**Documentation: Analysis of Average Percentage of Deaths Attributable to Indoor House Pollution**

**1. Introduction**

This document outlines the scope and potential insights derived from the provided dataset concerning the average percentage of deaths in various countries and regions, with a specific focus on their potential attribution to indoor house pollution. The analysis aims to identify trends, geographical correlations, and demographic vulnerabilities related to this environmental health issue.

**2. Data Description**

The primary dataset includes the following columns:

* **Country/Region:** The geographical entity for which the data is reported
* **Avg Percent of Death:** The average percentage of deaths that are potentially attributable to indoor house pollution within the specified country/region.

**Planned Data Addition:**

* **Death by Age:** This additional dataset will include information on the distribution of deaths potentially attributable to indoor house pollution across different age groups within the studied countries/regions.

**3. Research Questions**

The following questions will guide the data analysis and interpretation:

**3.1. General Overview of Death Percentages**

1. **Which countries/regions exhibit the highest and lowest average percentage of deaths potentially attributable to indoor house pollution?**
   * *Objective:* To identify the top and bottom performers in terms of this health indicator based on the current data.
2. **How does the average percentage of deaths vary across different regions/countries?**
   * *Objective:* To describe the overall distribution and range of the "Avg Percent of Death" across all entities in the dataset, providing a broader context to the extreme values.

**3.2. Geographical Correlation**

1. **Is there a geographical correlation with the average percentage of deaths?**
   * *Objective:* To investigate whether countries/regions located in specific continents, climate zones, or with shared geographical characteristics demonstrate similar trends in the average percentage of deaths. This analysis may involve mapping the data or grouping countries by region.

**3.3. Age-Related Vulnerability (Post "Death by Age" Data Integration)**

1. **Are certain age groups more vulnerable to deaths related to indoor house pollution?**
   * *Objective:* To determine if specific age demographics (e.g., infants, young children, the elderly, or working-age adults) are disproportionately affected by indoor house pollution-related deaths across the surveyed regions.
2. **Does the average percentage of death due to indoor house pollution vary significantly among different age demographics across various countries/regions?**
   * *Objective:* To assess if the age-specific impact of indoor house pollution-related deaths shows different patterns or magnitudes across distinct countries or regions.
3. **Is there a relationship between the overall average percentage of death in a region and the age distribution of those deaths?**
   * *Objective:* To explore whether regions with a higher overall average percentage of indoor house pollution-related deaths also exhibit particular age-distribution patterns (e.g., if higher overall rates correlate with a higher burden on specific age groups).

**4. Expected Outcomes**

By answering these questions, the analysis aims to provide:

* A clear understanding of the global or regional hotspots and low-impact areas for indoor house pollution-related deaths.
* Insights into potential geographical factors influencing these death rates.
* A deeper understanding of the demographic groups most at risk, which can inform targeted public health interventions and policy development.