

빅데이터 분석 결과 보고서

신도시의 성공적인 병 · 의원 개원을 위한 입지 분석 서비스

IPAD

(Intelligent Passionate Active Developers)

장준호, 이다연, 박예린

목차

1. 분석 배경 및 기획의도

1-1 분석 배경

1-2 기획 의도

2. 분석 목표

3. 분석 데이터

3-1 데이터 수집

3-2 데이터 정합성 검증

3-3 데이터 전처리

4. 분석 과정 및 결과

5. 활용방안

6. 참고문헌

1. 분석 배경 및 기획 의도

1-1 분석 배경

코로나19라는 국가적 유행병 현상, 경제 및 부동산 환경, 환자의 의료 지식 향상과 소비 심리 등 다양한 요인으로 인해 개원 시장은 변화해 왔다. 치과 의원도 예외는 아닐 것이며, 치과의사는 개원과 폐업에 있어 수많은 고민을 거듭하고 있다. 객관적인 자료를 통해 치과 의원의 개·폐업 현황을 살펴보았다. 2021년 치과 의원의 개업률은 3.38%, 폐업률은 2.78%로 전년 대비 개업률은 0.38%p, 폐업률은 0.34%p 증가하였다. 2012년도를 제외하고 개업률이 폐업률 보다 높았다. 치과 의원의 개업률은 2012년도에 3.99%로 가장 높았고, 2020년도에 2.44%로 가장 낮았으며, 폐업률은 2004년도에 5.11%로 가장 높았고, 2020년도에 3.0%로 가장 낮았다. 2003년도부터 2021년도까지 치과 의원의 개·폐업 추세는 감소하는 추세이며 폐업률 보다 개업률의 감소추세가 더 가파르게 나타났다. 그럼에도 불구하고 전체 치과 의원의 절반이 위치한 수도권의 시·군·구별 개·폐업 현황을 보면 서울은 주요 지역의 폐업률이 증가한 반면 경기도는 신도시 지역의 개업률이 증가하였다.

1-2 기획 의도

병원 개원을 준비하는 의사를 위한 신도시 입지 분석 서비스를 개발하여 개원 지역 선택에 도움을 제공한다. 신도시의 주택 데이터, 인구 데이터 등을 분석해서 예상 주거 인구를 제공해 폐업하지 않고 안정적인 병원 운영을 할 수 있도록 도움을 제공한다.

2. 분석 목표

주택 데이터와 인구 데이터 등 빅데이터를 활용하여 위례신도시 3개의 시·구의 주거 인구 분석 및 예측

3. 분석 데이터

3-1 데이터 수집

■ 주택 데이터

주택 데이터는 주택 수와 가구 수를 사용했다. 데이터는 통계지리정보서비스에서 JSON 형식으로 수집해서 Excel 형식으로 변환했다. 데이터의 주기는 1년이고 2015년부터 2022년의 데이터를 포함하고 있다.

데이터: 주택 수, 가구 수

■ 인구 데이터

인구 데이터는 출생 건수와 사망 건수를 사용했다. 데이터는 통계청에서 Excel 형식으로 수집했다. 데이터의 주기는 1년이고 2015년부터 2022년의 데이터를 포함하고 있다.

데이터: 출생 건수, 사망 건수

■ 사업체 데이터

사업체 데이터는 사업체 수를 사용했다. 데이터는 통계지리정보서비스에서 JSON 형식으로 수집해서 Excel 형식으로 변환했다. 데이터의 주기는 1년이고 2015년부터 2021년의 데이터를 포함하고 있다.

데이터: 사업체 수, 종사자 수

3-2 데이터 정합성 검토

데이터 분석을 위해 데이터 정합성 검정을 한 결과 시·구의 사업체 수, 종사자수 데이터에서 문제를 발견했다.

송파구 위례동, 성남시 위례동, 하남시 위례동 3개의 2022년 사업체 수, 종사자수에 대한 정보가 결측 값을 가지고 있었다.

결측 값은 NULL 값으로 처리하고 데이터 분석 단계에서는 제외하고 진행했다.

3-3 데이터 전처리

데이터를 분석에 사용하기 위해 R을 이용해서 데이터를 matrix 형태로 저장했다. 3개의 시·구를 비교하기 위해 총 3개의 데이터 파일을 만들었다.

데이터 분석 전 사전에 회귀 분석을 진행한 결과 사업체 수는 주거 인구에 영향이 적은 것으로 확인되었다. 그래서 회귀 분석을 진행할 때 영향이 가지 않도록 사업체 수를 matrix에서 삭제했다.

4. 분석 과정 및 결과

주거 인구를 예측하기 위해 다중 선형 회귀 분석을 했다. 회귀 분석을 위한 변수 선택법으로 단계적 선택법을 사용해 종속변수에 기여도가 높은 독립 변수를 선별했다.

위례 신도시의 주거 인구를 예측하기 위해 공공데이터를 수집하고 R을 이용해 분석했다. 주거 인구를 종속변수로 정하고 출생 건수, 사망 건수, 주택 수, 가구 수를 독립변수로 지정하여 회귀분석 모델을 구했다.

4-1 송파구 위례동 주거 인구 분석 및 시각화

	B	t value	p
intercept	624.7812	0.927	0.422
출생 건수	1.4892	0.921	0.425
사망 건수	5.4639	0.333	0.761
주택수	2.0687	1.925	0.150
가구수	0.9076	0.973	0.402
Adjusted R-squared	0.9997		
p-value	4.524e-06		

[표 1-1 송파구 위례동 주거 인구 회귀분석]

송파구 위례동 회귀분석 결과를 stepAIC() 함수로 단계적 선택법을 진행해 독립변수를 선별했다.

	B	t value	p
intercept	225.06441	1.369	0.22
주택수	3.12072	172.799	2.53e-12
Adjusted R-squared	0.9998		
p-value	2.534e-12		

[표 1-2 단계적 선택법 진행 후 송파구 위례동 주거 인구]

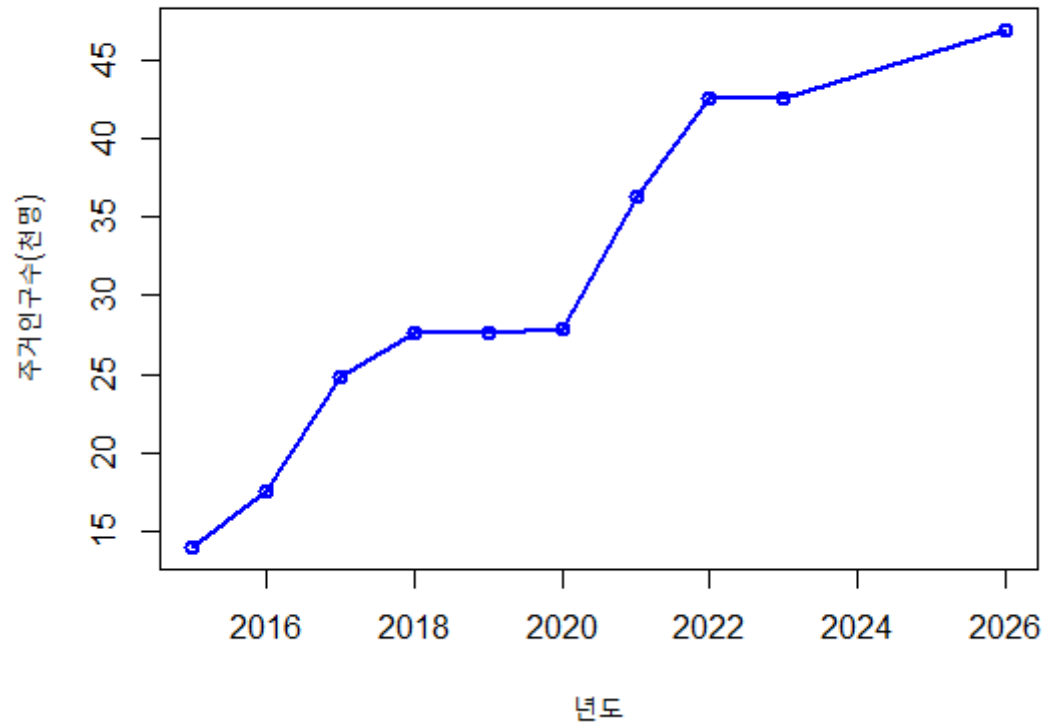
단계적 선택법을 통해 주택 수가 독립변수로 선별됐다.

송파구 위례동 모델의 회귀 식은 다음과 같다.

$$\text{송파구 위례동 주거 인구} = 225.06441 + (3.12072 * \text{주택 수})$$

독립변수 주택 수에 입주 예정인 주택의 수를 더하고 나머지 독립변수는 최근 연도와 같다고 가정하면, 송파구 위례동 주거 인구 추이는 다음과 같다.

송파구 위례동 주거인구 변화



[그래프 1-1 송파구 위례동 주거 인구 추이]

4-2 성남시 위례동 주거 인구 분석 및 시각화

	B	t value	p
intercept	146.1135	0.302	0.7825
출생 건수	-8.0886	-1.190	0.3196
사망 건수	-57.8077	-5.717	0.0106
주택수	4.2751	4.791	0.0173
가구수	-0.2439	-0.347	0.7515
Adjusted R-squared	0.9991		

p-value	1.802e-05
----------------	-----------

[표 2-1 성남시 위례동 주거 인구 회귀분석]

성남시 위례동 회귀분석 결과를 stepAIC() 함수로 단계적 선택법을 진행해 독립변수를 선별했다.

	B	t value	p
intercept	160.4309	0.377	0.72561
출생 건수	-6.8545	-1.340	0.25126
사망 건수	-59.0073	-7.030	0.00216
주택수	3.9788	17.386	6.42e-05
Adjusted R-squared	0.9993		
p-value	2.825e-07		

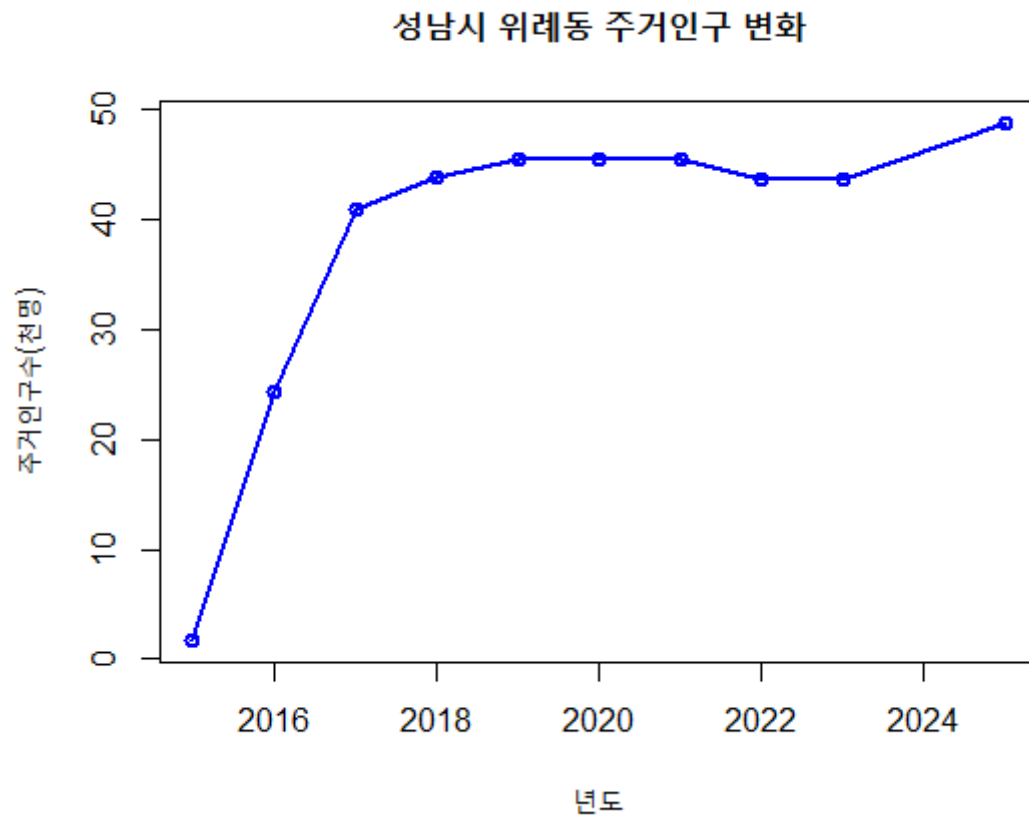
[표 2-2 단계적 선택법 진행 후 성남시 위례동 주거 인구]

단계적 선택법을 통해 출생 건수, 사망 건수, 주택 수가 독립변수로 선별됐다.

성남시 위례동 모델의 회귀 식은 다음과 같다.

성남시 위례동 주거 인구 = 160.4309 + (-6.8545 * 출생 건수) + (-59.0073 * 사망 건수) + (3.9788 * 주택 수)

독립변수 주택수에 입주 예정인 주택의 수를 더하고 나머지 독립변수는 최근 연도와 같다고 가정하면, 성남시 위례동 주거 인구 추이는 다음과 같다.



[그래프 2-1 성남시 위례동 주거 인구 추이]

4-3 하남시 위례동 주거 인구 분석 및 시각화

	B	t value	p
intercept	711.3380	1.833	0.1642
출생 건수	-8.6942	-1.360	0.2669
사망 건수	8.7123	0.456	0.6794
주택수	1.9374	3.243	0.0477
가구수	1.5263	3.181	0.0500
Adjusted R-squared	0.998		

p-value	6.477e-05
----------------	-----------

[표 3-1 하남시 위례동 주거 인구 회귀분석]

하남시 위례동 회귀분석 결과를 stepAIC() 함수로 단계적 선택법을 진행해 독립변수를 선별했다.

	B	t value	p
intercept	652.4941	1.991	0.1174
출생 건수	-7.9297	-1.439	0.2236
주택수	1.8553	3.637	0.0220
가구수	1.6320	4.339	0.0123
Adjusted R-squared	0.9984		
p-value	1.644e-06		

[표 3-2 단계적 선택법 진행 후 하남시 위례동 주거 인구]

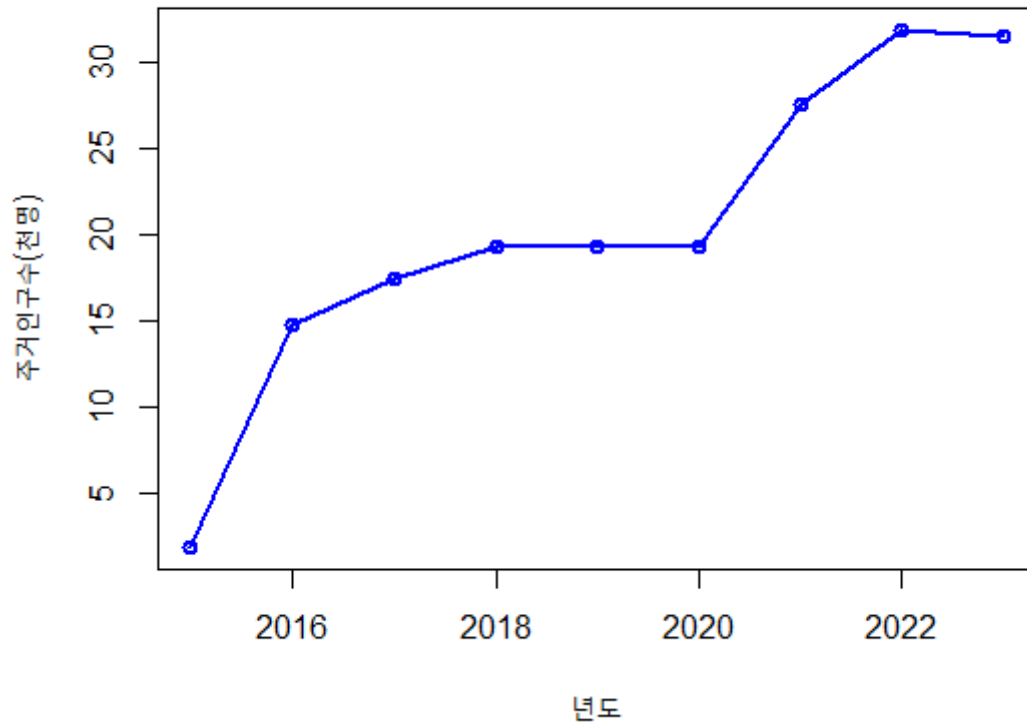
단계적 선택법을 통해 출생 건수, 주택 수, 가구 수가 독립변수로 선별됐다.

하남시 위례동 모델의 회귀 식은 다음과 같다.

하남시 위례동 주거 인구 = 652.4941 + (-7.9297 * 출생 건수) + (1.8553 * 주택 수) + (1.6320 * 가구수)

독립변수 주택수에 입주 예정인 주택의 수를 더하고 나머지 독립변수는 최근 연도와 같다고 가정하면, 하남시 위례동의 주거 인구 추이는 다음과 같다.

하남시 위례동 주거인구 변화



[그래프 3-1 하남시 위례동 주거 인구 추이]

회귀 분석에서 p-value(유의수준) 값은 구한 회귀모델이 의미 있는 모델인지를 나타내는 것으로, 이 값이 작을수록 의미 있는 모델이다. p-value 값이 유의 수준인 0.05보다 작으면 해당 독립 변수는 종속변수에 통계적으로 유의한 영향을 미치는 것으로 간주된다.

Adjusted R-squared (조정된 R제곱)값은 모델의 설명력을 나타내며 0~1 사이의 값을 갖고 이 값이 클수록 회귀모델이 현실을 잘 나타낸다.

결론적으로 p-value와 Adjusted R-squared 값에 의해 3개의 시·구의 회귀분석 모델은 유의하게 도출 되었다고 할 수 있다.

5. 활용방안

개원을 준비하는 의사는 본 프로젝트를 활용하여 서울과 경기 신도시 지역 주거 인구를 예측할 수 있으며, 예측 결과를 활용함으로써, 적합한 지역을 선택할 수 있습니다. 예측된 주거 인구는 해당 지역의 인구 동향과 수요를 파악하는데 도움을 주고, 이를 통해 개원한 후에도 지속적으로 활발한 환자 유입을 기대할 수 있는 장소를 선정할 수 있게 됩니다. 따라서 본 프로젝트는 개원 준비 단계에서 의사에게 중요한 역할을 보여줍니다.

6. 참고문헌

“치과의원 개·폐업 현황과 지역별 특성 이슈 레포트” 치과의료정책연구원(2022년 10월 제2022-09호)

홍성효 “세대별 시·군·구 간 인구이동 결정요인에 관한 실증분석” 서울도시연구 (2012 13권 1호 1~19페이지)

“인구 주택 총 조사”, “전국 사업체 조사” 통계지리정보 서비스

“인구 동향 조사” 국가 통계 포털