

# 프로젝트 결과 보고서

GG조 (이찬호, 정소현, 정용주) | 2019.12.05~2019.12.18 (14일)

주제 선정	독립 및 자취를 위한 사람들에게 합리적인 집 값 정보를 제공하기 위해 전국의 모든 주거형태의 매매, 전세, 월세 가격 데이터를 수집 후 집 값을 분석
담당 분야	Python 크롤링, 분석 - 팀 전원 HTML - 정용주
사용 환경	Jupyter notebook (Python), Visual Studio Code (HTML)
프로젝트 내용	<ol style="list-style-type: none"><li>데이터 수집<ul style="list-style-type: none"><li>- 네이버 부동산 크롤링</li><li>- KB 부동산 주택가격동향 (19.11 기준)</li><li>- KB 부동산 시세통계 면적당 평균가 (19.11 기준)</li></ul></li><li>데이터 전처리<ul style="list-style-type: none"><li>- 네이버 부동산 크롤링 자료의 현재 층 컬럼을 생성하기 위해 층 컬럼에서 현재 층과 전체 층으로 분류 후 각 컬럼을 생성</li><li>- 생성된 현재 층 컬럼을 최저층은 범주 1, 저층은 범주 2, 중층은 범주 3, 고층은 범주 4, 최고층은 범주 5로 분류 후 현재 층 범주 컬럼 생성</li><li>- 필요 없는 컬럼인 층, 현재 층, 전체 층을 삭제 후 새로운 csv 파일로 저장</li></ul></li><li>데이터 분석<ul style="list-style-type: none"><li>- 서울특별시와 6개의 광역시(부산, 대구, 인천, 광주, 대전, 울산)의 아파트 크기별로 가격을 비교 분석</li><li>- 서울특별시의 25개 지역 구의 매물가를 boxplot을 통해 비교 분석</li><li>- 서울특별시의 25개 지역 구의 매매 건수와 가격을 시계열 추세를 이용하여 분석</li><li>- 시계열 추세를 이용하여 데이터, 추세요인, 계절요인, 불규칙요인을 분석</li><li>- 자기상관함수, 부분자기상관함수, 1차 차분을 이용하여 분석 법의 신뢰성 확인</li><li>- 서울특별시의 25개 지역 구와 층 범주로 다항회귀분석(면적, 매매가격)을 진행</li></ul></li><li>시각화<ul style="list-style-type: none"><li>- 서울특별시와 6개의 광역시의 아파트 크기별 가격 비교 (Axes3D)</li></ul></li></ol>

- 서울특별시의 25개 지역 구의 매물 가격 비교 분석 (Boxplot)
- 서울특별시의 25개 지역 구의 지난 10년간 매매 건수 그래프
- 서울특별시의 25개 지역 구의 매매 건수 시계열 추세 그래프
- 매매 건수의 ACF (자기상관함수), PACF (부분자기상관함수) 그래프
- 매매 건수의 1차 차분 그래프
- 매매 건수의 ARIMA 모형 그래프
- 서울특별시의 25개 지역 구의 지난 3년간 매매 가격 그래프
- 서울특별시의 25개 지역 구의 매매 가격 시계열 추세 그래프
- 매매 가격의 ACF (자기상관함수), PACF (부분자기상관함수) 그래프
- 매매 가격의 1차 차분 그래프
- 매매 가격의 ARIMA 모형 그래프
- 서울특별시의 25개 지역 구와 총 범주 다항회귀분석(면적, 매매가격) 그래프

#### 5. 분석 예측

- ARIMA 모형을 사용하여 수집 한 데이터 학습 후 미래의 매매 건수와 매매 가격을 예측

#### 6. 분석 결과

- 아파트의 면적이 커질수록 매매가가 상승하는 것으로 결과 도출
- 서울특별시와 6개 광역시 중 서울특별시의 가격이 가장 높은 것으로 결과 도출
- 서울특별시의 전체 지역 구의 매물가중 상위 3개의 지역이 강남구, 서초구, 송파구이며 하위 3개의 지역이 금천구, 관악구, 도봉구로 확인
- 서울특별시의 전체 지역 구의 매매 건수가 매년 비슷한 형태를 띄는 것을 확인
- 서울특별시의 전체 지역 구의 3년간 매매 가격이 비슷한 형태로 상승
- 서울특별시의 전체 지역 구는 높은 층일수록 가격이 상승하며 가격 상승의 폭이 큰 것으로 확인이 되었고 25개 구 중 매물가가 상위일수록 대부분의 가격이 상승하며 하위일수록 중층 이하의 층일수록 매매 가격이 하락하는 것을 볼 수 있음

#### 7. 후기

- 학부 때 배웠던 다양한 통계적인 내용을 바탕으로, 프로젝트에 무슨 통계 분석 방법을 써야 적합할지 고민하고 적용시키는 과정에서 다양한 통계 분석 방법들을 다시 되새길 수 있었다. 또한 파이썬으로 분석한 건 처음이었는데, 어려웠지만 파이썬을 더 열심히 공부해서 파이썬으로 데이터 분석하는 데 능숙해지고 싶다고 생각했다.

