# Project Summary – CMS Nursing Home COVID-19 Dashboard

This project is a comprehensive data cleaning and analysis exercise using publicly available data from the **Centers for Medicare & Medicaid Services (CMS)**, which tracks weekly COVID-19 statistics across U.S. nursing home facilities. The dataset includes critical metrics such as resident and staff vaccination rates, confirmed COVID-19 cases, and quality assurance check outcomes — all vital for monitoring healthcare facility compliance and pandemic control efforts.

## Objective

To transform a messy, inconsistent dataset into a clean, insight-driven Excel dashboard that can support public health decision-making, highlight facility-level performance, and expose the implications of data quality on reported outcomes.

## Process Overview

### 1. Data Acquisition

- Sourced raw data from CMS Nursing Home COVID-19 reporting files
- Total initial dataset: 14,791 rows (excluding headers), with state, facility, and vaccination data

### 2. Data Cleaning

- Performed deep cleaning using **Power Query** in Excel:
    - Removed 253 records with missing federal provider numbers
    - Filtered out facilities that did not **submit data** or failed the **Quality Assurance Check**
    - Replaced or excluded abnormal provider IDs (e.g., 8-digit or invalid 6-digit codes)
    - Removed or handled columns with significant null values (e.g., resident weekly confirmed COVID cases)
    - Reformatted percentage columns and recalculated "recent %" fields to validate logic
    - Created a new column: **"Provider State (Full Name)"** for enhanced visualization
- Final retained dataset: **62%** of the original rows

### 3. Data Quality Notes

- Facilities included were only those with valid submission status and QA approval
- As a result, many vaccination-related metrics showed **100% compliance**, signaling **reporting bias** introduced through QA-based filtering

## Dashboard and Visualizations

Created an interactive **multi-page Excel dashboard** that includes:

- **Top 20 States** by:
  - Total COVID-19 cases
  - Staff vaccination rates
  - Resident vaccination rates
- **Facility-level view** for exploring individual performance
- Dynamic filters using **PivotTables, Slicers, and Measures**
- Tooltips, data labels, and conditional formatting for clarity

### Key Calculated Measure Example:

Vaccination Rate (%) = SUM(Number of Up to Date) / SUM(Number Eligible)

## Key Insights

- **Data completeness significantly affects the reliability of national trends**
  - Only 66% of retained facilities passed the CMS QA check
- **Most retained facilities reported 100% vaccination** — indicating strong compliance or denominator inconsistencies
- **State-level analysis reveals disproportionate case burdens**, with variation in staff vaccination vs resident coverage
- Filtering only QA-passed data may exclude facilities with greater risk or lower compliance — skewing insights toward "best performers"

## Tools & Skills Used

| Tool / Skill | Use Case |
|---|---|
| Excel | Dashboard development, data modelling |
| Power Query | Data cleaning and transformation |
| PivotTables | Aggregated reporting |
| DAX Measures | Dynamic metric calculation |
| Data QA | Ensuring report integrity and reliability |

## Limitations

- Final analysis is based on QA-passed facilities only; may not reflect underperforming or non-compliant nursing homes.
- Percentages near 100% should be interpreted in light of retained data only.

## Outcome & Value

This project demonstrates my ability to:

- Handle real-world messy healthcare data

- Apply business logic filters with purpose
- Build interactive, presentation-ready Excel dashboards
- Translate public health data into practical, visual insights

.