Data Documentation

Data Source:

The data has been obtained from the Center for Disease Control and Prevention (CDC) website and it represents the overall US COVID-19 Vaccine administration and vaccine equity data at county level. The 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. Primary key elements contained in the dataset are the percentages of first dose, second dose, and booster dose vaccinations by different age groups of people as per county and state in the US.

Accessing the data:

The dataset has been obtained from the CDC website and includes COVID-19 Vaccinations in the United States by County which can be accessed through an API. The dataset has been obtained from the website: https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh and the endpoint of the API is: https://data.cdc.gov/resource/8xkx-amqh.json.

The CSV format of the file has been downloaded from the link provided above and is used for processing. The data set consists of about 1526072 rows [1] and has various fields such as county of residence, recipient state, and the proportion of fully vaccinated people whose Federal Information Processing Standards (FIPS) code is reported and matches a valid county FIPS code in the jurisdiction.

The dataset has also been accessed through web scraping https://data.cdc.gov/browse?category=Vaccinations. After web scraping this link, the dataset's information page has been accessed through web crawling the URL extension. This helped me in retrieving the API endpoint.

Data Dictionary:

Amongst 66 variables, only few have been described which have been used for analysis and user interaction in this project.

- Completeness_pct: Represents the proportion of fully vaccinated people whose Federal Information Processing Standards (FIPS) code is reported and matches a valid county FIPS code in the jurisdiction.
- Recip_County: County of Residence
- Recip_State: State of Residence
- Booster_Doses: Total number of people who are fully vaccinated and have received a booster (or additional) dose.
- SVI_CTGY: SVI Category, Social Vulnerability Index (SVI) refers to the potential negative effects on communities caused by external stresses on human health. Such stresses include natural or human-caused disasters, or disease outbreaks. Reducing social vulnerability can decrease both human suffering and economic loss.
- Metro_status: Metro/Non-metro, refers to the type of county

References:

[1] NCIRD. (2020, December 13). COVID-19 Vaccinations in the United States, County. Centers for Disease Control and Prevention. Retrieved March 22, 2022, from https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh