Project Objective:

The objective of this project is to collect, merge, and analyze COVID-19 vaccination data from various sources, including Our World in Data GitHub repository, to gain insights into the global vaccination progress. The project aims to provide a comprehensive dataset that includes country-level vaccination information, such as the total number of vaccinations, people vaccinated, people fully vaccinated, daily vaccinations, and more. This dataset will be used for research and analysis to understand vaccination trends and their impact on public health.

Design Thinking Process:

1. Understanding the Problem: The project starts with a clear understanding of the problem: the need for a consolidated and reliable dataset on COVID-19 vaccinations. This involves identifying the key data sources, data fields, and the format in which data is available.

2. Data Cleaning and Preprocessing:
   - Data quality is assessed to handle missing values, outliers, and inconsistencies.
   - Data preprocessing techniques are applied to ensure that the data is in a consistent format for analysis.

3. Data Analysis and Visualization:
   - The cleaned dataset is analyzed to extract meaningful insights.
   - Visualizations and statistical analysis are performed to understand trends, distribution, and correlations within the data.

4. Development Phases:
   a. Data Cleaning and Preprocessing:
      - Address missing values by imputation or removal, depending on the nature of the data.
      - Handle outliers and inconsistencies in the data.
      - Convert data types to ensure consistency.

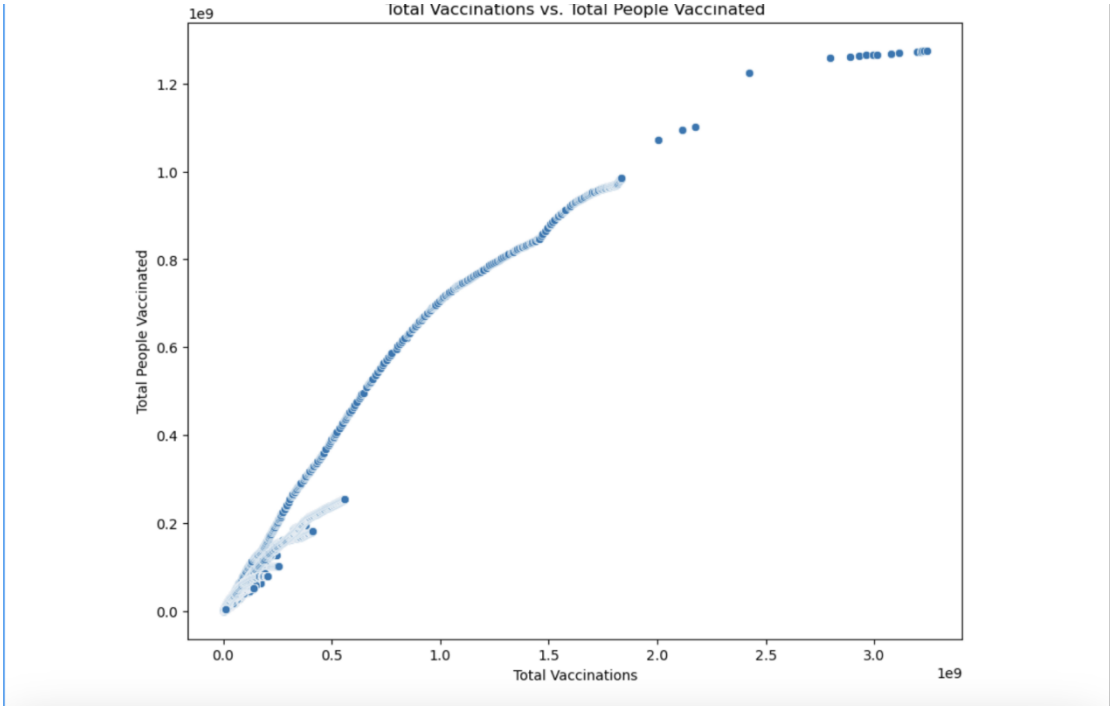
   b. Data Analysis and Insights:
      - Explore the dataset to identify vaccination trends, such as the daily vaccination rate, cumulative total, and coverage per hundred.
      - Generate visualizations, including time series charts, maps, and statistical summaries.
      - Identify patterns, anomalies, and key statistics.

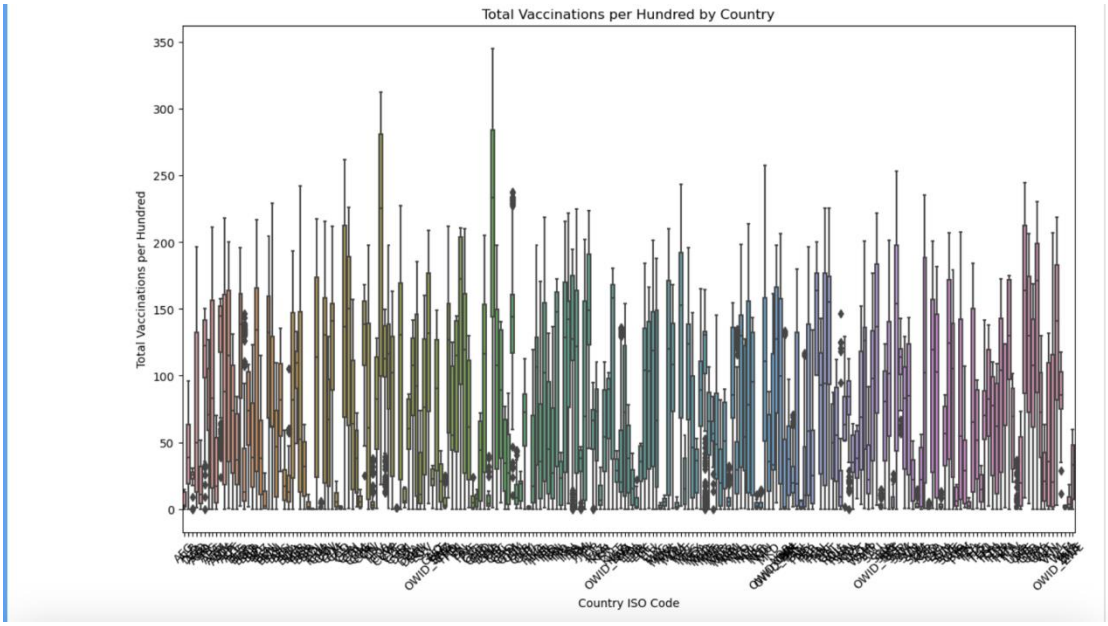
   c. Documentation and Reporting:
      - Create documentation that explains the data sources, cleaning processes, and analysis methods used.
      - Generate reports or dashboards to present the findings in an accessible and informative manner.

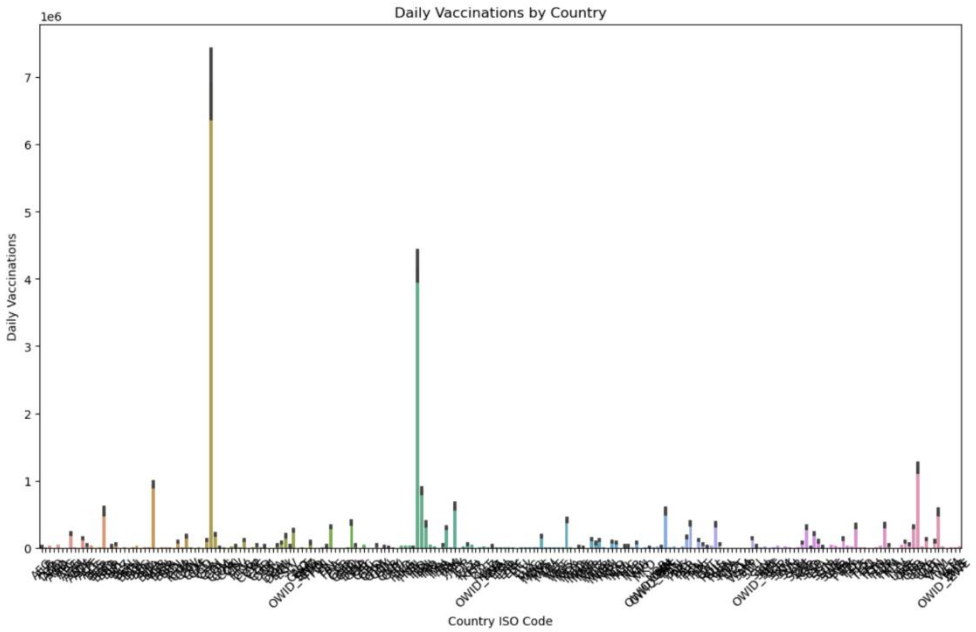
   d. Maintenance and Updates:
      - Establish a process for regular data updates to keep the dataset current.
      - Implement version control to track changes and updates to the dataset.

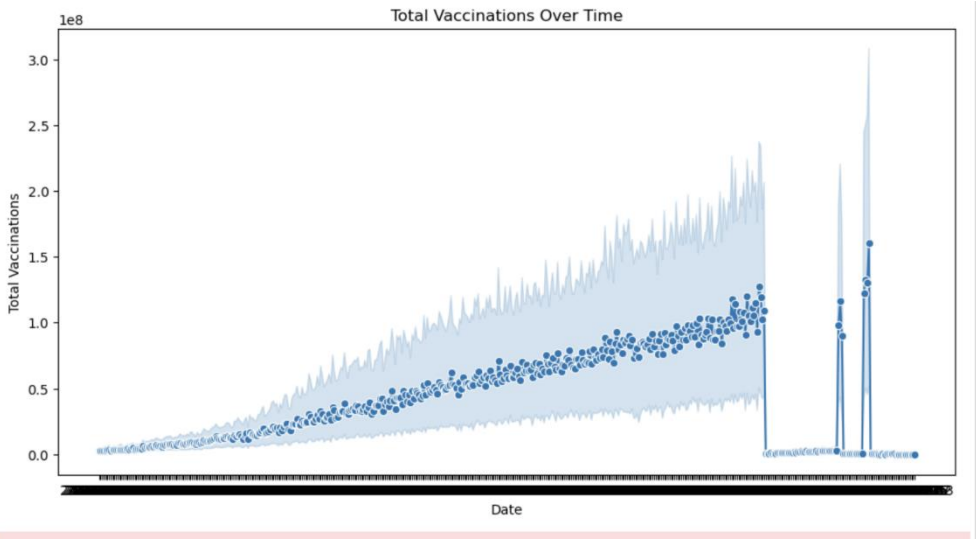
By following these design thinking principles and development phases, the project aims to provide valuable insights into the global COVID-19 vaccination efforts and contribute to informed decision-making and public health strategies.

GRAPHICAL VISUALIZATION






Daily Vaccinations by Country
1e6
Daily Vaccinations
Country ISO Code


Total Vaccinations Over Time
1e8
Total Vaccinations
Date

STATISTICAL REPRESENTATION

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   country                       86512 non-null  object
 1   iso_code                      86512 non-null  object
 2   date                          86512 non-null  object
 3   total_vaccinations            43607 non-null  float64
 4   people_vaccinated             41294 non-null  float64
 5   people_fully_vaccinated       38802 non-null  float64
 6   daily_vaccinations_raw        35362 non-null  float64
 7   daily_vaccinations            86213 non-null  float64
 8   total_vaccinations_per_hundred    43607 non-null  float64
 9   people_vaccinated_per_hundred     41294 non-null  float64
 10  people_fully_vaccinated_per_hundred  38802 non-null  float64
 11  daily_vaccinations_per_million    86213 non-null  float64
 12  vaccines                      86512 non-null  object
 13  source_name                   86512 non-null  object
 14  source_website                86512 non-null  object
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
None
      total_vaccinations  people_vaccinated  people_fully_vaccinated  \
count        4.360700e+04       4.129400e+04             3.880200e+04
mean         4.592964e+07       1.770508e+07             1.413830e+07
std          2.246004e+08       7.078731e+07             5.713920e+07
min          0.000000e+00       0.000000e+00             1.000000e+00
25%          5.264100e+05       3.494642e+05             2.439622e+05
50%          3.590096e+06       2.187310e+06             1.722140e+06
75%          1.701230e+07       9.152520e+06             7.559870e+06
max          3.263129e+09       1.275541e+09             1.240777e+09

      daily_vaccinations_raw  daily_vaccinations  \
count            3.536200e+04        8.621300e+04
mean             2.705996e+05        1.313055e+05
std              1.212427e+06        7.682388e+05
min              0.000000e+00        0.000000e+00
25%              4.668000e+03        9.000000e+02
50%              2.530900e+04        7.343000e+03
75%              1.234925e+05        4.409800e+04
max              2.474100e+07        2.242429e+07

      total_vaccinations_per_hundred  people_vaccinated_per_hundred  \
count                    43607.000000                   41294.000000
mean                        80.188543                      40.927317
std                         67.913577                      29.290759
min                          0.000000                       0.000000
25%                         16.050000                      11.370000
50%                         67.520000                      41.435000
75%                        132.735000                      67.910000
max                        345.370000                     124.760000

      people_fully_vaccinated_per_hundred  daily_vaccinations_per_million
```

count	38802.000000	86213.000000
mean	35.523243	3257.049157
std	28.376252	3934.312440
min	0.000000	0.000000
25%	7.020000	636.000000
50%	31.750000	2050.000000
75%	62.080000	4682.000000
max	122.370000	117497.000000