# **Project 2 Group 12: Transformers**

# **Project Overview**

Team selected the economic data, and wanted to review and analyze the economic trends of certain countries over a time period of 25 years. We used the data source as Worldbank.org for this analysis because there’s access to databases for countries with large sets of data that goes as far back as 1960 and mostly it's the source of truth. We want to look at: population growth, human capital index, gdp, gdp growth, unemployment, inflation, carbon emissions, Individuals using the internet, and foreign direct investment. We are looking at the USA and China and will compare the difference in data between these 2 countries.

### **Contributors**

Yi Lu

Bose Ravula

Victor Pang

### **Project Objective**

Our objective for this project was to compare the emission vs. gdp growth to see whether decoupling or no decoupling emission. change in the unemployment rate and output growth rate to test Okun’s Law and inflation rate vs. unemployment rate to test the Phillips Curve to show our findings.

### **Project Report**

### **Extract**

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

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

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

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.

Graphical user interface, application, table

Description automatically generated with medium confidenceGraphical user interface, text, application

Description automatically generated

### **Transform**

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 used PgAdmin to create the tables that we wanted to insert our data frames into. Then we used jupyter to put the dataframe information into those tables. We used SQL to join those tables based on their indicator code for our final table

### **Text Description automatically generated with medium confidence**

### 

### 

### **Load**

Our final merged table displays 50 rows, 25 for China and 25 for the USA. The Columns consist of Countries, indicators, and years. The table was merged on the indicator code. For each indicator code, it displays China and USA in 2 rows, making it easier to compare the change in data over time. With it displayed in this manner, it makes it easier to manipulate the table to find the economic data and trends in China and USA.

Graphical user interface, application

Description automatically generated

### **Improvements for the Future**

Some improvements we could make on this would be the ability to filter which years we’d like to view. For some data, there are a lot of null values. We could also have merged other indicator codes.