

# 1.Chest X-rays(indiana university) 분석하기

1-1. projection파일과 reports 파일을 이용하여 분석해보았습니다

| uid  |      | filename                  | projection |
|------|------|---------------------------|------------|
| 0    | 1    | 1_IM-0001-4001.dcm.png    | Frontal    |
| 1    | 1    | 1_IM-0001-3001.dcm.png    | Lateral    |
| 2    | 2    | 2_IM-0652-1001.dcm.png    | Frontal    |
| 3    | 2    | 2_IM-0652-2001.dcm.png    | Lateral    |
| 4    | 3    | 3_IM-1384-1001.dcm.png    | Frontal    |
| ...  | ...  | ...                       | ...        |
| 7461 | 3997 | 3997_IM-2048-1002.dcm.png | Lateral    |
| 7462 | 3998 | 3998_IM-2048-1001.dcm.png | Frontal    |
| 7463 | 3998 | 3998_IM-2048-1002.dcm.png | Lateral    |
| 7464 | 3999 | 3999_IM-2049-1001.dcm.png | Frontal    |
| 7465 | 3999 | 3999_IM-2049-2001.dcm.png | Lateral    |

7466 rows × 3 columns

| reports_df |      |   |   |   |   |                |   |   |
|------------|------|---|---|---|---|----------------|---|---|
| uid        |      | MeSH  | Problems  | image   | indication  | comparison     | findings  | impression  |
| 0          | 1    | normal  | normal  | Xray Chest PA and Lateral                         | Positive TB test                                  | None.          | The cardiac silhouette and mediastinum size ar... | Normal chest x-XXXX.                              |
| 1          | 2    | Cardiomegaly;borderline;Pulmonary Artery/enlarged | Cardiomegaly;Pulmonary Artery                     | Chest, 2 views, frontal and lateral               | Preop bariatric surgery.                          | None.          | Borderline cardiomegaly. Midline sternotomy XX... | No acute pulmonary findings.                      |
| 2          | 3    | normal  | normal  | Xray Chest PA and Lateral                         | rib pain after a XXXX, XXXX XXXX steps this XX... | NaN            | NaN   | No displaced rib fractures, pneumothorax, or p... |
| 3          | 4    | Pulmonary Disease, Chronic Obstructive;Bullous... | Pulmonary Disease, Chronic Obstructive;Bullous... | PA and lateral views of the chest XXXX, XXXX a... | XXXX-year-old XXXX with XXXX.                     | None available | There are diffuse bilateral interstitial and a... | 1. Bullous emphysema and interstitial fibrosis... |
| 4          | 5    | Osteophyte/thoracic vertebrae/multiple/small;T... | Osteophyte;Thickening;Lung                        | Xray Chest PA and Lateral                         | Chest and nasal congestion.                       | NaN            | The cardiomeastinal silhouette and pulmonary...   | No acute cardiopulmonary abnormality.             |
| ...        | ...  | ...   | ...   | ...   | ...   | ...            | ...   | ...   |
| 3846       | 3995 | Lung/hyperdistention/mild;Diaphragm/bilateral/... | Lung;Diaphragm;Cicatrix;Pulmonary Atelectasis     | Xray Chest PA and                                 | Nausea, vomiting x2 weeks.                        | XXXX, XXXX     | The cardiomeastinal silhouette and                | 1. Interval resolution of bibasilar               |

## projection의 항목

uid (Unique Identifier)

:각 이미지 또는 기록에 대한 고유 식별자.

filename

:파일 이름을 나타내며, 파일 시스템 내에서 파일을 식별하는 데 사용됩니다.

projection

:이미지를 촬영한 투영(projection) 또는 뷰(view)를 나타냅니다.

## report의 항목

uid (Unique Identifier)

각 이미지 또는 기록에 대한 고유 식별자.

MeSH (Medical Subject Headings)

의학 용어 및 주제 태그.

Problems (문제)

환자의 주요 건강 문제 또는 진단 요약.

image(이미지)

X-ray 이미지 파일 경로 또는 URL.

indication (적응증)

X-ray 촬영의 이유.

comparison (비교)

이전 X-ray와의 비교 결과.

findings (소견)

의료 전문가가 이미지에서 관찰한 결과.

impression (판단)

진단 결과의 요약 또는 결론.

1.정보들을 분석할때 효과적인 분류를 위해 df 2개 항목 정보에 공통적으로 들어있는 frontal(앞면)과 lateral(옆면)으로 기준으로 함.

두개를 중심으로 딥러닝 또는 머신러닝에 활용됨.

2.항목마다 비교하여 항목에 있는 정보들을 분석 할 경우, 데이터 표본 수를 쌓기 위해 problem,impression을 중요 항목으로 선정.

## 1-2. 두개의 데이터 프레임을 하나의 데이터 프레임(merge\_df)로 합침(uid항목을 공통항목으로)

- Frontal과 Lateral의 분포, 주요 진단(Problems)별 빈도를 통계를 냄

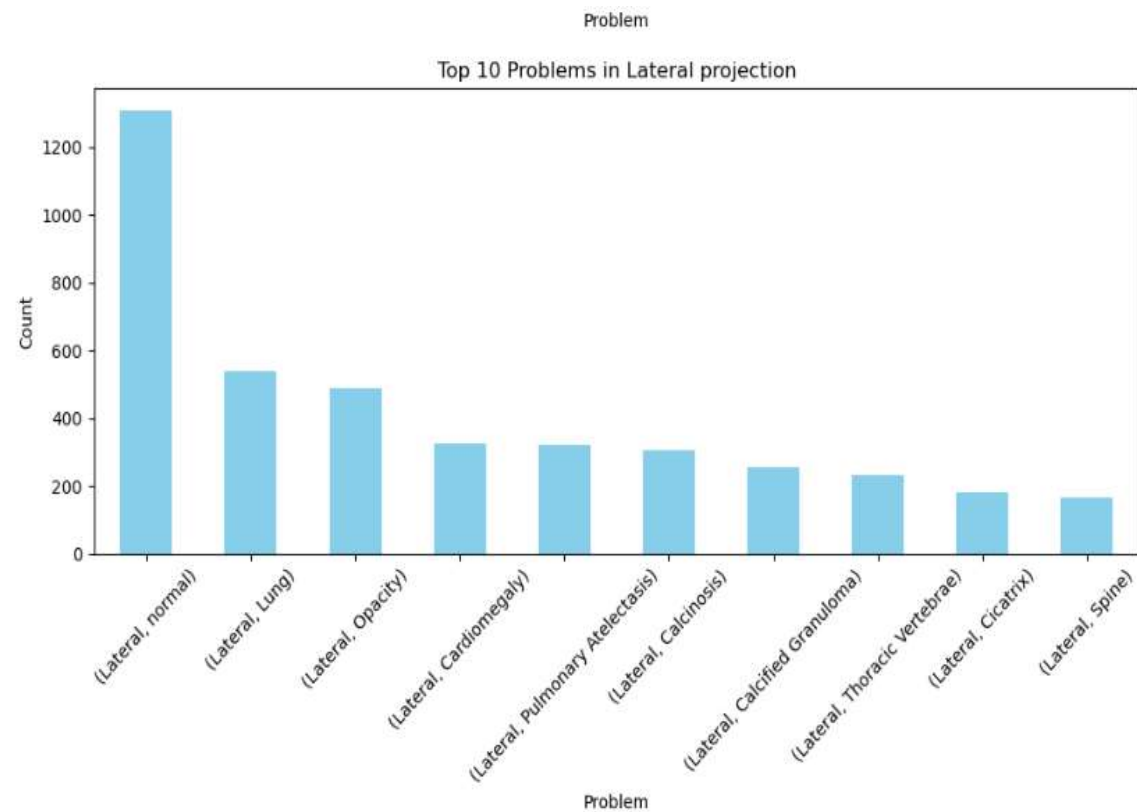
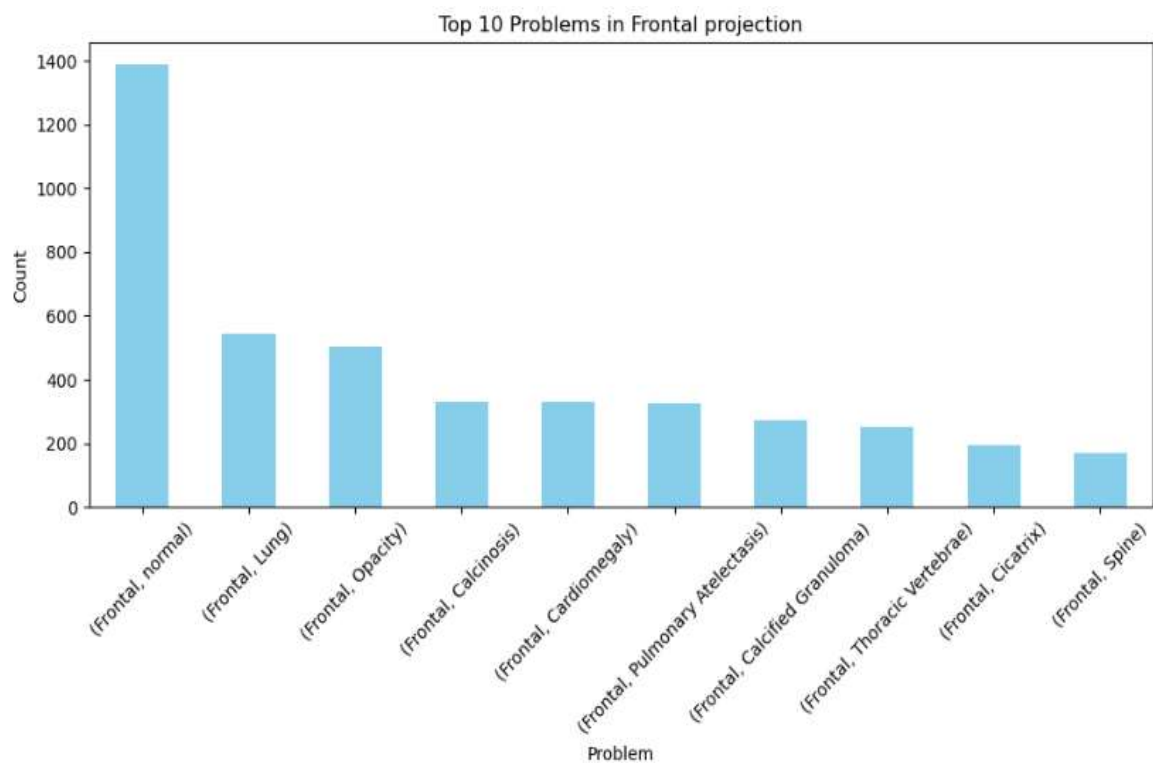
-Frontal과 Lateral 이미지에서 발견된 문제의 차이점을 분석함

```
1.      : (projection
      Frontal    3818
      Lateral    3648
      Name: count, dtype: int64,
```

```
2.      Problems
      normal          2695
      Lung            1083
      Opacity          988
      Cardiomegaly     655
      Pulmonary Atelectasis  645
      Name: count, dtype: int64,
```

```
3.      projection
      Frontal    normal          1387
                Lung            545
                Opacity          501
                Calcinosis       332
                Cardiomegaly     331
                ...
      Lateral    Thoracic vertebrae    1
                Bronchitis           1
                Trachea, Carina       1
                Colonic Interposition  1
                Hemothorax           1
      Name: Problems, Length: 239, dtype: int64)
```

## frontal,lateral 에 따른 problems 분포(막대그래프)



## 투영 방식별 분포:

앞면 이미지: 3818개

옆면 이미지: 3648개

## 주요 문제(Problems):

normal(정상): 2695건

Lung(폐): 1083건

Opacity(불투명도): 988건

Cardiomegaly(심비대): 655건

Pulmonary Atelectasis(폐기종): 645건

## 방식에 따른 문제 분포:

앞면 이미지에서는 정상, 폐, 불투명도가 주요 문제

옆면 이미지에서는 앞면 이미지와 같은 분포를 이룸. 그러나 주요 문제에서는 적은 분포를 이룸.

thoracic vertebare(흉추),bronchitis(기관지염),trachea(기관)와 carina(기관 분기),colonic interposition(결장 이식),hemothorax(혈흉)등이 1건

## 결론

앞면: 일반적인 폐 및 심장 질환을 관찰하고 진단하는 데 유용.

옆면: 보완적인 정보를 제공하며, 앞면에서 관찰하기 어려운 특수 문제(척추, 기관지 등)를 확인.

1-3.

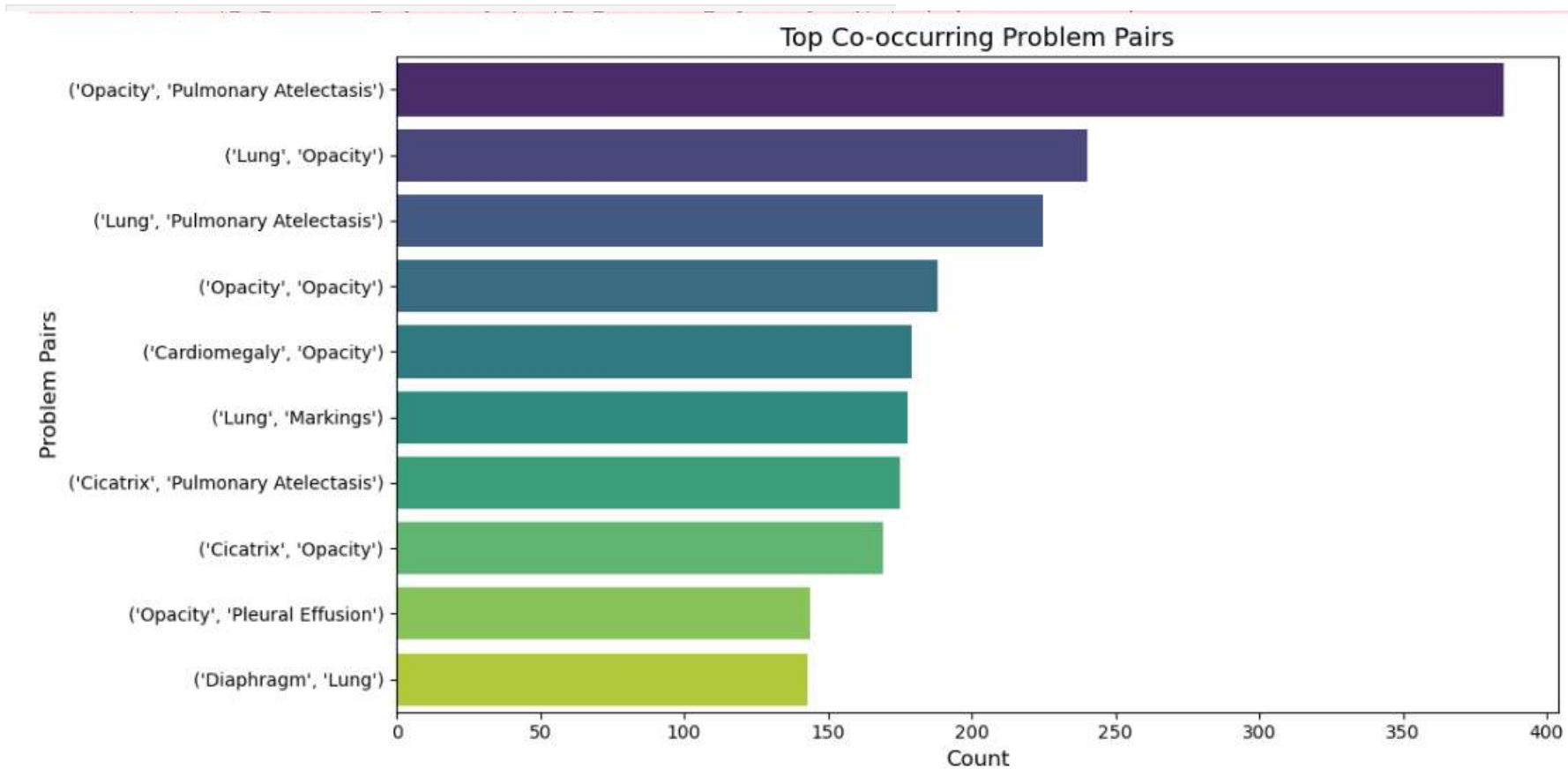
## 문제 상관관계 분석:

problem 항목에서의 데이터 끼리의 연관성을 알아보려함.

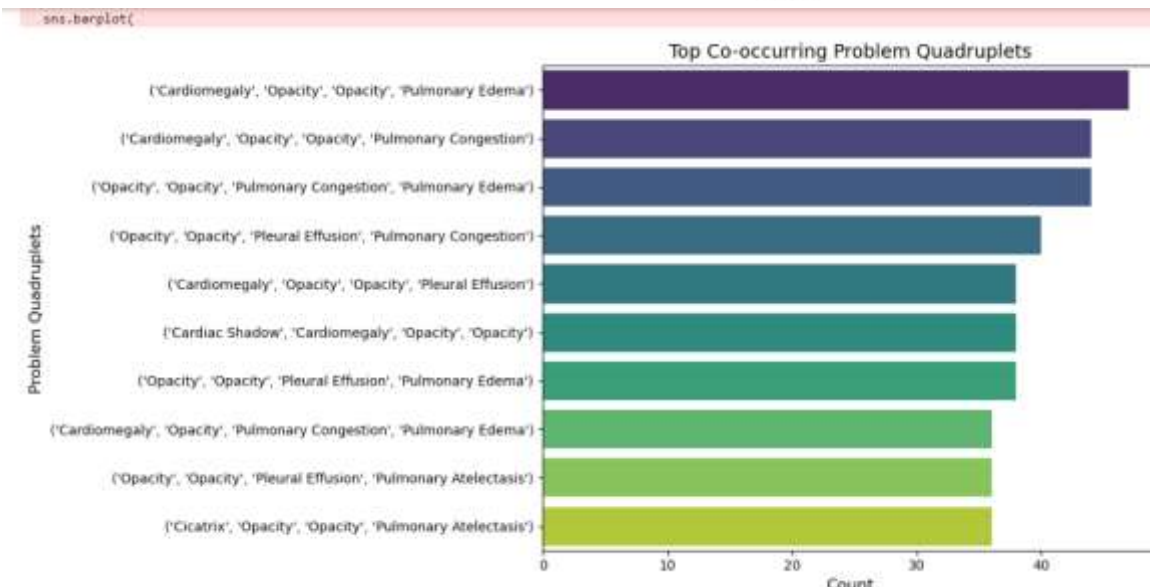
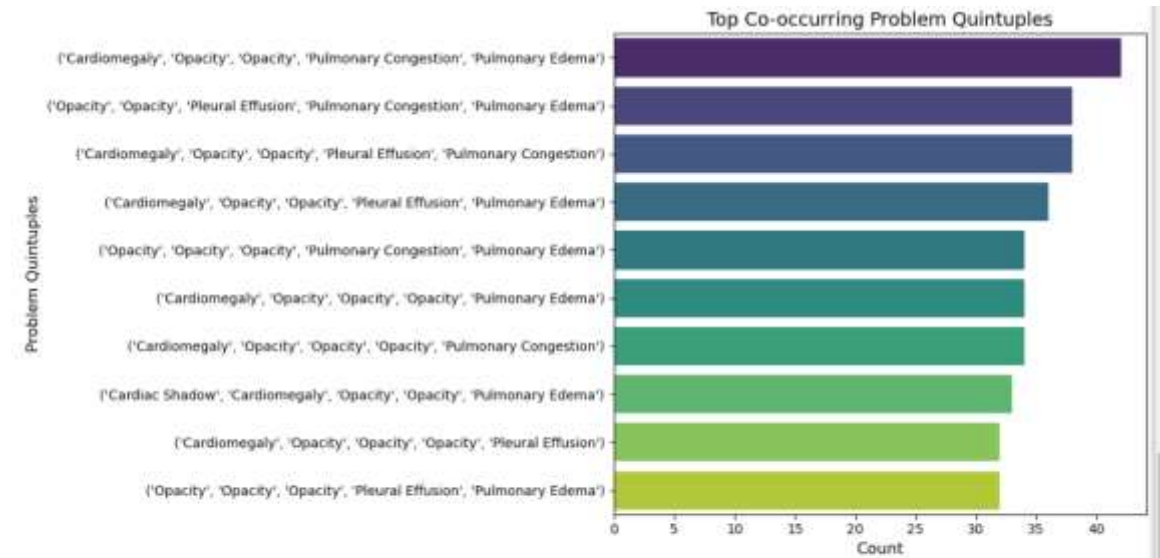
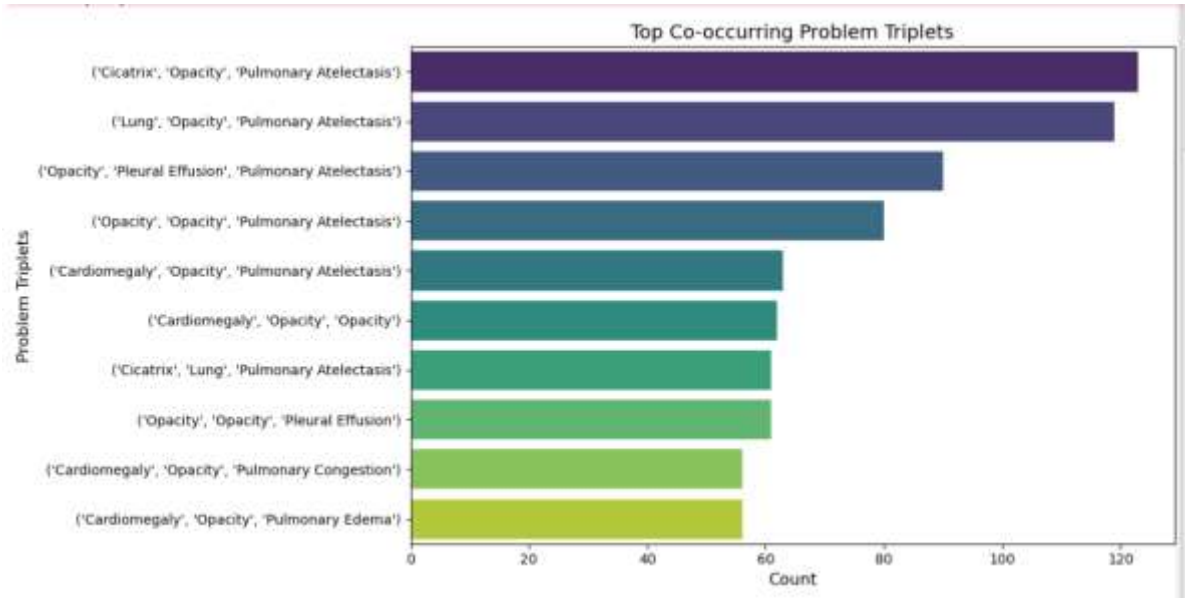
자주 함께 발견되는 문제들을 분석하여 상관관계를 시각화.

## 최종 진단(Impression) 빈도 분석 및 시각화:

최종 진단의 상위 빈도를 시각화.



# - 3개 이상의 상관관계 분석

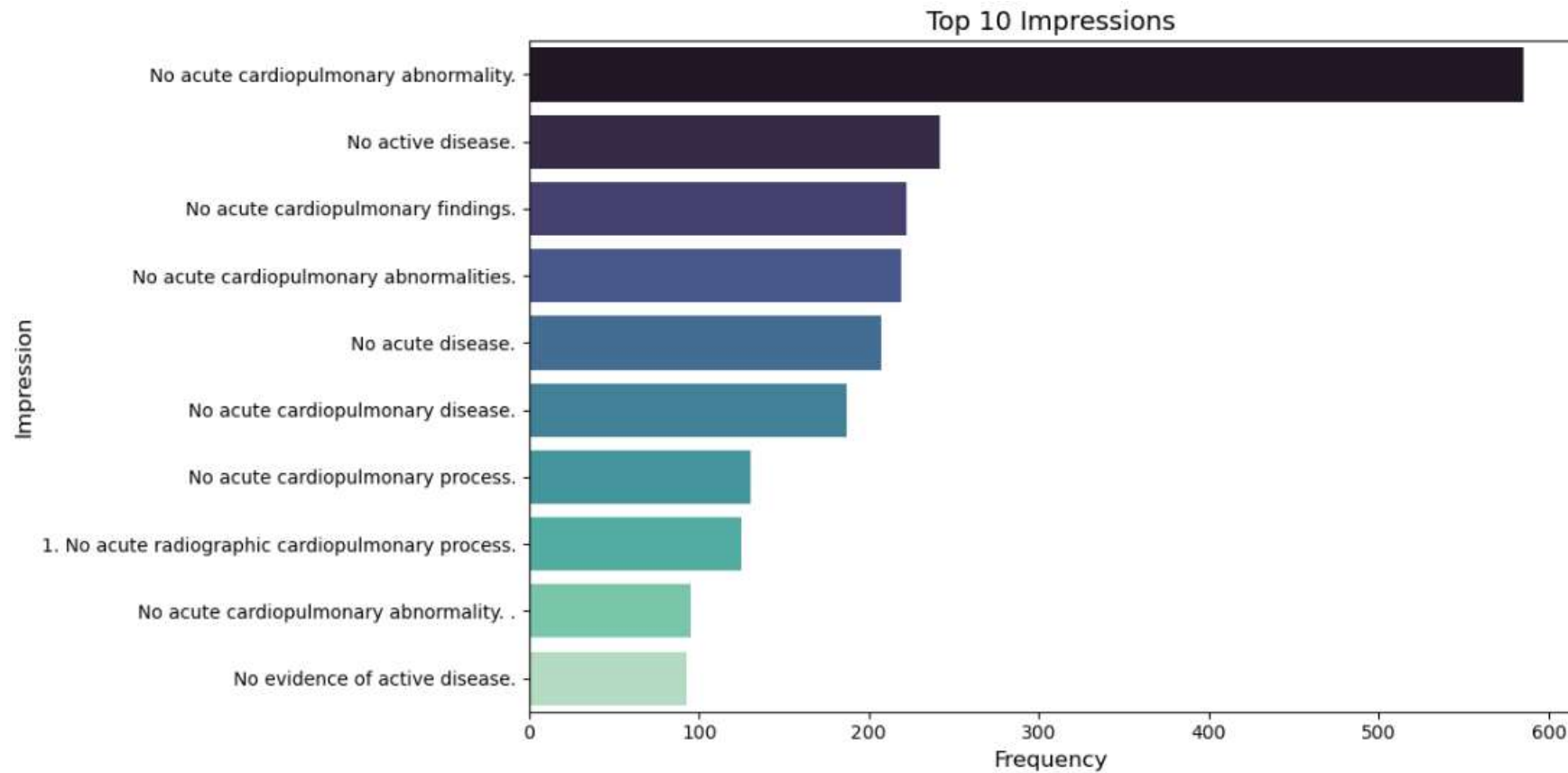




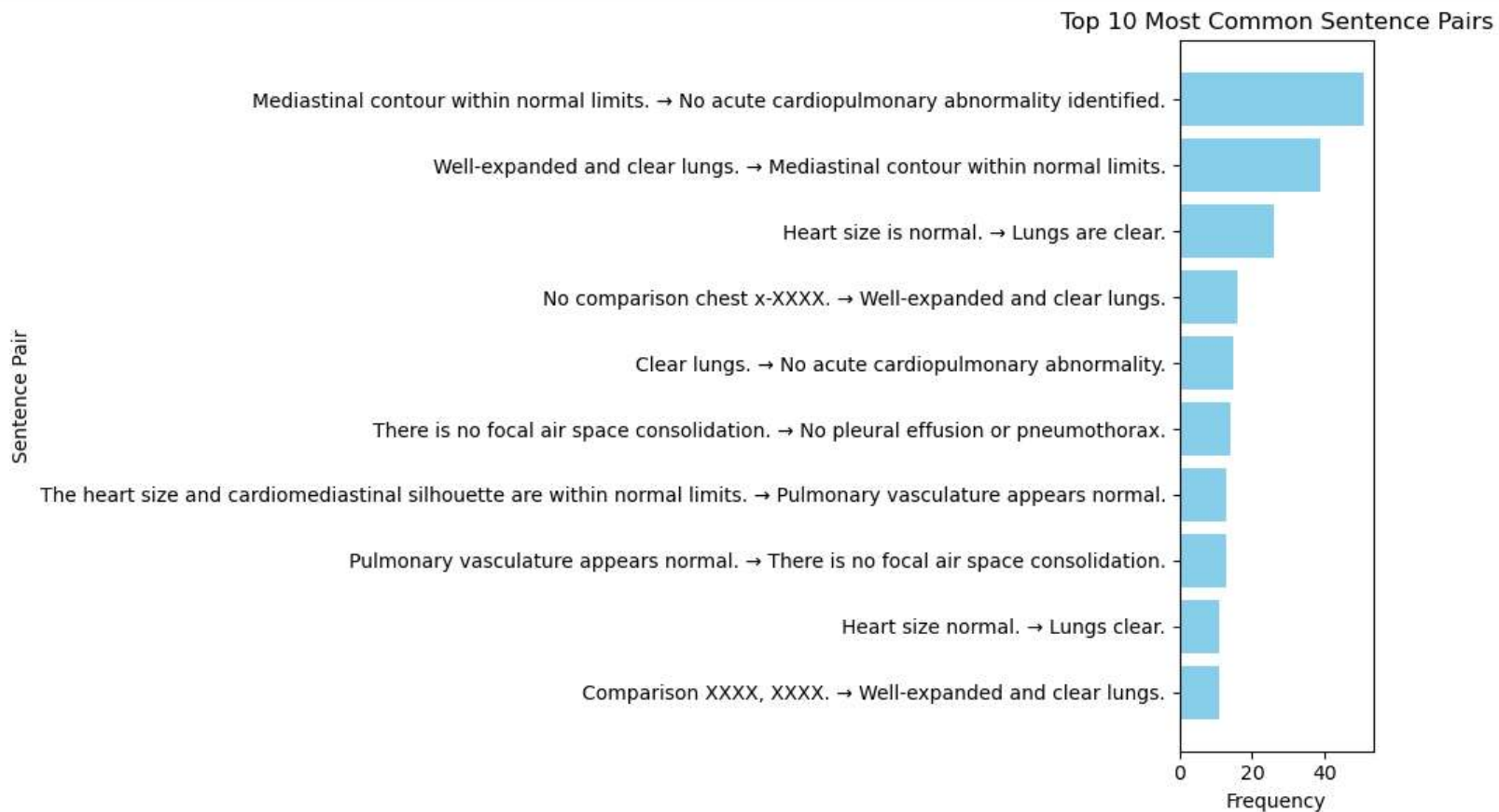
결론:최대 5개로 묶어서 상관관계 분석한 결과 두개에 비해 세개 이상으로 묶어서 상관관계를 분석한 표본의 수가 현저히 적었다.

->두개로 묶었을때가 상관성을 명확히 보여주고 직관적인 것을 알 수 있습니다.

## Impression 항목에서의 top10



# Impression 항목에서의 상관관계 분석



## impression 항목의 상관관계 분석한 결과

-> 표본이 40개이상인 항목은 mediastinal contour within normal limits.->no acute cardioplumonary banormality identified.

(종격동 윤곽은 정상 범위 내에 있습니다. -> 급성 심폐 이상은 확인되지 않았습니다.)

그래서 imprssion은 상관관계 분석에 적절치 않다.

## 분석결과:

### 문제의 상관관계 분석:

자주 함께 발견되는 문제 쌍의 상위 10개를 분석한 결과, 특정 문제들이 서로 연관되어 발생하는 경향이 있습니다.

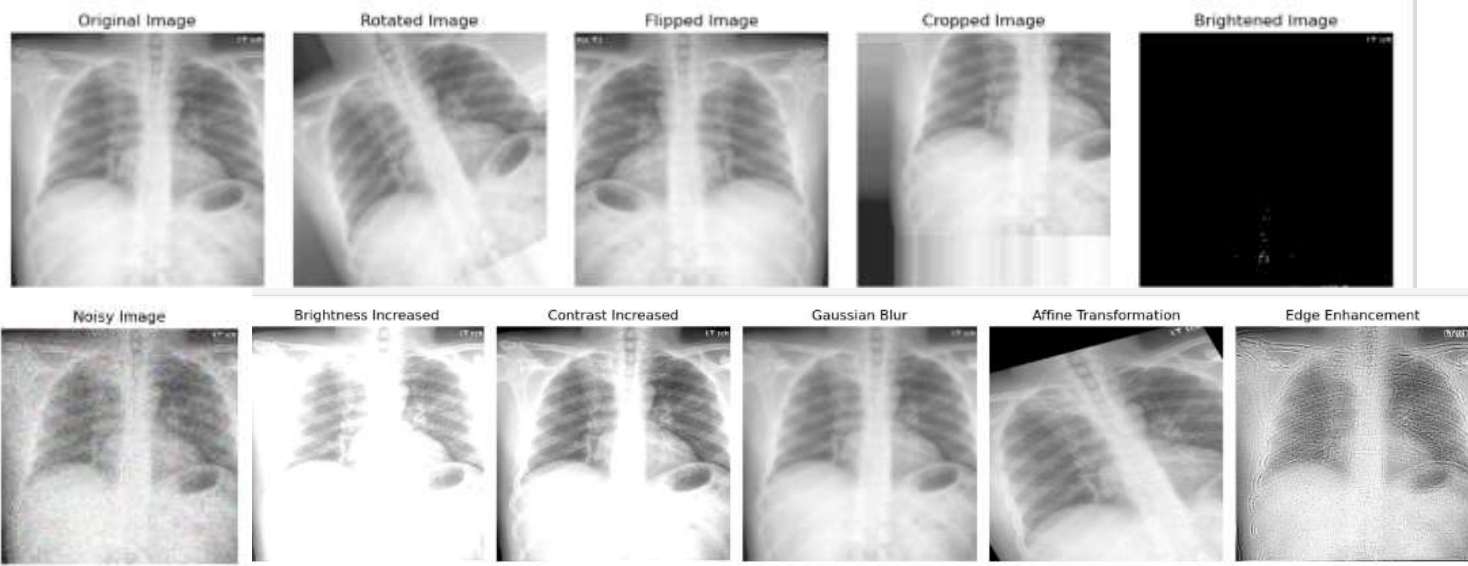
예: Lung(폐)과 Opacity(불투명도)의 빈도가 2번째로 높습니다.

### 최종 진단(Impression) 빈도:

상위 10개의 진단 결과 중 대부분이 "정상"에 해당하거나 경미한 소견을 나타냅니다.

이는 주로 "Normal chest x-ray"(정상)와 같은 진단이 빈번하다는 점을 보여줍니다.

## 2. Chest X-rays(indiana university) 이미지 데이터 전처리 방법



@ 가우시안 노이즈:이미지, 신호 또는 데이터를 왜곡시키기 위해 추가되는 랜덤 노이즈

.원본

.회전(rotation)

원하는 각도 이내로 이미지를 랜덤 회전.(40도로 설정)

.수평 및 수직 플립 (Flip)

이미지를 랜덤으로 수평 및 수직 반전.

. 무작위 자르기 (Random Cropping)

이미지의 위치를 랜덤으로 이동하여 자르기 효과.

. 밝기 조정 (Brightness Adjustment)

이미지의 밝기를 조정.

. 노이즈 추가 (Noise Addition)

가우시안 노이즈 추가.

. 이미지 밝기 증가/감소 (Brightness Adjustment)

밝기를 증가시키거나 감소시켜 데이터의 다양성을 높입니다.

. 이미지 대비 증가/감소 (Contrast Adjustment)

이미지의 밝고 어두운 영역 간의 차이를 조정합니다

. 이미지 가우시안 블러 (Gaussian Blur)

이미지를 흐리게 만들어 노이즈를 제거하거나 특정 특징을 약화시킵니다.

. 이미지 모핑 (Affine Transformation)

이미지를 평행 이동, 회전, 스케일 조정 등으로 변환합니다.

. 이미지 엣지 강조 (Edge Enhancement)

이미지에서 가장자리나 경계선을 강조하여 중요한 특징을 부각시킵니다.

# refernce

- Machine Learning & Deep Learning 0310 - 이미지 데이터 전처리 2020. 3. 10. 15:29 <https://ivo-lee.tistory.com/91>
- OpenCV Development Team. "Image Processing in OpenCV." OpenCV Official Documentation, 2025. [https://docs.opencv.org/4.x/da/d6e/tutorial\\_py\\_image\\_arithmetics.html](https://docs.opencv.org/4.x/da/d6e/tutorial_py_image_arithmetics.html)