**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

1. **분석 목표**
2. **분석 데이터**

3-1 데이터 수집

3-2 데이터 정합성 검정

3-3 데이터 전처리

1. **분석 과정 및 결과**
2. **활용방안**
3. **참고문헌**

**1. 분석 배경 및 기획 의도**

**1-1 분석 배경**

2022년 개업 치과 병·의원은 800개이며 폐업한 치과 병·의원은 536개이다. 이는 10년 전인 2013년 대비 개업률 29.5%와 폐업률 30.4% 를 기록했다. 개업률이 가파르게 상승하던 10년 전을 제외하고는 폐업률이 높다는 사실을 알 수 있다. 그렇기 때문에 시장 경쟁을 이겨내지 못하면 폐업을 선택할 수밖에 없는 현실이 바로 치과 시장의 한계이다. 폐업하는 병원이 증가한다는 것은 경영이 쉽지 않은 것을 의미한다. 하지만 심평원이 발간한 손에 잡히는 의료 심사 평가 길잡이에 따르면 전국 치과의 절반 이상이 서울을 벗어나 수도권에 밀집해 있는 것으로 나타났다. 2021년과 2022년에는 신도시의 치과 개원이 활발하게 나타나 2021년 대비 2022년의 개업 수가 32.5% 증가한 것을 보여 주고 있다.

**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**

* 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

* 1. **데이터 정합성 검정**

데이터 분석을 위해 데이터 정합성 검정을 한 결과 사업체 데이터에서 2022년 사업체 수가 결측 값을 가진 것을 발견했다. 결측 값은 NULL 값으로 처리하고 데이터 분석 단계에서는 제외하고 진행했다.

* 1. **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 출산 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 |
| **출산 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 | | |

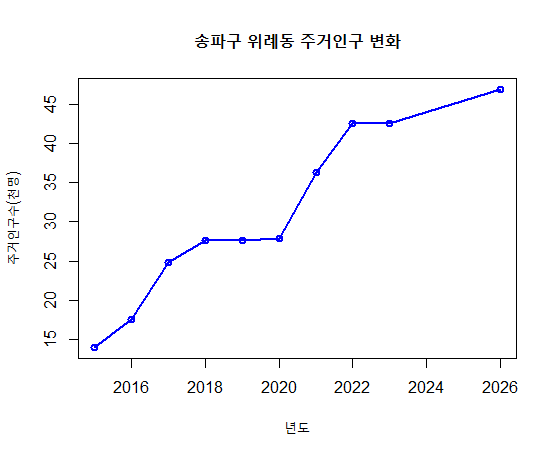
**[표 1-2 단계적 선택법 진행 후 송파구 위례동 주거 인구]**

단계적 선택법을 통해 주택 수가 독립변수로 선별됐다.

송파구 위례동 모델의 회귀 식은 다음과 같다.

**송파구 위례동 주거 인구 = 225.06441 + (3.12072 \* 주택 수)**

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



**[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 |
| **출산 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 |
| **출산 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 | | |

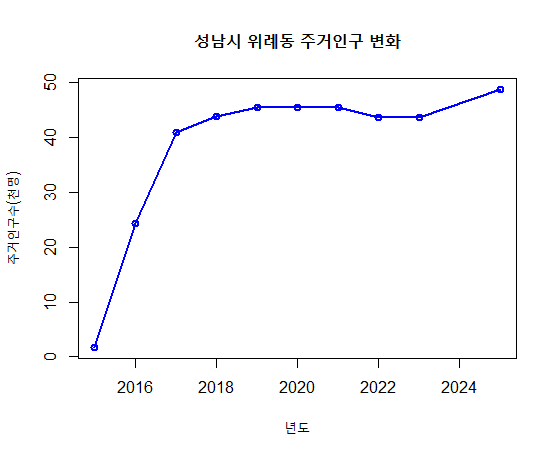
**[표 2-2 단계적 선택법 진행 후 성남시 위례동 주거 인구]**

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

성남시 위례동 모델의 회귀 식은 다음과 같다.

**성남시 위례동 주거 인구 = 160.4309 + (-6.8545 \* 출산 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 |
| **출산 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 |
| **출산 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 | | |

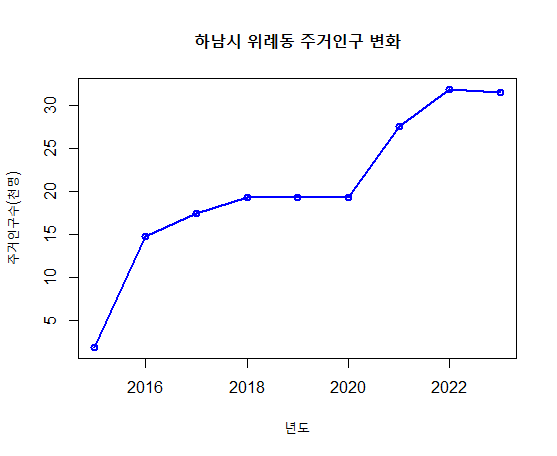
**[표 3-2 단계적 선택법 진행 후 하남시 위례동 주거 인구]**

단계적 선택법을 통해 출산 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 \* 주택수) + (1.6320 \* 가구수)**

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



**[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 해당 지역의 주거 인구를 예측할 수 있으며, 위 결과를 활용하여 적합한 지역을 선정할 수 있다.

**6. 참고문헌**