

1. 노인 파트 데이터 column 설명

- 건강보험심사평가원_요양기관 개설 현황
 - 원 데이터 column
 - ◆ 암호화된 요양 기호, 요양 기관명, 요양종별, 시도명, 시군구명, 도로명주소, 표시과목명, 개설일자
 - ◆ 암호화된 요양 기호는 요양 시설에 부여된 기호에 대해 암호화된 값
- 행정안전부_지역별 연령별 주민등록 인구현황
 - 원 데이터 column
 - ◆ 행정기관, 총 인구수, 연령구간 인구수, 0~9세, 10~19세, 20~29세, 30~39세, 40~49세, 50~59세, 60~69세, 70~79세, 80~89세, 90~99세, 100세 이상
 - ◆ 연령구간 인구수 == 총 인구수

2. 노인 파트 데이터 전처리

- 건강보험심사평가원_요양기관 개설 현황

- 관심 있는 정보: 년도별 요양 기관수
- 필요한 column: 시도명, 년도, 누적기관수

| | 시도명 | 연도그룹 | 누적기관수 |
|---|---------|--------|-------|
| 0 | 강원특별자치도 | 2008이전 | 1443 |
| 1 | 강원특별자치도 | 2009 | 1504 |
| 2 | 강원특별자치도 | 2010 | 1562 |
| 3 | 강원특별자치도 | 2011 | 1627 |
| 4 | 강원특별자치도 | 2012 | 1698 |

- 연도그룹: 특정 년도
- 누적 기관수: 특정 년도까지 설립된 기관 수

- 행정안전부_지역별 연령별 주민등록 인구현황

- 관심 있는 정보
 - ◆ 행정기관별 총 인구수와 행정기관별 노인(60세 이상) 인구수
- 필요한 column
 - ◆ 행정기관, 총 인구수, 60~69세, 70~79세, 80~89세, 90~99세, 100세 이상, 년도
 - ◆ -> 행정기관, 년도 노인 인구수,
 - ◆ 년도는 데이터를 저장하는 과정에서 추가

| | 행정구역 | 년도 | 노인인구수 |
|---|-------|------|---------|
| 0 | 전국 | 2008 | 7110229 |
| 1 | 서울특별시 | 2008 | 1342702 |
| 2 | 부산광역시 | 2008 | 538619 |
| 3 | 대구광역시 | 2008 | 331818 |
| 4 | 인천광역시 | 2008 | 311800 |

- 지역별, 년도별 노인 인구수 대비 기관수 비율

- column: 시도명, 년도, 노인인구 기관수 비율

| | 시도명 | 년도 | 노인인구_기관수_비율 |
|---|---------|------|-------------|
| 0 | 강원특별자치도 | 2008 | 0.005100 |
| 1 | 강원특별자치도 | 2009 | 0.005194 |
| 2 | 강원특별자치도 | 2010 | 0.005223 |
| 3 | 강원특별자치도 | 2011 | 0.005346 |
| 4 | 강원특별자치도 | 2012 | 0.005325 |

- 노인인구 기관수 비율: 기관수/60세 이상의 인구수

3. 노인 파트 데이터 EDA

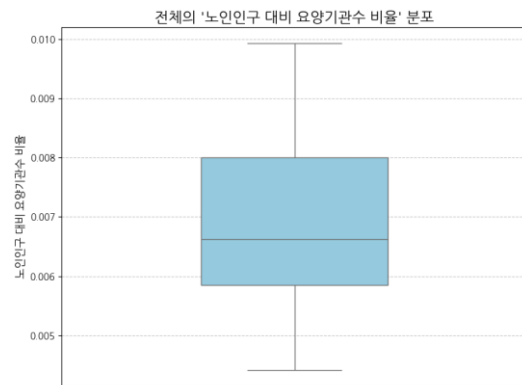
- 데이터에 대한 이해

- 열: 시도명, 년도, 노인인구_기관수 비율
- 결측치 확인 및 처리
 - ◆ 세종특별자치시의 경우 2012년부터 설립되었기 때문에 11년까지의 인구 데이터가 존재하지 않음 따라서 비율이 NAN이 됨
 - ◆ 따라서 해당 부분 제거, 즉 세종특별자치시의 경우 2012~2024년 데이터를 이용

- 단일 변수 분석(노인인구_기관수_비율)

- 노인인구_기관수_비율(시간을 2008~2024년 모두 고려할 때)

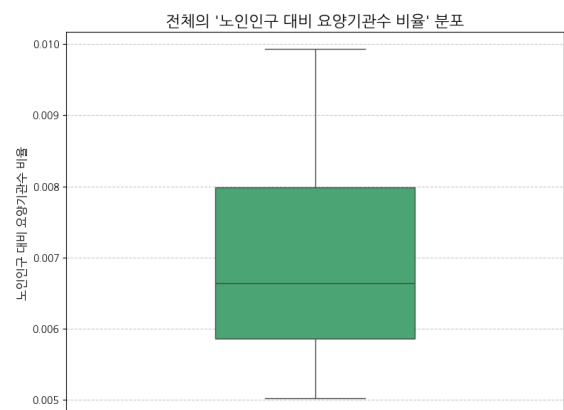
| | |
|-------|------------|
| count | 285.000000 |
| mean | 0.006867 |
| std | 0.001320 |
| min | 0.004417 |
| 25% | 0.005852 |
| 50% | 0.006620 |
| 75% | 0.007998 |
| max | 0.009930 |



- ◆ 평균적으로 0.006867의 비율로 요양기관 1개당 약 145명을 커버해야 함
- ◆ 제일 요양기관이 많은 지역도 0.009930으로 요양기관 1개당 약 100명을 커버해야 함

- 노인인구_기관수_비율(시간을 2024년으로 고정할 때)

| | |
|-------|-----------|
| count | 17.000000 |
| mean | 0.006960 |
| std | 0.001488 |
| min | 0.005024 |
| 25% | 0.005858 |
| 50% | 0.006633 |
| 75% | 0.007988 |
| max | 0.009930 |



- ◆ 평균적으로 0.006960의 비율로 요양기관 1개당 약 144명을 커버해야 함
- ◆ 제일 비율이 높은 지역도 요양기관 1개당 약 100명을 커버해야 함
- ◆ 그래도 min값이 0.004417에서 0.005024로 비율이 늘어나 제일 부족했던 지역이 약 226명을 커버해야 했던 것에 비해 200명을 커버하게 됨

- 단일 변수 분석(시도명, 년도)

■ 시도명(17개의 시도명)

- ◆ 강원특별자치도, 경기도, 경상남도, 경상북도, 광주광역시, 대구광역시, 대전광역시, 부산광역시, 서울특별시, 세종특별자치시, 울산광역시, 인천광역시, 전라남도, 전북특별자치도, 제주특별자치도, 충청남도, 충청북도

◆ 년도

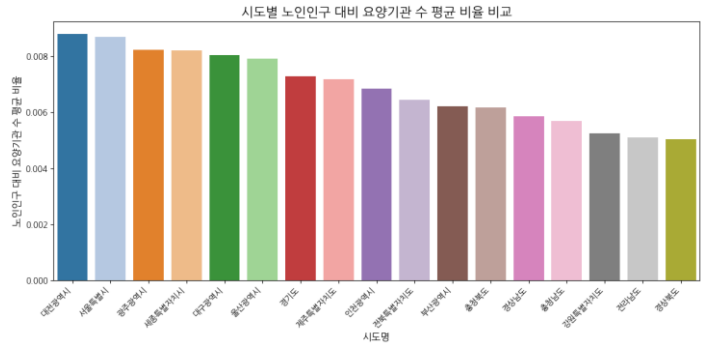
- 2008년~2024년
- 단, 세종특별자치시의 경우에만 2012년~2024년

- 변수 간 관계 분석

■ 지역별 비율

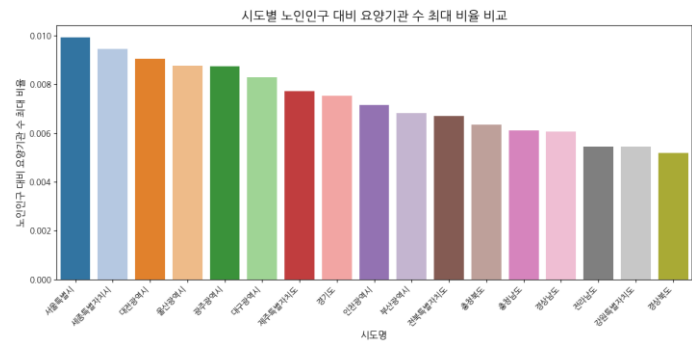
◆ 지역별 평균 비율 비교

| 시도명 | |
|---------|----------|
| 대전광역시 | 0.008812 |
| 서울특별시 | 0.008694 |
| 광주광역시 | 0.008247 |
| 세종특별자치시 | 0.008214 |
| 대구광역시 | 0.008041 |
| 울산광역시 | 0.007921 |
| 경기도 | 0.007301 |
| 제주특별자치도 | 0.007182 |
| 인천광역시 | 0.006851 |
| 전북특별자치도 | 0.006446 |
| 부산광역시 | 0.006210 |
| 충청북도 | 0.006169 |
| 경상남도 | 0.005859 |
| 충청남도 | 0.005705 |
| 강원특별자치도 | 0.005249 |
| 전라남도 | 0.005114 |
| 경상북도 | 0.005042 |



◆ 지역별 최고 비율 비교

| 시도명 | |
|---------|----------|
| 서울특별시 | 0.009930 |
| 세종특별자치시 | 0.009451 |
| 대전광역시 | 0.009036 |
| 울산광역시 | 0.008758 |
| 광주광역시 | 0.008748 |
| 대구광역시 | 0.008290 |
| 제주특별자치도 | 0.007719 |
| 경기도 | 0.007537 |
| 인천광역시 | 0.007144 |
| 부산광역시 | 0.006824 |
| 전북특별자치도 | 0.006707 |
| 충청북도 | 0.006348 |
| 충청남도 | 0.006114 |
| 경상남도 | 0.006052 |
| 전라남도 | 0.005451 |
| 강원특별자치도 | 0.005438 |
| 경상북도 | 0.005186 |

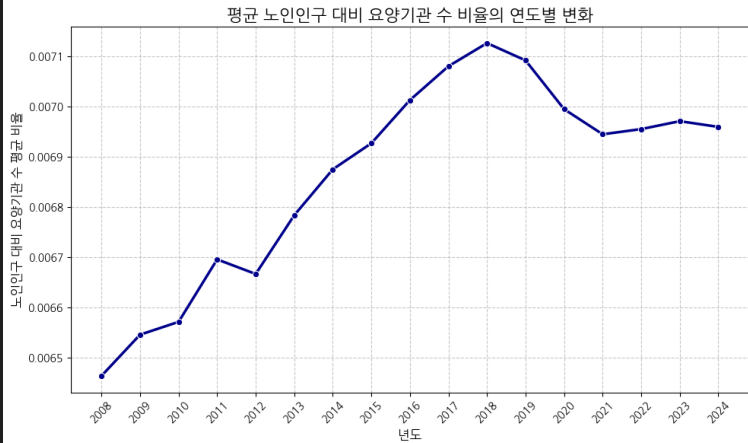


◆ 서울특별시, 대전광역시, 세종특별자치시, 광주광역시는 평균과 최고 비율 둘 다 top 5에 들었음

◆ 경상북도, 강원특별자치도, 전라남도, 충청남도, 경상남도는 평균과 최고 비율 둘 다 bottom 5에 들었음

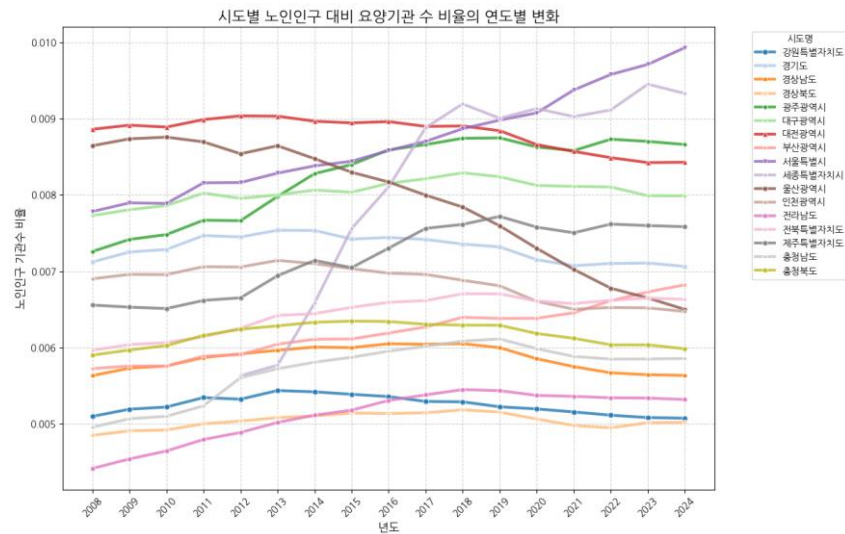
■ 시간의 흐름에 따른 비율 변화

| 년도 | |
|------|----------|
| 2008 | 0.006464 |
| 2009 | 0.006546 |
| 2010 | 0.006571 |
| 2011 | 0.006696 |
| 2012 | 0.006667 |
| 2013 | 0.006784 |
| 2014 | 0.006875 |
| 2015 | 0.006927 |
| 2016 | 0.007013 |
| 2017 | 0.007081 |
| 2018 | 0.007127 |
| 2019 | 0.007092 |
| 2020 | 0.006995 |
| 2021 | 0.006945 |
| 2022 | 0.006956 |
| 2023 | 0.006971 |
| 2024 | 0.006960 |



◆ 2018년까지는 증가하다가 2019년부터는 일정비율을 유지하고 있음

■ 지역 및 시간 교차 분석



◆ 증가하는 경우

- 서울특별시, 부산광역시, 전북특별자치도 경우 꾸준히 증가
- 세종특별자치시, 광주광역시, 제주특별자치도, 충청남도, 전라남도의 경우 증가하다가 일정 비율 유지

◆ 일정 비율을 유지

- 대구광역시, 충청북도, 경상남도, 강원특별자치도, 경상북도

◆ 감소하는 경우

- 대전광역시, 경기도, 인천광역시

◆ 급격히 감소하는 경우

- 울산광역시

ANOVA를 돌리려고 함

- ➔ 데이터가 시도별로 그룹했기 때문에 그룹 내 데이터 수가 1, 따라서 ANOVA 못함

Outlier Detection

- ➔ IQR 기반으로 Outlier 탐지하기, box-plot 그리기
- ➔ Grubbs' Test 등을 이용해서 단일 값 이상치 검정하기
 - 정규성 검정을 꼭 통과해야 사용할 수 있음
 - 정규성 검정을 통과하지 못했다면 Dixon's Q Test 사용

ANOVA

- ➔ 설립년도와 년도를 이용해서 시도별 년도별 요양시설 수를 계산할 수 있음
- ➔ 그룹 내의 데이터 수가 1보다 커짐
- ➔ 하지만 같은 그룹 내에서 시간에 따라 변하는 값이기 때문에 단순 ANOVA를 사용하면 X
 - 같은 그룹 내의 데이터들이 서로 독립적이지 않음

반복측정 ANOVA

- ➔ 한 시도에서 연도별로 반복 측정된 값 -> 시간에 따라 변화하는 패턴 비교 가능
- ➔ 시도 간 차이를 비교할 수 없음

혼합설계 ANOVA(mixed ANOVA)

- ➔ between-subject factor: 시도명, 서로 독립인 그룹
 - 독립성: 시도 간 관측치는 서로 독립적이어야 함 -> 가정으로 깔고 가기
 - 정규성: 시도별 평균값의 오차가 정규분포를 따라야 함
 - 등분산성: 시도별 누적기관수의 분산이 비슷해야 함
- ➔ within-subject factor: 연도그룹, 각 시도에서 반복 측정된 조건
 - 정규성: 시도 내부에서 연도별 측정값의 오차가 정규분포를 따라야 함
 - 구형성: 모든 연도쌍의 차이의 분산이 동일해야 함
 - ◆ 보통 깨지니까 보정해서 사용할 것