# Stratified Sampling Analysis Report

To: Senior Data Analyst

From: Ovili Ifeakachukwu

Date: 2024-11-02

Subject: Analysis of Stratified Sampling for the SHS Dataset

--------------------------------------------------

## Objective

The objective of this analysis was to perform stratified sampling on the SHS dataset to derive population-representative insights across different provinces. By applying stratification, this approach aims to ensure each province's representation, enabling a more accurate understanding of survey data trends across diverse regions.

## Methodology

1. Data Preparation  
 The SHS dataset was structured to include a new column, `Stratum`, which segments the data by province (`PROVINCP`). This stratification allows for precise sample allocation across provinces.

2. Stratified Sampling Implementation  
 - Predefined sample sizes (`nh`) were assigned to each stratum (province) based on desired representation.

- A Stratified Simple Random Sampling Without Replacement (STSRSWOR) approach was applied to create a sample that reflects population proportions across the provinces.

- The sample data was further refined by merging specific columns of interest and removing rows with missing values.

## Real-Life Application

This analysis simulates a structured approach to survey sampling, where stratification ensures equitable representation from multiple population segments (provinces). Such a method is valuable in governmental and socio-economic studies to understand trends within various demographic segments while maintaining sample accuracy and representation.

## Conclusion

Through this analysis, a proportionally representative sample of the SHS dataset was achieved using stratified sampling methods. The results can be reliably used to infer conclusions across the entire population with minimized bias, thus ensuring accurate and relevant insights for decision-making.