Factors influencing Inflation and Recession

## 

## Team Members:

Group 1

* Debanjan Saha, MS in Data Analytics Engineering, Northeastern University, Boston
* Zeyuan Liu, MS in Data Analytics Engineering, Northeastern University, Boston
* Guangyu Han, MS in Data Analytics Engineering, Northeastern University, Boston

## Objective:

Using the different indicators released on the World Bank's website, we are attempting to identify the many aspects that have a significant influence in socio-economic development. We choose the top 100 indicators that play a crucial part in defining the movement of the global economy and the numerous elements that influence it.

The data is first obtained from the publicly accessible World Bank Data website. The data are then processed using wide-form transformation and combined into a single dataset including all metrics. Various supervised, unsupervised and hierarchical clustering approaches are then used to the data in order to infer the effect of the indicators in determining the socio-economic conditions of the nation. Also, we analyze the effect of these indicators across various income-levels.

## User Stories:

User Story 1:

[edited by Zeyuan]

As an international citizen or employee of an international corporation , they need to visualize the effect of various topics worldwide geographically so that they can make a wise region choice or adjust the international strategy.

Our web application should ensure:

1. They can filter based on countries so that they can explore the potential regions chosen on both country-level and continent-level.
2. They can filter based on metrics so that they can focus on the metrics they care about and ignore the metrics they don’t care about among 100+metrics from the World Bank.
3. They can scroll bars based on time so that they can know whether the region that they are interested in is a good place in any year since 1970.

User Story 2:

[edited by Zeyuan]

As a statistician, socio-economic researchers or college students, they need to visualize time-series plots of various topics so that they can know the description of statistical features of these data for papers or class projects.

Our web application should ensure:

1. They can choose specific statistical graph so that they can know about the statistical features deeply
2. They can filter based both countries and time so that they can focus their research on any subset of the whole world

User Story 3:

[edited by Zeyuan]

As an international investor, whoever cares about the future instead of history and present. So they want to predict the specific topics using past data so that they can adjust their investing activities.

Our web application should ensure:

1. They can get clear prediction results without any operation. All the work is done by the machine learning models.
2. They can do some advanced settings if they have a related background of technology.
3. They can predict based on countries and metrics so that they can focus on their investing activities on any subset of the whole world.

[edited by Guangyu]

User Story 1:

Geographically visualize our topics, such as Economics Growth, Climate, and Education, in a worldwide map to present the data in a straightforward way

User Story 2:

Apply Time-Series Analysis on our topics as well as indicators, and forecast the future trend based on the previous patterns

User Story 3:

Build machine learning models (basically classification and clustering) to predict the future trends on each of our individual topic

[added by Debanjan]

1. A comprehensive overview of the different environmental elements that impact economic development and inflation.
2. Examine a few of the elements that impact the inflation rate in an economy.
3. Population's influence on the rate of inflation change
4. Influence of government expenditures, costs, debt, and other market factors on inflation rate increases

[edited by Guangyu]

1. Compare various economic indicators, such as unemployment rate and GDP growth, and their trends now with those in the previous times, especially last recession ones.
2. High tech export

## Previous Works: (reference to articles – dont post link- write what you understood from the articles)

[edited by Zeyuan]

1. World band deflater

<https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG?end=2021&start=2021&view=map>

1. If we are interested in recession judgment

http://stickyslides.blogspot.com/search?q=financial+crisis

## Contributions:

## Technical Design

### Data Extraction

First we start by collecting historical time series data released by the World Bank based on various indicators for certain environmental factors (or domains) such as economy and growth, financial sector, private and public sector, population metrics, and many more.

A data dictionary containing all the various indicators used for this project can be found here:

[World Bank Selected Indicators](https://docs.google.com/spreadsheets/d/1oFzUxV16f8TkZ_04179REckPsG69E6OoED1TLfZu6l0/edit?usp=sharing)

## Data Transformation

The data collected is in wide-form with countries as index and the various years as columns. We pivot the data such that we have the dates as our index and the various countries as columns.

## Data Loading

Once the transformations are completed, all the various metrics are combined together into a single dataset which contains data for all the indicators

## Data Wrangling

Various data checks are put into place, and various data wrangling techniques are applied

Modeling

Unsupervised Hierarchical clustering methods are used to cluster the data into various categories based on the risks associated with it.

## UI (User Interface) Front End

A dashboard has been built to reflect the trends and factors which lead to inflation.

## Data Visualization with Interactivity

This dashboard contains plots for various factors and can be filtered on various levels such as countries, continents, etc.

[Edited by Zeyuan] Another potential topics

Sports player body research (from UCI machine learning)

<https://archive.ics.uci.edu/ml/datasets/daily+and+sports+activities#>

Factors influenced higher education student’s (from UCI machine learning)

<https://archive.ics.uci.edu/ml//datasets/Higher+Education+Students+Performance+Evaluation+Dataset>

I think this dataset is more understandable for both of us. It is easier to visualize this information.

For example,

User story 1:

If I am a student who wants to play sports before midterm, Then I want to check whether this behavior will influence my grades.

User story 2:

If I am a parents who wants to know whether brothers and sisters influence their kids’ academic performace, Then I can check this information on our dashboard

User story 3:

If I am a professor who wants to lease some reading assignments of their students, then I can check this information on our dashboard

We want to find the countries with less than 45 indicators which have less than 10 percent missing values