

Hands-on: Preprocessing with MIMIC-III

Python Code Example

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

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

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

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

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

✓ PHI Detection and Removal

✓ Abbreviation Expansion

✓ Text Normalization

✓ Negation Detection

✓ Concept Mapping

✓ Quality Validation