

# Datasheet for web scraped data (GDP)

## Motivation for Dataset Creation

- Created to calculate the annual percentage growth rate of GDP at market prices based on the constant local currency by the World Bank.
- The data can also be used to assess the GDP performance of the different countries in different time frames.

## Dataset Composition

- The dataset features 3 columns of data i.e. GDP Growth %, Annual change in GDP, and year in question
- Relations between years and GDP data are explicit.
- The entire dataset published consists of GDP data of many more countries and the extracted data on the USA was directly observable as a table.

## Data Collection Process

- The World Bank collects most of its data through surveys and data collection efforts and majorly uses sources such as the UN to directly collect data.
- We sampled only data on the USA from the Data Bank of World Development through a third-party website [macrotrends.net](https://www.macrotrends.net).
- The sampling done in this dataset would be deterministic.
- The larger population of the dataset consists of GDP data on several countries as a ranking.

## Data Preprocessing

- For our use case, we used a web scraper to extract table data using Requests and BeautifulSoup.
- Relevant text elements were extracted to create a clean list of 'GDP yearly data'.

- Pairs of data points were created to extract GDP growth % and YoY change. The final data frame is then formatted with a year in DateTime and float values for percentages.

## Dataset Distribution

- The dataset is distributed via the Data Bank and Open data site of the World bank.
- The dataset is governed by Creative Commons Attribution 4.0 International License and additional terms on acknowledgment and use mentioned here: <https://www.worldbank.org/en/about/legal/terms-of-use-for-datasets>

## Dataset Maintenance

- The open-source data set was published by the World Bank which in turn is maintained and distributed by [macrotrends.net](https://www.macrotrends.net).
- Once the dataset becomes obsolete, it would likely be updated by [macrotrends.net](https://www.macrotrends.net). However, it's important to note that they store up to 100 years of economic data with suggested use cases.

## Legal & Ethical Considerations

- The table extracted is owned and protected by third-party copyright law. Since no personal data is stored, GDPR does not apply.
- Disputes arising from the usage of data may be resolved under guidelines provided by the World Bank in a specific time period.
- An ethical constraint must be to verify the accuracy of data about countries on a third-party website. This is addressed by our source through citations. It is not clear which particular source the World Bank collected this data from.