

# **빅데이터 분석 결과 보고서**

## **신도시의 성공적인 병원 개원을 위한 입지 분석 서비스**

**IPAD-1**

**곽동영, 권현준, 송도훈**

## 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 검사
  - 3.3 데이터 전처리
4. 분석 과정 및 결과
  - 4.1 서울 연도별 매출 분석 결과 및 시각화
  - 4.2 서울 지역 매출 분석 결과
  - 4.3 지역 매출 분석 결과 및 시각화
5. 활용방안
6. 참고문헌

# 1. 분석 배경 및 기획의도

## 1.1 분석 배경

전국 15,000여개의 치과 의원 중 33%에 해당하는 5,000개의 치과 의원이 서울에 밀집되어 있는 것으로 나타났다. 인구 10만명 대비 치과의원 수에서도 2023년 12월 기준 서울은 10만명 당 51.95개를 기록하여 45.08개의 광주, 40.54개의 부산과 큰 차이를 보이며 이미 포화된 상태라고 할 수 있다. 이에 따라 서울에서의 개원이 아닌 다른 지역에서 개원을 하려는 의사들이 증가하고 있다.

개원을 하기 위해서는 많은 자금이 들어가기 때문에 여러 사항들을 고려하지 않고 개원하였다가 폐원할 시 손해가 상당히 크다고 할 수 있다. 따라서 개원 시 주변 환경에 대한 조사는 필수라고 볼 수 있다.

## 1.2 기획 의도

서울을 벗어나 신도시에서 개원하려는 의사들에게 매출 분석 서비스를 개발하여 도움을 줄 수 있도록 한다. 입지에 대한 정보와 진료 과목의 평균 데이터를 분석하여 예상 매출을 제공하여 개원하는 데 도움을 제공한다.

# 2. 분석 목표

-서울의 행정동별/연도별 데이터와 신도시의 지역 정보 데이터를 활용하여 해당 지역의 치과 의원 개원 시의 매출 분석 및 예측

### 3. 분석 데이터

#### 3.1 데이터 수집

##### - 서울 지역 데이터

서울의 여러 가지 데이터 중에서 개원 입지와 관련되어 있다고 생각하는 데이터를 수집하였다. 총인구, 연령별(20대 미만, 20대, 30대, 40대, 50대, 60대, 70대 이상) 인구, 유동 인구, 상주 인구, 치과 의원 개수, 세대 수, 버스 정류장 수, 지하철 역 수의 데이터를 환경에 관한 데이터로 구성하였고 매출, 아파트 매매 가격, 월 평균 소득을 경제와 관련된 데이터로 구성하였다. 위의 데이터는 행정동별로 2019년부터 4개년의 데이터를 사용하였다.

- 인구, 연령별 인구, 세대 수 (통계청) 서울 행정동별로 정리된 csv 파일 사용

- 매출, 유동 인구, 상주 인구, 치과 의원 수, 아파트 매매 가격 (서울데이터광장) 서울 행정동별로 정리된 csv 파일 사용

##### -개원 지역 데이터

여러 신도시 중에서 위례 신도시를 대상으로 하였다. 서울 지역의 데이터와 비교해야 하기 때문에 매출을 제외한 서울과 같은 데이터를 수집하였다. 위례가 서울, 성남, 하남 세 도시에 걸쳐져 있어 각 시에 해당하는 데이터를 수집하였고 성남과 하남의 경우 몇몇 데이터가 없는 관계로 다른 지역과 비교하여 임의로 사용하였다.

- 인구, 연령별 인구 (통계청) csv 파일 사용

- 세대 수, 월 평균 소득, 상주 인구, 유동인구 (소상공인마당) 해당 동 데이터 기입

- 치과 의원 개수 \_ 보건의료빅데이터개방시스템 참고하여 기입

- 지하철 역 수, 버스 정류장 수 대략적인 수치 기입

## 3.2 데이터 정합성 검사

분석을 위해 각 데이터의 정합성을 검사한 결과 연도에 따른 지역 구역의 변화에 따라 결측값이 존재하였다.

총인구 데이터에서 향동의 경우 2019년에 결측값이 존재했다.

연령별 데이터에서 향동, 일원2동, 상일동의 경우 결측값이 존재하는 연도가 있었다.

매출 데이터에서 결측값이 존재하는 행정동이 여러 개 발견되었다.

치과 의원 수 데이터에서 결측값이 존재하는 행정동이 여러 개 발견되었다.

발견된 결측값은 모두 NULL 값으로 처리하였고 분석 시에 제외하고 진행하였다.

## 3.3 데이터 전처리

데이터들을 분석하기 위해서는 모든 데이터들을 같은 형태로 만들 필요가 있었다. 마지막 매출을 예상하는 지역의 단위가 행정동이기 때문에 서울의 자치구 데이터를 세분화 시킬 필요가 있었다. 행정동 코드별로 정렬하여 모든 데이터가 맞는 코드에 들어가도록 하였다.

연도별 데이터를 데이터프레임 형태로 정리하여 총 4개의 연도별 데이터 파일을 생성하였다. 또한 데이터를 수집할 때 각 데이터마다 조사 기간이 달랐기 때문에 한달을 기준으로 데이터들을 조정하였다. 연도별 데이터를 수집할 때 연도가 분기별로 나뉘져 있는 데이터는 해당 연도의 4분기 데이터를 사용하였다. 데이터의 일관성을 위해 소수점이 나오는 데이터는 반올림하여 모두 정수의 형식으로 통일하였다.

## 4. 분석 과정 및 결과

회귀 분석을 활용하여 개원 시 매출을 예측하려 한다. 2019년부터 2022년까지 4개년의 매출을 통해 매출 추이를 분석하고 한 지역에 대한 계산식을 구하기 위하여 서울 행정동 425개, 4개년 총 1700개의 매출과 관련된 데이터를 회귀분석을 사용한다. 각 연도별로 회귀식을 구한 후 연도별 매출 추이를 확인하고 전체 데이터에 대한 회귀식을 구하여 지역 예상 매출에 대한 회귀식을 설정한다.

종속변수를 매출로 설정하고 독립변수(인구, 연령별 인구, 유동 인구, 상주 인구, 세대 수, 아파트 매매 가격, 월 평균 소득, 치과 의원 수, 버스 정류장 수, 지하철 역 수) 변화에 따른 모델을 구하기 위해 다중선형 회귀분석을 사용한다. 앞서 실행한 회귀분석의 적합성을 높이기 위한 변수 선택법으로는 단계적 선택법(stepwise selection)을 사용한다.

식을 구한 후 독립 변수들 간의 단위 차이가 큰 점을 고려하여 각 독립 변수마다 상용 로그를 사용하여 변수 간의 차이를 작게 만들고 식에 대입한다.

위와 같은 과정은 수집한 데이터들을 기반으로 R프로그램을 사용하여 진행한다.

### 4.1 서울 연도별 매출 분석 결과

종속변수: 연도별 매출(sale)

독립변수: 총인구(Population), 20대 미만 인구(Under20), 20대 인구(S20), 30대 인구(S30), 40대 인구(S40), 50대 인구(S50), 60대 인구(S60), 70대 이상(Over70), 유동인구(Float), 상주인구(Resident), 세대 수(Households), 아파트 매매 가격(HousePrice), 월 평균 소득(Income), 치과 의원 수(DentalClinic), 버스 정류장 수(Bus), 지하철 역 수(Subway)

#### 4-1-1. 2019년 서울 매출 분석 결과

[표1-1]

독립 변수	B	t-value	p-value
상수항(Constant term)	4.254e+08	1.487	0.1383
Population	-1.958e+05	-1.340	0.1816
Under20	3.297e+04	0.161	0.8725
S20	3.366e+05	2.265	0.0243
S30	5.067e+04	0.434	0.6647
S40	5.116e+05	1.935	0.0540
S50	1.440e+04	0.082	0.9344
S60	4.482e+05	1.661	0.0979
Over70	NA	NA	NA
Float	-1.496e+01	-0.674	0.5010
Resident	6.317e+03	0.279	0.7803
Households	-9.519e+03	-0.104	0.9174
Houseprice	1.646e-01	0.242	0.8087
Income	-1.597e+02	-1.558	0.1204
dentalClinic	1.022e+08	12.078	<2e-16
Bus	-3.368e+05	-0.113	0.9104
Subway	5.890e+07	1.177	0.2401
Adjusted R <sup>2</sup>	0.5072		
p-value	<2.2e-16		

\* B: 가중치, t-value: 추정치/표준 오차 p-value: 유의수준, Adjusted R<sup>2</sup>: 조정 결정 계수

[표1-1]의 결과를 단계적 선택법을 사용하여 변수를 다시 선택한다.

[표1-2]

독립 변수	B	t-value	p-value
상수항(Constant term)	3.432e+08	1.386	0.1670
population	-1.376e+05	-2.110	0.0358
S20	2.541e+05	2.469	0.0142
S40	4.146e+05	2.012	0.0452
S60	3.230e+05	1.851	0.0652
Income	-1.272e+02	-1.885	0.0606
DentalClinic	1.023e+08	13.903	<2e-16
<b>Adjusted R<sup>2</sup></b>	0.5193		
<b>p-value</b>	<2.2e-16		

단계적 선택법을 사용한 후 독립 변수는 인구, 20대 인구, 40대 인구, 60대 인구, 월 평균 소득, 치과 의원 수가 선택되었다.

[표1-2] 모델의 회귀식은 다음과 같다.

2019년 서울 지역 예상 매출 = (3.432e+08) - (1.376e+05 \* 총인구) + (2.541e+05 \* 20대 인구) + (4.146e+05 \* 40대 인구) + (3.230e+05 \* 60대 인구) - (1.272e+02 \* 월 평균 소득) + (1.023e+08 \* 치과 의원 수)

Adjusted R<sup>2</sup>: 0.5193

p-value: <2.2e-16

2019년 독립변수들의 평균으로 구한 2019년 지역 매출은 47,087,861원이다.



#### 4-1-2. 2020년 서울 매출 분석 결과

[표2-1]

독립 변수	B	t-value	p-value
상수항(Constant term)	6.274e+08	2.254	0.0251
Population	-8.379e+03	-0.054	0.9573
Under20	-1.178e+04	-0.059	0.9528
S20	2.799e+05	2.068	0.0397
S30	5.760e+04	0.522	0.6023
S40	5.193e+05	2.017	0.0448
S50	-4.391e+04	-0.253	0.8003
S60	3.768e+05	1.549	0.1226
Over70	NA	NA	NA
Float	-8.779e+00	-0.380	0.7045
Resident	-1.291e+05	-1.581	0.1151
Households	-6.629e+04	-0.755	0.4509
Houseprice	1.752e-01	0.266	0.7902
Income	-1.985e+02	-1.998	0.0468
dentalClinic	9.974e+07	12.194	<2e-16
Bus	-9.458e+05	-0.322	0.7474
Subway	7.627e+07	1.561	0.1197
<b>Adjusted R<sup>2</sup></b>	0.5024		
<b>p-value</b>	<2.2e-16		

[표2-1]의 결과를 단계적 선택법을 사용하여 변수를 다시 선택한다.

[표2-2]

독립 변수	B	t-value	p-value
상수항(Constant term)	4.850e+08	1.982	0.04856
S20	1.918e+05	2.382	0.01792
S40	4.693e+05	2.765	0.00609
S60	2.656e+05	2.001	0.04641
Resident	-1.327e+05	-2.629	0.00907
Income	-1.537e+02	-2.318	0.02120
DentalClinic	9.813e+07	13.300	<2e-16
Subway	6.756e+07	1.425	0.15547
Adjusted R <sup>2</sup>	0.5127		
p-value	<2.2e-16		

단계적 선택법을 사용한 후 독립 변수는 20대 인구, 40대 인구, 60대 인구, 상주 인구, 월 평균 소득, 치과 의원 수, 지하철 역 수가 선택되었다.

[표2-2] 모델의 회귀식은 다음과 같다.

2020년 서울 지역 예상 매출 = (4.850e+08) + (1.918e+05 \* 20대 인구) + (4.693e+05 \* 40대 인구) + (2.656e+05 \* 60대 인구) - (1.327e+05 \* 상주 인구) - (1.537e+02 \* 월 평균 소득) + (9.813e+07 \* 치과 의원 수) + (6.756e+07 \* 지하철 역 수)

Adjusted R<sup>2</sup>: 0.5127

p-value: <2.2e-16

2020년 독립변수들의 평균으로 구한 2020년 지역 매출은 61,499,386원이다.

#### 4-1-3. 2021년 서울 매출 분석 결과

[표3-1]

독립 변수	B	t-value	p-value
상수항(Constant term)	4.550e+08	1.456	0.146
Population	-7.038e+04	-0.431	0.667
Under20	6.732e+04	0.295	0.768
S20	1.135e+05	0.733	0.464
S30	1.515e+04	0.117	0.907
S40	3.539e+04	0.120	0.905
S50	3.925e+04	0.187	0.852
S60	9.266e+04	0.352	0.725
Over70	NA	NA	NA
Float	-2.533e+01	-0.987	0.324
Resident	1.924e+04	0.387	0.699
Households	1.137e+04	0.114	0.909
Houseprice	-4.001e-01	-0.855	0.394
Income	-8.892e+01	-0.831	0.407
dentalClinic	1.177e+08	13.532	<2e-16
Bus	-3.50e+06	-1.126	0.261
Subway	5.908e+07	1.065	0.288
<b>Adjusted R<sup>2</sup></b>	0.512		
<b>p-value</b>	<2.2e-16		

[표3-1]의 결과를 단계적 선택법을 사용하여 변수를 다시 선택한다.

**[표3-2]**

독립 변수	B	t-value	p-value
상수항(Constant term)	4.999e+08	2.752	0.00632
Income	-1.532e+02	-2.776	0.00588
DentalClinic	1.177e+08	17.434	<2e-16
<b>Adjusted R²</b>	0.5258		
<b>p-value</b>	<2.2e-16		

단계적 선택법을 사용한 후 독립 변수는 월 평균 소득, 치과 의원 수가 선택되었다.

[표3-2] 모델의 회귀식은 다음과 같다.

2021년 서울 지역 예상 매출 = (4.999e+08) – (1.532e+02 \* 월 평균 소득) + (1.177e+08 \* 치과 의원 수)

Adjusted R²: 0.5258

p-value: <2.2e-16

2021년 독립변수들의 평균으로 구한 2021년 지역 매출은 64,138,145원이다.

#### 4-1-4. 2022년 서울 매출 분석 결과

**[표4-1]**

독립 변수	B	t-value	p-value
상수항(Constant term)	7.089e+08	2.291	0.0228
Population	4.548e+04	0.732	0.4651

Under20	-1.938e+05	-0.999	0.3188
S20	1.217e+05	1.026	0.3059
S30	6.662e+04	0.621	0.5352
S40	1.119e+05	0.459	0.6464
S50	3.149e+04	0.196	0.8447
S60	-4.592e+04	-0.278	0.7814
Over70	4.286e+04	0.270	0.7876
Float	-3.743e+01	-1.458	0.1460
Resident	6.856e+03	0.457	0.6480
Households	-1.396e+05	-1.512	0.1318
Houseprice	-1.191e-01	-0.259	0.7960
Income	-1.964e+02	-1.862	0.0637
dentalClinic	1.147e+08	13.677	<2e-16
Bus	-2.356e+06	-0.715	0.4754
Subway	5.245e+07	0.961	0.3376
<b>Adjusted R<sup>2</sup></b>	0.507		
<b>p-value</b>	<2.2e-16		

[표4-1]의 결과를 단계적 선택법을 사용하여 변수를 다시 선택한다.

**[표4-2]**

독립 변수	B	t-value	p-value
상수항(Constant term)	6.543e+08	3.690	0.000271
Income	-1.907e+02	-3.528	0.000491
DentalClinic	1.118e+08	17.200	<2e-16
<b>Adjusted R²</b>	0.5175		
<b>p-value</b>	<2.2e-16		

단계적 선택법을 사용한 후 독립 변수는 월 평균 소득, 치과 의원 수가 선택되었다.

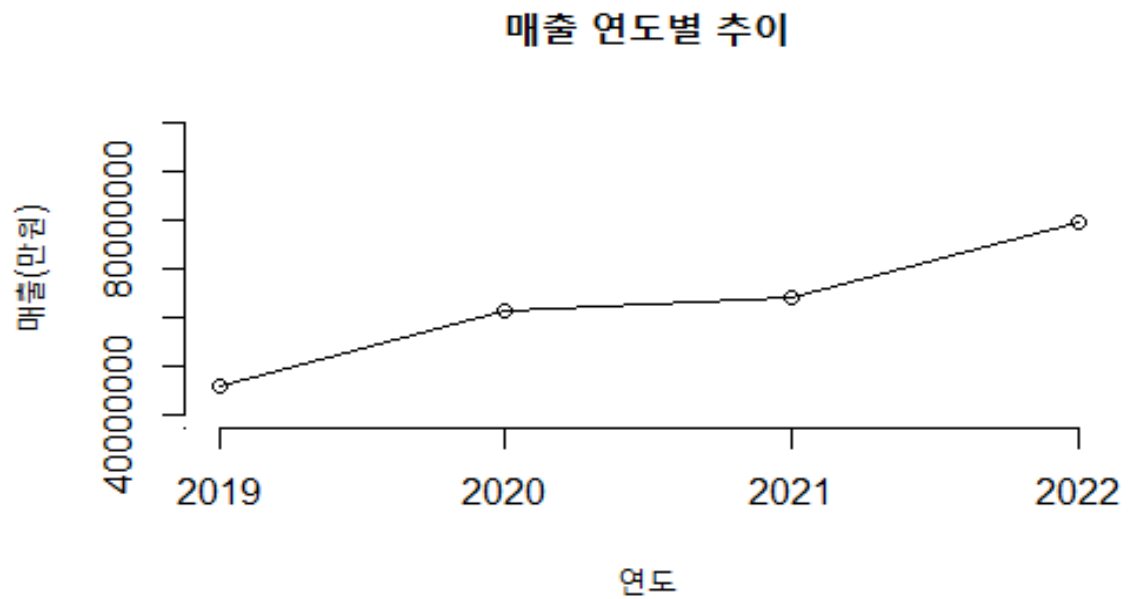
[표4-2] 모델의 회귀식은 다음과 같다.

2022년 서울 지역 예상 매출 = (6.543e+08) – (1.907e+02 \* 월 평균 소득) + (1.118e+08 \* 치과 의원 수)

Adjusted R²: 0.5175

p-value: <2.2e-16

2022년 독립변수들의 평균으로 구한 2022년 지역 매출은 79,553,724원이다.



## 4.2 서울 지역 매출 분석 결과

[표5-1]

독립 변수	B	t-value	p-value
상수항(Constant term)	4.907e+08	3.447	0.000589
Population	2.445e+04	0.442	0.658923
Under20	-1.465e+05	-1.495	0.135232
S20	1.1081e+05	1.488	0.136913
S30	-8.184e+04	-1.219	0.223269
S40	1.732e+05	1.439	0.150418
S50	-1.241e+05	-1.475	0.140405
S60	1.394e+05	1.539	0.124021
Over70	-1.148e+05	-1.311	0.189968
Float	-2.394e+01	-2.028	0.042818

Resident	7.756e+03	0.738	0.460783
Households	-4.259e+04	-0.969	0.332789
Houseprice	3.729e-02	0.157	0.875111
Income	-1.542e+02	-3.156	0.001644
dentalClinic	1.089e+08	26.206	<2e-16
Bus	-1.895e+06	-1.221	0.222396
Subway	6.301e+07	2.450	0.014425
<b>Adjusted R<sup>2</sup></b>	0.5212		
<b>p-value</b>	<2.2e-16		

[표5-1]의 결과를 단계적 선택법을 사용하여 변수를 다시 선택한다.

**[표5-2]**

독립 변수	B	t-value	p-value
상수항(Constant term)	5.196e+08	4.545	6.11e-06
S20	7.990e+04	2.930	0.00345
S30	-3.855e+04	-1.443	0.14926
S60	9.193e+04	1.909	0.05658
Over70	-1.002e+05	-1.844	0.06540
Float	-2.391e+01	-2.150	0.03176
Income	-1.665e+02	-5.874	5.64e-09
DentalClinic	1.092e+08	26.882	<2e-16
Subway	5.767e+07	2.269	0.02344
<b>Adjusted R<sup>2</sup></b>	0.5214		
<b>p-value</b>	<2.2e-16		



단계적 선택법을 사용한 후 독립 변수는 총인구, 20대 미만 인구, 20대 인구, 40대 인구, 유동 인구, 세대 수, 아파트 매매 가격, 월 평균 소득, 치과 의원 수가 선택되었다.

[표5-2] 모델의 회귀식은 다음과 같다.

서울 지역 예상 매출 =  $(5.196e+08) + (7.990e+04 * 20대\ 인구) - (3.855e+04 * 30대\ 인구) + (9.193e+04 * 60대\ 인구) - (1.002e+05 * 70대\ 이상\ 인구) - (2.391e+01 * 유동\ 인구) - (1.665e+02 * 월\ 평균\ 소득) + (1.092e+08 * 치과\ 의원\ 수) + (5.767e+07 * 지하철\ 역\ 수)$

Adjusted R<sup>2</sup>: 0.5214

p-value: <2.2e-16

4개년 독립변수들의 평균으로 구한 서울 지역 매출은 63,870,153원이다.

### 4.3 지역 매출 분석 결과 및 시각화

#### 4-3-1. 서울시 송파구 위례동 매출 분석 결과

[표6-1]

서울시 송파구 위례동			
20대 인구	30대 인구	60대 인구	70대 이상 인구
4,241	6,812	3,796	2,201
유동 인구	월 평균 소득	치과 의원 수	지하철 역 수
3,050,220	3,920,833	6	0
예상 월 매출			
100,784,840			

#### 4-3-2. 성남시 수정구 위례동 매출 분석 결과

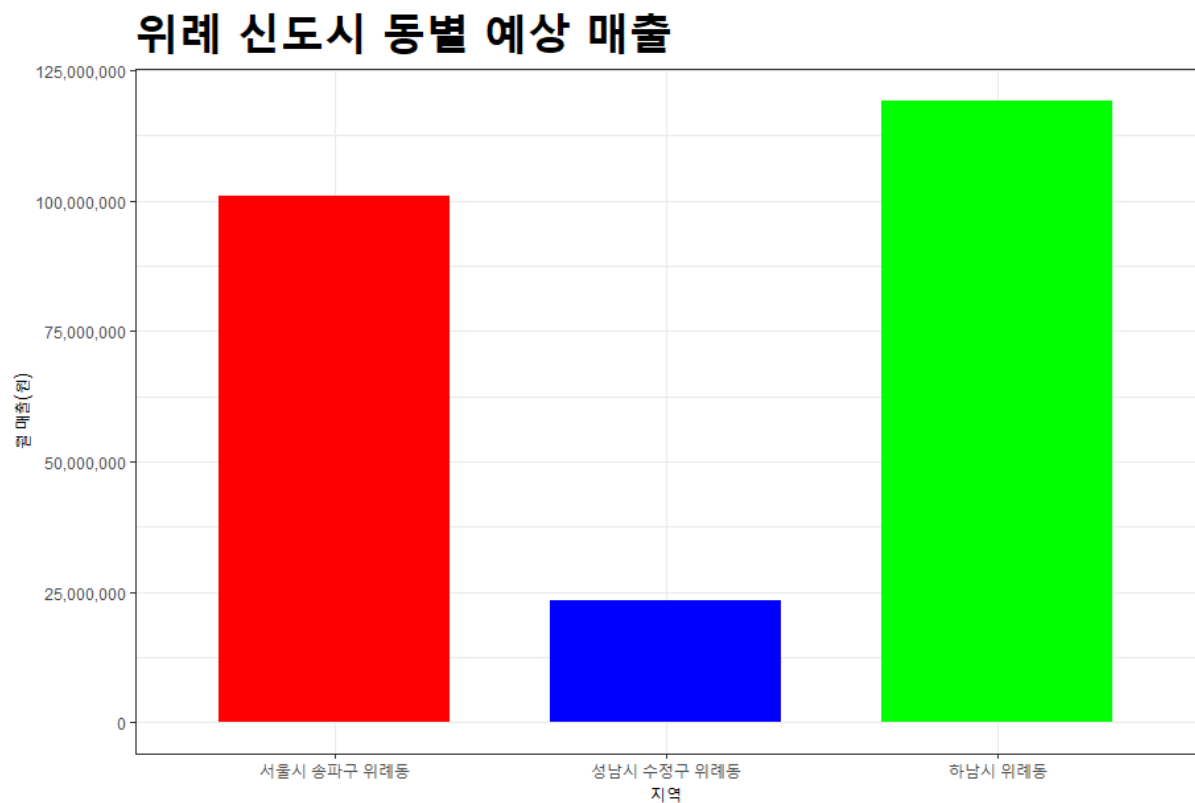
[표6-2]

성남시 수정구 위례동			
20대 인구	30대 인구	60대 인구	70대 이상 인구
4,664	7,231	5,724	3,434
유동 인구	월 평균 소득	치과 의원 수	지하철 역 수
2,087,940	3,802,500	29	1
예상 월 매출			
23,428,556			

#### 4-3-3. 하남시 위례동 매출 분석 결과

[표6-3]

하남시 위례동			
20대 인구	30대 인구	60대 인구	70대 이상 인구
3,484	5,151	3,472	1,623
유동 인구	월 평균 소득	치과 의원 수	지하철 역 수
646,680	4,405,000	5	0
예상 월 매출			
119,214,002			



지역에 대한 회귀분석을 한 결과 조정 결정 계수(Adjusted R<sup>2</sup>) 값이 0.5214로 약 52%의 설명력을 가지고 있는 회귀식이 도출되었다. p-value의 값은 2.2e-16보다 작다고 나온다. 이 값이 0.05보다 작을 시 의미 있는 모델이라 할 수 있으므로 의미 있는 모델이라 할 수 있다.

매출의 마이너스 요소인 주요 변수로는 월 평균 소득이 있고 매출의 플러스 요소인 주요 변수로는 20대 인구 수가 있다. 치과 의원의 수는 플러스 지역 전체로는 플러스 요인이긴 하지만 지역 치과 의원 하나의 수익을 계산할 때 마이너스 요소가 되는 것을 볼 수 있다.

지역 데이터에 따라 예상 매출에 큰 차이가 존재하기 때문에 개원 시 주변 환경에 대한 고려는 필수라는 결론이 도출되었다.

## 5. 활용방안

본 입지분석서비스로 병원입지선정 정보에 대한 병원들의 미래 예상 매출 정보의 산출이 가능하도록 병원입지선정 정보의 구성요소를 추출하고 분석 모형을 설계하였다. 병원 입지 정보수요자인 개원 예정 병원과 병원입지선정 정보 공급자 간의 정보비대칭으로 병원들은 개원 장소를 선택하기 쉽지 않다.(김기현 외, 2020)

미래경영예측이 가능한 정보를 개원을 생각하고 있는 의료인들에게 지원함으로써 본 서비스는 안정적인 의료 수급 체계를 조성하는 것에 기여하고 이를 통하여 국민보건향상에 기여하고자 한다.

본문에서는 치과만 대상으로 하였지만 다른 진료 과목의 데이터를 갖고 식을 새로 도출할 시 다른 과목의 매출도 예상할 수 있을 것이다.

## 6. 참고문헌