**COVID-19 Data Analysis Project Stage – I**

**Report on COVID-19 Dataset (Task 2 Part -1)**

**Table of contents**

* Introduction
* Overview
* Dataset Description
* Datatype – variable dictionary
* Conclusion
* References

**Team Members**

Keerthi Devireddy

Sony Annem

Supriya Kottam

**Introduction**

This report provides an overview of the covid- 19 dataset available on usfacts.org, along with the data overview, categories, and a datatype – variable dictionary which will be helpful in understanding, maintaining and communicating the structure of the underlying data in the dataset. The dataset provided is a valuable resource for keeping track of accurate covid cases information across all the 50 states of the United States.

**Overview**

Source: usfacts.org

URL: <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>

Dataset overview:

Provides information like number of cases reported, number of deaths recorded, average cases and deaths and hospitalizations per week. It includes data at the national, state and county levels).

**Dataset Description**

The covid-19 dataset provides comprehensive information about the spread and impact of the covid-19 virus in the United States. It includes data at the national, state and county levels. It indulges information like number of cases reported, number of deaths recorded, average cases and deaths and hospitalizations per week.

**Datatype – variable dictionary**

The data dictionary provides a clear understanding of the variables present in the COVID-19 dataset, their definition, datatypes, possible values, and whether they include null data, or the field is required for analysis or not. It serves as a valuable reference for data analysis and interpretation.

**Geographic Information:** State, StateFIPS, County, CountyFIPS

**COVID Information:** Cases, Deaths

**COVID-19 Datasets**

**1.COVID Cases Dataset (covid\_confirmed\_usafacts.csv)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Definition** | **Data type** | **Possible values** | **Any missing values?** | **Required?** |
| State | The US state for which covid-19 data is reported. | String object | Names of US states. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| StateFIPS | The FIPS (Numbers which uniquely identify geographic areas) for which covid-19 data is reported. | Integer | US State FIPS. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| County | The county of a US state for which covid-19 data | String object | Names of US Sates counties | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| CountyFIPS | The FIPS (unique 5-digit code representing specific county) for which covid-19 data is reported. | Integer | US State County FIPS. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| Date | Cases confirmed on that specific date. | Integer | Any positive number representing cases count. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |

Datatype check screenshot

A screenshot of a computer

Description automatically generated

**2.COVID Deaths Dataset (covid\_deaths\_usafacts.csv)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Definition** | **Data type** | **Possible values** | **Any missing values?** | **Required?** |
| State | The US state for which covid-19 data is reported. | String object | Names of US states. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| StateFIPS | The FIPS (Numbers which uniquely identify geographic areas) for which covid-19 data is reported. | Integer | US State FIPS. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| County | The county of a US state for which covid-19 data | String object | Names of US Sates counties | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| CountyFIPS | The FIPS (unique 5-digit code representing specific county) for which covid-19 data is reported. | Integer | US State County FIPS. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| Date | The total number of covid deaths recorded in county or state level on a specific date. | Integer | Positive integers representing  number of deaths. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |

A screenshot of a computer

Description automatically generatedDatatype check screenshot

**3.COVID Population Dataset (covid\_county\_population\_usafacts.csv)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Definition** | **Data type** | **Possible values** | **Any missing values?** | **Required?** |
| State | The US state for which covid-19 data is reported. | String object | Names of US states. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| County Name | The county of a US state for which covid-19 data | String object | Names of US Sates counties | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| CountyFIPS | The FIPS (unique 5-digit code representing specific county) for which covid-19 data is reported. | Integer | US State County FIPS. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| Population | Number of people residing in a county. | Integer | Integer | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |

Datatype check screenshot

A screenshot of a computer screen

Description automatically generated

**Conclusion**

The covid-19 dataset provided by usafacts.org is a valuable resource for understanding the spread and impact of the covid-19 pandemic in the United States. It includes a wide range of variable related to cases and deaths. The data dictionary provided gives an overview of the dataset’s variables and their data types, making it easier for the future data analysis.

**References**

<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>

<https://analystanswers.com/what-is-a-data-dictionary-a-simple-thorough-overview/>