Analysis Approach

At first, we tried to pivot the problem by identifying current issues in both Latin America and South East Asia.

We gathered the indicators on the country level from the OECD website. These indicators are directed towards the social and economic development of the country. The data available on this website is on year level, going back to 1961 till 2022.

We picked 4 major issues faced by these regions and tried to map how the indicators behave with the involvement of US Aid. The purpose here is to which aspect of US Aid will help us in improving the indicators.

Each team member picked 1 topic of their interest and their analysis is towards the specific topic.

**Kushal**

**Problem statement**: To reduce unemployment rate within Latin America using US AID

**Approach**:

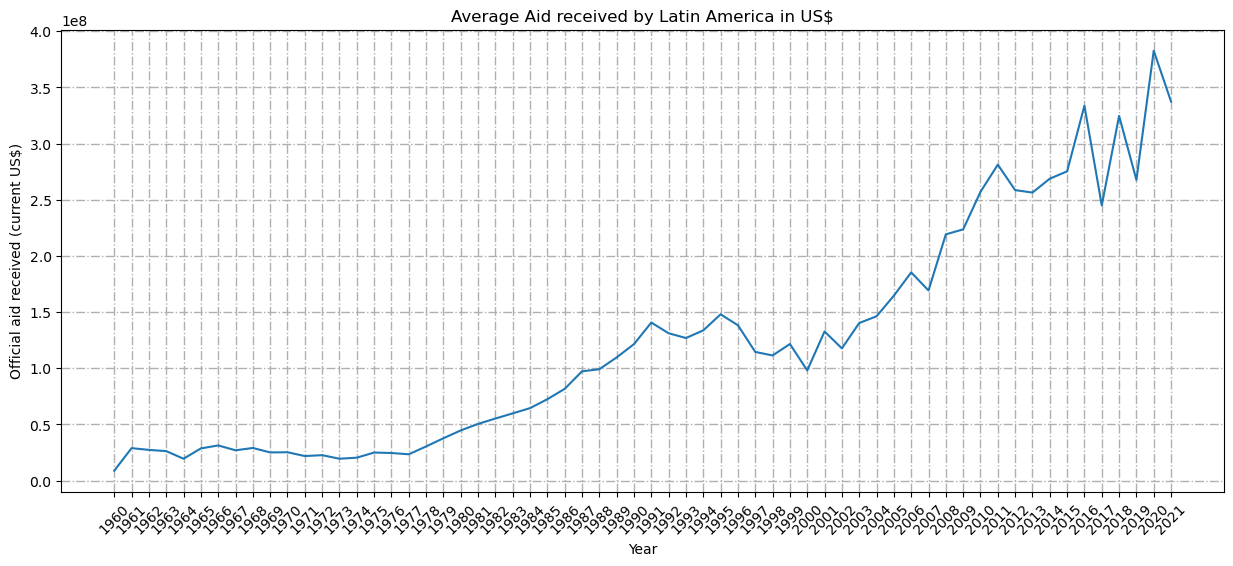
EDA on different indicators with US Aid will help us in identifying if there’s a significant impact on US Aid on such indicators. Our assumption here is the by providing targeted US Aid, we see an improvement in the unemployment rate.

Here we will use data from OECD and foreignassistance.gov to see the trends in US Aid and the indicators.

**Assumption:**

Foreign aid -> f(GDP, crime rate, corruption, employment rate, political stability, poverty rate)

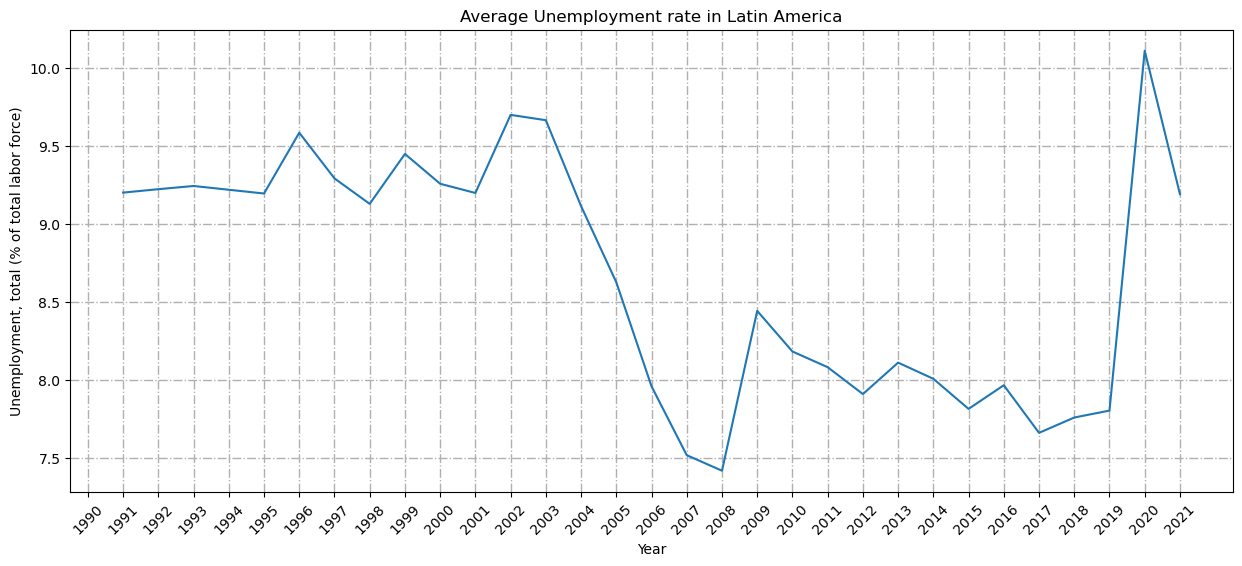
**EDA**:



Above chart shows the average US Aid provided to Latin America from 1960 - 2021.

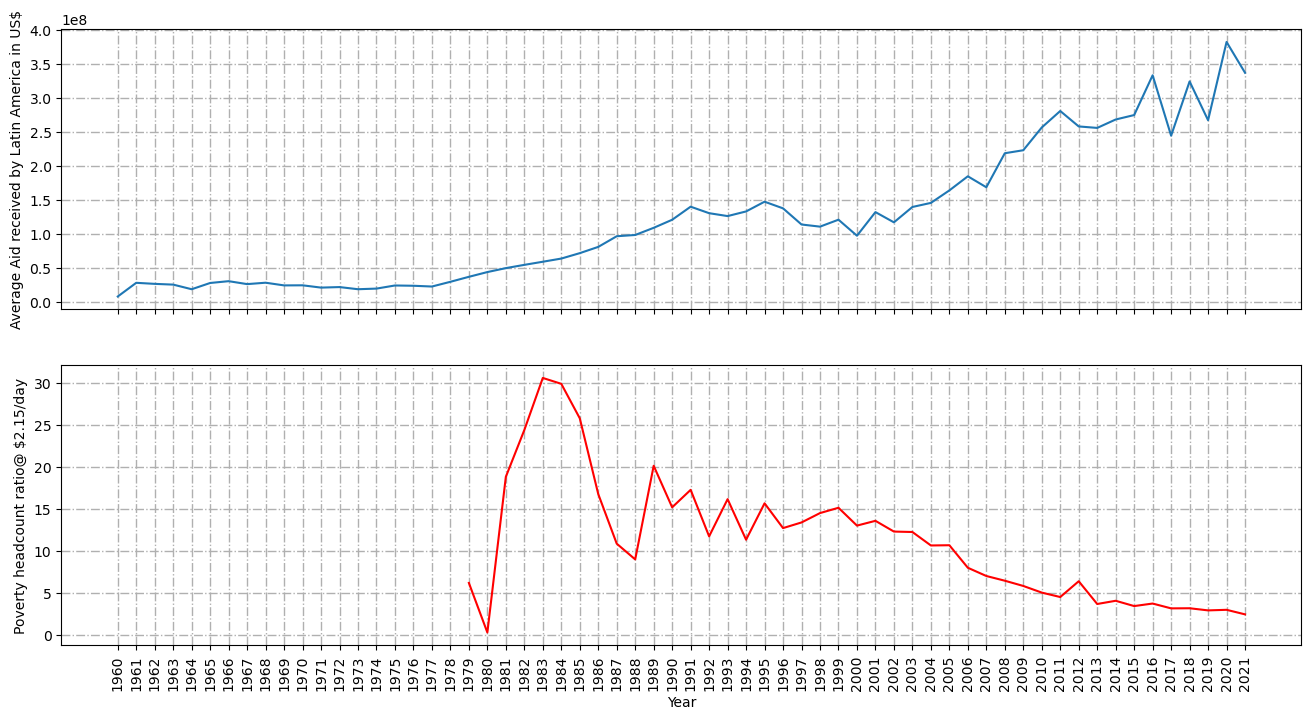
As seen in the chart above, there has been a significant raise in US Aid from 2001.

Lets look at the indicator: Unemployment rate



This chart in Average Unemployment rate in Latin America from 1990 - 2021.

If we look at both the graphs together below:



If we look at this trend, we see increase in US Aid from 2001, there has been a significant drop in the unemployment rate from 2002.

**Hypothesis: Could this possibly mean that US Aid towards Latin America had a significant impact on reducing the Poverty rate?**

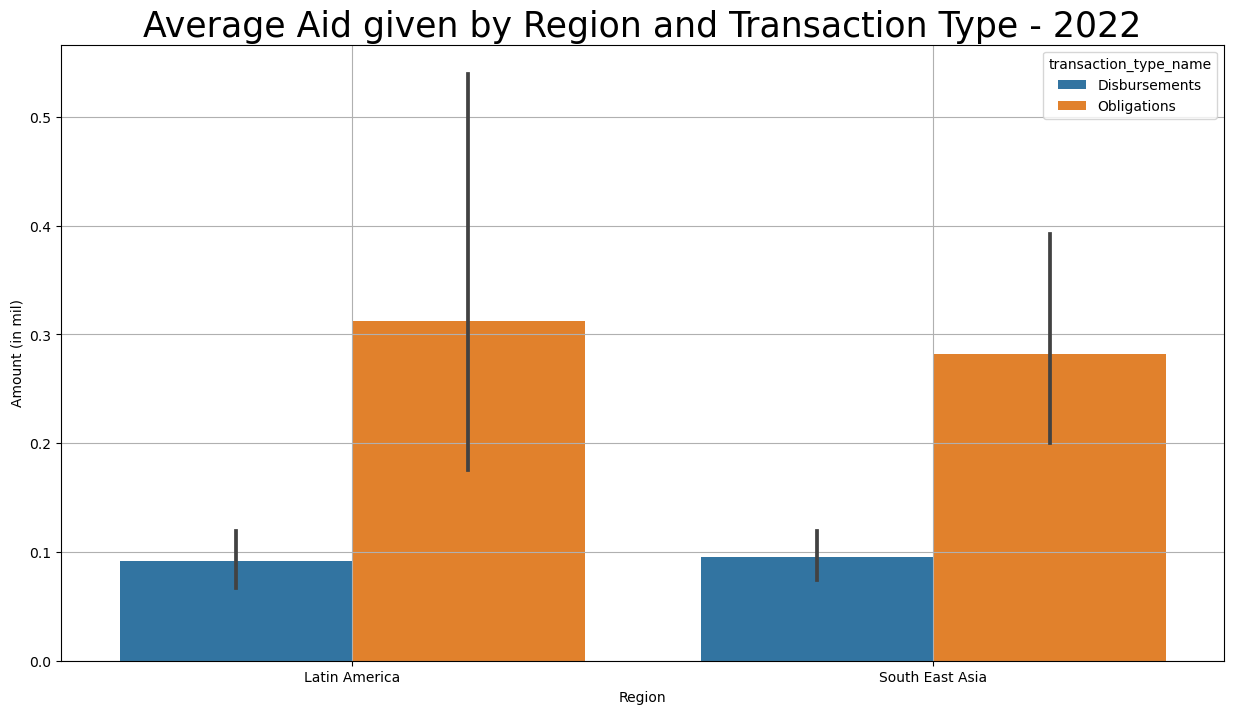
Let’s understand more about US Aid and its distribution across different aspects.

There are two major types of transactions within US Aid

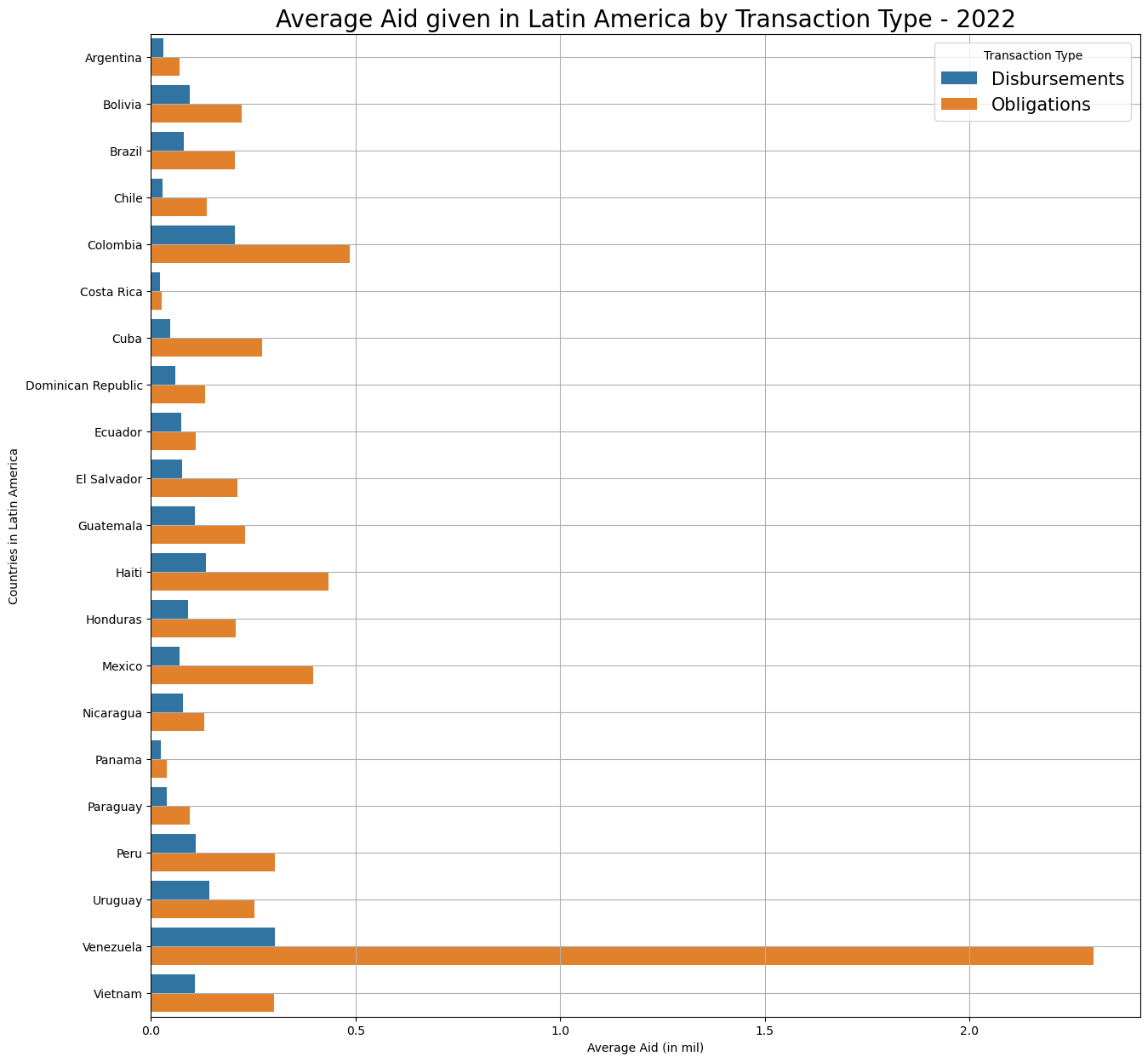
**Disbursements** - transfers of goods, services, or cash for which the recipient incurs no legal debt.

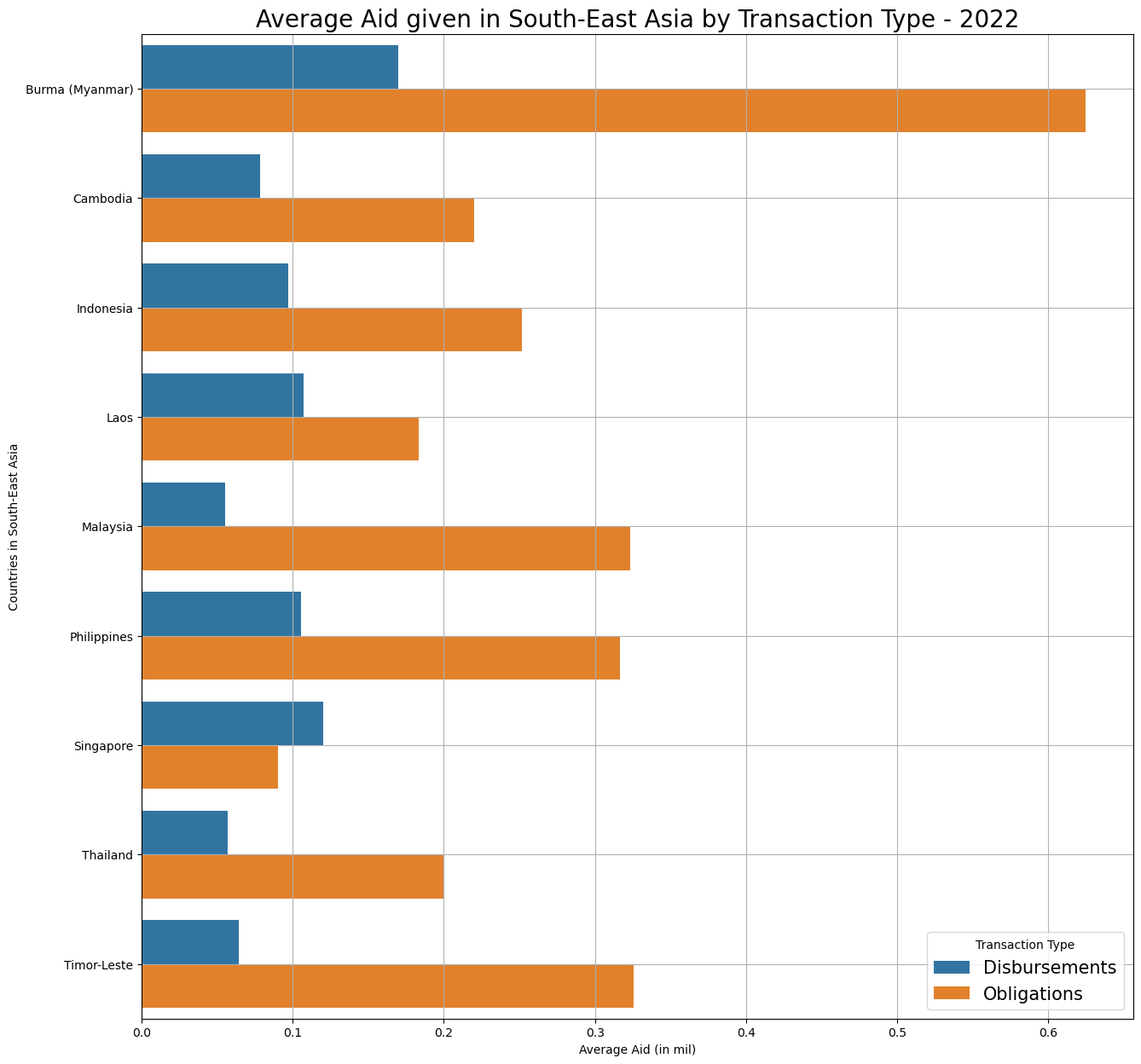
**Obligation** - a binding agreement that will result in outlays, immediately or in the future. Budget resources must be available before obligations can be legally incurred.

Distribution of these two types in prioritized 2 region is as below:

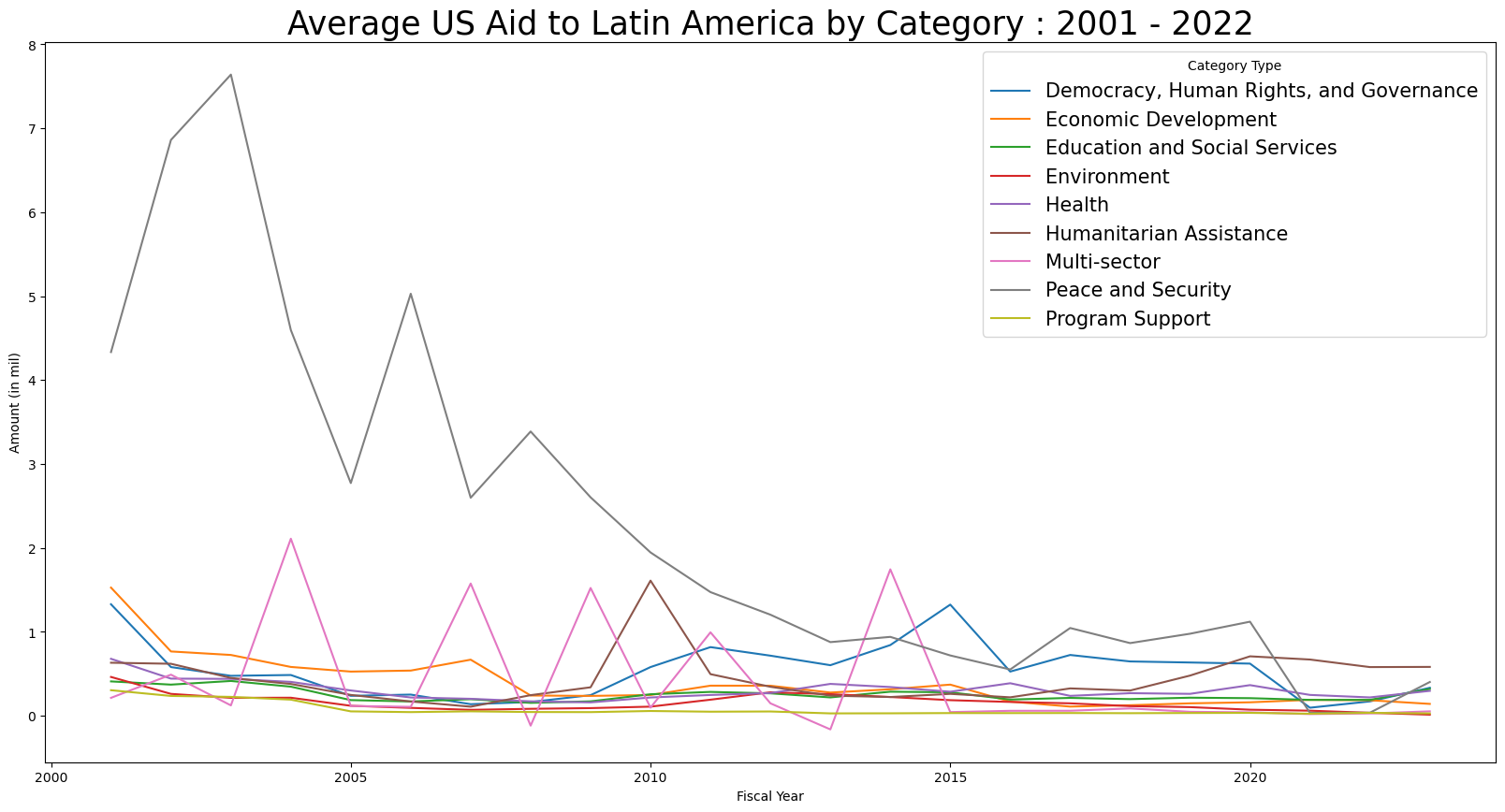


Within each region, the US aid is as below:





Further we will look into how US Aid is divided into different categories defined by US Embessy.

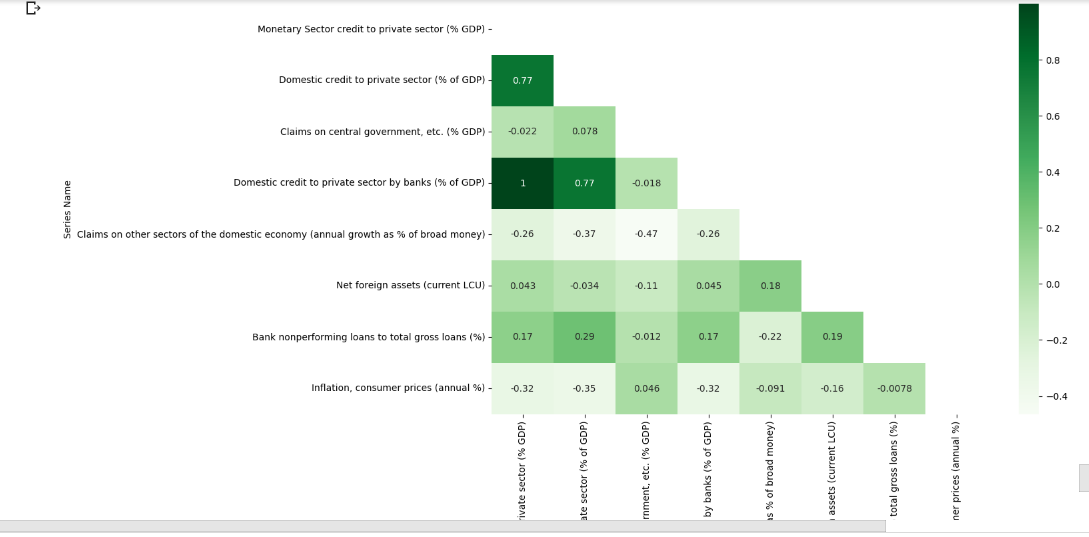


**Next Steps:**

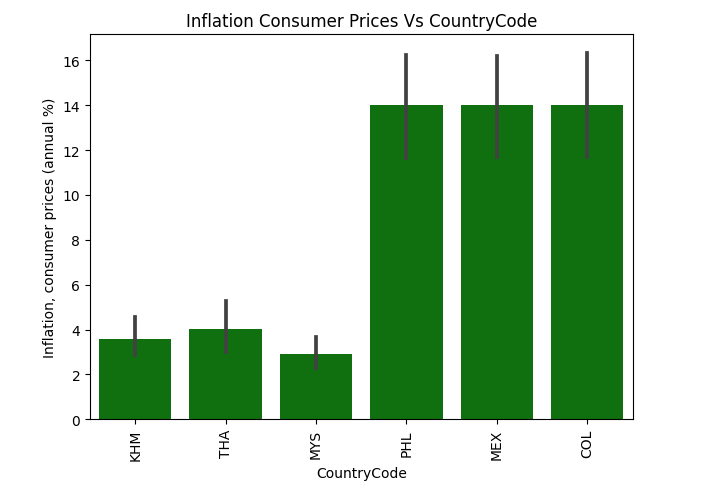
* Identify categories of US Aid and compare impact on unemployment rate
* Prove whether ‘US Aid impacts Unemployment rate’ is correlation or causation.
  + If yes, measure impact in terms of ‘how much US mil Aid will reduce the unemployment rate by 1%’
* Product design

Ron

I did some analysis on how Inflation is affecting South East Asia and Latin America by identifying relationships between Inflation and other indicators that affect it.



We notice that there is a strong relationship between Domestic credit to private sector and Monetary sector credit to private sector.



We notice that Phillipines, Mexico and Colombia are experiencing the highest inflation.

Because of time, I plan to spend sometime doing more EDA in the coming days.

**Jimin**

Visualize the phenomenon of decreasing fertility rate among SEA countries

EDA was conducted by extracting data from the keywords "Population Growth Rate" from The World Bank data.

* From 1960 to 1999, the null value was more than half, so we removed it all. Only data from 2000 to 2022 were used for analysis.
* only the country of SEA , Latin America was considered
* found out all of SEA countries suffers decreasing fertility rate
* Compared to SEA countries, Latin America suffers less from decreasing population growth

|  |
| --- |
| **Population Growth rate of SEA countries (y: population growth rate)** |

|  |  |  |
| --- | --- | --- |
| **Myanmar**  y=-0.012171x+(25.277301) | **Cambodia**  y=-0.035844x+(73.575046) | **Timor-Leste**  y=-0.021336x+(44.810789) |
|  |  |  |
| **Indonesia**  y=-0.027414x+(56.300127) | **Lao PDR**  y=-0.005464x+(12.483275) | **Malaysia**  y=-**0.073933x**+(150.488973  suffering most rapid decline |
|  |  |  |
| **Philippines**  y=-0.025878x+(53.850689) | **Thailand**  y=-0.035659x+(72.313878) | **Vietnam**  y=-0.005749x+(12.564092) |

1. we can find that Malaysia suffers the most rapid decline in population growth rate by looking at the value of the trend line.
2. while suffering most lowest population growth rate in Thailand, it suffers 2nd most rapidest decreasing population growth rate

|  |
| --- |
| **Population Growth rate of Latin American countries** |

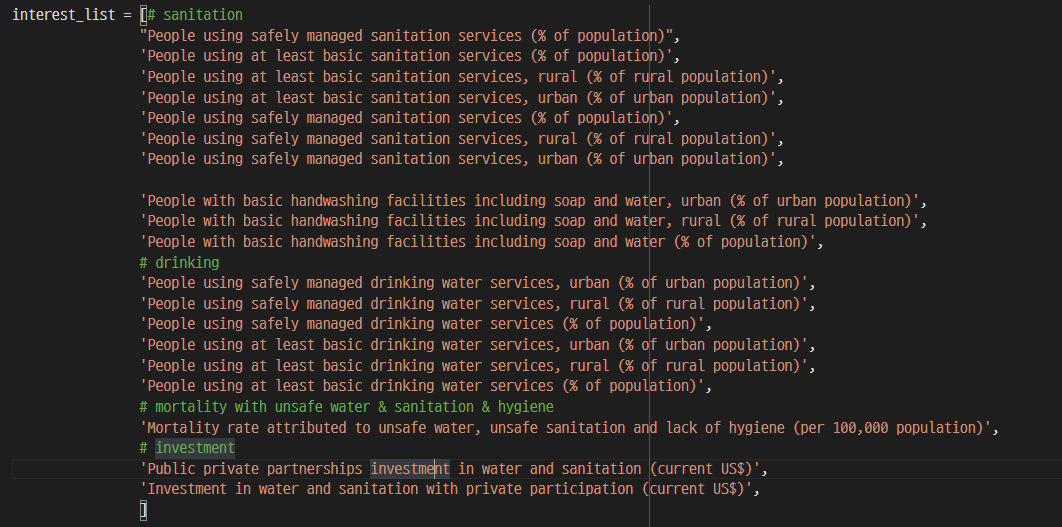
**Seunggil**

Hypothesis that I selected is “Water and Sanitation”/

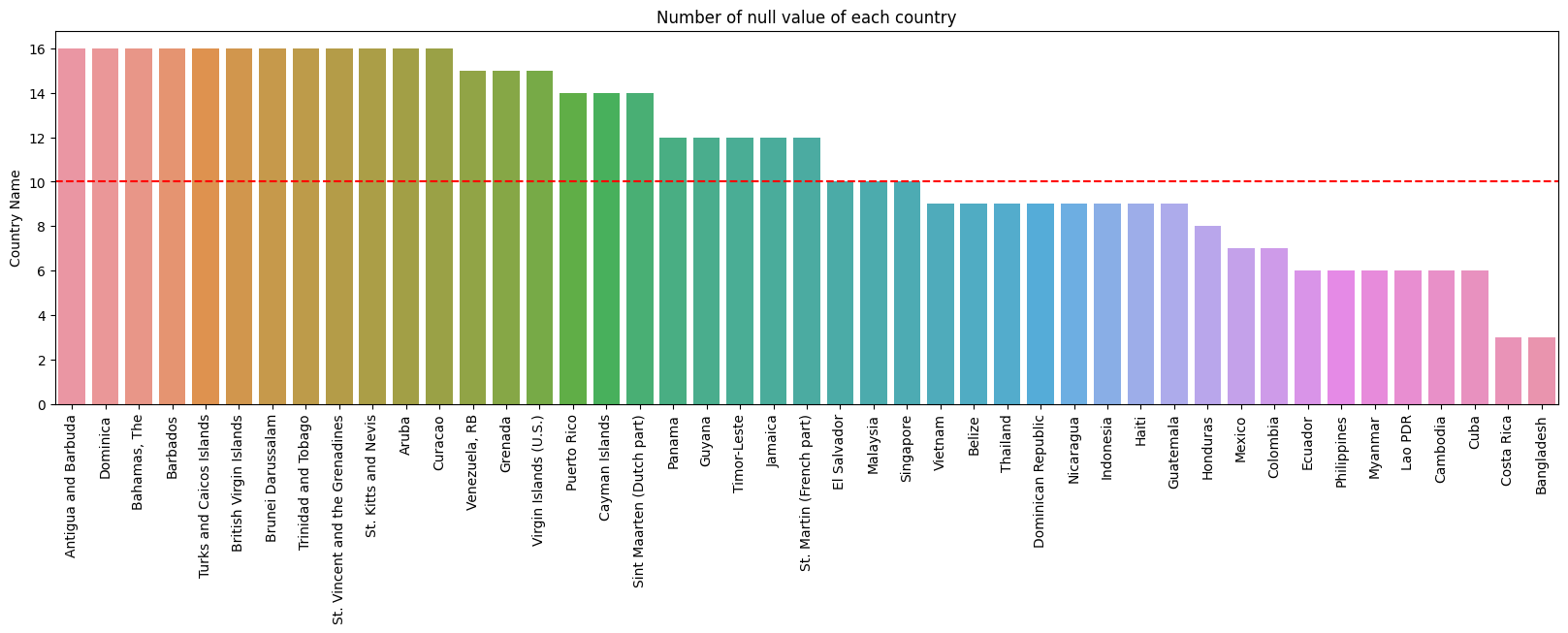
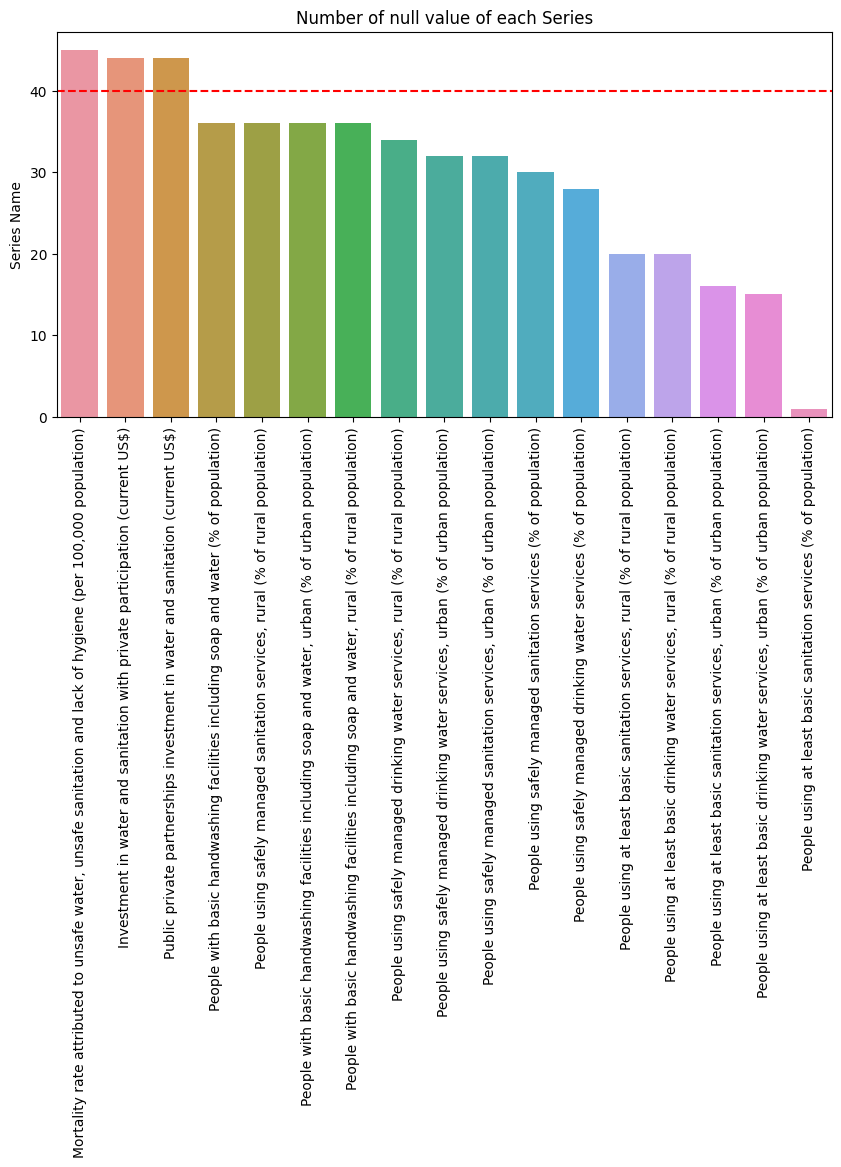
EDA was conducted by extracting data from the keywords "Water", "Hygiene" and "Sanitation" in The World Bank Open Data.

* From 1960 to 1999, the null value was more than half, so we removed it all. Only data from 2000 to 2022 were used for analysis.
* When importing raw data, only data belonging to Latin America, South Asia, and East Asia Category were imported. Before starting the EDA, there were a total of 89 countries.

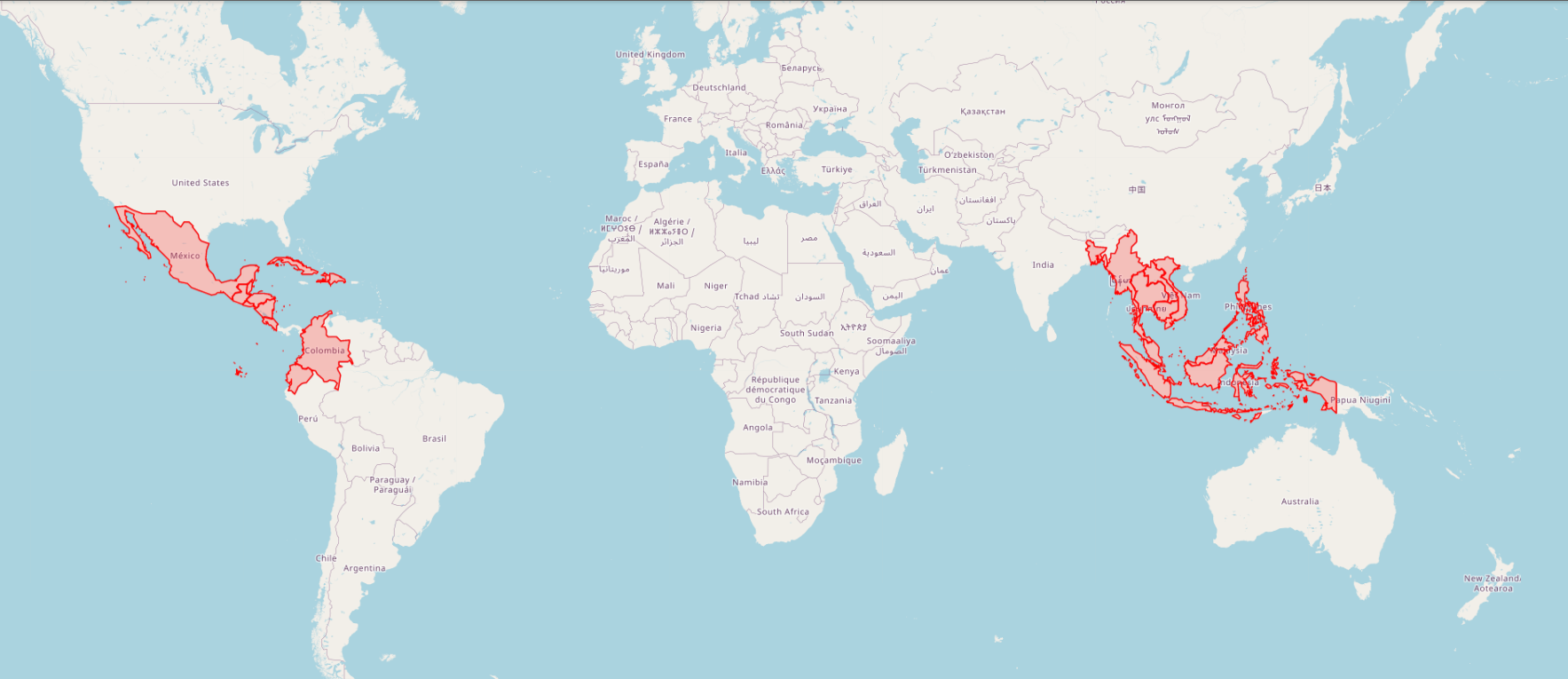
Only the columns related to Foreign Aid were intended to be used for analysis, and the corresponding Series Name is as shown below.



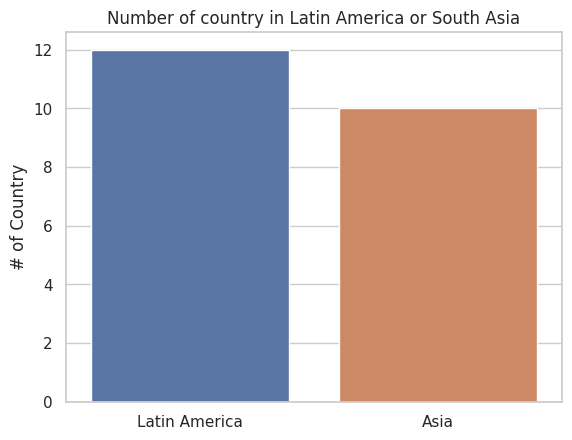
Since data with too many null values adversely affect data analysis and modeling, a specific criterion was created and feature selection was conducted.

* First, it is the number of null values of data for each country. If there were more than 10 (red lines), all data from the country were dropped.  
  
* After that, we focused on Series data from unremoved data. If each series has more than 40 null values (red lines), the entire series has been dropped.
* As a result of this process, a total of 22 countries were selected and the remaining 60 countries were removed.

Countries selected without removal are shown in the figure below. Most of them include Latin America and South-East Asia.



To find out the distribution of data by country, we visualized the number of countries in Latin America and South-East Asia.



It can be seen that the two data are balanced, with 12 Latin America and 10 South Asia countries.

Finally, we visualized all the trends from 2000 to 2022 of all the data. Each column represents a Series Name, and each row represents a country that fits the Country Code. The full data is shown in the figure below.

