Group 12: Transformers

## Sources:

<https://data.worldbank.org/country/united-states?view=chart>

<https://data.worldbank.org/country/china?view=chart>

**Data source**: https://data.worldbank.org/

## Final Production Database:

We are going to merge these 2 csv’s based on indicator names and compare the results between the USA and China. This will be a Postgres relational database. The final database should be in first normal form.

## Description of findings:

By looking at the merged file based on indicator names, we can compare the emission vs. gdp growth to see whether decoupling or no decoupling emission. We can also use the data to view the change in the unemployment rate and output growth rate to test Okun’s Law and we can view inflation rate vs. unemployment rate to test the Phillips Curve to show our findings.

## Project Report

* The report describes the original data sources and how the data were formatted. (10 points)
* The report explains what data cleaning or transformation was performed and why it was needed. (20 points)
* The report describes the structure of the final database along with any improvements that could be made in the future. (20 points)

The original data source was found on data.worldbank.org. This website gives data of many countries and the data relating to its economy with some countries' data going as far back as 1960. From there we downloaded the csv’s for China and USA specifically. The csv’s are organized with countries, indicators, and years as the columns. The rows are organized by each indicator per row, giving over 1000 rows.

The data that we transformed was GDP, Inflation, Unemployment rate, and Emissions. We cleaned this data by finding the rows that we needed, putting them in a list, and then forming a dataframe with those rows. Then we dropped the columns we didn’t need because we only wanted information from 1990-2021. We went into PgAdmin to create the tables that we wanted to insert our dataframes into. Then we used jupyter to put the dataframe information into those tables.

We wanted to transform this data because this data will allow us to understand the relationships between the rows and compare the two countries.

The final structure for this database is…. It could be improved by…..

**Data Cleansing:**

Change the Format of the year column

Dropped the null values