**Data Sources**

COVID-19 Vaccination and Case Trends by Age Group, United States Metadata Updated: [September 28, 2021](https://catalog.data.gov/dataset/covid-19-vaccinations-in-the-united-statescounty#sec-dates)

Trends in vaccinations and cases by age group, at the US national level. Data is stratified by at least one dose and fully vaccinated.

|  |  |
| --- | --- |
| **Resource Type** | Dataset |
| **Metadata Created Date** | July 1, 2021 |
| **Metadata Updated Date** | September 28, 2021 |
| **Publisher** | CDC |

Explain the attributes of the data you have used :

The data that we have used is up to date data coming from the most trustful federal agency in the united states (Centers of Disease Control and Prevention the CDC) works under the Department of Health and Human Services.

Data represents all vaccine partners including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities.

**Data Preparation**

Our data consist of 5 columns and 2573 rows, presents the vaccination trend by age over 10 months recordings period. It’s very well structured and organized data, the only issue that we may need to deal with is the outliers because they would significantly affect our analysis. In order to prevent that we need to :

1. Detect the outliers in the data using boxplot or scatter plot.
2. Change the values of the outliers using trimming function.