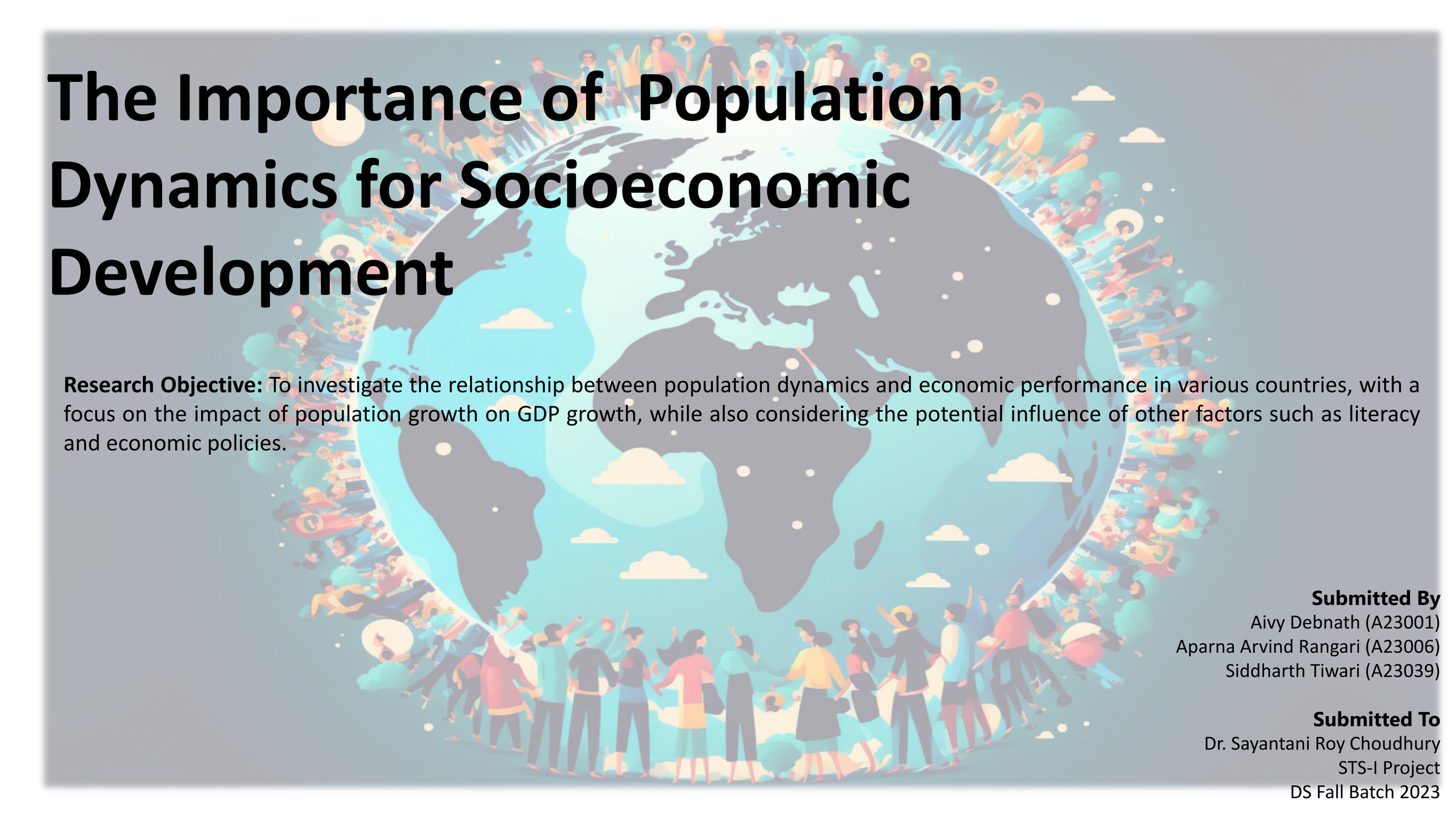
The Importance of Population Dynamics for Socioeconomic Development



Research Objective: To investigate the relationship between population dynamics and economic performance in various countries, with a focus on the impact of population growth on GDP growth, while also considering the potential influence of other factors such as literacy and economic policies.

Submitted By

Aivy Debnath (A23001)

Aparna Arvind Rangari (A23006)

Siddharth Tiwari (A23039)

Submitted To

Dr. Sayantani Roy Choudhury

STS-I Project

DS Fall Batch 2023

Source

Data has been collected from world bank-(data bank ,world development indicators)

Dataset Link: ***https://databank.worldbank.org/source/world-development-indicators***

Timeframe: **2020**
Variables include Population, GDP(Gross Domestic Product) ,GNI(Gross National Income), Urban population, Rural population and Literacy Rate.

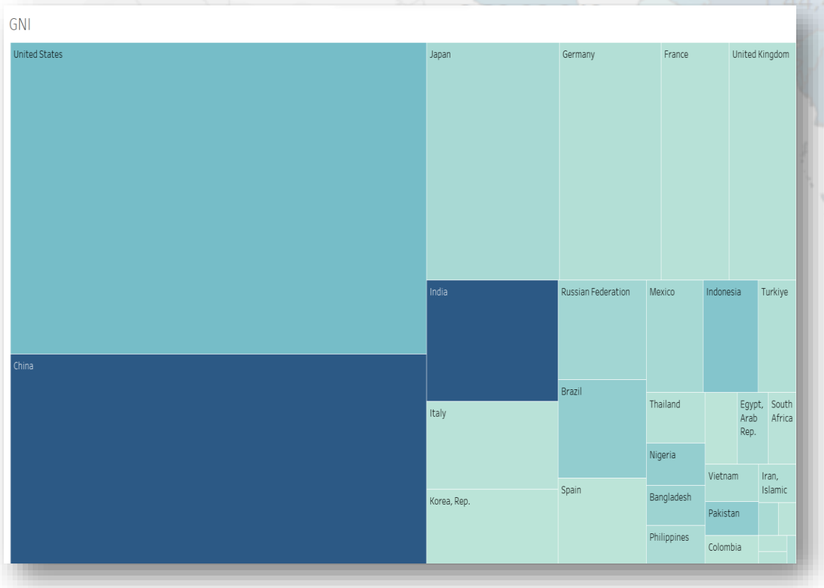
From the huge data we have collected 32 diverse and populous countries for analysis on the relationship between population dynamics and economic performance, with a focus on the impact of population growth on GDP growth, while also considering the potential influence of other factors such as literacy and economic policies. Here is the links of our excel and collab files to get further insights of our project.

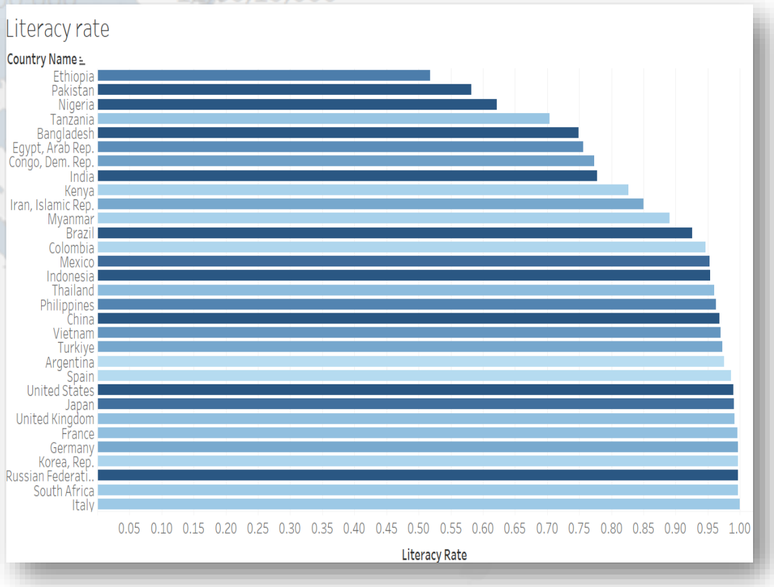
Collab: ***stats project.ipynb - Colaboratory (google.com)***

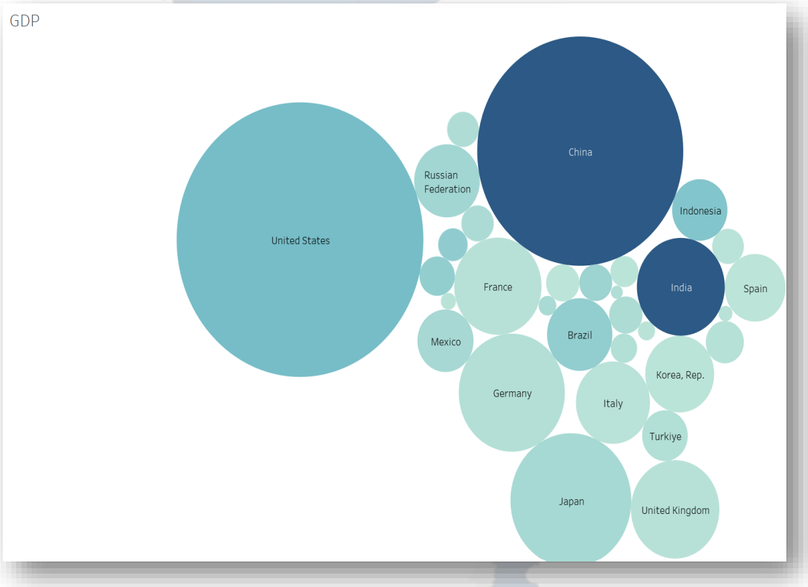
Excel:
https://docs.google.com/spreadsheets/d/1yWZr927URNNGKeVTNCDNmdGf9C5PRLX/edit?usp=drive_link&ouid=115499780536300477217&rtpof=true&sd=true

Methodology

- **Univariate Analysis:** Descriptive/Summary statistics and data exploration
- **EDA:** Visualization of relationships
- **Regression Analysis:** Modelling economic performance as a function of population dynamics
- **Bivariate Analysis**: Examining correlations between variables.







Analysis and Result

Bivariate Analysis: Correlation

```
df_filtered.corr()
```

```
<ipython-input-26-6754dbd00055>:4: FutureWarning: The default value of numeric_only in DataFrame.corr is d
  df_filtered.corr()
```

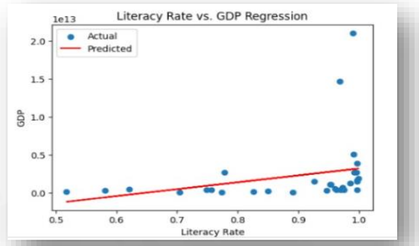
	Literacy Rate	GNI	GDP	Rural Population	Urban Population	Total Population
Literacy Rate	1.000000	0.282068	0.282029	-0.211437	0.074424	-0.070490
GNI	0.282068	1.000000	0.999875	0.262547	0.647319	0.480166
GDP	0.282029	0.999875	1.000000	0.270408	0.657119	0.489605
Rural Population	-0.211437	0.262547	0.270408	1.000000	0.789669	0.944291
Urban Population	0.074424	0.647319	0.657119	0.789669	1.000000	0.947483
Total Population	-0.070490	0.480166	0.489605	0.944291	0.947483	1.000000

Univariate Analysis : Mean, Median, Mode, Skewness, range, Standard deviation

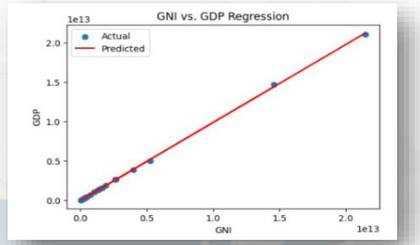
GNI		*GDP*		*Rural Population*		*Urban Population*		*Total Population*	
Mean	2.1965E+12	Mean	2.1837E+12	Mean	87006309.3	Mean	109485865	Mean	196463418
Standard Error	8.064E+11	Standard Error	7.966E+11	Standard Error	32432778.7	Standard Error	30215177.7	Standard Error	59666166.9
Median	4.8948E+11	Median	5.0046E+11	Median	36372705	Median	54715704	Median	89561404
Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard Deviation	4.4898E+12	Standard Deviation	4.4353E+12	Standard Deviation	180578069	Standard Deviation	168230990	Standard Deviation	332207158
Sample Variance	2.0159E+25	Sample Variance	1.9672E+25	Sample Variance	3.2608E+16	Sample Variance	2.8302E+16	Sample Variance	1.1036E+17
Kurtosis	12.8922142	Kurtosis	12.6412138	Kurtosis	15.8236781	Kurtosis	14.5055341	Kurtosis	11.363921
Skewness	3.52987242	Skewness	3.5076783	Skewness	3.90028688	Skewness	3.63994829	Skewness	3.46227932
Range	2.1425E+13	Range	2.1012E+13	Range	905105186	Range	852257089	Range	1399204223
Minimum	4.7449E+10	Minimum	4.8717E+10	Minimum	3579773	Minimum	14553419	Minimum	45195777
Maximum	2.1472E+13	Maximum	2.106E+13	Maximum	908684959	Maximum	866810508	Maximum	1444400000
Sum	6.8092E+13	Sum	6.7694E+13	Sum	2697195587	Sum	3394061825	Sum	6090365963
Count	31	Count	31	Count	31	Count	31	Count	31

Regression Analysis:

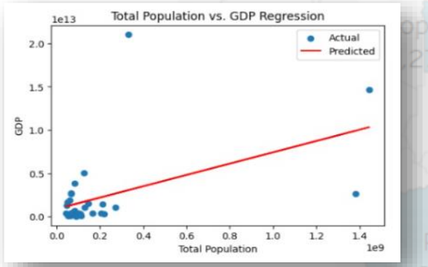
Literacy Rate vs GDP



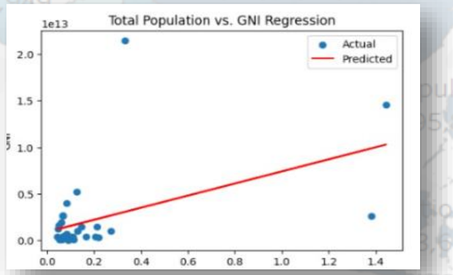
GNI vs GDP



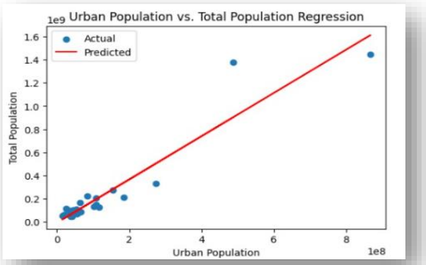
Total population vs GDP



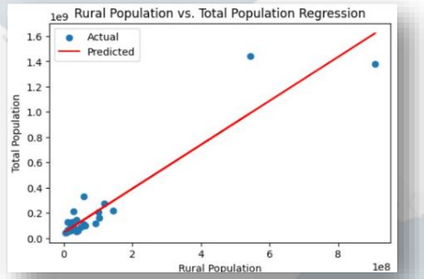
Total population vs GNI



Urban Population vs Total Population



Rural Population vs Total Population



Population vs GNI

Regression Statistics								
Multiple R	0.480166021							
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.21561E+11	8.39441E+11	1.097827	0.281312	-8E+11	2.63841E+12	-7.95289E+11	2.63841E+12
Total Population	6489.505705	2201.449577	2.947833	0.006261	1987.036	10991.97564	1987.035774	10991.97564

Population vs GDP

Regression Statistics								
Multiple R	0.48960512							
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	8.99442E+11	8.24298E+11	1.091161	0.284185	-7.9E+11	2.58532E+12	-7.86437E+11	2.58532E+12
Total Population	6536.709663	2161.737613	3.023822	0.005181	2115.46	10957.95951	2115.459818	10957.95951

Literacy vs GDP

Multiple R	0.282028968							
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-5.9029E+12	5.16704E+12	-1.1424	0.262631	-1.6E+13	4.66494E+12	-1.64706E+13	4.66494E+12
Literacy Rate	9.0894E+12	5.74176E+12	1.583034	0.12426	-2.7E+12	2.08326E+13	-2.65381E+12	2.08326E+13

Population vs Literacy

Multiple R	0.070489571							
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.895401429	0.02925987	30.60169	1.26E-23	0.835558	0.955244582	0.835558276	0.955244582
Total Population	-2.9201E-11	7.67346E-11	-0.38054	0.706315	-1.9E-10	1.27739E-10	-1.86141E-10	1.27739E-10

Interpretation

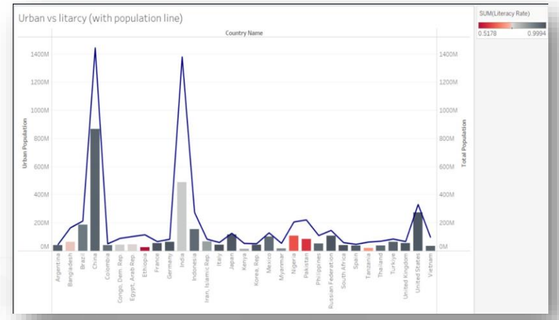
- Both GDP, GNI and Population(Urban, Rural, and total) has a positive skew:
  - This means here are a few countries with very **high** GNI and GDP, while the majority of countries have much lower. This is a reflection of the fact that global wealth is highly unequally distributed
  - A small number of countries account for a **large** share of global **wealth** and majority of countries are not benefiting fully from **globalization**.

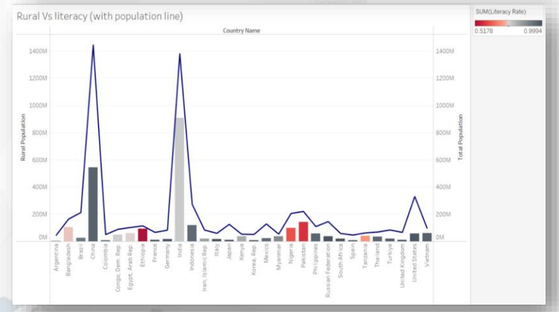
- Correlation Matrix:
  - It shows that the **urban population** is negatively correlated with **rural population** (-0.789). This means that countries with higher urban populations tend to have lower rural populations and population densities. Which supports our migration theory
  - It does not show a strong correlation between **literacy rate** and **total population** (-0.070). This means that there is no clear relationship between literacy rate and total population.

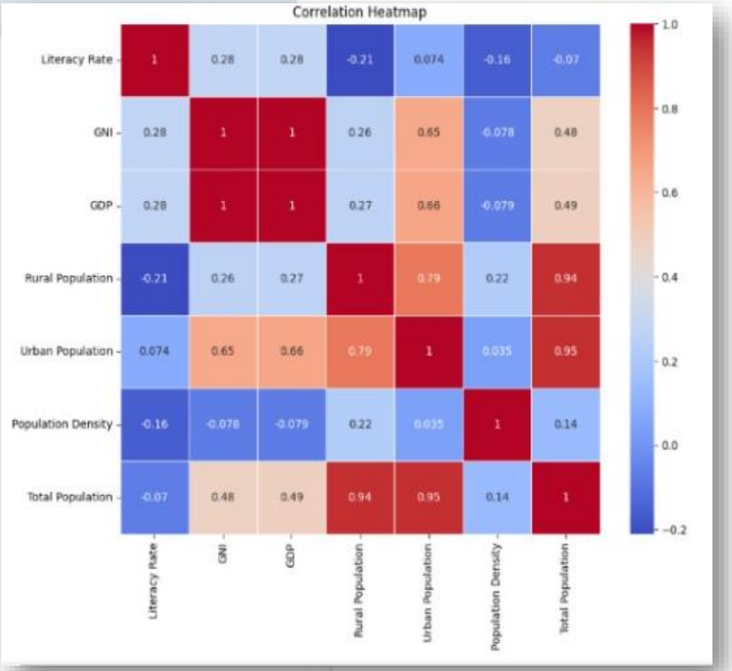
- The **regression** analysis with **GNI**:
  - The multiple R for 'GNI' and 'GDP' is 0.9999, indicating an extremely strong positive correlation between Gross National Income (GNI) and GDP. This suggests a near-perfect linear relationship between GNI and GDP, which is expected, as GNI is a major component of GDP
- The **regression** analysis with **Literacy Rate**:
  - The multiple R for **Literacy Rate** and **GDP** is 0.2820, indicating a weak positive correlation between literacy rate and GDP. This suggests that countries with higher literacy rates tend to have slightly higher GDPs

- Total Population: This **coefficient** (6489.51) represents the estimated change in the dependent variable for a one-unit increase in the total population, holding all other variables constant. The associated p-value (0.006) is less than 0.05, indicating that the total population is statistically significant in predicting the GNI

- Insights from the correlation matrix the **positive** correlation between **GNI** and **Urban Population** (0.6473), indicating that as GNI increases, urbanization tends to rise. The strong positive correlation between **Urban Population** and **Rural Population** (0.7897), suggesting that countries with higher urban populations also tend to have larger rural populations. The **strong positive** correlation between **Rural** Population and **Total** Population (0.9443), indicating that as the rural population increases, so does the total population.







Conclusions

- Population dynamics and socioeconomic development are closely **intertwined**. Our analysis shows that a country's economic progress is influenced by factors such as population size and literacy rate
- There is a positive, albeit weak, correlation between literacy rate and GDP, suggesting that literacy can have a modest impact on economic performance
- The relationship between **GNI** and **GDP** is exceptionally **strong**, underscoring the fundamental role of GNI in determining a country's economic output
- The size of the total population shows a moderate positive correlation with GDP, indicating population size does play a role in economic performance
- **Urbanization** is positively associated with **GNI**, reflecting the trend of higher income levels in urban areas
- The strong **correlation** between rural and urban populations implies that changes in one often correspond with changes in the other.


Population
14,62,38,042
Population
6,70,62,000
Population
8,39,92,949
Population
1,44,44,00,000
Population
12,60,10,000
Population
33,10,02,651
Population
10,27,80,145
Population
1,38,00,04,385
Population
10,95,81,078
Population
12,89,32,753
Population
20,61,39,589
Population
11,49,63,588
Population
6,30,32,400
Population
27,35,28,615
Population
5,08,82,884
Population
21,25,59,417
Population
5,93,10,635
Population
4,51,95,777

Thank You