

## NPHA Doctor Visits

---

- Alize Ispahani
- Manasi Acharya
- Namrata De
- Sahar Saiyed

# Introduction

## **Objective:**

Our goal is to deliver meaningful insights into the health and well-being of older adults, utilizing data from NPHA dataset.

We aspire for our app to aid policymakers and insurance providers in making well-informed decisions concerning the well-being of older adults in the USA.






## **Intended Audience:**

- **Healthcare Policymakers:** Develop informed healthcare policies.
- **Insurers:** Understand healthcare utilization patterns and optimize insurance offerings.



# Project Goals

## Primary Goals:

-  Explore the data on National Poll on Healthy Aging
-  Understand some of the basic traits and trends.
-  Train a customized Machine Learning Model to predict the number of doctor visits.
-  See predictions from the model with your own data.
-  See how changing inputs of one parameter affects your mode.

# About the Dataset

Attribute	Description	Importance
Age	Categorized into 50-64 years and 65-80 years	There is only one age group categorized.
Number of Doctors Visited	Count of different doctors seen in a year (Target variable)	Helps in understanding healthcare utilization patterns.
Physical, Mental and Dental Health	Rated on a scale from Excellent to Poor, with an option for "Refused"	Provides insights into the overall health status and its impact on doctor visits.
Various sleep disturbance indicators	Binary indicators for various sleep disturbances	Explores the impact of sleep issues on healthcare utilization.
Employment	Employment status such as full-time, part-time, retired, or not working	Impact of employment status on healthcare utilization.
Gender	Categorized as Male, Female, with options for "Refused" or "Not asked"	Examines gender-based differences in healthcare behaviors.
Race	Categorized as White, Black, Other, Hispanic, 2+ Races, with options for "Refused" or "Not asked"	Assess racial disparities in healthcare utilization.

# The Data

- **Data Processing:** Dropped 'Age' feature before training since there is only one age group.
- **Data Quality:**
  1. No missing values or known errors, clearly defined attributes names.
  2. Attribute explanations were incorrect on the UCI repository page.

## **Metadata and Protected Attributes:**

- **Documentation:** Detailed variable descriptions on the UCI repository.
- **Protected Attributes:** Race, Gender, Age.

## **Design Decisions:**

- **Visualization Tools:** Balsamiq (prototyping), Streamlit (Dashboard), Python (Matplotlib, Plotly)
- **Machine Learning Models:** Predict doctor visits and analyze feature importance. (Random Forest)
- **Bias Analysis:** Ensuring fair representation of demographic features.



# Challenges Faced

---

**Data Imbalance:** Imbalance in demographic features, particularly race.

**Sensitive Information Handling:** Ethical use and analysis of protected attributes.

**Computationally Exhaustive:** LIME and SHAP graphs take ample amount of time to reload, so optimization was needed.

# Reflection on the Process

## **Achievements:**

- Created a dashboard with actionable insights.
  - Identified key factors influencing doctor visits and health status.
- 
- Addressed potential biases and evaluated the impact of sensitive features.

## **Lessons Learned:**

- Importance of data quality.
- Ethical considerations in handling sensitive information.
- Iterative approach to refining analysis and models.

## **Future Improvement:**

- Enhanced Models: Developing more sophisticated models to improve prediction accuracy.
- Broader Applications: Applying similar methodologies to other age groups or health datasets.
- Policy Impact: Collaborating with policymakers to translate insights into tangible healthcare improvements.



Thank you

---