**A recommended data citation**

In APA format

Anandhu, H. (2021). Latest Worldwide Vaccine Data [Data set]. Kaggle. <https://www.kaggle.com/datasets/anandhuh/latest-worldwide-vaccine-data?resource=download>

**A description of any considerations for long-term preservation. This might include**

**Whether the file formats are proprietary or might be in danger of going obsolete**

**Whether any specific software is needed to open any of the files**

The dataset is in a CSV format and since CSV is a widely used and well-established file format, it is not likely to go obsolete in the foreseeable future. Furthermore, the use of a plain text file format makes it relatively easy to read and work with the data in the file, regardless of the specific software used to open it. In terms of long-term preservation, it is generally recommended to store datasets in a format that is as open and widely supported as possible, in order to ensure that the data remains accessible over time. In this case, the use of CSV format is a good choice in this regard. Additionally, it is important to ensure that the dataset is properly documented, including information about the data collection process, data cleaning and processing steps, and any other relevant information. This documentation should be stored along with the dataset itself, in order to provide context and enable reproducibility. Overall, as long as the dataset is stored in an open and widely supported format, and is accompanied by adequate documentation, it should be well-suited for long-term preservation

**A statement about a copyright license that would be appropriate for this data set**

The dataset on Kaggle comes with ​​"Database Contents License”. The Kaggle database contents license applies to the datasets that are available in the Kaggle database and are governed by the following terms:

* The datasets are provided under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).
* This means that the datasets can be freely used, shared, and adapted, as long as appropriate credit is given to the original source of the data, and any derivative works are shared under the same license terms.
* The license also requires that users of the dataset do not imply endorsement or association with the original data source, or use the data in a way that could be construed as harmful or defamatory.

**A statement about any human subject considerations that apply to this data set. This might include:**

**Whether the data have personally identifiable data about people**

**Whether any steps were taken to anonymize or otherwise adjust the data for privacy or ethical reasons**

The dataset is related to worldwide vaccine data, and it is not explicitly stated whether the data contain personally identifiable information or human subject data. However, it is important to consider ethical and privacy concerns when working with any type of data that may include personal information or sensitive data about individuals. Depending on the specific data and context, this may require taking steps to anonymize or de-identify the data, or obtaining appropriate consent or ethical clearance from relevant authorities. It is important to note that different countries and jurisdictions may have different laws and regulations regarding the use of personal data and human subject research, and it is the responsibility of the researcher or user to ensure that they are complying with relevant regulations and ethical principles. In the case of the Kaggle dataset, it is not clear whether any specific steps were taken to anonymize or otherwise adjust the data for privacy or ethical reasons. Therefore, it is important for users of the dataset to carefully review the data and assess any potential ethical or privacy concerns, and take appropriate steps to address these concerns as needed.