

# 데이터 사이언스 2차 개인보고서

2020.12.11

정규형

# Contents

---

I. Data Structure

II. Preprocessing & EDA

III. Research Question

IV. Conclusion

# I. Data Structure

## 데이터 확인 및 수집

### Additional Data integration

사업장명	병원 기본 속성	병원 세부 속성	최종보정리뷰	진료만족도	의료진친절도	시설만족도
비에스종합병원 이화여자대학교의과대학부속서울병원 학교법인연세대학교의과대학세브란스병원 의료법인동신의료재단 동신병원 서울특별시서울의료원 성애의료재단 성애병원 의료법인 청구성심병원 순천향대학교 부속 서울병원 강북삼성병원 의료법인 힐링의료재단 옥천성모병원 의료법인인화재단한국병원 정주성모병원 의료법인정산의료재단효성병원 한림대학교동탄성심병원 당진종합병원 재단법인아산사회복지재단보령아산병원	병원 인허가 데이터 정보 (도로명전체주소, 입원실 수, 병상 수, 좌표 등)	건강보험심사평가원 병원정보 (진료과목, 각 과목별 전문의 수, 의료장비 정보 등)	네이버 평점 크롤링 데이터 (평점, 건수) + 구글 평점 크롤링 데이터 (평점, 건수) + 다음 평점 크롤링 데이터 (평점, 건수)	굿닥 평점 크롤링 데이터 (평점, 건수)		
				NA		

〈팀 프로젝트를 통해 생성한 최종 데이터〉

시도별 의료인력현황,  
연령별 인구현황

출처: <https://kosis.kr>

상급병실료  
최저 및 최고비용

출처: <http://www.hira.or.kr>

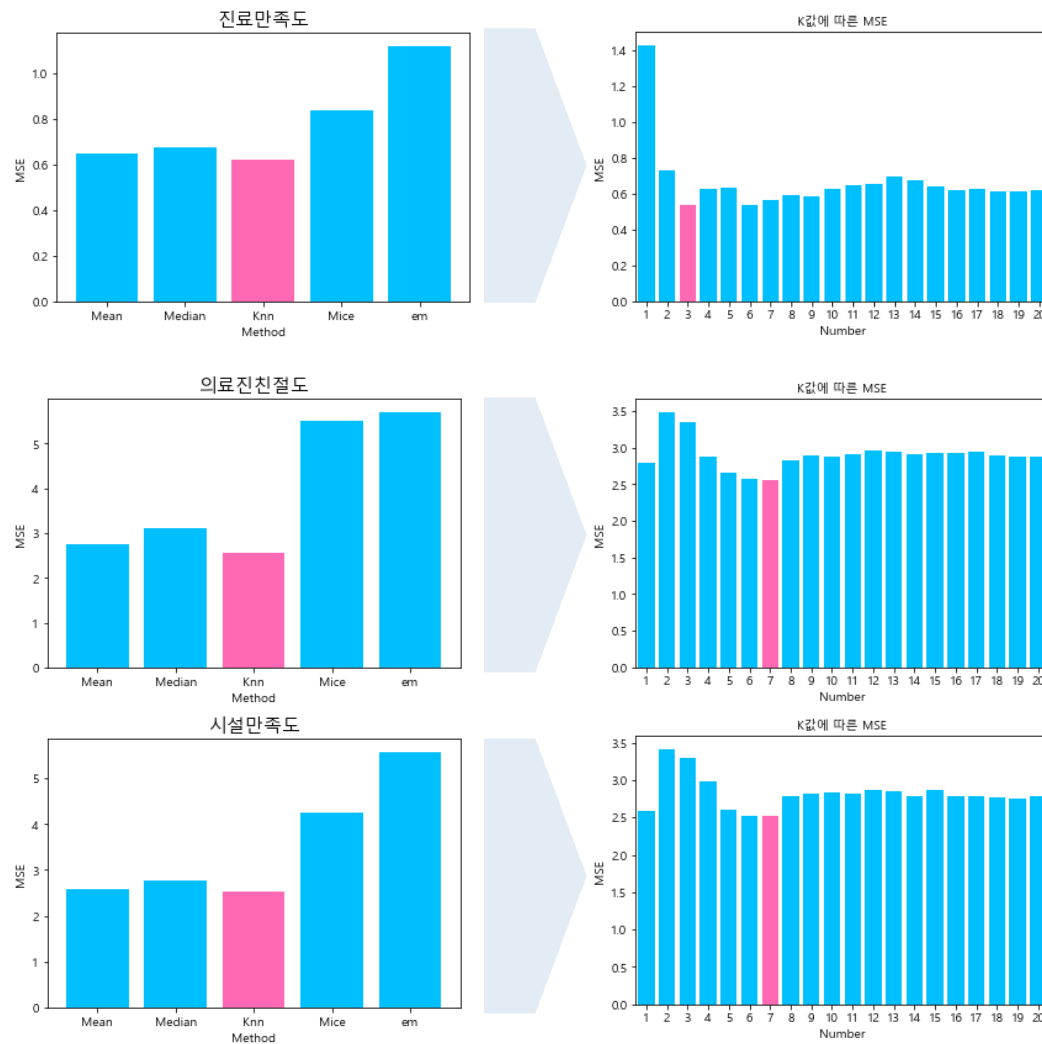
## • R.Q 관련 데이터 수집 및 통합

- 상급병실료 데이터는 기존 데이터에 통합하는 과정을 수행했으며 인력 및 인구 현황 데이터는 참고자료로 사용

## II. Preprocessing & EDA

팀 프로젝트에서 정하지 못한 결측치 처리

### 굿닥 평점 NA imputation



- 팀 프로젝트의 경우 다양한 기법을 소개하는 것에 그쳐, 가장 타당한 결측치 처리 방안에 대해 찾아보고자 아래와 같은 과정을 진행함

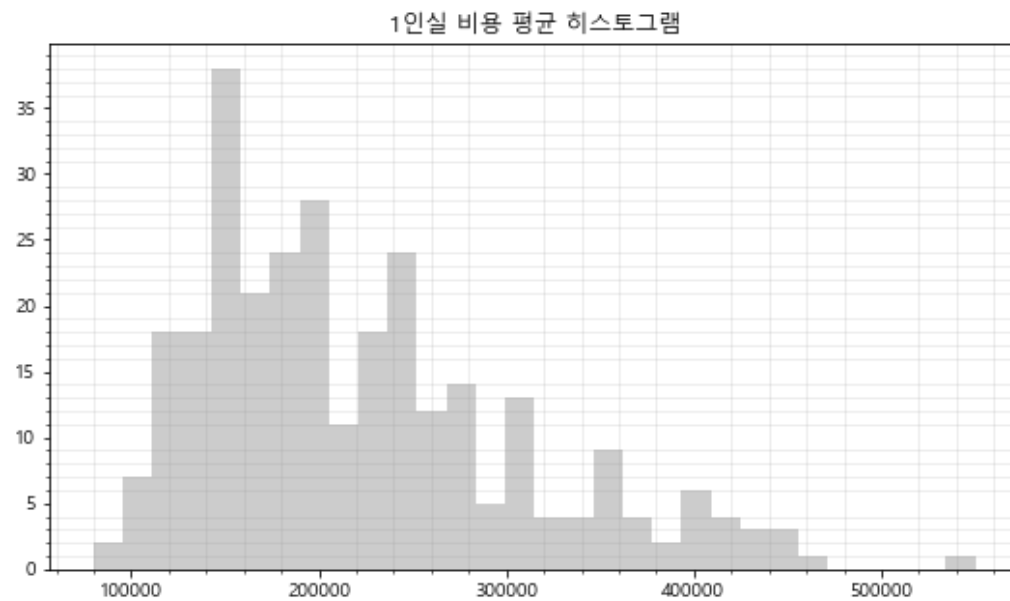
	의료인 수	입원 실수	병상수	의사총 수	일반 의 의사 수	인턴 의사 수	레지던트 의사 수	전문 의합 계	내과	신경 과	...	MRI	초음 파영 상진 단기	종양치료 기 (Gamma Knife)	종양치료 기 (Cyber Knife)	종양치 료기 (양 성자치 료기)	체외 충격 파쇄 석기	혈액투 석을위 한인공 신장기	시설 만족 도	총 건 수	최종보정 점수
0	14.0	55.0	213.0	20.0	1.0	0.0	0.0	18.0	4.0	0.0	...	1.0	4.0	0.0	0.0	0.0	1.0	23.0	NaN	27	2.652909
1	796.0	303.0	653.0	182.0	1.0	0.0	0.0	181.0	43.0	8.0	...	2.0	46.0	0.0	0.0	0.0	1.0	48.0	10.0	147	3.244260
2	4758.0	823.0	2455.0	1183.0	7.0	132.0	212.0	828.0	215.0	26.0	...	13.0	114.0	1.0	1.0	0.0	1.0	78.0	9.6	528	3.567129
3	112.0	59.0	207.0	21.0	3.0	0.0	0.0	18.0	4.0	0.0	...	1.0	4.0	0.0	0.0	0.0	0.0	14.0	10.0	40	2.721797
4	840.0	168.0	623.0	224.0	1.0	26.0	71.0	124.0	35.0	5.0	...	3.0	35.0	0.0	0.0	0.0	1.0	55.0	7.8	147	3.221450

- 최종데이터 50개 행의 만족도를 NA로 대입하여 생성된 데이터에 대한 NA imputation을 각 만족도별로 처리한 뒤 예측 값과 실제 값을 비교하여 MSE값을 도출함
- 그 결과 모든 만족도에서 Knn을 이용한 MSE값이 가장 낮게 나타남
- 추가적으로 K에 따라 다른 결과가 나오기 때문에 20까지의 값을 대입하여 최적의 K를 도출하고 이를 통해 최종데이터를 생성함

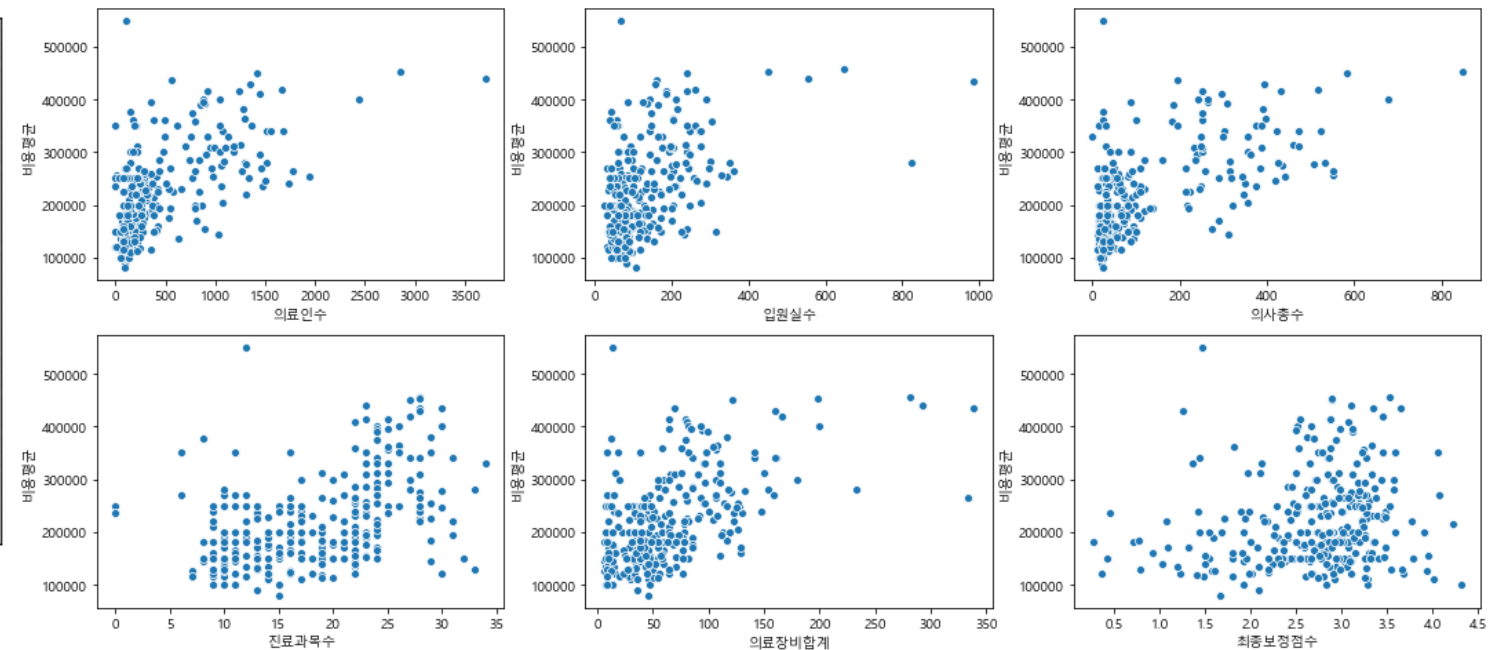
## II. Preprocessing & EDA

### 추가데이터에 대한 EDA

#### 상급병실료 데이터



주요 변수와 비용간의 Scatter Plot

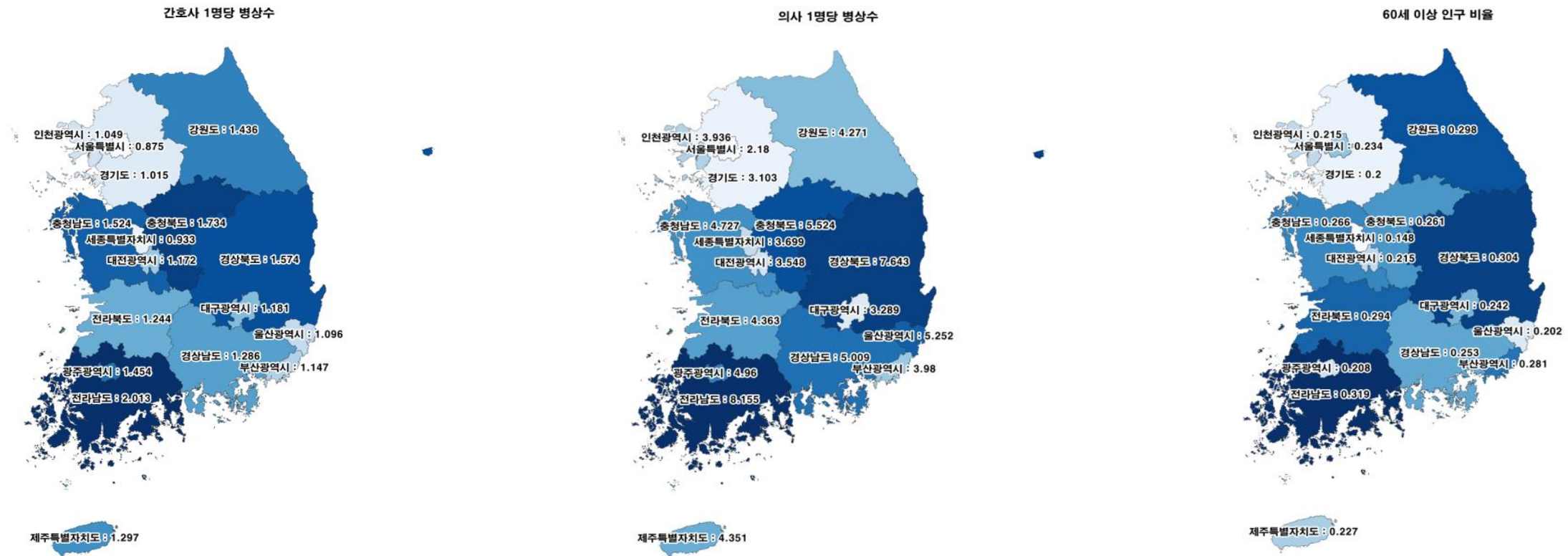


- 병원 별 1인실 비용의 평균은 10만원~20만원 대 값의 빈도수가 높게 나타나고 있으며 가격이 오를수록 병원의 수가 줄어들고 있다.
  - 1인실 비용의 평균과 기존 데이터의 주요 변수 간의 산점도를 통해 여러 변수에서 어느정도 상관관계가 있음을 확인할 수 있다.
- ▶ 최종적으로 추가자료통합을 통해 상급병실료 별 병원 특성을 파악해보고자 한다.

## II. Preprocessing & EDA

### 추가데이터에 대한 EDA

#### 시도별 의료인력현황 & 연령별 인구현황 데이터



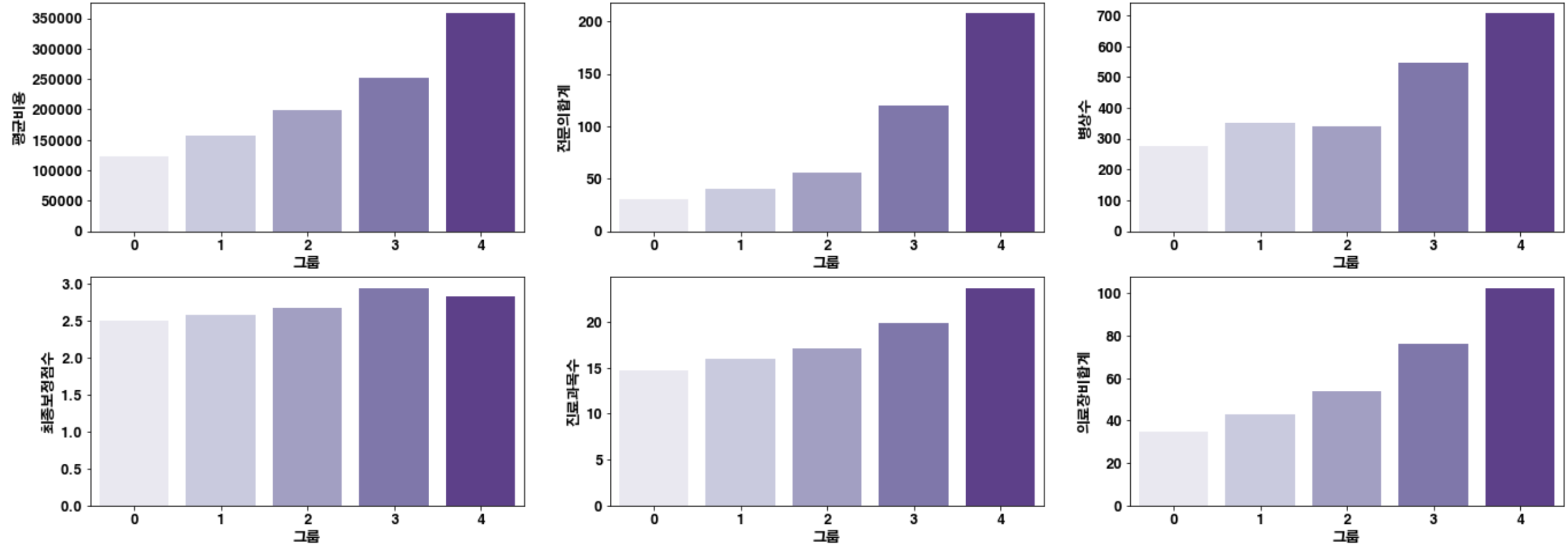
- 의료인력 1명당 병상 수 확인 결과 수도권에 비해 지방에서 인력의 부담이 클 것으로 예상된다.
  - 60세 이상 인구 비율 또한 지방에서 높게 나타나고 있다.
- ▶ 통합 데이터의 지방 종합병원 현황을 살펴보고 인력이 적절하게 배치되어 있는지 확인해보고자 한다.

### III. Research Question

#### 상급병실료 별 병원의 특성의 파악

- 상급병실료 평균비용을 기준으로 5구간 나누어 그룹을 부여함 (그룹0: 가장 저렴한 가격대 병원 ~ 그룹4: 가장 비싼 가격대 병원)
- 이를 통해 얻은 그룹별 주요변수의 시각화 결과는 아래와 같다.

#### 그룹별 주요변수 평균 값 barplot



### III. Research Question

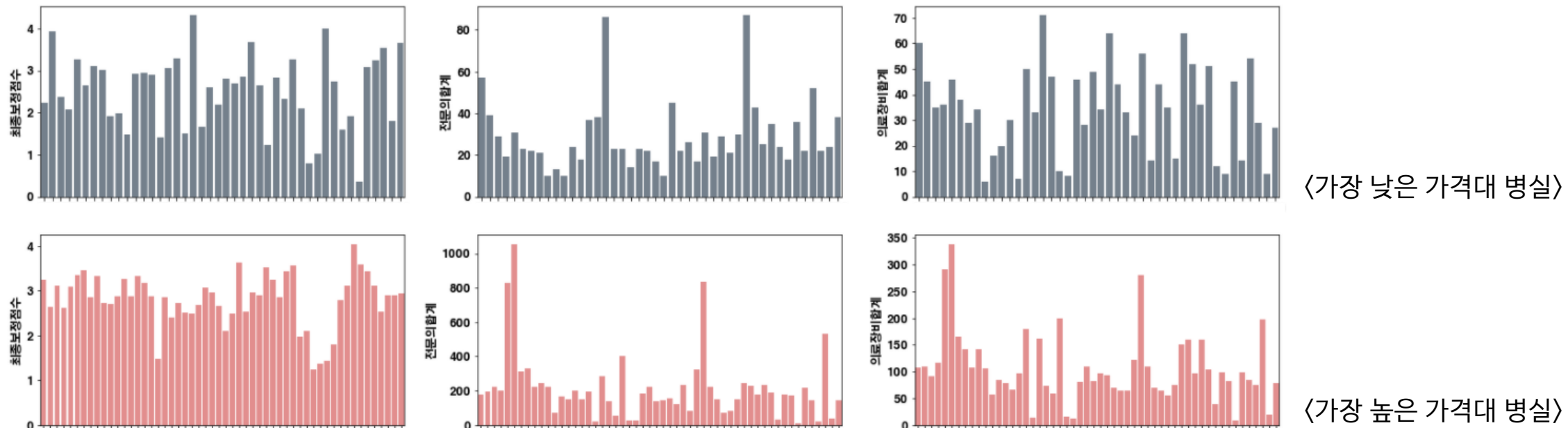
#### 상급병실료 별 병원의 특성의 파악

##### ● 그룹별 특성 파악결과

- 전문의, 병상, 진료 과목 등 대부분의 항목에서 병실 가격이 높아질수록 수가 증가함을 확인함
- 특히 최종점수에서도 그룹 간의 차이가 나타나 높은 가격을 지불할수록 어느정도 보장된 서비스를 받을 수 있을 것으로 예상됨
- 하지만 저렴한 가격대의 병원에서도 높은 점수, 많은 의료진을 가진 곳이 존재했으며 반대로 비싼 가격대의 병원에서도 낮은 점수와 적은 의료진을 가진 곳이 존재함

▶ 즉, 환자에게 병원의 병실료가 적절하게 책정된 것인지 객관적인 정보를 제공하여 정보 불균형 문제를 해결할 수 있을 것이다.

#### 최저가, 최고가 병원그룹의 주요변수 barplot

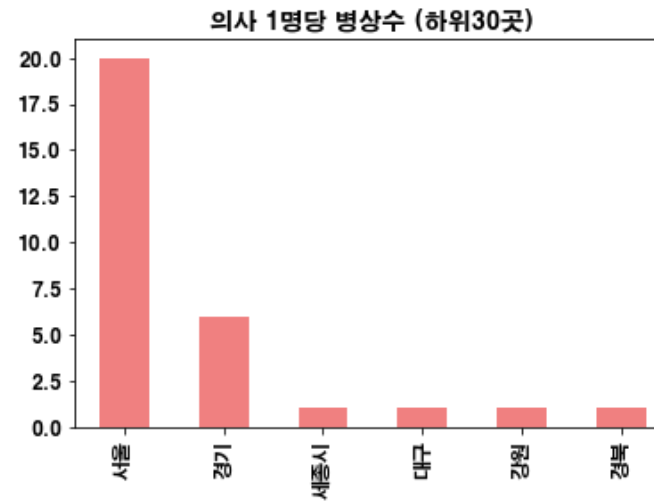
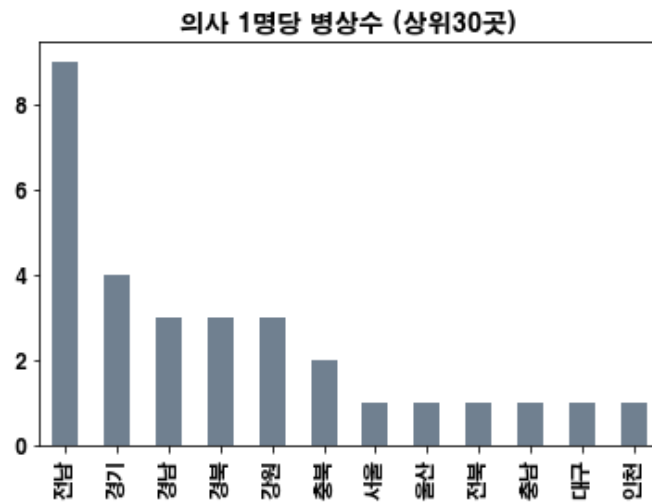
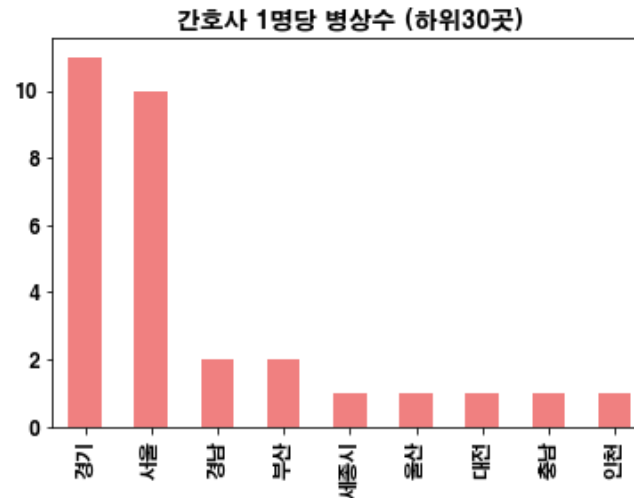
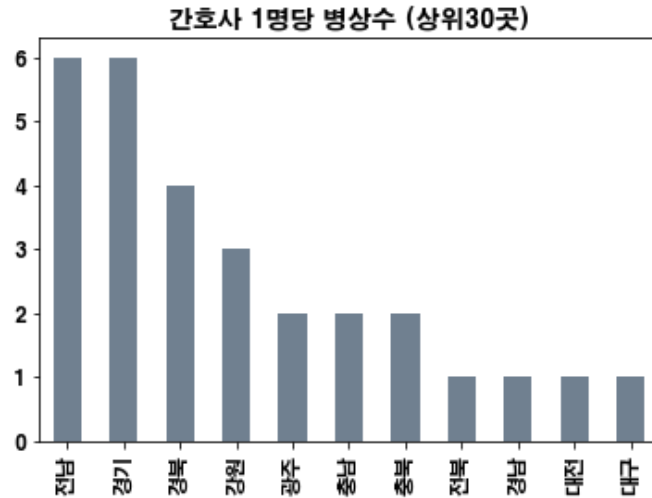




### III. Research Question

#### 인력 배치 적절여부 파악

#### 의료인력 배치현황



#### 지역별 의료인력 1명당 병상 수 현황

	간호사	의사
상위30곳 중간값	6.02	15.63
하위30곳 중간값	0.77	2.01

단위: 명

- EDA에서 언급한 바와 같이 의료인력 1명당 병상 수가 많은 곳은 대부분 지방에 위치 함  
(상위30곳: 1명당 부담해야하는 병상 수가 많음을 의미함)
- 하위30곳은 대부분 서울, 경기에서 차지하고 있으며 의사의 경우 두 지역을 제외한 다른 지역에서 비율이 매우 적음
- 특히, 전남지역이 다른 지역에 비해 의료인력 부족현상이 극심하게 나타나고 있음

### III. Research Question

#### 인력 배치 적절여부 파악

##### 종합병원 기준 요건

제3조의3(종합병원) ① 종합병원은 다음 각 호의 요건을 갖추어야 한다. <개정 2011. 8. 4.>

1. 100개 이상의 병상을 갖춘 것
  2. 100병상 이상 300병상 이하인 경우에는 내과 · 외과 · 소아청소년과 · 산부인과 중 3개 진료과목, 영상의학과, 마취통증의학과와 진단검사의학과 또는 병리과를 포함한 7개 이상의 진료과목을 갖추고 각 진료과목마다 전속하는 전문의를 둘 것
  3. 300병상을 초과하는 경우에는 내과, 외과, 소아청소년과, 산부인과, 영상의학과, 마취통증의학과, 진단검사의학과 또는 병리과, 정신건강의학과 및 치과를 포함한 9개 이상의 진료과목을 갖추고 각 진료과목마다 전속하는 전문의를 둘 것
- ② 종합병원은 제1항제2호 또는 제3호에 따른 진료과목(이하 이 항에서 "필수진료과목"이라 한다) 외에 필요하면 추가로 진료과목을 설치 · 운영할 수 있다. 이 경우 필수진료과목 외의 진료과목에 대하여는 해당 의료기관에 전속하지 아니한 전문의를 둘 수 있다.

출처: <https://www.law.go.kr/법령/의료법>

- 위 종합병원 요건에서 확인할 수 있듯이 병상, 진료과목에 대한 기준은 존재하지만 인력 기준이 구체적으로 명시되어 있지 않은 것을 확인할 수 있다.
  - 매년 병원 전체 병상 수는 지속적으로 증가하여 OECD국가 중 높은 순위를 차지하는 반면 인구 천명 당 간호인력은 6.9명으로 OECD평균(9명)에 비해 크게 못 미치고 있으며 의사 또한 같은 상황이다.
  - 지방의 경우 위 수치보다 더 낮을 것이며 앞서 확인한 것처럼 종합병원 또한 인력 부족 현황이 나타나고 있다.
- ▶ 코로나19가 유행하고 있는 현시점에서 의료인력 부족은 큰 문제이다. 지방의 의료 인력 확보를 통해 안정적인 의료서비스를 제공할 수 있어야 할 것이다. 또한, 인력에 대한 구체적인 법적 기준을 마련하여 인력난으로 인해 의료시스템 붕괴되지 않도록 대비해야 한다. 혹은 현재의 간호등급 제도를 보완할 필요성이 있다.

## IV. Conclusion

### 결론 및 한계점

#### ● 결론

- 의료서비스와 병실가격 간의 관계가 존재하나 병실가격에 비해 부족한 의료서비스를 제공하는 병원이 존재함.
  - ▶ 의료분야는 정보 불균형이 크기에 지속적으로 병원에 대한 양적, 질적 데이터 수집 및 공급을 통해 환자가 적절한 병원을 찾을 수 있도록 노력해야함.
- 수도권과 지방 간의 의료인력 및 장비의 격차가 크게 나타나고 있으며 광역시를 제외한 지역에서 더욱 극심하게 나타남
  - ▶ 의료인력에 대한 법적 근거 마련, 간호등급제도 등을 통해 지방의 종합병원 인력 및 시설을 재정립하여 격차를 해결해야함.

#### ● 한계점

- 간호사 수를 구하는 과정(의료인 수 - 의사 수)에서 데이터의 품질이 좋지 않아 음의 값을 갖는 병원이 존재함
  - ▶ 해당병원의 경우 추가 정보 수집을 통해 해결했지만 최종데이터 기준월과 다를 수 있다는 문제점 발생. 또한, 일반병원으로 범위를 넓힌다면 수작업을 통해 해결이 불가함.
- 상급 병실 가격 데이터 활용 및 주관적 정보 부족
  - ▶ 대부분의 사람이 이용하는 일반 병실 가격을 적용하지 못해 대표성이 떨어질 수 있으며 환자들의 리뷰 데이터 혹은 병원 내 의사에 따라 다른 리뷰 정보를 추가하지 못해 질적 판단이 불가함.

