

Hands-on: Preprocessing with MIMIC-III

Python 코드 예제

```
import pandas as pd
from presidio_analyzer import AnalyzerEngine
from presidio_anonymizer import AnonymizerEngine

# PHI 제거
analyzer = AnalyzerEngine()
anonymizer = AnonymizerEngine()

text = "Patient John Doe, MRN 123456"
results = analyzer.analyze(text, language='en')
anonymized = anonymizer.anonymize(text, results)

# 약어 확장
abbrev_dict = {'BP': 'blood pressure', 'HR': 'heart rate'}
text = text.replace('BP', abbrev_dict['BP'])

# LOINC 매핑
loinc_code = '2339-0' # Glucose [Mass/volume] in Blood
```

✓ PHI 탐지 및 제거

✓ 텍스트 정규화

✓ 약어 확장

✓ 부정 탐지

✓ 개념 링크

✓ 품질 검증