

# Data Report: Analysis of COVID-19 Death Ratios by Age Group

By: Sumon Kazi(23293505)

## Question:

### Research Question

Which age group suffered the most during COVID-19?

## Objectives

1. Identify age groups most affected by COVID-19.
2. Understand age-related vulnerabilities during the pandemic.
3. Provide data-driven insights for public health strategies.

## 1. Chicago COVID-19 Dataset

- **Why Chosen:** Provides detailed city-level COVID-19 data, including deaths and cases categorized by age groups, ideal for analyzing urban trends.
- **Source:** City of Chicago Open Data Portal.
- **URL:** "<https://data.cityofchicago.org/api/views/6irb-gasv/rows.csv?accessType=DOWNLOAD>"
- 
- **Data Content:** Columns include "Age Group," "Cases," "Deaths," "Vaccinated," and "Week End Date."
- **Data Quality:**
  - Missing values were filled using the median for numeric fields.
  - Duplicate entries and inconsistent date formats were resolved.
- **License:** Openly accessible under the Open Government Data Act.

## 2. CDC COVID-19 Vaccination Dataset

- **Why Chosen:** Offers national-level vaccination data segmented by age groups, enabling analysis of vaccination impact on mortality trends.

- **Source:** Centers for Disease Control and Prevention (CDC).
- **URL:** "<https://data.cdc.gov/api/views/hk9y-quqm/rows.csv?accessType=DOWNLOAD>"
- 
- **Data Content:** Includes columns like "Age Group," "Vaccination Completed," "Population," and "Week Ending Date."
- **Data Quality:**
  - Missing data replaced with default values or category medians.
  - Unified inconsistent "Age Group" labels across datasets.
- **License:** Available under the Open Data Commons Public Domain Dedication License (PDDL)

## Methodology

1. **Data Gathering:**  
Collected datasets from Chicago's data portal and the CDC, documenting their structure and coverage.
2. **Data Preparation:**
  - Handled missing data using medians for numeric fields and placeholders for text fields.
  - Standardized age group labels (e.g., unified "65+" across datasets).
  - Converted date columns to a consistent format using `pandas.to_datetime`.
3. **Analysis:**
  - Examined death trends by age group and vaccination coverage.
  - Calculated critical metrics:
    - **Death-to-Case Ratio** = Deaths / Cases.
    - **Vaccination Coverage** = Vaccinated Population / Total Population.
4. **ETL Pipeline:**
  - Built an automated pipeline using Python, Pandas, and SQLite.
  - Filtered essential columns: "Age Group," "Deaths," "Cases," "Vaccination Completed."

- Stored cleaned and unified data in an SQLite database for efficient querying.
5. Documentation:
- Generated visual charts for death ratios and vaccination rates by age group.  
Summarized results for actionable insights.
- 

## Results

- Deaths:
    - The 65+ age group had the highest death ratios.
    - Death rates were minimal in the under-18 group.
  - Vaccinations:
    - Highest vaccination rates observed in the 65+ group.
    - Lower vaccination coverage in younger populations (18–24), correlating with lower death counts.
  - Trends:
    - Weekly death spikes coincided with major COVID-19 surges.
    - Vaccination substantially reduced death ratios across all age groups.
- 

## Limitations

- Geographic Scope: Chicago data is local; CDC data reflects national trends, leading to potential inconsistencies.
  - Reporting Delays: Death and vaccination data may include lags.
  - Generalization: Results may not fully account for healthcare disparities or population density differences.
- 

## Conclusion

Older adults (65+) faced the highest COVID-19 mortality, emphasizing their vulnerability. Vaccinations proved critical in reducing deaths, with lower ratios among vaccinated groups. This analysis highlights the importance of targeted vaccination campaigns and data-driven public health strategies.