거래가 데이터를 기반으로 하고, 네이버 부동산 데이터는 현재 시장의 매물을 반영한다. 때문에 과거 시세와 현재 시세를 비교하여 가격 변화 경향을 확인할 수 있기에 시장의 안정성과 변동성을 파악할 수 있다. 국토교통부 데이터는 위치의 세부 정보가 부족하고, 네이버 부동산 데이터는 시간적 정보를 제공하지 않는다. 때문에 두 데이터를 적절히 활용하여 시세 변동 및 현재 매물 상태를 모두 고려한 데이터 분석을 진행하고자 한다.

2. 데이터 수집 과정

2.1 데이터 수집 방법

Selenium과 BeautifulSoup 라이브러리를 활용하여 국토교통부 실거래가 데이터와 네이버 부동산 데이터를 자동으로 크롤링하고, CSV 파일로 저장하는 파이썬 코드를 작성하여 강원대학교 인근 지역(효자동, 후평동, 석사동, 퇴계동)의 부동산 데이터를 수집했다.

- 국토교통부 부동산 데이터 ([국토교통부 실거래가 공개시스템 - 단독/다가구](#))



<그림1 - 국토교통부 실거래가 공개시스템 페이지 상단>

[단독다가구] 강원특별자치도 춘천시 효자동

매매

전월세

법정동/리

기준년도

월

면적

금액

전월세구분

효자동

2024

전채

--전체--

--전체--

--전체--

11월

법정동	지번	주택유형	연면적 (㎡)	계약일	계약기간	보증금 (만원)	중전보증금 (만원)	전산공부
	도로명		계약구분		갱신요구권사용	월세 (만원)	중전월세 (만원)	
효자동	3**	단독	30	20	24.11~26.11	500	-	Q 보기
	백령로		신규		-	40	-	
효자동	1**	다가구	29.7	19	25.01~25.12	200	-	Q 보기
	돌담길		신규		-	30	-	
효자동	6**	다가구	21	15	24.11~25.11	200	-	Q 보기
	서부대성로173번길		신규		-	30	-	
효자동	7**	다가구	33	14	24.12~26.12	6,000	-	Q 보기
	백령로119번길		신규		-	0	-	
효자동	4*	다가구	164.83	11	24.11~27.11	15,000	-	Q 보기
	성심로		신규		-	50	-	
효자동	3**	다가구	120	08	-	3,000	-	Q 보기
	공지로252번길		-		-	75	-	
효자동	2**	다가구	45	07	24.11~26.11	9,000	-	Q 보기
	공지로		신규		-	0	-	
효자동	7**	다가구	25.85	06	-	4,000	-	Q 보기
	성심로47번길		-		-	0	-	
효자동	1**	다가구	41	05	24.11~25.11	200	-	Q 보기
	백령로75번길		신규		-	30	-	
효자동	6**	단독	28	02	24.11~25.11	500	-	Q 보기
	두하길		신규		-	43	-	

*도로조건, 전산공부는 공개시점 기준으로 제공되어 조회시점의 현황, 전산공부 등과 상이할 수 있으니 참고하시기 바랍니다.

〈그림2 - 춘천시 효자동 2024년도 전월세가 정보〉

국토교통부 페이지에서 연도와 주소를 선택하여 돋보기 버튼을 누르면 해당 지역의 매매가와 전월세가 정보를 확인할 수 있다. (그림 1, 2 참고) 강원대학교 인근 지역의 부동산 정보를 수집하기 위해 크롤링을 활용하였으며, 필요한 정보는 면적 60㎡ 이하의 전월세 데이터를 대상으로 한다. 연도와 월별로 데이터를 수집(HTML 데이터)한 뒤, BeautifulSoup을 사용하여 파싱하고, 필요한 정보를 DataFrame으로 정리하여 CSV 파일로 저장하였다. 해당 데이터 수집 코드는 [molit_capture.py](#)에 구현되어 있으며, 이를 통해 손쉽게 원하는 데이터를 자동으로 추출할 수 있도록 설계했다.

데이터는 2014년 1월부터 2024년 9월 ‘석사동, 퇴계동, 효자동, 후평동’의 정보를 수집하였으며 [법정동/지번/도로명/주택유형/연면적(㎡)/계약구분/계약일/계약기간/갱신요구권사용/보증금(만원)/월세(만원)/중전보증금(만원)/중전월세(만원)/전산공부]를 헤더로 한다. 최종적으로 수집된 데이터는 [CSV 파일](#)로 저장되었으며, (22153,14)의 형태를 가진다. 22153개의 행은 각각 하나의 부동산 데이터를 나타낸다.

- 네이버 부동산 데이터 ([네이버 부동산 - 원룸/투룸](#))

<그림3 - 네이버 부동산 원룸/투룸 매물 페이지>

네이버 부동산의 원룸/투룸 매물 페이지에 접속하면 중간 부분에 지역을 선택할 수 있다. 지역을 선택하면 좌측에 해당 구역의 부동산 매물 정보를 확인할 수 있다. (그림 3 참고) 크롤링을 활용하여 원하는 지역을 선택하고, 좌측에 뜨는 팝업의 스크롤을 내리면서 모든 매물 정보를 수집하였다. 국토교통부 데이터를 추출할 때와 동일하게 Selenium으로 로드된 HTML 데이터를 BeautifulSoup으로 파싱하여 필요한 정보를 DataFrame으로 정리하여 CSV 파일에 저장하였다. 해당 데이터 수집 코드는 [naver_capture.py](#)에 구현되어 있으며, 이를 통해 손쉽게 원하는 데이터를 자동으로 추출할 수 있도록 설계했다.

데이터는 '석사동, 퇴계동, 효자동, 후평동'의 정보를 수집하였으며 [매물 유형/매물 제목/소재지/가격/면적/층수/방향/태그/중개 플랫폼/중개사/등록일]를 헤더로 가진다. 이 데이터는 실시간 부동산 매물 정보를 나타내기 때문에 정확도를 위하여 2024/10/17~2024/11/10 동안 4번의 데이터 수집을 완료하였다. 최종적으로 수집된 데이터는 [CSV 파일](#)로 저장되었으며, (1200,11)의 형태를 가진다. 1200개 행은 각각 하나의 부동산 데이터를 나타낸다.

2.2 데이터의 적합성 및 신뢰성

본 데이터는 국토교통부 실거래가 시스템 및 네이버 부동산 플랫폼에서 수집된 정보로, 신뢰할 수 있는 공인된 출처를 기반으로 한다. 데이터는 2014년부터 2024년까지 강원대학교 인근 지역(효자동, 후평동, 석사동, 퇴계동)을 대상으로 하며, 면적 60㎡ 이하의 전월세 데이터를 중심으로 수집되었다.

수집된 데이터는 출처별로 특징이 다르다. 국토교통부 데이터는 기간별 부동산 정보를 확보할 수 있지만, 정확한 위치 정보를 제공하지 않는 한계가 있다. 반면, 네이버 부동산 데이터는 실시간 매물 정보와 함께 정확한 위치 정보를 제공하지만, 매물 정보가 사용자 입력에 의존하므로 일부 데이터의 신뢰성이 부족할 수 있다. 이러한 차이를 보완함으로써 두 데이터는 서로를 보완하며, 보다 정밀한 데이터 분석이 가능하다.

데이터 수집 과정에서는 Selenium과 BeautifulSoup을 활용해 정확도를 유지했다. 또한, 수집된 데이터는 Pandas를 이용해 중복 제거, 누락값(NaN) 처리, 불필요한 헤더 제거 등의 전처리 과정을 수행하여 데이터의 일관성과 신뢰성을 확보하고자 한다.

종합적으로, 본 데이터는 강원대학교 인근 지역 부동산 시장 분석에 적합하며, 신뢰할 수 있는 자료로 활용될 수 있다.

3. 데이터 수집 목적

3.1 프로젝트의 배경 및 필요성

강원대학교 인근 지역은 많은 학생과 교직원이 거주하며, 원룸 및 투룸 매물이 활발히 거래되는 주요 부동산 시장 중 하나이다. 그러나 학생들과 거주자들은 적절한 주거 환경을 선택하는 데 있어 부동산 가격 변동 및 시장 상황에 대한 정보를 얻기 어렵다.

특히, 전월세 시장의 경우 시기별로 가격 차이가 크고, 지역별로 임대료가 상이하게 나타나므로 이를 체계적으로 분석하는 것이 필요하다. 본 프로젝트는 강원대학교 인근 지역의 부동산 데이터를 체계적으로 수집하여, 주거 선택 시 실질적인 도움을 줄 수 있는 데이터를 제공하는 것을 목표로 한다.

3.2 데이터 수집의 구체적 목표

- 부동산 시세 변동 추이 분석

국토교통부 실거래 데이터를 활용하여 월별 및 연도별 가격 변화를 파악하고, 부동산 시장에서 계절적 요인(성수기/비수기)이 존재하는지 확인한다. 이를 통해 가격 상승 또는 하락이 주로 발생하는 시기를 분석하며, 과거 시세와 현재 시세를 비교하여 가격 변동성을 도출하고 향후 몇 개월간의 가격 변화를 예측한다.

- 거리와 부동산 가격 간 상관관계 분석

강원대학교와의 거리, 건물 면적, 예상 거래가 등의 데이터를 수집하고 이를 3차원 그래프로 시각화하여 대학과의 거리 변화에 따른 가격 차이를 도출한다. 추가로 주변 상권과 대중교통 정보를 고려하여 현재 시점에서 최적의 매물을 선택할 수 있는 근거를 제공한다.

- 주거 선택 지원

분석 결과를 기반으로 강원대학교 인근에서 이사를 계획 중인 학생들에게 효율적인 주거 선택지를 제공함으로써 참고할 수 있는 구체적이고 실질적인 데이터를 확보한다.

4. 데이터 전처리

4.1 데이터 정제(cleaning) 및 정규화(normalization)

데이터 전처리 과정은 수집된 원시 데이터를 분석에 적합한 형태로 변환하는 데 초점을 두고 진행되었다. 주요 작업은 데이터 클리닝, 정규화, 계약일 형식 변환 및 전세 보증금 데이터를 월세로 변환하는 작업으로 구성되었다.

- 데이터 정제

수집된 모든 데이터에서 불필요한 열(중개 플랫폼, 중개사, 등록일)을 제거하여 데이터를 간결화하였다. 매물 제목에 따른 정확한 소재지를 반영하기 위해 미리 정의된 매물 제목-소재지 정보를 사용하여 소재지를 업데이트하였다. 중복된 매물 제목을 제거하고, 불필요한 카테고리(도시형생활주택, 다가구, 빌라 등)를 제외하여 데이터의 정확성과 일관성을 확보하였다.

- 전세 보증금을 월세로 변환

전세 보증금 데이터를 월세로 변환하기 위해 전세 전환율 계산 공식을 적용하였다.

$$\text{전세 전환율 공식 : 월세} = \text{전세보증금} \times (\text{기준금리} + \text{이율}) \div 12$$

기준금리는 계약일에 따라 적용되며, get_base_rate 함수를 통해 계산되었다. 이율은 대통령령에 의거하여 2021년 이후 2%, 이전에는 3.5%로 설정하였다. (주택임대차보호법 시행령 제9조 - 월차임 전환 시 산정률) 전세 데이터를 월세로 변환하여 분석의 통일성을 확보하고, 전세와 월세 간 비교를 가능하게 하였다.

- 데이터 정규화

소재지 정보를 기반으로 Google Maps API를 활용하여 각 매물의 위도와 경도를 추가하였다. 매물의 소재지 정보를 정확히 분석하고 이를 활용한 공간적 시각화를 가능하게 하기 위해, Google Maps API를 통해 위도와 경도 좌표 데이터를 획득하였다. 이를 통해 모든 매물 데이터에 일관된 위도와 경도 정보를 부여함으로써, 지역별 매물의 공간적 분포를 명확히 분석하고 다양한 형태의 지도 시각화를 수행할 수 있는 기반을 마련하였다.

Google Maps API는 일정 요청 수 초과 시 추가 비용이 발생하고, 일일 요청 제한이 존재하므로, 일일 요청 수를 관리하고 효율적으로 데이터를 처리하기 위한 방법을 적용하였다. 요청 데이터를 기록하고 관리하여 API 요청량 이 초과되지 않도록 제한을 설정하였으며, 중단 시에도 작업을 재개할 수 있도록 처리 상태를 저장하는 기능을 포함하였다.

위도와 경도 정보를 포함한 데이터셋([naver_data_updated](#))은 공간적 분석과 시각화를

기반으로 강원대학교 인근 부동산 시장의 특성을 심층적으로 파악할 수 있는 준비를 완료하였다. ([DataCleaning_2.py](#), [daily_usage.json](#), 그림 4 참고)

```

문제  출력  디버그 콘솔  터미널  로트

처리 중 (47/57): 강원도 춘천시 효자동 129-169
주소 처리 중: 강원도 춘천시 효자동 129-169
요청량 데이터 저장 완료: {'request_count': 94, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8737575, 경도: 127.737885
처리 중 (48/57): 강원도 춘천시 효자동 18-28
주소 처리 중: 강원도 춘천시 효자동 18-28
요청량 데이터 저장 완료: {'request_count': 95, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8763572, 경도: 127.7442797
처리 중 (49/57): 강원도 춘천시 효자동 751-16
주소 처리 중: 강원도 춘천시 효자동 751-16
요청량 데이터 저장 완료: {'request_count': 96, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8791408, 경도: 127.7482811
처리 중 (50/57): 강원도 춘천시 효자동 651
주소 처리 중: 강원도 춘천시 효자동 651
요청량 데이터 저장 완료: {'request_count': 97, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8746442, 경도: 127.7358463
처리 중 (51/57): 강원도 춘천시 효자동 590-50
주소 처리 중: 강원도 춘천시 효자동 590-50
요청량 데이터 저장 완료: {'request_count': 98, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8689793, 경도: 127.7259663
처리 중 (52/57): 강원도 춘천시 효자동 757-2
주소 처리 중: 강원도 춘천시 효자동 757-2
요청량 데이터 저장 완료: {'request_count': 99, 'date': '2024-11-17'}
변환 결과 - 위도: 37.87854129999999, 경도: 127.7411618
처리 중 (53/57): 강원도 춘천시 효자동 129-1
주소 처리 중: 강원도 춘천시 효자동 129-1
요청량 데이터 저장 완료: {'request_count': 100, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8746893, 경도: 127.7363164
처리 중 (54/57): 강원도 춘천시 효자동 176-11
주소 처리 중: 강원도 춘천시 효자동 176-11
요청량 데이터 저장 완료: {'request_count': 101, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8666334, 경도: 127.7370479
처리 중 (55/57): 강원도 춘천시 효자동 613-7
주소 처리 중: 강원도 춘천시 효자동 613-7
요청량 데이터 저장 완료: {'request_count': 102, 'date': '2024-11-17'}
변환 결과 - 위도: 37.875409, 경도: 127.7400761
처리 중 (56/57): 강원도 춘천시 효자동 650-2
주소 처리 중: 강원도 춘천시 효자동 650-2
요청량 데이터 저장 완료: {'request_count': 103, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8745335, 경도: 127.7355552
처리 중 (57/57): 강원도 춘천시 효자동 176-11
주소 처리 중: 강원도 춘천시 효자동 176-11
요청량 데이터 저장 완료: {'request_count': 104, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8666334, 경도: 127.7370479
위도와 경도를 한 줄로 합치기
업데이트된 CSV 파일 저장 중: /Users/jang-yeong-ung/Documents/real_estate-data-analysis/naver_data_updated/2024-10-20-04:35-효자동.csv
파일 저장 완료: /Users/jang-yeong-ung/Documents/real_estate-data-analysis/naver_data_updated/2024-10-20-04:35-효자동.csv
CSV 파일 처리 시작: /Users/jang-yeong-ung/Documents/real_estate-data-analysis/naver_data/2024-10-17-21:08-퇴계동.csv
CSV 파일 로드 완료, 총 58개의 행.
처리 중 (1/50): 강원도 춘천시 퇴계동 369-9
주소 처리 중: 강원도 춘천시 퇴계동 369-9
요청량 데이터 저장 완료: {'request_count': 105, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8631809, 경도: 127.7299503
처리 중 (2/50): 강원도 춘천시 퇴계동 369-9
주소 처리 중: 강원도 춘천시 퇴계동 369-9
요청량 데이터 저장 완료: {'request_count': 106, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8631809, 경도: 127.7299503
처리 중 (3/50): 강원도 춘천시 퇴계동 369-9
주소 처리 중: 강원도 춘천시 퇴계동 369-9
요청량 데이터 저장 완료: {'request_count': 107, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8631809, 경도: 127.7299503
처리 중 (4/50): 강원도 춘천시 퇴계동 369-9
주소 처리 중: 강원도 춘천시 퇴계동 369-9
요청량 데이터 저장 완료: {'request_count': 108, 'date': '2024-11-17'}
변환 결과 - 위도: 37.8631809, 경도: 127.7299503
처리 중 (5/50): 강원도 춘천시 퇴계동 1006-2
주소 처리 중: 강원도 춘천시 퇴계동 1006-2

```

〈그림4 - Google Maps API(Geocoding API) : 소재지에 따른 위도 경도 값 업데이트〉

4.2 결측치 처리

- 결측치 탐지 : 주요 필드(전용면적, 위도, 경도, 10㎡당 월세)에서 결측치를 탐지하였다.
 - > [DataAnalysis_2.py](#)에서 dropna()를 사용하여 전용면적의 결측치를 확인하고 처리했다.
 - > df = df.dropna(subset=['전용면적'])
 - > [DataAnalysis_2_2.py](#)에서도 동일하게 결측치를 처리하였다.
 - > df = df.dropna(subset=['전용면적', '위도', '경도'])
- 결측치 처리 : 필수 필드에서 결측치가 발견될 경우 해당 데이터를 삭제하였다.
- 결측치 대체 : 일부 누락된 값(소재지)은 매물 제목과의 매핑 정보를 활용하여 보완하였다
 - > [data_cleaning.py](#)의 함수 참고

4.3 이상치 처리

- **이상치 탐지** : 면적이 60㎡ 이상인 데이터, 면적당 월세 값이 1 이하 이거나 면적당 월세 값이 비정상적으로 높은 경우는 이상치라고 판단하였다.
- **이상치 처리** : 데이터 분석 시각화 시에 표시되지 않도록 처리했다.
 - > [DataAnalysis_1.py](#)에서 10㎡당 월세가 1 이하인 데이터 제외
 - > [git commit history](#) 참고 (면적당 월세가 비정상 적으로 큰 경우 제거)

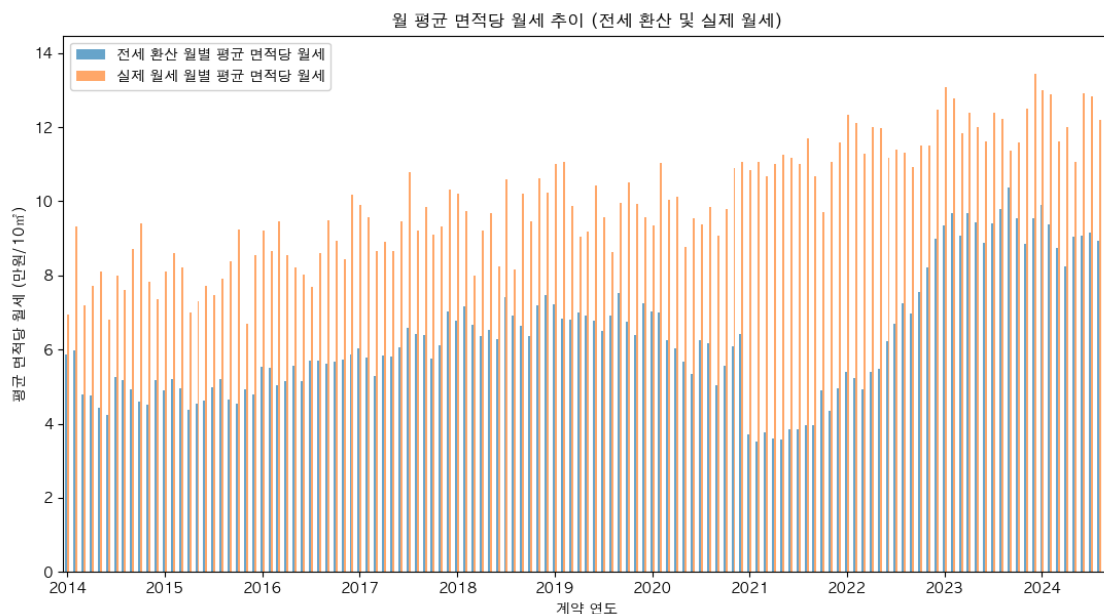
5. 데이터 시각화

5.1 데이터 탐색적 시각화 (Exploratory Data Analysis; EDA)

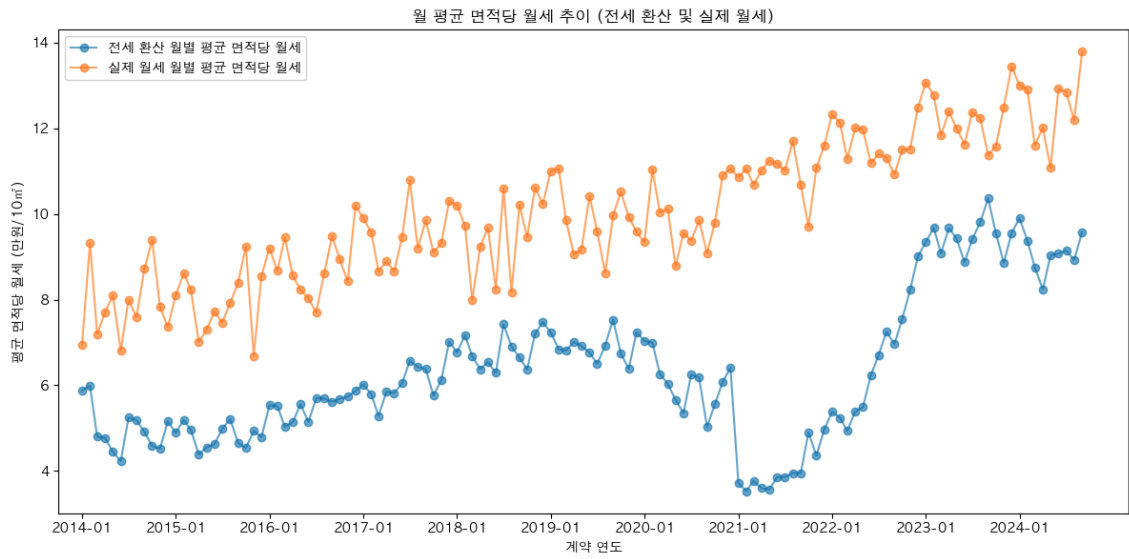
데이터 탐색적 시각화(EDA)는 수집된 데이터의 패턴, 경향, 이상치를 식별하고, 심층 분석을 위한 기반을 마련하기 위해 수행되었다. 본 분석에서는 다음과 같은 시각화를 통해 강원대학교 인근 부동산 시장의 특성을 파악하였다.

- 연도-월별 평균 면적당 월세 추이 분석

월별 평균 면적당 월세 변화를 시각화하여 특정 시기의 가격 상승 및 하락 패턴을 확인해보았다. 데이터의 전반적인 트렌드를 파악하기 위해 막대 그래프와 꺾은선 그래프를 활용했다. 편의를 위해 y 축 평균 면적당 월세의 단위를 (만원/10㎡)으로 설정하였고 월세와 전세를 구분하여 색을 구분하여 표현하였다. 또한 산점도가 아닌 평균값으로 시각화를 진행했는데, 그 이유는 부동산 시장의 경향성을 파악하기 위해서이다. (그림 5, 6 참고)



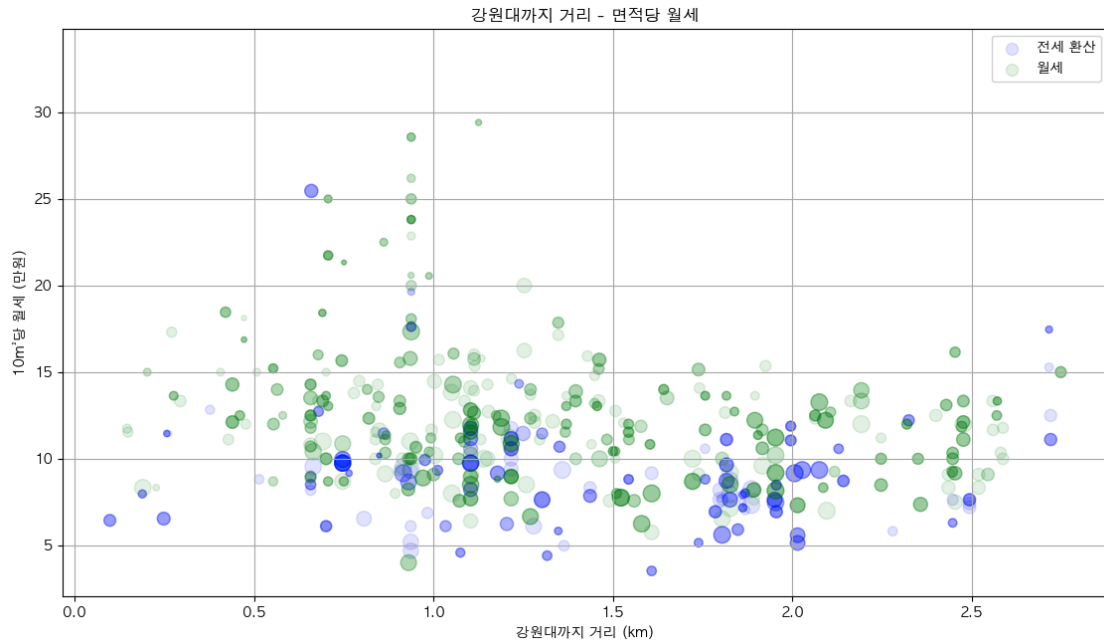
<그림 5 - 월 평균 면적당 월세 변화 추이 확인 : 막대 그래프>



<그림 6 - 월 평균 면적당 월세 변화 추이 확인 : 꺾은선 그래프>

- 강원대학교로부터의 거리와 면적당 월세 간의 관계

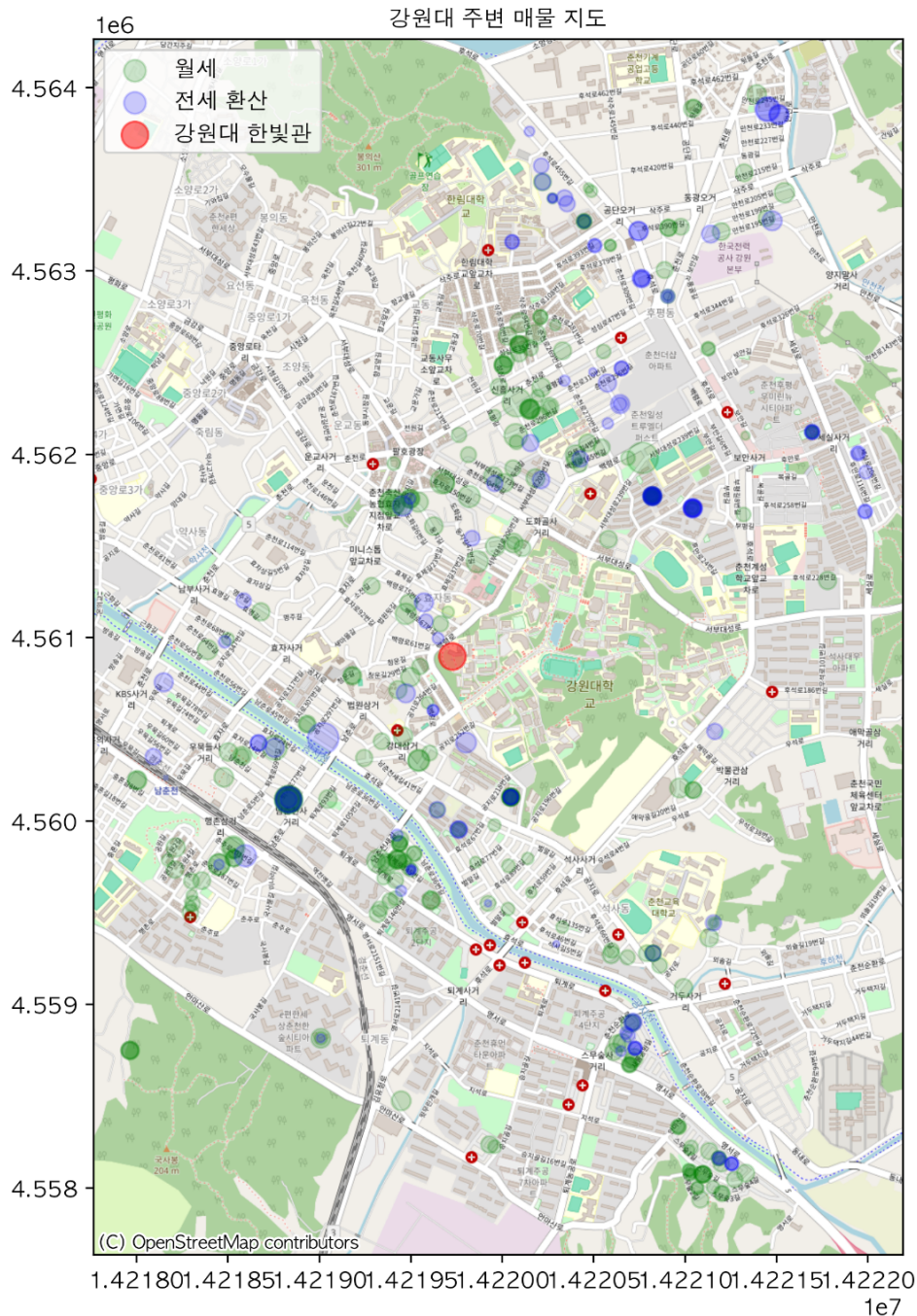
강원대학교로부터 매물까지의 거리와 면적당 월세 간의 상관관계를 시각적으로 확인하기 위해 산점도를 활용하였다. X축은 강원대학교와 매물 사이의 거리(단위: km), Y축은 면적당 월세(만원/10㎡)로 설정하였으며, 각 데이터 포인트는 매물을 나타낸다. 월세와 전세 환산 데이터를 구분하여 색상으로 표현하여 각 유형의 관계를 명확히 했다. (그림 7 참고)



<그림 7 - 강원대와 매물 사이 거리에 따른 면적당 월세 확인 : 산점도 그래프>

- 지도 기반 매물 분포 시각화

강원대학교를 중심으로 주변 매물의 공간적 분포를 분석하기 위해 지도 기반 시각화를 수행하였다. 매물의 위치를 위도와 경도를 기반으로 지도 상에 표시하고, 월세와 전세 환산 매물을 색상으로 구분하였다. 각 점의 크기는 면적당 월세 금액(만원/10㎡)에 비례하여 설정하여, 금액이 높은 매물은 더 큰 점으로 표시하였다. 지도 기반 시각화를 통해 특정 지역에서의 월세와 전세 환산 매물의 분포를 파악할 수 있었으며, 공간적 집중도와 지역별 특징을 확인할 수 있었다. (그림 8 참고)



<그림 8 - 강원대 인근 매물 확인 : 위도, 경도, 점의 크기(면적당월세)>

5.2 주요 그래프 및 통계적 분석 결과

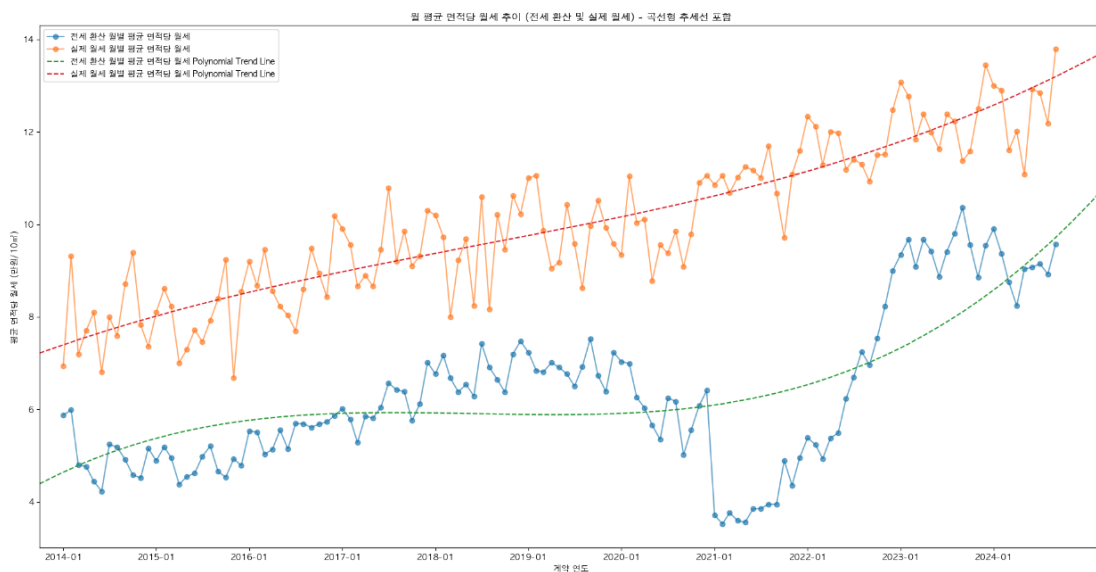
- 연도-월별 평균 면적당 월세 추이 분석

연도-월별 평균 면적당 월세 꺾은선 그래프에 3차 다항 회귀 분석(Polynomial Regression)을 적용하여 장기적인 상승 또는 하락 추세를 나타내는 추세선을 그래프에 추가하였다.

추세선 분석 : 다항 회귀 분석(3차)을 적용하여 비선형적인 데이터의 변화를 모델링.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

(y : 평균 면적당 월세, x : 시간(연도-월), ϵ : 오차)



<그림 9 - 월 평균 면적당 월세 그래프에 3차 다항 회귀 추세선 추가>

월세와 전세 환산 월세 모두에서 시간이 지남에 따라 비선형적인 상승 추세를 보였으며, 특히 2021년 이후 급격히 증가하는 경향이 나타났다. 월세 상승 추세는 강원대학교 인근 부동산 시장의 수요 증가와 함께 물가 상승, 전세자금대출 금리 인상 등의 외부 요인이 영향을 미친 것으로 보인다. 다항 회귀 분석을 통해 이러한 비선형적인 경향을 더욱 명확히 파악할 수 있었다. 외부 요인 분석은 인사이트 도출 부분에서 자세히 다루도록 하겠다.

- SARIMA 모델을 활용한 월세 예측

SARIMA(Seasonal Autoregressive Integrated Moving Average) 모델을 활용하여 월 평균 면적당 월세의 계절적 및 비계절적 요소를 모두 고려한 예측 결과를 시각화하였다. 실제 월세 데이터와 SARIMA 모델로 예측된 1~2년간의 월세를 꺾은선 그래프로 구분하여 표시했다. (그림 10, 11 참고)

＞ SARIMA 모델 분석 :

Seasonal AutoRegressive Integrated Moving Average 모델은 계절성이 포함된 시계열 데이터를 분석하고 미래 값을 예측하는 데 사용되는 통계적 모델이다. 강원대학교 인근 부동산의 월 평균 면적당 월세 데이터를 기반으로 SARIMA 모델을 적용하여 다음 1~2년간의 월세를 예측하였다.

＞ SARIMA 모델 구성 요소 :

SARIMA 모델은 아래와 같은 파라미터로 구성되며, 각 요소는 데이터의 시계열적 특징을 반영한다.

$$SARIMA(p, d, q)(P, D, Q, s)$$

＞ p : 비계절적 자기회귀(AutoRegressive, AR) 차수

이전 관측치가 현재 값에 미치는 영향을 나타내는 차수. 예를 들어, p=1은 바로 직전 값이 현재 값에 영향을 미친다는 것을 의미한다.

＞ d : 비계절적 차분(Integration, I) 횟수

데이터의 추세(Trend)를 제거하기 위해 몇 번의 차분을 적용할지를 나타낸다. 예를 들어, d=1은 데이터의 차이를 한 번 계산하여 안정성을 확보한다.

＞ q : 비계절적 이동평균(Moving Average, MA) 차수

과거의 오차가 현재 값에 미치는 영향을 나타내는 차수.

＞ P,D,Q : 계절적 ARIMA 모델의 차수

각각 자기회귀, 차분, 이동평균을 계절성(Seasonality)에 기반하여 계산하는 차수. 예를 들어, P=1, D=1, Q=1 은 계절적 성분에서도 이전 관측치와 오차를 한 번씩 고려한다.

＞ s : 계절 주기

계절성 주기를 나타내며, 본 분석에서는 데이터가 월별로 기록되었으므로 s=12로 설정하였다(1년 단위 계절성).

＞ SARIMA 모델 설정 :

본 분석에서는 아래와 같은 SARIMA 파라미터를 사용하였다 :

$$SARIMA(1, 1, 1)(1, 1, 1, 12)$$

＞ p=1, d=1, q=1 : 월별 데이터에서 비계절적 요인을 반영.

＞＞ p= 1 : 직전 월의 평균 면적당 월세가 현재 월세에 영향을 미침.

＞＞ d= 1 : 차분을 통해 데이터를 안정화하여 추세를 제거.

＞＞ q= 1 : 직전 오차가 현재 월세에 영향을 미침.

- > P=1, D=1, Q=1, s=12 : 계절성을 반영하기 위해 1년 주기로 계절적 요인 설정.
 - >> P= 1 : 1년 전의 관측치가 현재 월세에 영향을 미침.
 - >> D= 1 : 계절적 차분을 통해 계절적 패턴 제거.
 - >> Q= 1 : 계절적 오차가 현재 월세에 영향을 미침.
 - >> s= 12 : 계절 주기를 1년 단위(12개월)로 설정.

> SARIMA 모델 적용 과정 :

1. 월별 평균 면적당 월세 데이터 생성 : SARIMA 모델 적용을 위해 시계열 데이터가 필요하며, 이 데이터는 월별로 집계된 평균 면적당 월세를 기반으로 한다.
2. 차분(First Differencing) 적용 : 시계열 데이터의 정상성을 확보하기 위해 차분을 적용한다. (정상성은 데이터의 평균, 분산, 자기공분산이 시간에 따라 일정한 성질을 가지는 것을 의미) 코드에서 차분은 SARIMAX 모델 내부에서 자동으로 처리된다.
 - > order=(1, 1, 1)에서 d=1은 첫 번째 차분(First Differencing)을 의미하며, 이는 데이터의 비정상성을 제거하여 안정적인 시계열 모델링을 가능하게 한다.
3. SARIMA 모델 구성 : SARIMAX(ts_data_scaled, order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))를 사용하여 SARIMA 모델을 구성한다.
 - > order=(1, 1, 1)은 비계절적 자기회귀(AR), 차분(I), 이동평균(MA) 차수를 나타낸다.
 - > seasonal_order=(1, 1, 1, 12)은 계절적 자기회귀(AR), 계절적 차분(I), 계절적 이동평균(MA) 차수와 계절 주기(12개월, 1년)를 나타낸다.
4. 스케일링 : ARIMA 모델 적용 전 데이터를 표준화(Standardization)하여 모델의 수렴 속도를 높이고 안정성을 확보한다.
 - > scaler = StandardScaler()를 사용하여 데이터를 표준화한 뒤,
 - > scaler.fit_transform()을 통해 스케일링된 데이터를 생성한다.
5. 모델 피팅 및 적합도 평가 : SARIMA 모델을 학습 데이터에 맞춰 피팅한다.
 - > model.fit(dispatch=False)를 통해 SARIMA 모델을 학습하며, 이는 최적의 파라미터를 찾아 모델을 훈련한다.
 - > 모델 적합도를 평가하기 위해 AIC(Akaike Information Criterion) 및 BIC(Bayesian Information Criterion)를 사용한다. model.fit() 과정에서 내부적으로 계산된다.
6. 예측 데이터 생성 : 학습된 SARIMA 모델을 사용하여 향후 12개 (forecast_period=12) 동안의 월세를 예측한다.
 - > model_fit.forecast(steps=forecast_period)를 통해 예측값을 생성한다.
 - > 예측값은 스케일링된 상태이므로, 이를 원래 단위로 복원해야 한다. scaler.inverse_transform()을 사용하여 예측값을 원래 값으로 변환한다.

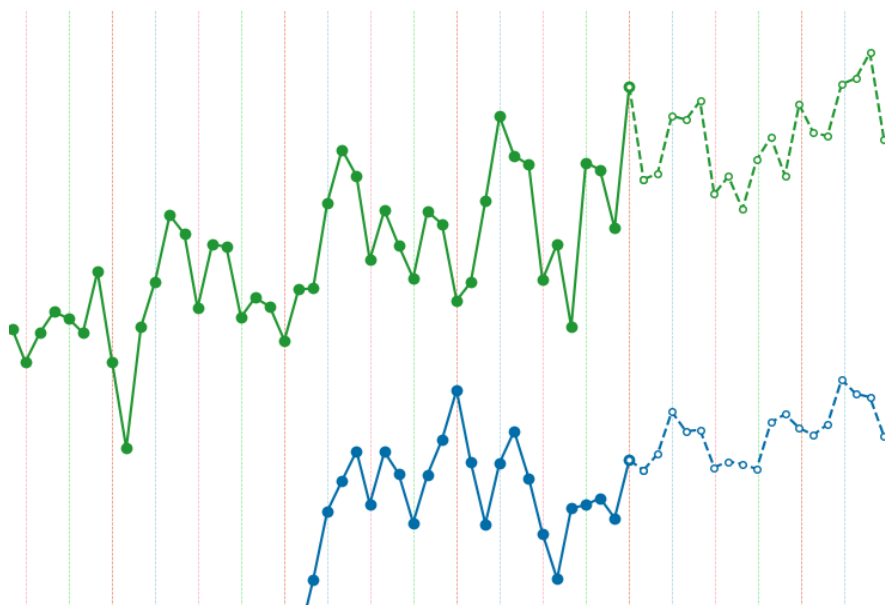
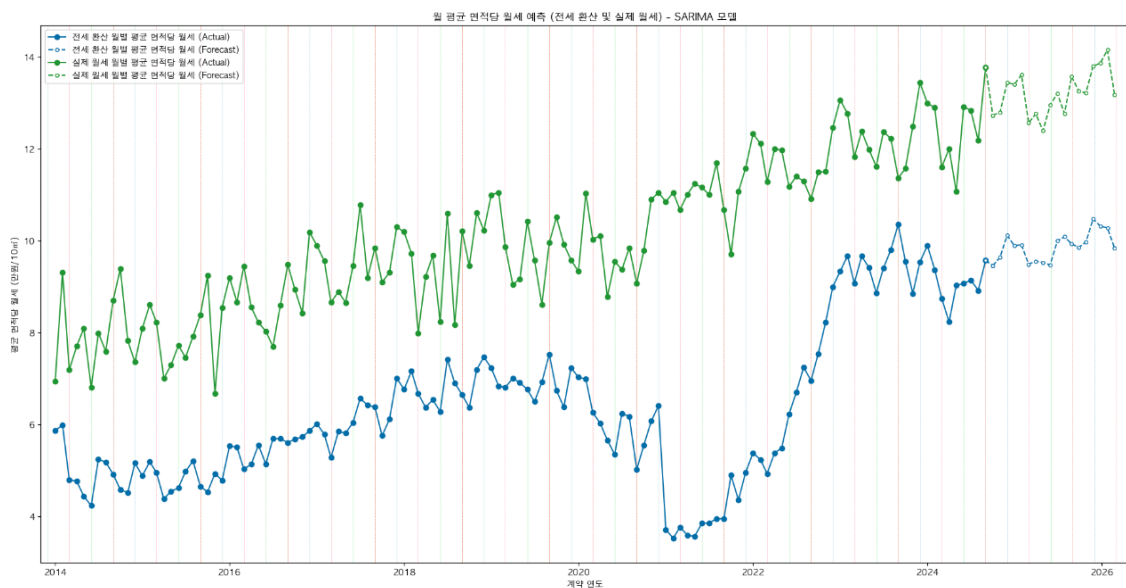
7. 날짜 생성 : 예측값에 해당하는 날짜를 생성하여 예측 결과를 시계열 데이터로 나타낸다.

> forecast_dates = [last_date + timedelta(days=30 * i) for i in range(1, forecast_period + 1)]는 마지막 관측일 이후의 날짜를 1개월 간격으로 생성한다

8. 실제 데이터와 예측 데이터 표시 : 실제 데이터와 예측 데이터를 한 그래프에 표시하여 비교한다.

> ax.plot(monthly_avg_df.index.to_timestamp(), monthly_avg_df[column], label=f'{column} (Actual)', ...)는 실제 데이터를,

> ax.plot(forecast_results[column].index, forecast_results[column], label=f'{column} (Forecast)', ...)는 예측 데이터를 나타낸다.



<그림 10, 11 - SARIMA 모델을 활용한 월평균 면적당 월세 예측 그래프 & 확대>

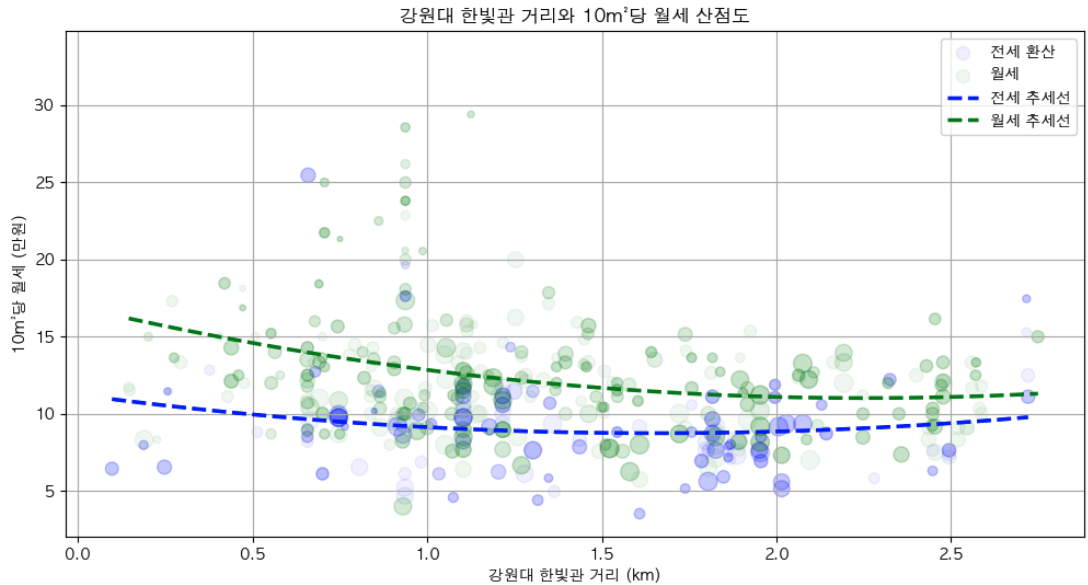
SARIMA 모델은 계절적 변동성과 장기적인 상승 추세를 모두 반영하여 강원대학교 인근 월세 데이터를 효과적으로 예측하였다. 예측 결과에 따르면, 2024년부터 2026년까지 평균 면적당 월세는 지속적인 상승이 예상된다. 또한, 봄, 가을 시즌에 부동산 가격이 상대적으로 저렴하다는 것을 확인해 볼 수 있다.

- 강원대학교와의 거리와 면적당 월세 간 상관관계

강원대학교로부터의 거리와 면적당 월세 간의 상관관계를 산점도로 표현했다. 월세와 전세 환산 데이터를 색상으로 구분하고, 각 데이터 포인트는 개별 매물을 나타냈다. 추가로 다항 회귀 분석을 적용하여 추세선을 그렸다. (기준은 강원대 한빛관 위도, 경도)

추세선 분석 : 다항 회귀 분석을 적용하여 거리와 월세 간의 비선형적인 관계를 시각적으로 나타냈다. 추세선은 거리가 가까울수록 월세가 더 높으며, 거리가 증가함에 따라 월세가 완만히 감소하는 경향을 확인할 수 있었다.

거리가 1km 이하일 때 월세가 가장 높으며, 1~2km 구간에서 월세가 감소하는 경향을 보였다. 다만, 특정 지역에서는 이와 다른 패턴이 관찰되었다. (그림 12 참고)



<그림 12 - 강원대와 매물 사이 거리에 따른 면적당 월세 확인 : 산점도 및 추세선>

6. 데이터 분석 및 인사이트 도출

6.1 분석 방법 및 기법

본 분석에서는 강원대학교 인근 부동산 데이터를 기반으로 월세 및 전세 시장의 트렌드를 파악하고, 미래의 가격 변화를 예측하며, 지역적 특성 분석을 위해 다양한 기법을 활용했다.

- 연도-월별 평균 면적당 월세 추이 분석

월별 데이터를 기반으로 평균 면적당 월세(단위: 만원/10㎡)를 계산하고, 이를 꺾은선 그래프로 시각화하였다. 데이터의 전반적인 경향성을 파악하기 위해 다항 회귀 분석을 적용하여 비선형적인 트렌드 라인을 도출하였다. 월세의 시간에 따른 변화를 파악하여 부동산 시장에서 발생하는 계절적 요인, 외부 요인의 영향을 분석하고, 장기적인 트렌드를 확인하고자 했다.

- SARIMA 모델을 활용한 월세 예측

시계열 분석 기법인 SARIMA(Seasonal AutoRegressive Integrated Moving Average) 모델을 사용하여 과거 데이터를 기반으로 향후 2년(2024~2026년)의 월세를 예측하였다. 모델 파라미터는 AIC 및 BIC 값을 통해 최적화했으며, 계절적 및 비계절적 요인을 모두 고려하였다. 계절적 요인과 장기적 트렌드를 반영한 예측 결과를 통해 부동산 시장의 안정성 및 중장기적 전망을 도출하고, 학생 및 임차인들에게 유용한 정보를 제공하고자 했다.

- 강원대학교와의 거리와 면적당 월세 간 상관관계 분석

산점도를 활용하여 강원대학교와의 거리(단위: km)와 면적당 월세(단위: 만원/10㎡) 간의 관계를 시각적으로 표현하였다. 데이터의 경향성을 명확히 하기 위해 다항 회귀 분석을 적용하여 추세선을 도출했다. 강원대학교와의 접근성이 월세 책정에 미치는 영향을 분석하고, 공간적 특성과 상관관계를 파악하여 부동산 시장의 특성을 이해하고자 했다.

6.2 주요 인사이트 및 해석

- 연도-월별 평균 면적당 월세 추이 분석

평균 면적당 월세는 전반적으로 상승하는 추세를 보이며, 특히 2020년 이후 큰 폭의 변동성을 보였다.

> 2020년~2021년 초 : 월세는 하락세를 보이다가 2021년 초에 급격히 감소한 후 반등하였다. 이는 2020년 7월에 시행된 '임대차 3법'(전세 2+2년 계약 연장 정책)으로 인해 전세 시장이 불안정해지면서 전세 물량이 감소하고, 전세 가격이 급등한 데 기인한 것으로 분석된다.

> 2021년 이후 : 전세의 회소성 증가로 인해 소비자들이 월세로 전환하는 현상이 나타났다. 이로 인해 월세 시장의 수요가 증가하면서 가격도 동반 상승했다.

월세 상승은 단순히 지역적 요인뿐만 아니라 외부 경제 요인(물가 상승, 금리 인상 등)과 정책적 변화가 복합적으로 영향을 미쳤다. 다항 회귀 분석을 통해 비선형적인 가격 변동 경향을 파악할 수 있었으며, 이는 시장의 장기적 불안정을 시사한다. (그림 9 참고)

- SARIMA 모델을 활용한 월세 예측

SARIMA 모델은 월세의 계절적 변동과 장기적 추세를 반영하여 다음과 같은 예측 결과를 도출했다.

＞ **계절적 변동** : 대학 인근 매물 특성상 이사가 많은 여름과 겨울 시즌에 평균 면적당 월세가 높다는 것을 확인할 수 있다. 만약 봄, 가을 시즌에 이사 계획을 세운다면 조금이라도 낮은 가격에 거래를 진행 할 수 있게 될 것이다.

＞ **장기적 상승** : 2024년부터 2026년까지 평균 면적당 월세는 꾸준히 상승할 것으로 예상된다. 이는 지역 내 수요 증가 및 외부 경제적 요인(물가 상승, 금리 변화 등)에 기인한 것으로 보인다.

SARIMA 모델은 단기적 변동성뿐만 아니라 장기적인 월세 상승을 정확히 반영했다. 이러한 예측은 학생 및 임차인들에게 향후 월세 상승에 대비할 수 있는 정보를 제공하며, 부동산 투자자들에게는 안정적인 투자 기회를 제시한다. 계절적 변동을 통해 여름철과 같은 특정 시기의 수요 집중 현상을 예측하여 실질적인 시장 전략을 수립할 수 있다. (그림 10, 11 참고)

- 강원대학교와의 거리와 면적당 월세 간 상관관계 분석

강원대학교와의 거리와 면적당 월세 간의 관계를 분석한 결과, 다음과 같은 경향이 확인 되었다.

＞ **1km 이내** : 월세가 가장 높게 나타났으며, 이는 학교와의 근접성이 월세 책정에 중요한 요인임을 보여준다.

＞ **1~2km 구간** : 거리가 증가함에 따라 월세가 완만하게 감소하는 경향이 나타났다.

＞ **특정 지역** : 예외적으로 월세가 높은 매물이 분포한 지역도 존재하며, 이는 상권, 대중 교통 접근성 등 다른 요인의 영향을 받았을 가능성이 있다.

강원대학교와의 접근성은 학생 및 교직원의 수요를 집중시키는 주요 요인으로, 학교와 가까운 지역일수록 월세가 높게 책정된다. 일부 지역에서는 거리 외에도 상권과 교통 인프라가 월세 형성에 영향을 미치는 것을 확인할 수 있다. (그림 12 참고)

- 지도 기반 매물 분포 시각화

강원대학교 주변 매물의 공간적 분포를 분석한 결과, 월세와 전세 환산 매물의 분포에서 다음과 같은 특성을 확인하였다.

＞ 강원대학교와 가까운 지역(효자동, 석사동)에 월세 매물이 집중되었으며, 이 지역은 학생들의 주거 수요가 높은 것으로 분석된다.

＞ 특정 지역(후평동)에서는 면적당 월세가 높은 매물이 다수 확인되었으며, 이는 주변 상권과 교통 접근성에 따른 가격 형성 요인을 반영한다.

지도 기반 분석은 강원대학교 인근 지역의 부동산 시장 구조를 이해하는 데 중요한 정보를 제공하였다. 특정 지역에서의 높은 월세 매물 집중 현상은 부동산 수요가 지역 특성에 따라 얼마나 다르게 작용하는지를 명확히 보여준다. (그림 8 참고)

7. 결론 및 제언

7.1 분석 결과 요약

본 프로젝트는 강원대학교 인근 부동산 시장을 다각도로 분석하여 지역 내 부동산 가격 형성 요인과 변동성을 이해하고, 미래 시장의 예측 가능성을 도출하고자 하였다. 주요 분석 결과는 다음과 같다.

- 장기적 월세 상승 :

강원대학교 인근 부동산 시장은 시장 안정성과 강원대학교의 위치적 특성에 의해 월세가 지속적으로 상승하는 추세를 보였다. 이는 해당 지역의 주거 수요가 꾸준히 유지되며, 외부 요인(물가 상승, 대출 금리 변화 등)이 가격 상승에 영향을 미친 것으로 해석된다.

- 거리와 가격의 상관관계 :

학교와의 접근성이 월세 책정에 주요한 영향을 미쳤으며, 학교에 가까운 지역(효자동, 석사동)일수록 월세가 높게 나타났다. 반면, 특정 지역(후평동 등)에서는 상권, 대중교통 접근성 등이 추가적인 요인으로 작용하여 예외적인 가격 패턴이 확인되었다.

- 예측 가능성 :

SARIMA 모델을 통해 월세의 계절적 변동성과 장기적인 상승 추세를 모두 반영한 예측 결과를 얻었다. 이는 시장 참여자(학생, 임차인, 부동산 투자자 등)들에게 향후 월세 상승에 대비할 수 있는 실질적인 의사결정 자료를 제공할 수 있다.

이러한 분석 결과는 강원대학교 인근 지역에 이사를 계획 중인 학생들뿐만 아니라, 부동산 투자자들에게도 유용한 정보를 제공하며, 부동산 시장의 구조적 이해를 돕는다.

7.2 프로젝트의 의의 및 한계

본 프로젝트는 강원대학교 인근 지역 부동산 데이터를 체계적으로 분석하여 부동산 시장의 가격 형성 요인과 시계열적 특성을 명확히 이해하는 데 기여하였다. SARIMA 모델을 활용한 시계열 예측과 다항 회귀 분석을 통해 시장의 계절적 변동성과 장기적 추세를 효과적으로 파악할 수 있었다. 지도 기반 시각화는 부동산 매물의 공간적 분포와 지역별 특성을 직관적으로 이해할 수 있었다.

본 연구는 강원대학교 인근 일부 지역(효자동, 석사동, 퇴계동, 후평동)의 데이터를 기반으로 진행되었으며, 다른 지역으로의 일반화도 진행하고 싶다.

SARIMA 모델은 시간에 따른 계절적 및 비계절적 패턴을 잘 반영하였으나, 외부 경제 요인(금리 변화, 정책 변화 등)을 반영하지 못한다. 특정 지역적 특성과 같은 비정량적 데이터를 모델링에 포함하지 못한 점도 한계로 작용한다. 추가로 층수, 건물 방향, 대출 금리 등 매물

가격에 영향을 미칠 수 있는 다양한 요인을 추가적으로 분석하지 못하였다.

7.3 추가 연구 방향 및 활용 방안

- 추가 연구 방향 :

＞ 주변 상권 및 남춘천역과 매물 사이 관계성 분석 : 매물의 가격과 주변 상권, 남춘천역 등 주요 교통 시설과의 거리 및 접근성을 추가적으로 분석하여, 공간적 요인이 부동산 시장에 미치는 영향을 더욱 깊이 이해하고자 한다.

＞ 층수, 건물 방향, 금리 등의 추가 변수 분석 : 매물의 층수, 건물 방향, 금리 등 다양한 요인과 매물 가격 간의 상관관계를 분석하여, 부동산 가격 형성에 영향을 미치는 세부 요인을 도출하고자 한다.

＞ 다양한 모델 적용 : SARIMA 모델 외에도 LSTM(Long Short-Term Memory), Prophet, XGBoost 등 다양한 예측 모델을 학습하고 적용하여, 부동산 데이터의 비선형적 특성을 반영한 예측 결과를 비교하고 최적화하고자 한다.

＞ 지역 분석의 일반화 : 본 연구에서 사용된 분석 방법론과 모델을 확장하여 춘천시 전체로 데이터를 확대 분석하고, 나아가 전국 단위의 통계 자료를 구축할 수 있을 것이다. 이를 통해 지역별 부동산 시장의 특성과 공통적인 요인들을 비교 분석할 수 있으며, 특정 지역에만 국한되지 않는 일반화된 인사이트를 도출할 수 있다. 전국 단위 분석은 부동산 정책 수립 및 투자 전략에서 더 큰 의사결정 자료로 활용 가능하며, 지역 간 가격 격차의 원인을 파악하는 데에도 기여할 것이다.

- 활용 방안 :

＞ 학생 및 임차인 지원 : 강원대학교 인근 부동산 시장 정보를 기반으로 한 시계열 분석 결과는 이사를 계획 중인 학생들과 임차인들에게 효율적인 선택지를 제공할 수 있다.

＞ 부동산 투자 전략 : 예측된 월세 상승 추세와 지역별 특성을 활용하여 부동산 투자자들에게 신뢰할 수 있는 시장 분석 자료를 제공할 수 있다.

＞ 정책 수립 지원 : 강원대학교 인근 지역 부동산의 계절적 변동성과 가격 변동 요인에 대한 이해는 지역 정책 수립 및 공공 데이터 활용 전략에 참고자료로 활용 될 수 있다.

7.4 결론

본 프로젝트는 강원대학교 인근 부동산 시장을 다각도로 분석하여 가격 형성 요인과 시장 변동성을 이해하고, 미래 시장의 동향을 예측하는 데 의미 있는 결과를 도출하였다.

강원대학교와의 접근성은 부동산 가격 형성의 주요 요인으로 작용하며, 학교 근처 지역에서는 월세가 상대적으로 높게 책정되는 경향을 보였다. 월세는 계절적 요인(예: 이사철)과 외부 경제 요인(예: 물가 상승, 대출 금리 변화)으로 인해 단기적인 변동성을 보이면서도 장기적으로는 지속적인 상승 추세를 나타냈다. SARIMA 모델을 통해 월세의 계절적 패턴과 장기적

상승 흐름을 예측함으로써, 시장 참여자들이 미래 월세 상승에 대비할 수 있도록 실질적인 의사결정 자료를 제공하였다. 지도 기반 시각화를 통해 매물의 공간적 분포를 분석함으로써 지역별 특성과 시장 구조를 명확히 이해할 수 있었다.

본 프로젝트는 지역적 맥락에 기반한 데이터 분석과 미래 시장 예측의 가능성을 보여주었으며, 부동산 시장 분석의 실질적 가치를 입증하였다. 향후 추가 연구와 데이터 확장을 통해 보다 정교하고 포괄적인 분석이 이루어진다면, 강원대학교 인근뿐 아니라 다른 지역의 시장 참여자들에게도 유익한 자료로 활용될 수 있을 것이다.