



# Final PROJECT PROPOSAL



## **Prepared by**

Mohamed Ahmed  
Mohamed Hossen

In this project proposal we will go through the dataset we want to use for our project graduation summation, Behavioral Risk Factor Surveillance System.

## **Presented to**

Dr.Sherihan Ali

# Executive Summary

The objective of the BRFSS is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases in the adult population.

Factors assessed by the BRFSS include tobacco use, health care coverage, HIV/AIDS knowledge or prevention, physical activity, and fruit and vegetable consumption. Data are collected from a random sample of adults (one per household) through a telephone survey.



## 1

### Problem Statement

Our dataset presents a complex challenge due to its extensive 335-feature composition, which includes a mix of essential, moderate, and non-essential features.

This heterogeneous feature set introduces significant complexity, making the dataset difficult to manage, interpret, and maintain effectively.

The diverse range of features, coupled with their varying levels of importance, creates several key challenges:

- Excessive feature count complicates data preprocessing and model development
- Mixed feature relevance leads to potential noise and reduced predictive performance
- The dataset's complexity increases the risk of overfitting and reduces overall interpretability






# Project Overview

The project will involve the development of this data set to address multiple targets and only obtain the features that are either directly related to the target or have a high Correlation value with the target variable.

## KEY FEATURES TO BE INCLUDED IN THE PROJECT ARE :

- **Clear , readable and usable headers**

This will be our first challenge since our data set headers are code for features addressed in the [BRFSS survey code book](#) 

- **Clear, monitored target features**

Enabling the audience to easily understand what we are scraping, seeking , and conducting in this data set.

- **Visualization**

Addressing audience with charts and visualization techniques to address the point we are concerned about.





# about Dataset



The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.

## Dataset Link

[Behavioral Risk Factor Surveillance System](#)

[Public health surveys of 400k people from 2011-2015](#)

[From Centers for Disease Control and Prevention and 1 collaborator .](#)

More clear downloadable csv on kaggle

[Behavioral Risk Factor Surveillance System](#)



## Our Team

Omar Samy Mohsen

Nour Eldin Mamdouh Ahmed

Mina Wahid Youssef

Mohamed Ahmed Mohamed

Rahma salah

Eman Gamal Badry