DSCI 510

Predicting Chronic Diseases from Personal Wearable Devices

Danielle Louie
November 25, 2025

Introduction

Health

Increasing public accessibility, knowledge, understanding of medical data for a healthier future.
(ex: Apple Watches, Health App)

Tech

Harnessing advancing technology to provide medical insights and improve prevention care.
(ex: AI, Machine Learning)

Data Sources

Apple Watch / Fitbit Data

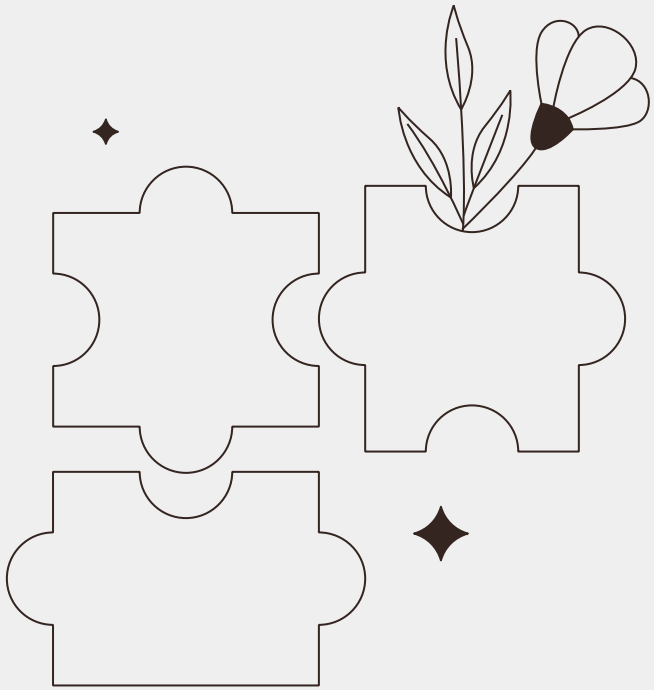
20 features and 6264 instances of physical health indicators

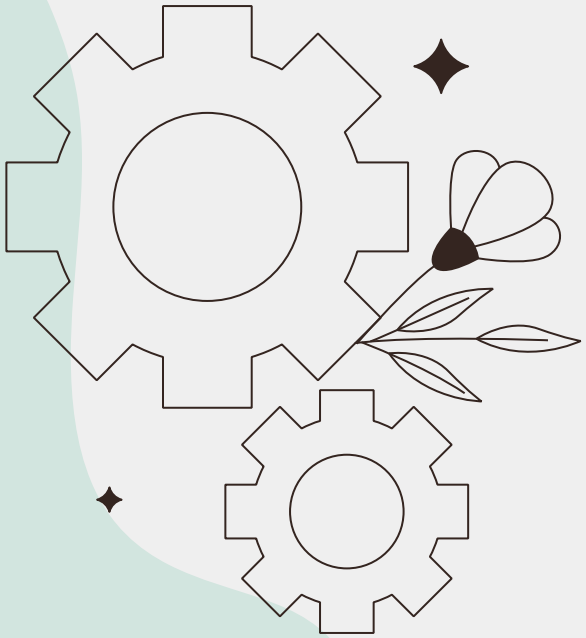
US Chronic Disease Indicators

34 public surveillance indicators and 271694 instances for chronic diseases

Behavioral Risk Factors

33 public surveillance indicators and 106260 instances for behavioral risk factors





Process and Results

EDA

Feature engineering and cleaning

Technique

Random Forest Classifier and one-hot-encoding

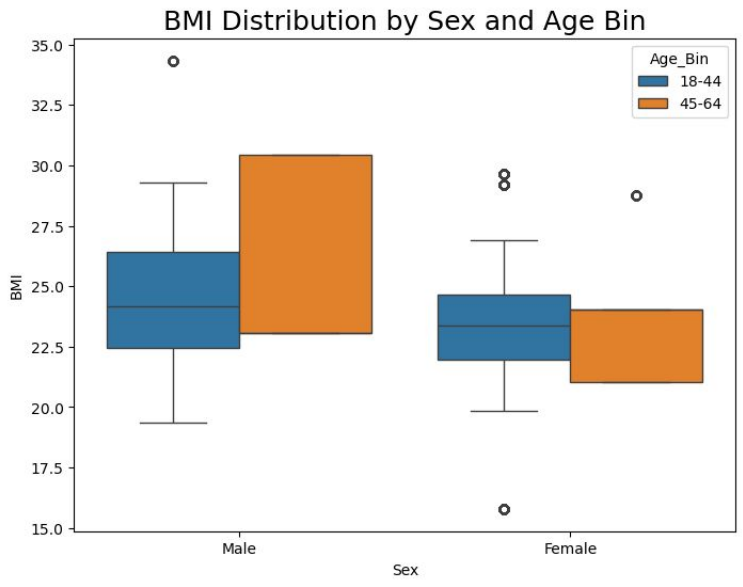
Results

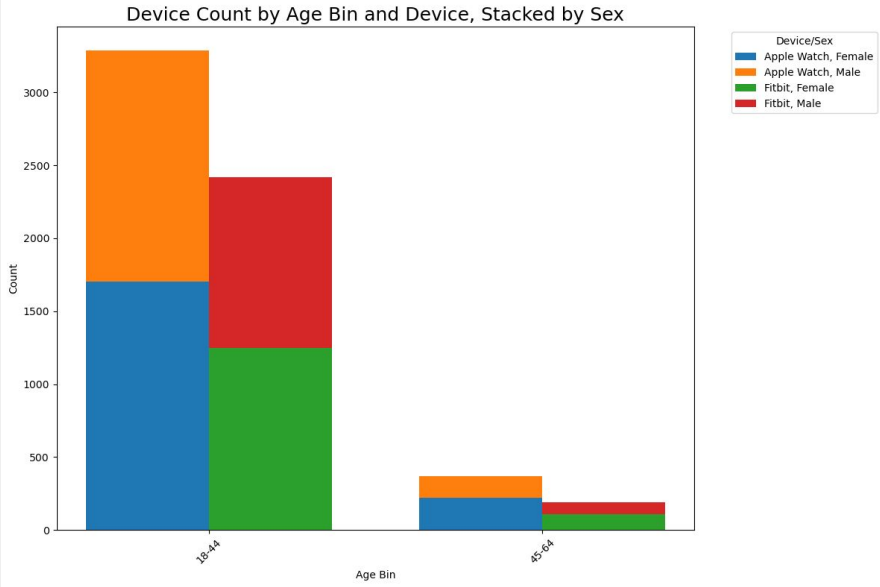
ML Algorithm bias identified through visualizations

Challenges

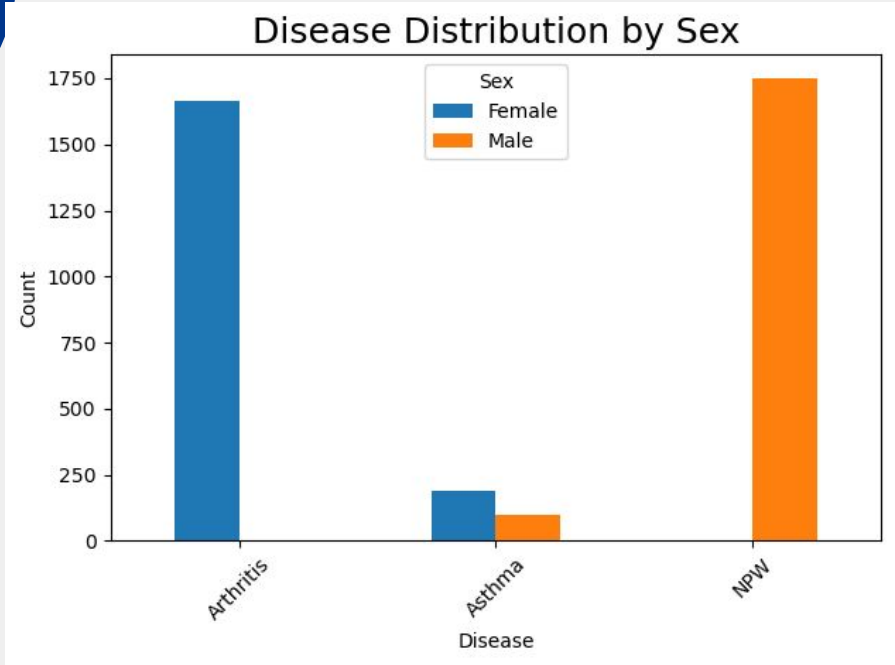
Difficulties and where improvements can be made

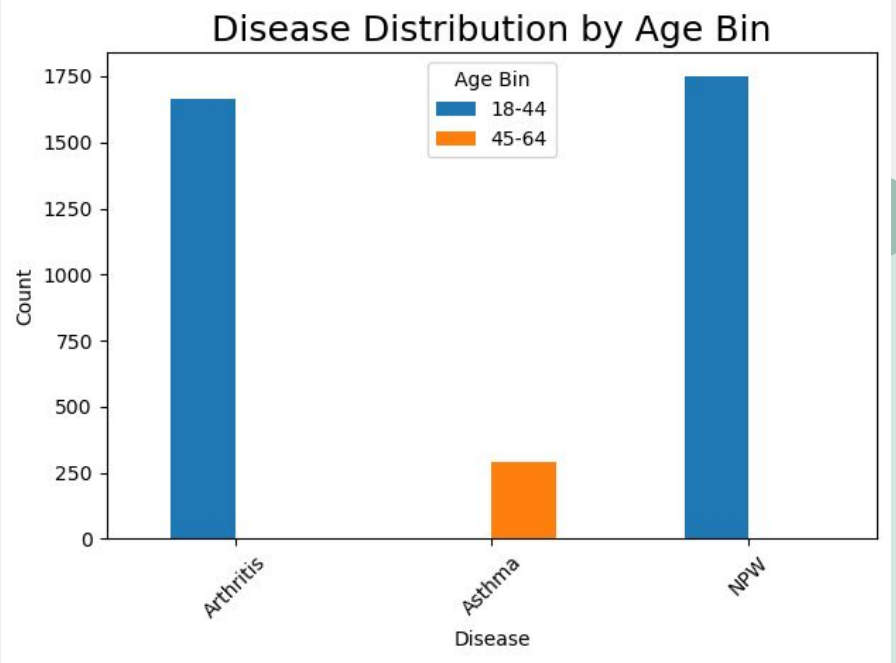
Process: EDA


BMI Distribution by Sex and Age Bin
Age_Bin
18-44
45-64
BMI
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
15.0
Male
Female
Sex

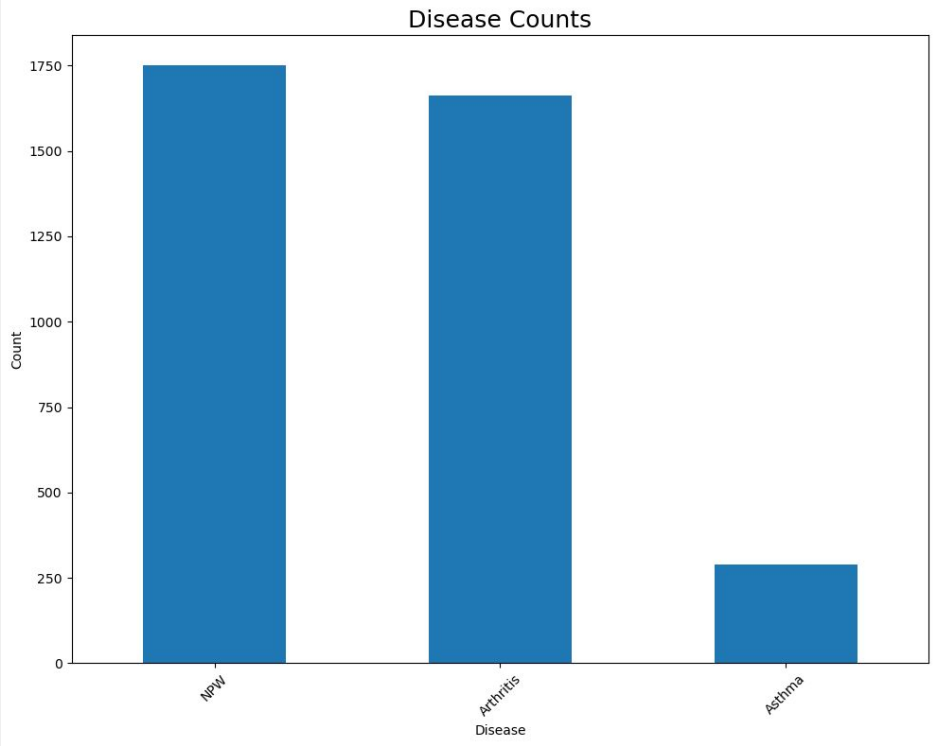

Device Count by Age Bin and Device, Stacked by Sex
Device/Sex
Apple Watch, Female
Apple Watch, Male
Fitbit, Female
Fitbit, Male
Count
3000
2500
2000
1500
1000
500
0
18-44
45-64
Age Bin

Results


Disease Distribution by Sex
Sex
Female
Male
Count
1750
1500
1250
1000
750
500
250
0
Arthritis
Asthma
NPW
Disease


Disease Distribution by Age Bin
Age Bin
18-44
45-64
Count
1750
1500
1250
1000
750
500
250
0
Arthritis
Asthma
NPW
Disease

Results


Disease Counts
1750
1500
1250
1000
750
500
250
0
Count
NPW
Arthritis
Asthma
Disease

Challenges and Improvements

Data Availability

Not much overlap, lots of assumptions made

ML Exploration

Test, explore, compare other algorithms

Features

Additional features, better engineering

Validation

Cross-validating with similar data/projects

Thank You!