

데이터 마이닝 7조 기획서



12132591 정동호

12131820 이건도

12161890 하나영

< 목 차 >

1. 주제 및 선정 이유

2. 분석방법

3. 데이터 조사

4. 변수 설명

5. 해결수단 선택

6. 실행방법 수립

1. 주제 및 선정 이유

- **주제**

인천광역시 아파트 단지 데이터를 이용한 아파트 제곱미터당 평균 매매 가격 예측 및 영향을 주는 변수 분석

- **분석 단위**

최근 3년(16년 1월부터 18년 12월까지의 월 기준) 인천광역시 단지별 아파트 평균 매매 가격 기준

- **선정 이유**

이 주제의 선택 이유는 두가지입니다. 먼저, 정부의 부동산 정책에도 불구하고 아파트 가격은 눈에 띄는 감소세를 보이지 않고 오히려 증가세를 보이기도 하면서 아파트 가격에 영향을 주는 요인이 무엇인지 궁금하게 되었습니다. 그리고 저희가 선정한 변수와 아파트 가격 간의 관계 분석을 통해서 각 변수 별로 아파트 가격에 얼마나 영향을 주는지, 해당 변수를 통제한다면 무슨 변화가 일어나는지, 그리고 현재 정부의 부동산대책의 방향성을 판단하거나 새로운 대책이 무엇이 될 수 있는지 이러한 분석을 통해 인천 아파트 가격 동향 예측을 통해 알고 싶었습니다.

2. 분석방법

저희가 생각한 분석 방법은 다음과 같습니다.

- 인천아파트 가격과 관련된 모든 독립변수들에 대한 관계를 분석할 수 있는 중회귀분석
- 각 독립변수와 인천 아파트 가격 간 상관 관계를 분석할 수 있는 회귀분석
- 중회귀분석으로 도출된 예측 가격과 실제 가격간 비교 분석
- 인천아파트 실제 가격의 시계열 분석
- 인천아파트 예측 가격의 시계열 분석
- 인천아파트 가격의 분포의 시각화 분석(지도)

3. 데이터 조사

국토교통부>실거래가 공개시스템 : 아파트 단지별 거래가격, 면적, 건축연도(노후도)

국토교통부>아파트 주거환경 통계 : 인천 대중교통 및 교육시설 인접

통계청>국가통계목록 : 미분양 규모, 혼인건수

한국감정원 : 실매매가격 변동률

부동산114 >REPS 3.0 : 단지비교, 거래건수, 단지규모

공공데이터 포털

인천광역시청> 통계정보 : 인천 구 동별 인구수 및 세대수

인천교육청> 학교현황 : 초,중,고 현황

그 외 구글링 등 인터넷 서치를 통해 정보 수집

4. 변수 설명

1) 아파트 매매 거래건수(거래량) – 월간 인천광역시 구별 거래규모 기준 건수 변화

~~2) 실거래 가격~~

~~3) 실매매가격 변동률 – 2번이상 거래된 아파트를 대상으로 한 가격 변동률~~

4) 단지 규모 – 아파트별 규모

5) 면적 – 아파트 면적 $16.5m^2$ (5평)기준

6) 아파트 노후도 – (2019년 – 건축연도)

7) 아파트 브랜드 가치(브랜드 순위에 따른 임의의 가중치)

- 2018년 베스트 아파트 브랜드 종합 순위 Top 10 매매가격 변화

8) 역세권 – 해당동에 지나는 지하철 노선도 개수

9)학군 -인천광역시 초, 중, 고 동 별 합산 갯수

10) 공원 및 문화시설 - 공원, 대형할인점, 영화관, 스타벅스 시설 동별 개수로 각각 설정

11) 전세가격 – 인천광역시 동별 아파트 단지 전세가격

~~12) 미분양 규모 - 준공 후 미분양규모(월별, 구별 파악)~~

13) 공시지가 변동률 - 전년 대비 공시지가 변동률들의 평균값

14) 기준금리- 한국은행이 금융기관과 환매조건부증권(RP) 매매, 자금조정 예금 및 대출 등의 거래를 할 때 기준이 되는 정책금리

(시중금리 - 한국은행의 기준금리가 시중은행의 대출금리에 영향)

~~15) 기준시가 - 토지와 건물 포함 재산가격~~

16) 서울 집값 - 서울 강남권 집값이 수도권 부동산 시장 흐름에 영향을 준다고 생각

17) 개발호재 유무 - 도시정비 및 택지개발사업, 교통시설 예정지역 존재 대기업 입주, 산업단지조성, 신설 도로, 도로확장.행정기관 이전, 아파트 신규 물량 등 -> 있으면 1 없으면 0을 이진변수로 구분

18) 인구 수(인구 증가율 포함) - 행정구역별, 성별, 나이별 인구수

19) 세대수 - 인천지역 구 동별로 세대수 조사 (가능하다면 단지별 세대수까지)

~~20) 유동인구 - 인구 전입 및 전출~~

21) 혼인건수 - 인천시 구별 월별 혼인건수

5. 해결수단 선택

- 인천 각 지역들의 아파트 가격에 영향을 끼치는 여러 변수들을 후에 배울 데이터 마이닝 기법과 통계적 분석들로 설명, 예측한다.
- 투자 수요와 실수요가 다를 수도 있는 집값의 특성들을 잘 이해하고 파악한다. 즉, 집값의 거품을 경계하며 분석한다.
- 같은 구 안에서도 특정 지역에 의해 평균 매매가의 변동률이 클 수 있다는 점 등을 생각한다.
- 서울과 다르게 상대적으로 저렴할 수 있는 인천의 특수성과 차이점을 인지한다.

➔ Presentation 및 Report의 결과물 형태.

6. 실행방법의 수립

누가?

- 팀원 세 명 모두 역할을 분담하여 예측 및 분석.

언제 무엇을?

- 주제에 대한 사전조사(개인 별로 계속 진행) ~ 3/26
- 데이터 준비(변수설정과 데이터수집, 데이터 전처리) ~ 3/28
- EDA ~ 4/24(중간보고1)
- 모델링 적용(변수선택과 분석기법사용) ~ 5월 초중순
- 검증(모델 적합도 분석) ~ 5월 중순
- 발표 준비 ~ 마감 전

어떻게?

- 인터넷 search, 통계청, 국토교통부 통계누리, '부동산114' 공모전 자료 등 사용

기대효과

- 분석을 통해 인천 아파트 매매가에 변동을 주는 요인들을 파악하여, 정부의 부동산 대책의 방향성 또는 새로운 정책 대안 등을 제시할 수 있을 것이다. 과제를 하며 집값에 대한 우리가 몰랐던 부분을 배울 수 있을 것 같다.

한계

- 한계점으로는 세세한 아파트 단지별로 분석이나, 조사된 변수들의 자료 부족으로 완벽한 예측이 힘들 것이다. 전문가도 예측하기 힘든 집값 예측 조건들의 다양한 외적 변수와 예상치 못한 변수들까지는 전부 포함시키지 못하는 한계점이 있다.
- 모델링 적용에 있어 학사 수준 분석 방법이 실제 적용하기에는 다소 무리가 있을 수 있다는 한계가 있을 수 있다.
- 실거래가 경우, 시차를 두고 거래를 하는데 거래건수마다 각각 다르기 때문에 수집하기 힘들.
- 최근 3년간 거래된 데이터이기 때문에 예상치 못한 미래의 사건, 사고에 대해서는 예측의 정확도의 한계가 존재.
- 단지 내 아파트에도 층 별, 방향 별 등 세부적인 사항보다는 단지를 기준으로 예측.