IMPACT OF GENDER EQUALITY ON ECONOMIC GROWTH

Presented by Rupashi Maurya

INTRODUCTION

Gender inequality is a critical issue that affects various dimensions of society, including economic growth and development. In many regions, including India, disparities in education, labor force participation, political representation, and health outcomes persist, limiting the potential for comprehensive economic advancement. This project aims to analyze the impact of gender inequality on economic growth across various countries by leveraging data analytics and machine learning techniques. By examining key indicators such as female labor force participation, secondary education attainment, maternal mortality rate, and political representation, we seek to understand how gender disparities influence economic performance, measured by GNI per capita.

OBJECTIVE

The primary objective of this project is to analyze socio-economic data to understand the impact of gender inequality on economic growth and to propose data-driven strategies to promote gender equality and enhance economic development. The specific objectives are:

- To collect and analyze socio-economic data from reliable sources related to gender inequality and economic growth.
- To identify key areas where gender disparities are most pronounced and how they affect economic performance.
- To understand the temporal and spatial trends of gender inequality in relation to economic indicators.
- To develop predictive models that illustrate the potential economic outcomes of reducing gender inequality.
- To propose actionable strategies and policy recommendations to address gender disparities and promote inclusive economic growth.

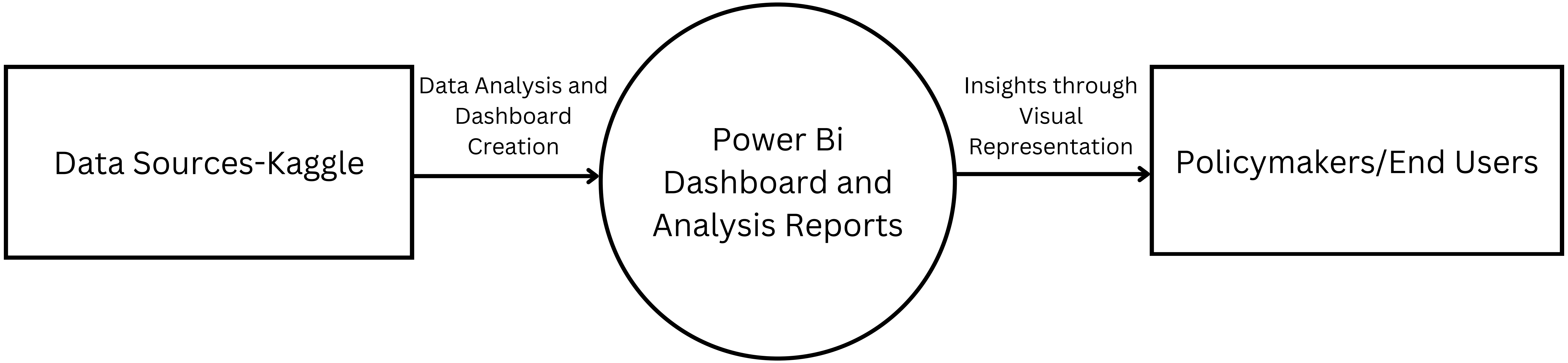
PROBLEM SIGNIFICANCE

Addressing gender inequality is crucial for unlocking the full potential of economic growth. When women have equal opportunities in education, employment, and political participation, economies benefit from increased productivity, innovation, and inclusive development. Reducing gender disparities can lead to higher household incomes, improved social well-being, and more resilient economies. Understanding and mitigating the economic impact of gender inequality is essential for achieving sustainable and equitable growth globally.

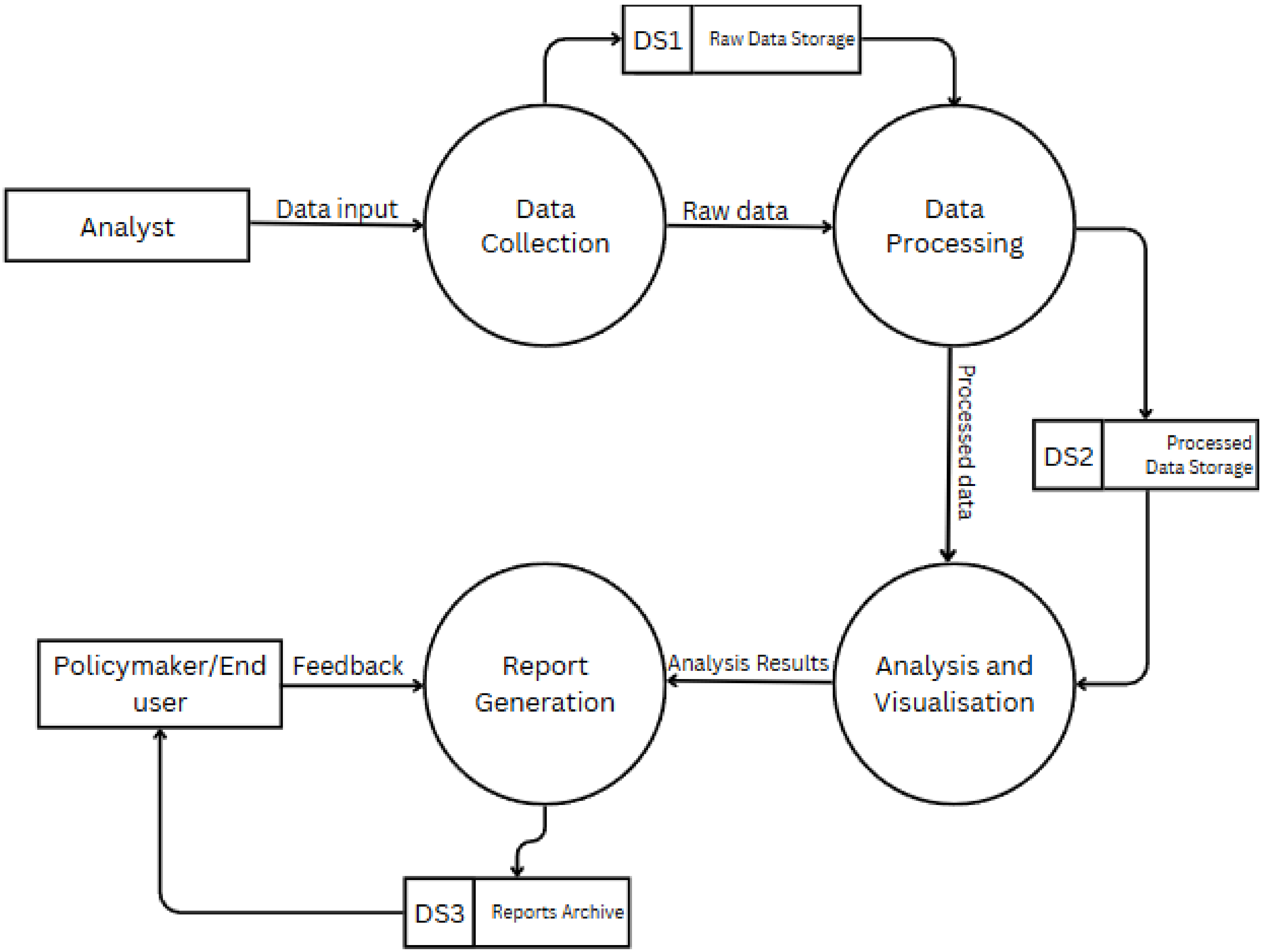
This project mainly addresses following Sustainable Development Goals (SDGs):

- SDG 5: Gender Equality - Promotes gender equality and empowers all women and girls.
- SDG 8: Decent Work and Economic Growth - Encourages inclusive and sustainable economic growth, full and productive employment, and decent work for all.

DFD LEVEL 0


Data Sources-Kaggle
Data Analysis and Dashboard Creation
Power Bi Dashboard and Analysis Reports
Insights through Visual Representation
Policymakers/End Users

DFD LEVEL 1


DS1
Raw Data Storage
Analyst
Data input
Data Collection
Raw data
Data Processing
Processed data
DS2
Processed Data Storage
Policymaker/End user
Feedback
Report Generation
Analysis Results
Analysis and Visualisation
DS3
Reports Archive

DATA COLLECTION

Data Source- Data is taken from Kaggle.

Link:
https://www.kaggle.com/datasets/gianinamariapetrascu/gender-inequality-index/data
and https://www.kaggle.com/datasets/iamsouravbanerjee/gross-national-income-per-capita

Data Description:
<u>The Gender Inequality Index (GII)</u> dataset provides a comprehensive measure of gender inequality across countries, capturing gender disparities in health, education, and economic opportunities. This dataset includes GII scores, as well as component scores for each indicator, for over 190 countries, in 2021.
<u>Gross National Income Per Capita</u> dataset contains Gross National Income Per Capita of all countries for year 2021.

FEATURES

```
#   Column
--- ------
 0   Country
 1   Human_development
 2   GII
 3   Rank
 4   Maternal_mortality
 5   Adolescent_birth_rate
 6   Seats_parliament
 7   F_secondary_educ
 8   M_secondary_educ
 9   F_Labour_force
 10  M_Labour_force
 11  HDI
 12  GNI
```

- **Country-** Names of Countries
- **GII-** Developed by the United Nations Development Programme (UNDP), the GII measures gender inequality by analyzing health, empowerment, and labor market participation indicators.
- **Rank-** Countries Rank according to GII.
- Maternal Mortality Ratio- Number of deaths due to pregnancy-related causes per 100,000 live births.
- **Adolescent Birth Rate-** Annual number of births to females aged 10-14 or 15-19 years per 1,000 females in the respective age group.
- **Human Development Index-** HDI is a composite index that measures average achievement in human development.
- **GNI Per Capita-** Gross National Income Per-Capita
- **Other-** Women Seats in Parliament, Female Secondary Education, Male Secondary Education, Female Labour Force Participation Rate, Male Labour Force Participation Rate.

DATA PREPROCESSING

DATA CLEANING

```python
# Inspect the dataset
print(df.info())
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Country                195 non-null    object 
 1   Human_development      191 non-null    object 
 2   GII                    170 non-null    float64
 3   Rank                   170 non-null    float64
 4   Maternal_mortality     184 non-null    float64
 5   Adolescent_birth_rate  195 non-null    float64
 6   Seats_parliament       193 non-null    float64
 7   F_secondary_educ       177 non-null    float64
 8   M_secondary_educ       177 non-null    float64
 9   F_Labour_force         180 non-null    float64
 10  M_Labour_force         180 non-null    float64
dtypes: float64(9), object(2)
memory usage: 16.9+ KB
None
```

HANDLING MISSING VALUES

Removing rows with Null Values:

```
# Handle missing values
df = df.dropna()  # For simplicity, dropping rows with missing values
```

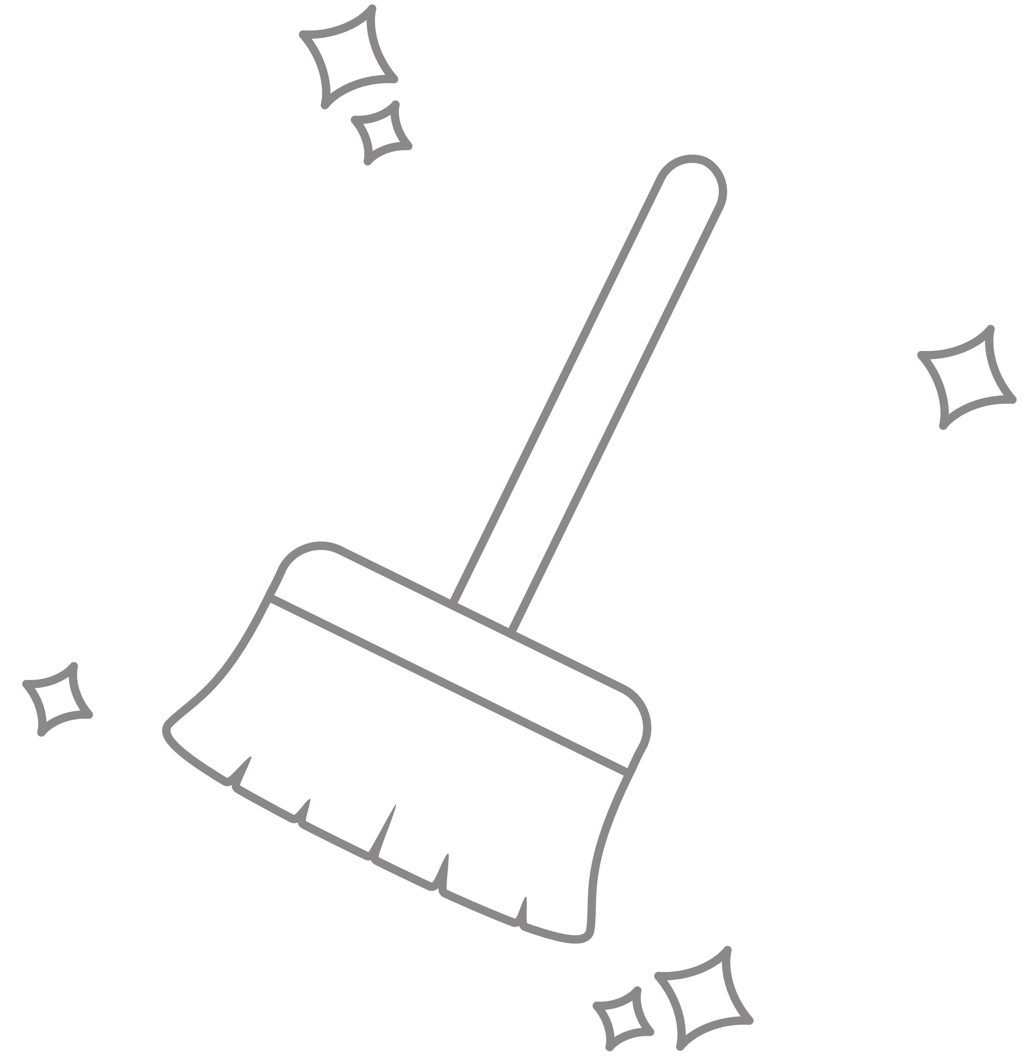
✓ 0.0s

DATASET AFTER HANDLING MISSING VALUES:

```
print(df.info())
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
Index: 170 entries, 0 to 190
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Country               170 non-null    object
 1   Human_development     170 non-null    object
 2   GII                   170 non-null    float64
 3   Rank                  170 non-null    float64
 4   Maternal_mortality    170 non-null    float64
 5   Adolescent_birth_rate 170 non-null    float64
 6   Seats_parliament      170 non-null    float64
 7   F_secondary_educ      170 non-null    float64
 8   M_secondary_educ      170 non-null    float64
 9   F_Labour_force        170 non-null    float64
 10  M_Labour_force        170 non-null    float64
dtypes: float64(9), object(2)
memory usage: 15.9+ KB
None
```



DATA TRANSFORMATION

MERGING DATASETS

```
# Merge the datasets on the 'Country' column
df_combined = pd.merge(df, df_gdp, on='Country', how='inner')
```

✓ 0.0s

STANDARDIZATION (For KNN Model)

```
# Split the data into training and testing sets
X1_train, X1_test, y1_train, y1_test = train_test_split(X1, y1, test_size=0.2, random_state=42)

# Standardize the features
scaler1 = StandardScaler()
X1_train_scaled = scaler1.fit_transform(X1_train)
X1_test_scaled = scaler1.transform(X1_test)
```

DATA TRANSFORMATION

FEATURE SELECTION

```python
df_gdp = pd.read_csv(r"C:\Users\AKSHAT\Downloads\Gross National Income Per Capita.csv")
```

✓ 0.0s

```python
df_gdp.head()
```

✓ 0.0s

	ISO3	Country	Continent	Hemisphere	Human Development Groups	UNDP Developing Regions	HDI Rank (2021)	Gross National Income Per Capita (1990)	Gross National Income Per Capita (1991)	Gross National Income Per Capita (1992)	...	Gross National Income Per Capita (2012)	Gross National Income Per Capita (2013)	Gross National Income Per Capita (2014)	Gross National Income Per Capita (2015)	Gross National Income Per Capita (2016)	Gross National Income Per Capita (2017)	Gross National Income Per Capita (2018)
0	AFG	Afghanistan	Asia	Northern Hemisphere	Low	SA	180.0	2684.550019	2276.289409	2059.868084	...	2125.862821	2193.553936	2178.507021	2101.589319	2077.566899	2085.487571	2054.939895
1	AGO	Angola	Africa	Southern Hemisphere	Medium	SSA	148.0	4845.706901	5405.349257	2073.902390	...	7280.845666	7478.104777	7704.231949	7652.656486	7189.426672	6861.575738	6381.521946
2	ALB	Albania	Europe	Northern Hemisphere	High	ECA	67.0	4742.215529	3358.087827	3080.746654	...	11146.263030	11552.982470	11691.648290	12016.297600	12484.624200	12802.148310	13302.705960
3	AND	Andorra	Europe	Northern Hemisphere	Very High	NaN	40.0	43773.146500	43175.147600	41935.787200	...	47126.814610	46385.095200	48483.720320	49936.874540	52267.738320	52650.225760	53483.306630
4	ARE	United Arab Emirates	Asia	Northern Hemisphere	Very High	AS	26.0	102433.136000	96250.290360	93043.477370	...	57445.954750	60005.695360	62573.505310	65577.512240	66881.329740	67667.508460	67195.095230

5 rows × 39 columns

```python
df_gdp  =df_gdp [['Country','HDI Rank (2021)', 'Gross National Income Per Capita (2021)']]
```

✓ 0.0s

3 columns

```python
df_gdp.head()
```

✓ 0.0s

	Country	HDI Rank (2021)	Gross National Income Per Capita (2021)
0	Afghanistan	180.0	1824.190915
1	Angola	148.0	5465.617791
2	Albania	67.0	14131.110390
3	Andorra	40.0	51166.626610
4	United Arab Emirates	26.0	62573.591810

(Reducing the input variable to the dataset by using only relevant data)

DATA ANALYSIS

DATA ANALYSIS

Descriptive Analysis Using describe() function

```
df_combined.describe()
```

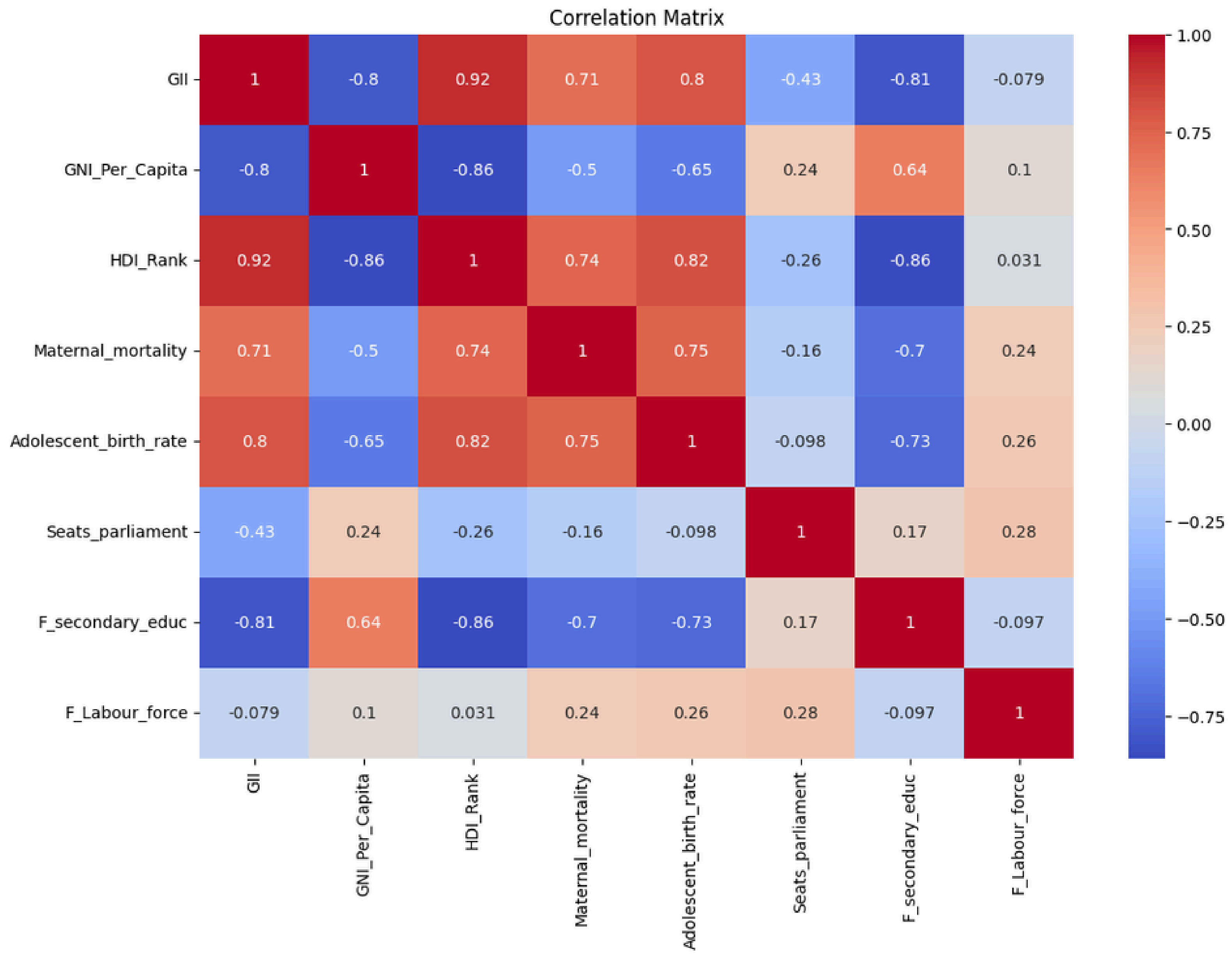
✓ 0.0s

	GII	Rank	Maternal_mortality	Adolescent_birth_rate	Seats_parliament	F_secondary_educ	M_secondary_educ	F_Labour_force	M_Labour_force	HDI Rank (2021)	Gross National Income Per Capita (2021)
count	166.000000	166.000000	166.000000	166.000000	166.000000	166.000000	166.000000	166.000000	166.000000	166.000000	166.000000
mean	0.342904	85.018072	155.855422	44.689759	25.527711	62.438554	66.782530	50.321084	69.996386	94.722892	19819.711796
std	0.197911	49.370642	233.686692	39.103850	12.393721	30.004300	26.882138	15.293991	8.539242	55.849541	19879.507589
min	0.013000	1.000000	2.000000	1.900000	0.000000	6.400000	13.000000	6.000000	43.900000	1.000000	731.786709
25%	0.169500	42.250000	10.250000	10.625000	16.875000	36.875000	44.250000	42.725000	65.200000	46.250000	4580.654231
50%	0.363000	85.500000	46.000000	34.650000	24.550000	69.700000	70.750000	52.150000	69.500000	93.000000	12463.753275
75%	0.505750	127.750000	181.500000	64.725000	34.675000	91.025000	92.700000	60.375000	75.375000	143.750000	30106.038932
max	0.820000	170.000000	1150.000000	170.500000	55.700000	100.000000	100.000000	82.500000	95.500000	191.000000	90918.644710

Key Findings :

The dataset summary reveals that the average maternal mortality is approximately 155.86 per 100,000 live births, and the mean adolescent birth rate is around 44.69 per 1,000 women. Additionally, on average, women hold 25.53% of parliamentary seats, and there is a notable disparity between female and male secondary education rates, with means of 62.44% and 66.78%, respectively.

HEATMAP CORRELATION OF EACH FEATURE IN THE DATASET (CORRELATION ANALYSIS)

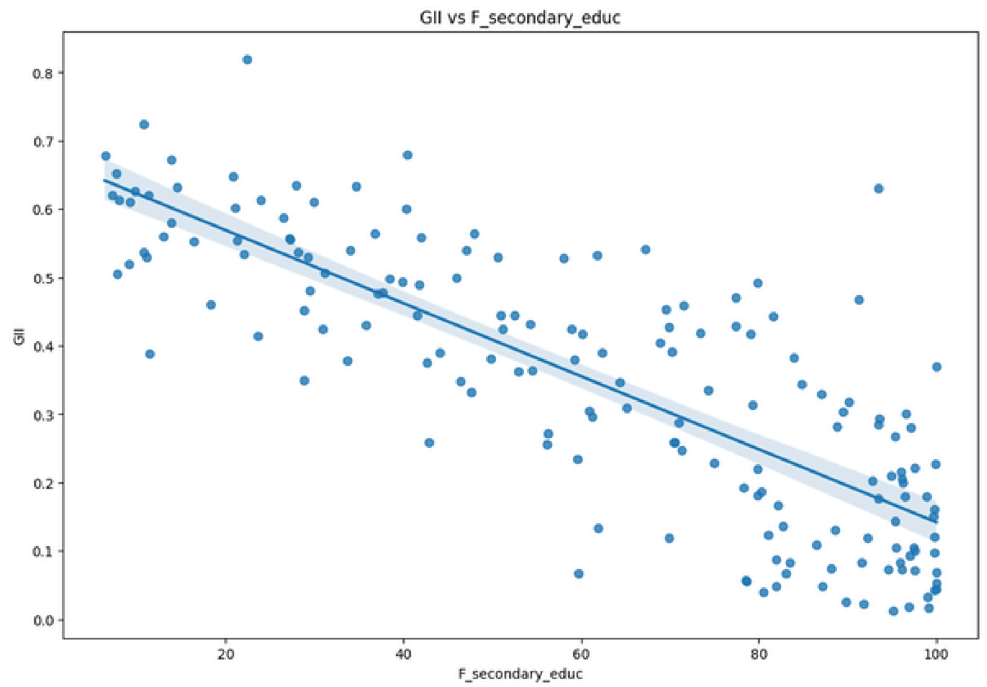


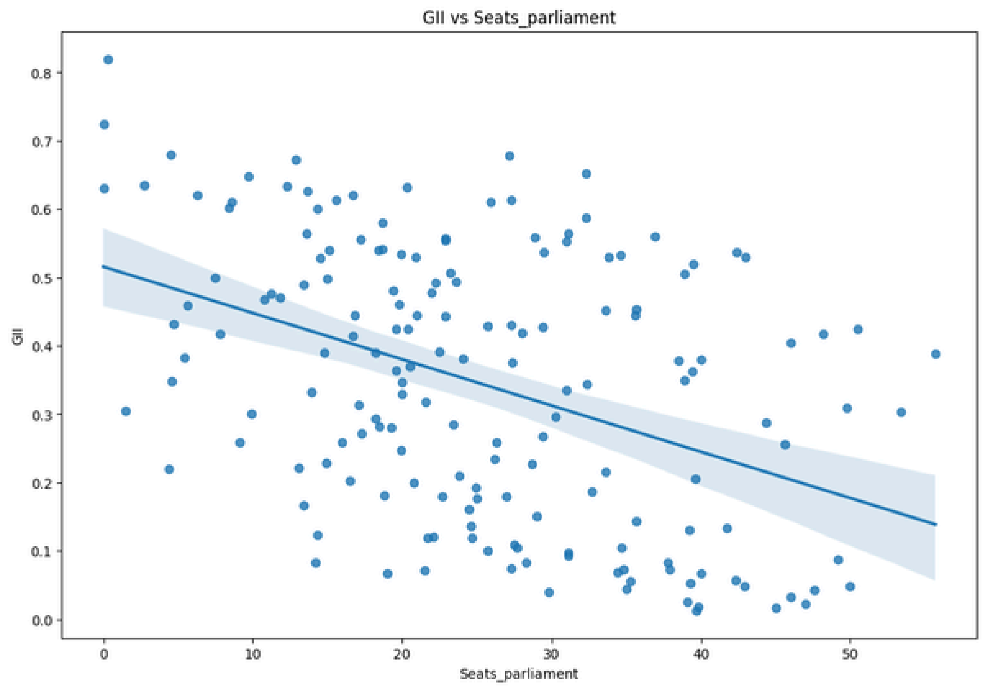
KEY FINDINGS

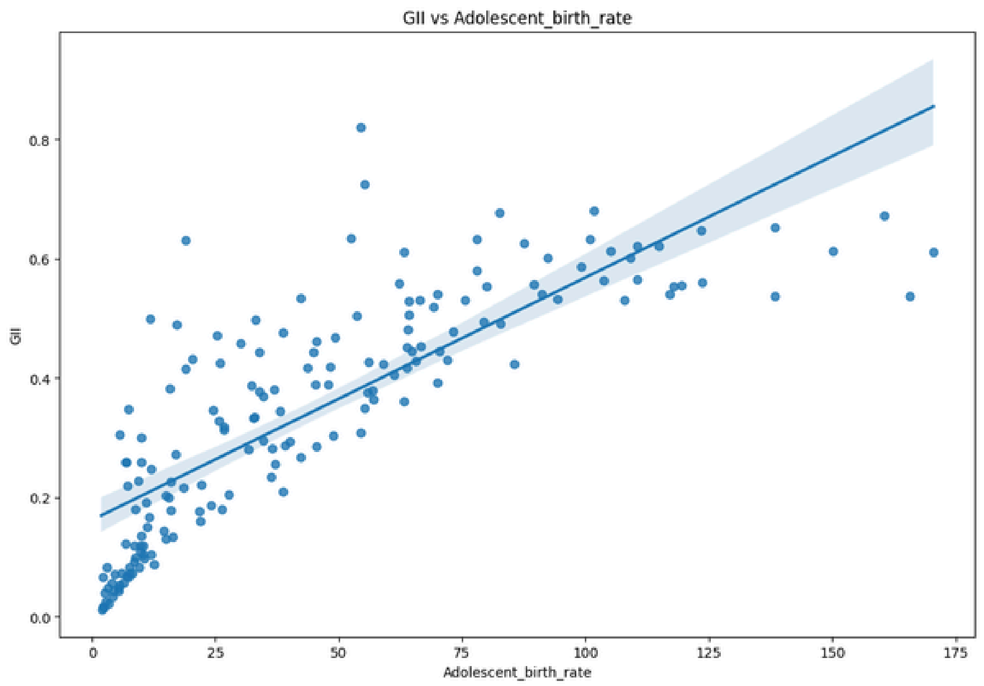
1. <u>Gender Inequality Index (GII) and Economic Indicators:</u>
   - GII vs. GNI Per Capita: Strong negative correlation (-0.81). Higher gender inequality is linked to lower GNI Per Capita.
   - GII vs. HDI Rank: Strong positive correlation (0.91). Higher gender inequality is linked to lower human development.
   - GII vs. Maternal Mortality: Strong positive correlation (0.78). Higher gender inequality is linked to higher maternal mortality.
2. <u>Education and Labour Force Participation:</u>
   - Female Secondary Education vs. GNI Per Capita: Strong positive correlation (0.61). Higher female education is linked to higher GNI Per Capita.
   - Female Labour Force Participation vs. GNI Per Capita: Positive but weak correlation (0.19). Some positive impact on GNI Per Capita.
3. <u>Adolescent Birth Rate and Economic Indicators:</u>
   - Adolescent Birth Rate vs. GNI Per Capita: Strong negative correlation (-0.69). Higher adolescent birth rates are linked to lower GNI Per Capita.
   - Adolescent Birth Rate vs. HDI Rank: Strong positive correlation (0.81). Higher adolescent birth rates are linked to lower human development.
4. <u>Political Participation:</u>
   - Seats in Parliament vs. GNI Per Capita: Weak positive correlation (0.28). Potential positive impact on GNI Per Capita.
   - Seats in Parliament vs. GII: Moderate negative correlation (-0.4). Higher political representation of women is linked to lower gender inequality.

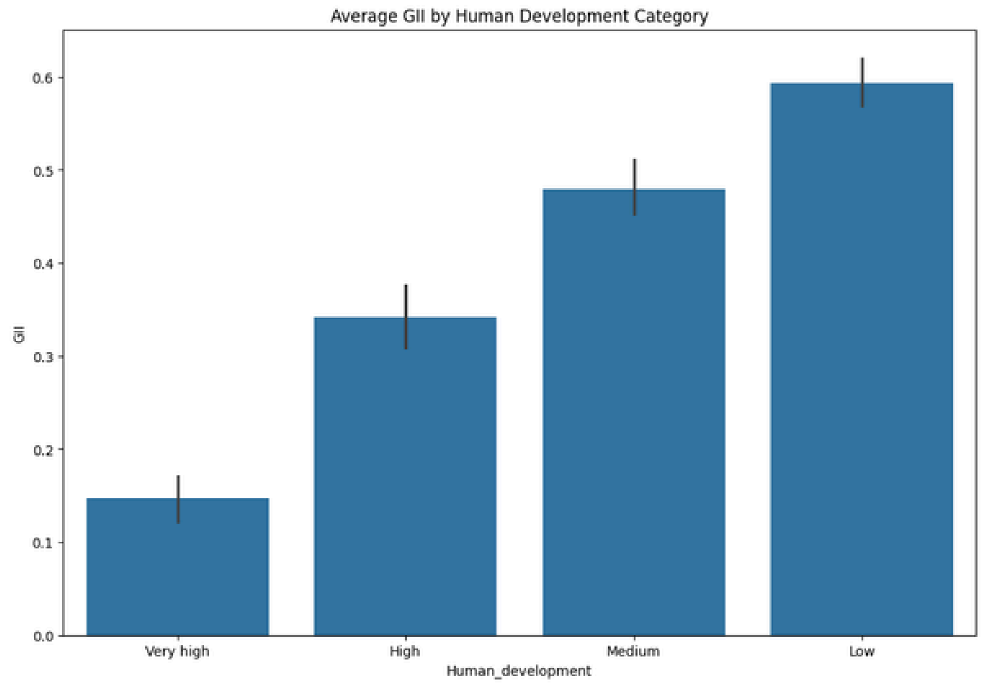
DATA ANALYSIS

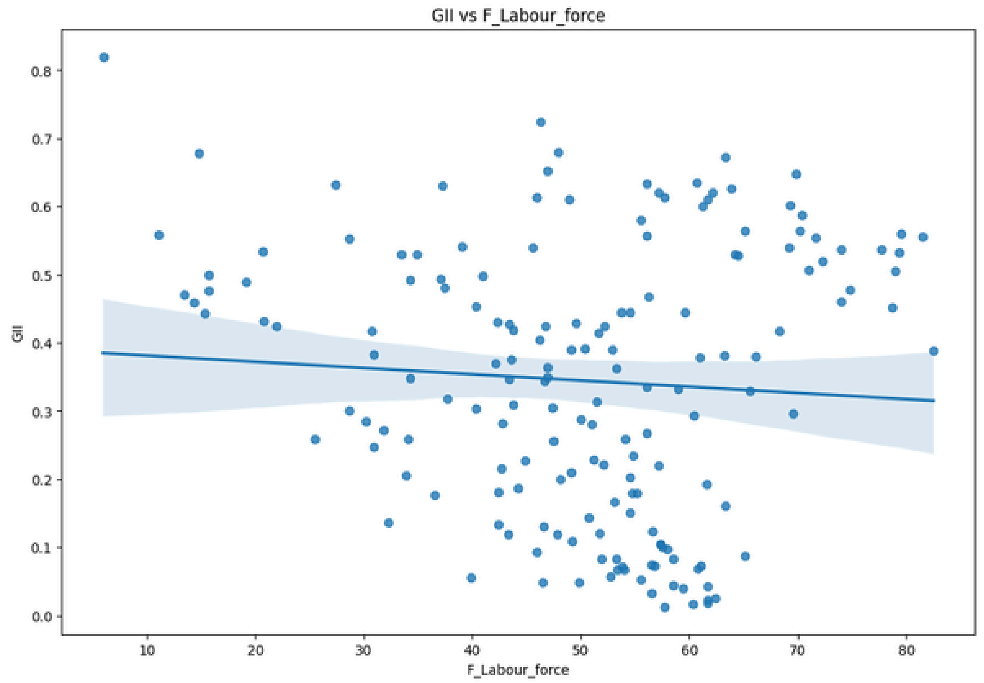
Factors Affecting GII (Using **Regression Analysis** and Bar Graph)-

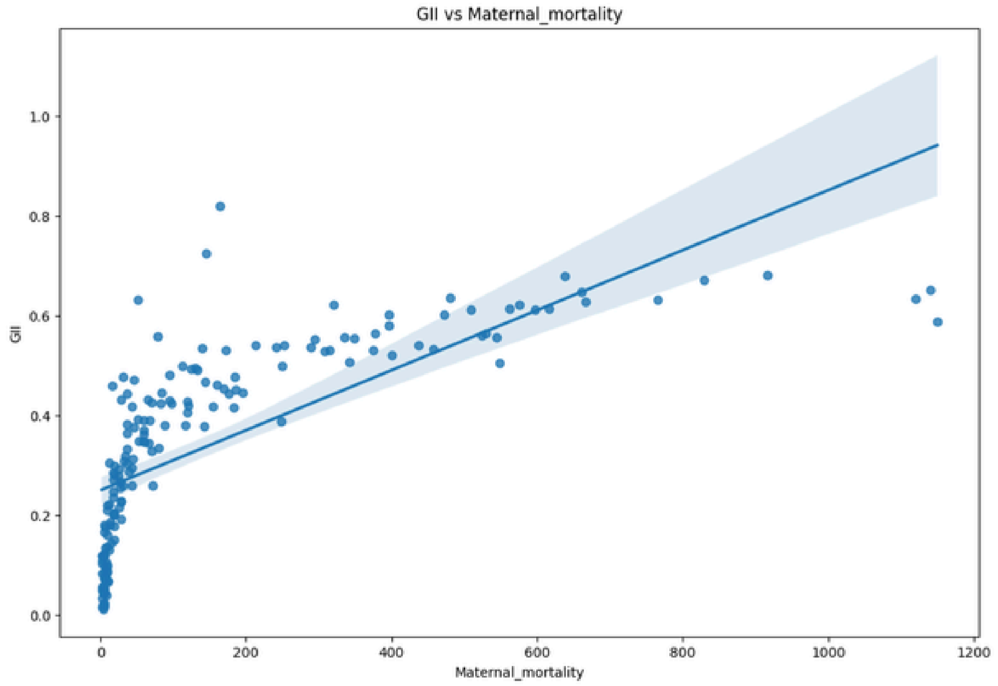












MACHINE LEARNING MODELS

MACHINE LEARNING MODEL (LINEAR REGRESSION)

```
# Select relevant columns
df_model = df_combined[['GII', 'GNI_Per_Capita', 'HDI_Rank']]  # Adjust columns as needed


# Define the feature matrix X and the target vector y
X = df_model[['GII', 'HDI_Rank']]
y = df_model['GNI_Per_Capita']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train the model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)
✓ 0.0s
```

```
# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R^2 Score: {r2}")
✓ 0.0s

Mean Squared Error: 44697439.65382441
R^2 Score: 0.8297065498094043
```

KEY FINDINGS

- Model Performance: The Linear Regression model achieves an **R^2 score** of approximately **0.83**, indicating that the model explains about 83% of the variance in GNI Per Capita based on GII and HDI Rank. This suggests a strong but slightly lower predictive power compared to the KNN model.
- Mean Squared Error: The mean squared error **(MSE) is approximately 44,697,349,** which indicates the average squared difference between the predicted and actual GNI Per Capita values. While this value is large due to the scale of GNI Per Capita, the R^2 score provides better context for the model's accuracy.
- Overall, while the Linear Regression model shows a **strong relationship** between the selected socio-economic indicators **(GII, HDI Rank) and economic output**.

MACHINE LEARNING MODEL (KNN)

```python
# Define the feature matrix X and the target vector y
X = df_knn[['GII', 'HDI_Rank']]
y = df_knn['GNI_Per_Capita']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Standardize the features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Initialize and train the KNN model
knn = KNeighborsRegressor(n_neighbors=5)
knn.fit(X_train_scaled, y_train)

# Make predictions
y_pred = knn.predict(X_test_scaled)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R^2 Score: {r2}")
```

✓ 0.0s

```
Mean Squared Error: 29240778.60148078
R^2 Score: 0.8885951161213896
```

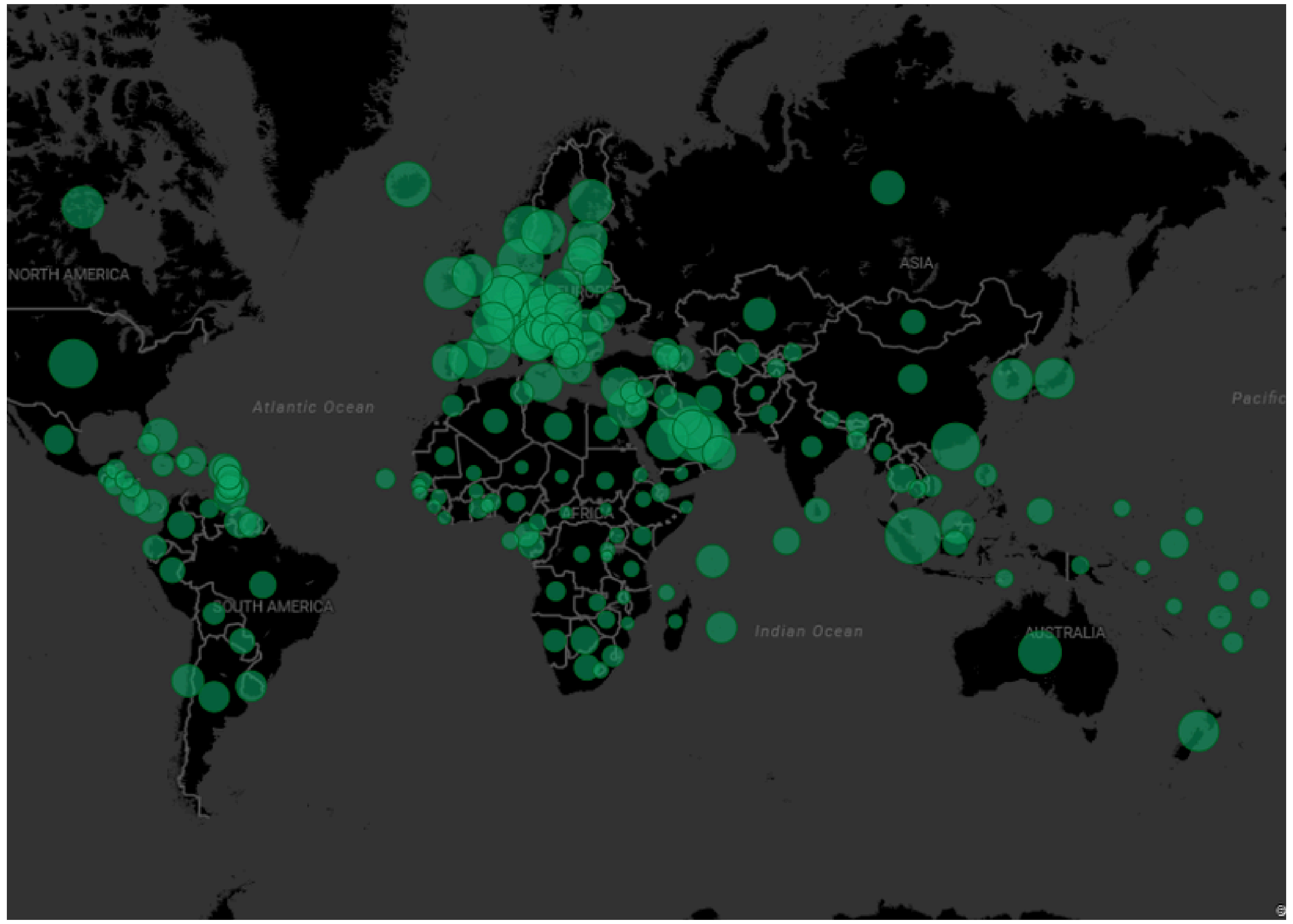
KEY FINDINGS

- Model Performance: The KNN regression model achieves an **R^2 score** of approximately **0.89**, indicating that the model explains about 89% of the variance in GNI Per Capita based on GII and HDI Rank. This suggests a strong predictive power of the selected features.
- Mean Squared Error: The mean squared error **(MSE) of approximately 29,240,778** indicates the average squared difference between the predicted and actual GNI Per Capita values. This value is higher compared to the KNN model's MSE, suggesting that the Linear Regression model is less accurate.
- Overall, the model shows a **strong relationship** between the selected socio-economic indicators **(GII, HDI Rank) and economic output**, the **KNN model performed better** in terms of predictive accuracy.

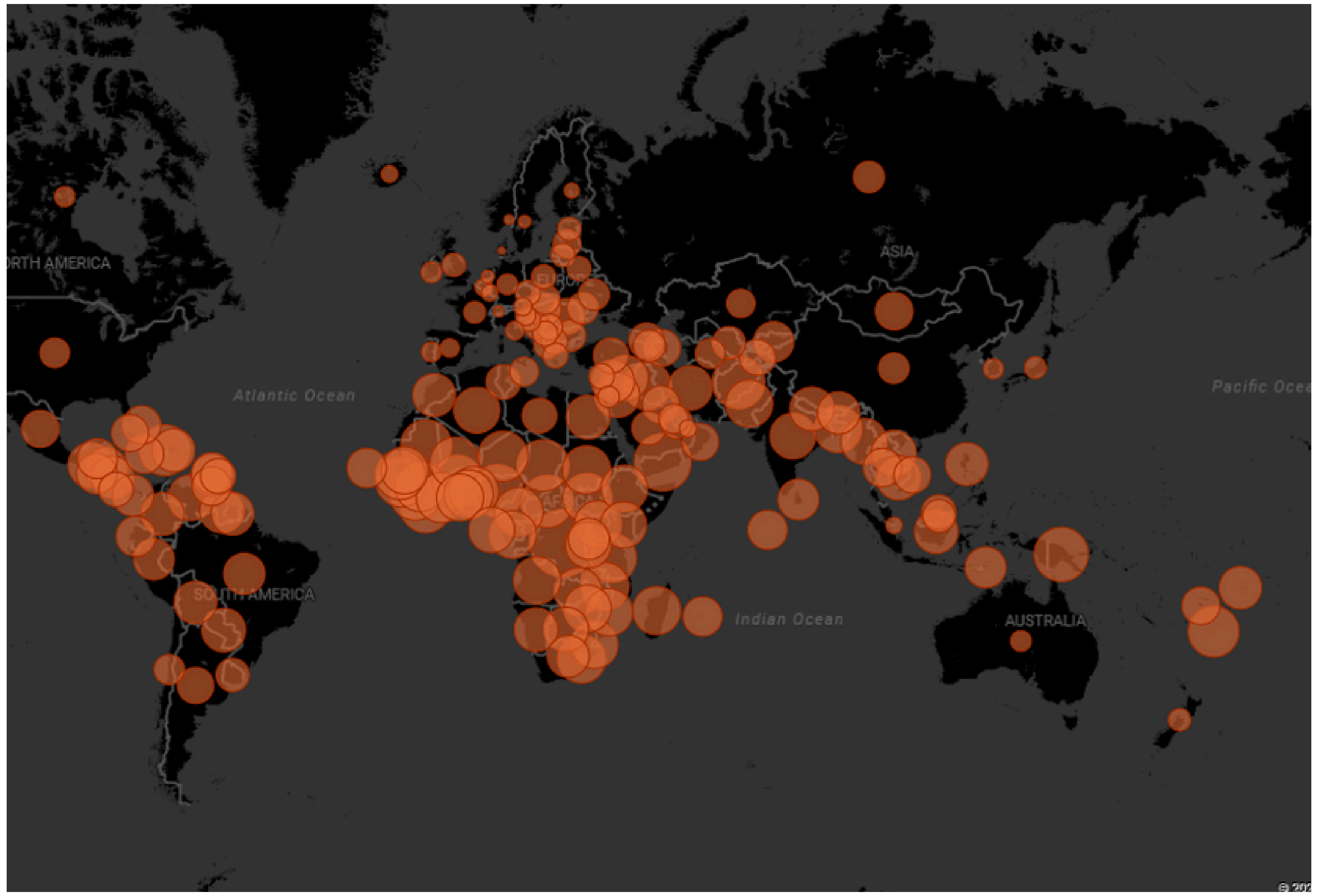
DATA VISUALISATION

DATA VISUALISATION

GNI Per Capita of all countries:



GII of all countries:



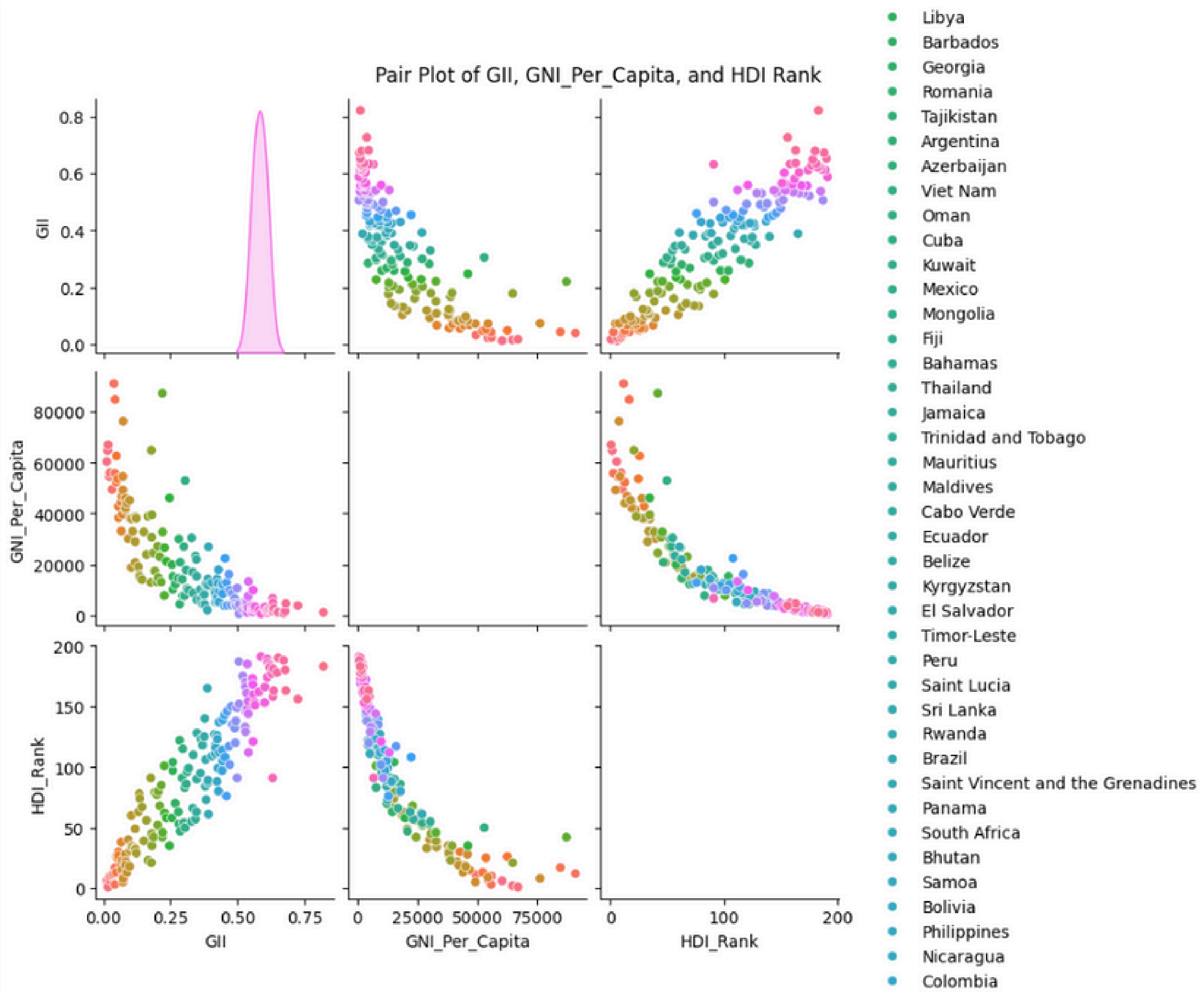
It is visible that countries with low GNI Per Capita have high GII and countries with high GNI Per Capita have low GII

DATA VISUALISATION

Pair Plot

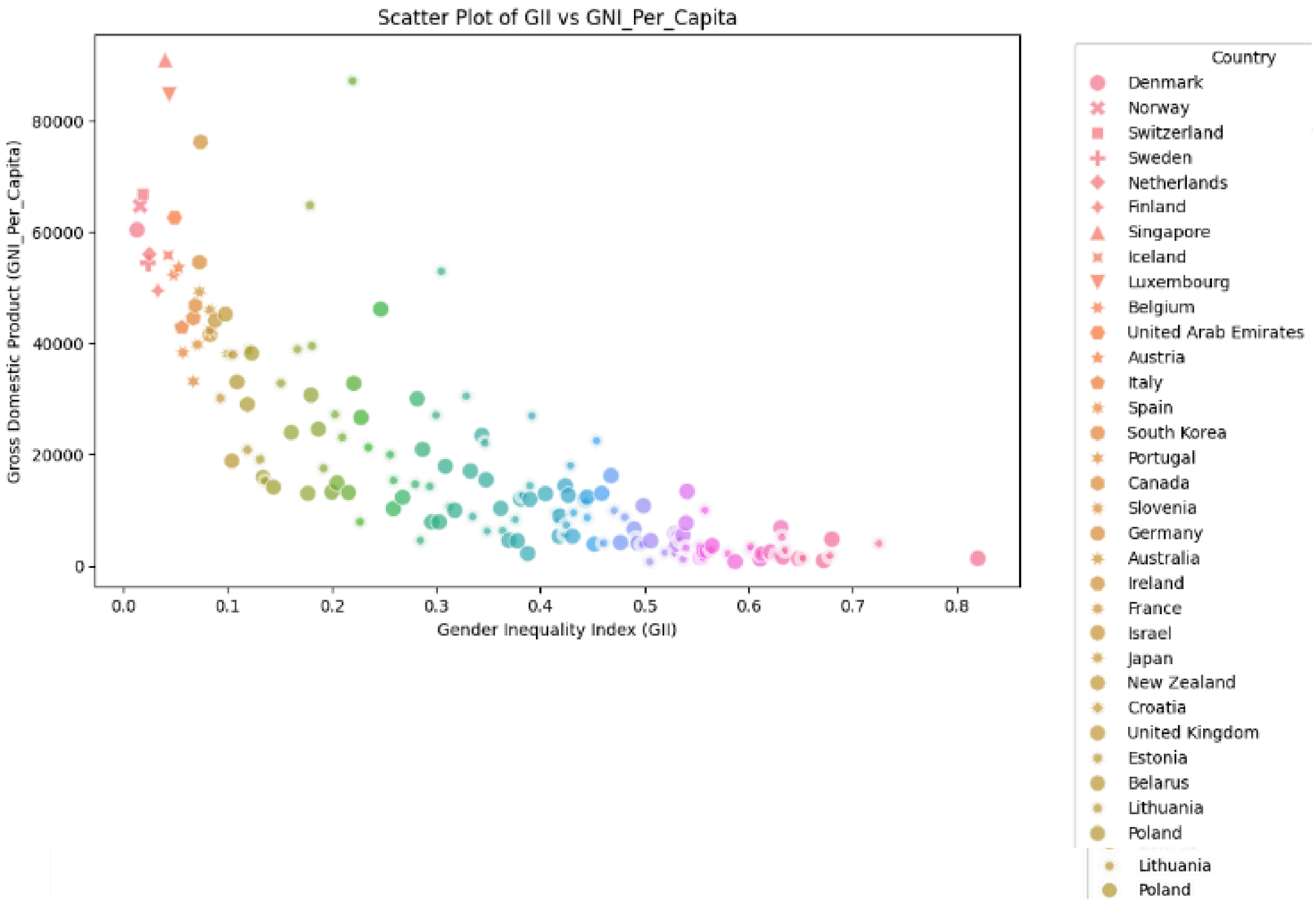
```python
# Pair plot for multiple variables
sns.pairplot(df_combined, vars=['GII', 'GNI_Per_Capita', 'HDI_Rank'], hue='Country')
plt.suptitle('Pair Plot of GII, GNI_Per_Capita, and HDI Rank', y=1.02)
plt.show()
```

✓ 4.7s



Scatter Plot

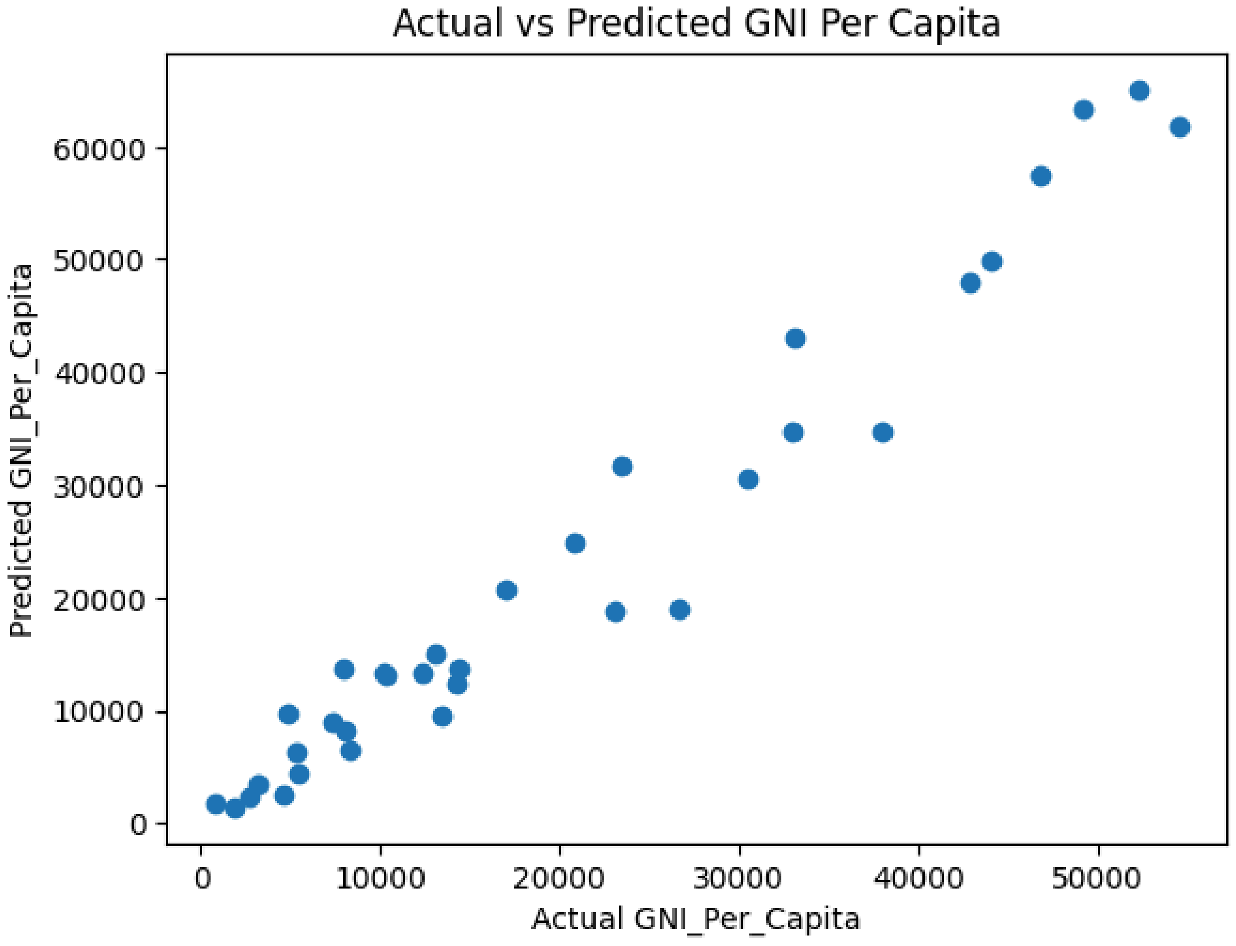
```python
# Scatter plot for GII vs GDP
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df_combined, x='GII', y='GNI_Per_Capita', hue='Country',style='Country', alpha=0.7, s=100)
plt.title('Scatter Plot of GII vs GNI_Per_Capita')
plt.xlabel('Gender Inequality Index (GII)')
plt.ylabel('Gross Domestic Product (GNI_Per_Capita)')
plt.legend(title='Country',bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```



DATA VISUALISATION

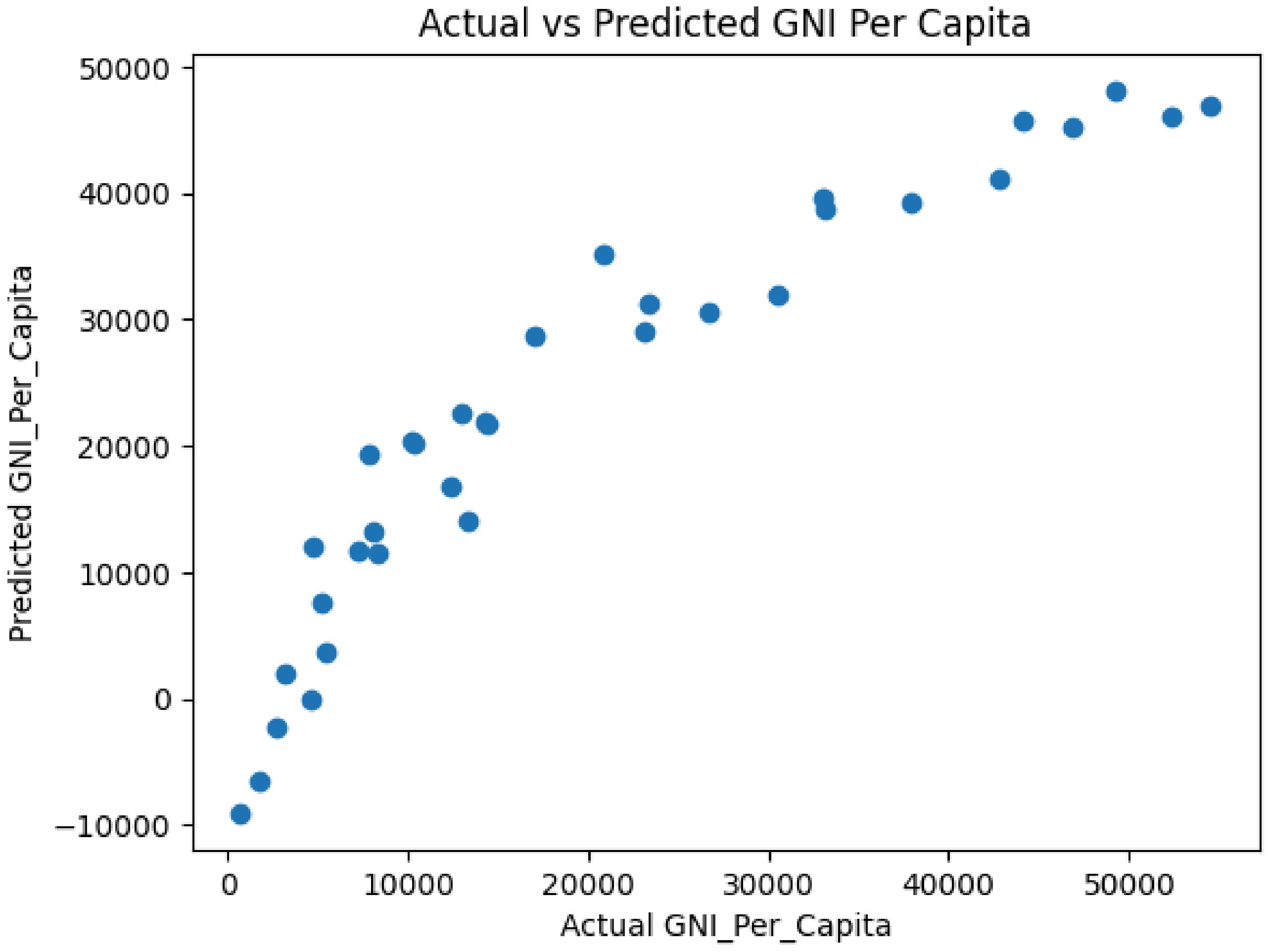
KNN Model

```python
# Plotting the results
plt.scatter(y_test, y_pred)
plt.xlabel('Actual GNI_Per_Capita')
plt.ylabel('Predicted GNI_Per_Capita')
plt.title('Actual vs Predicted GNI Per Capita')
plt.show()
```
✓ 0.2s

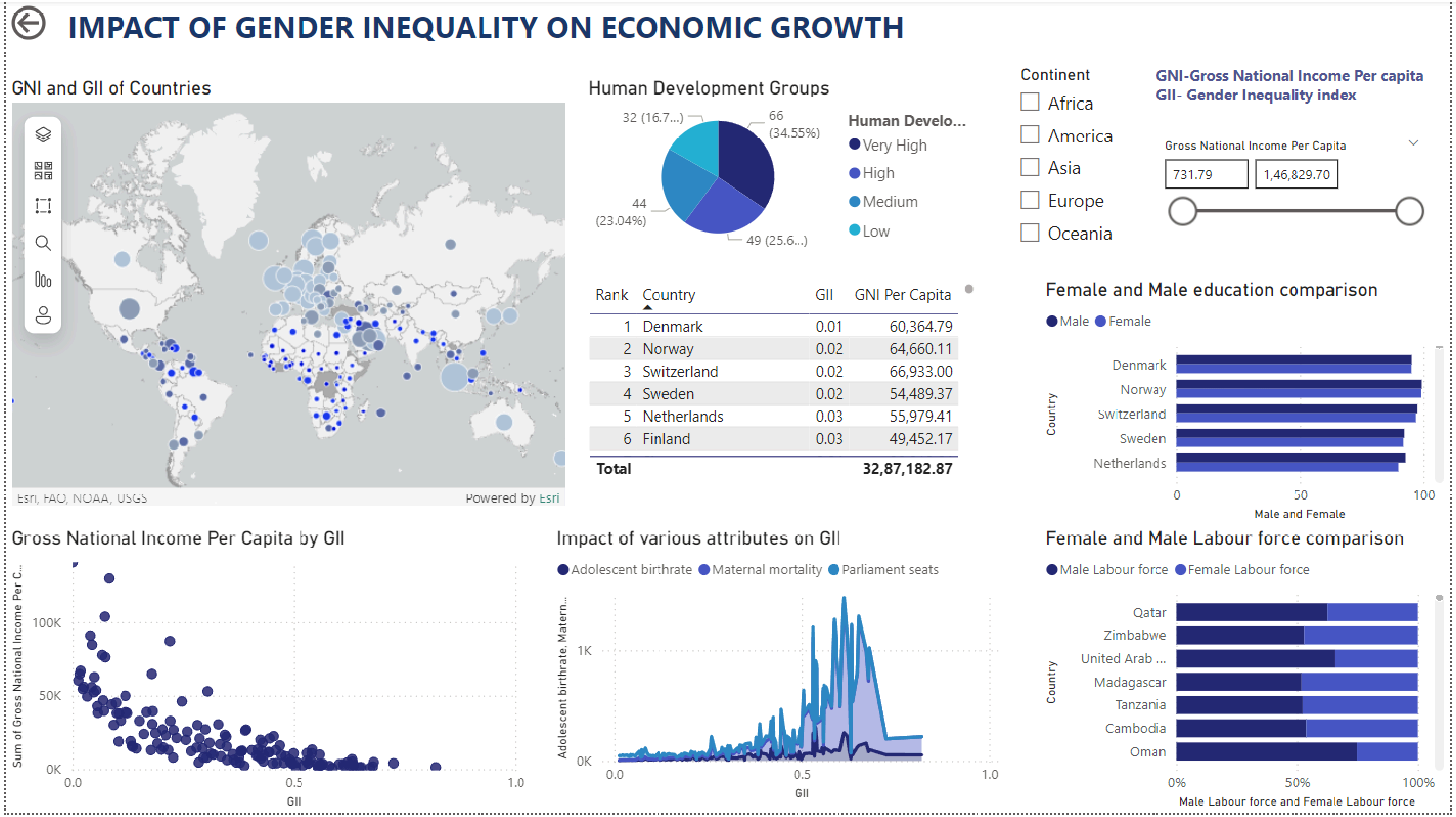


Linear Regression Model

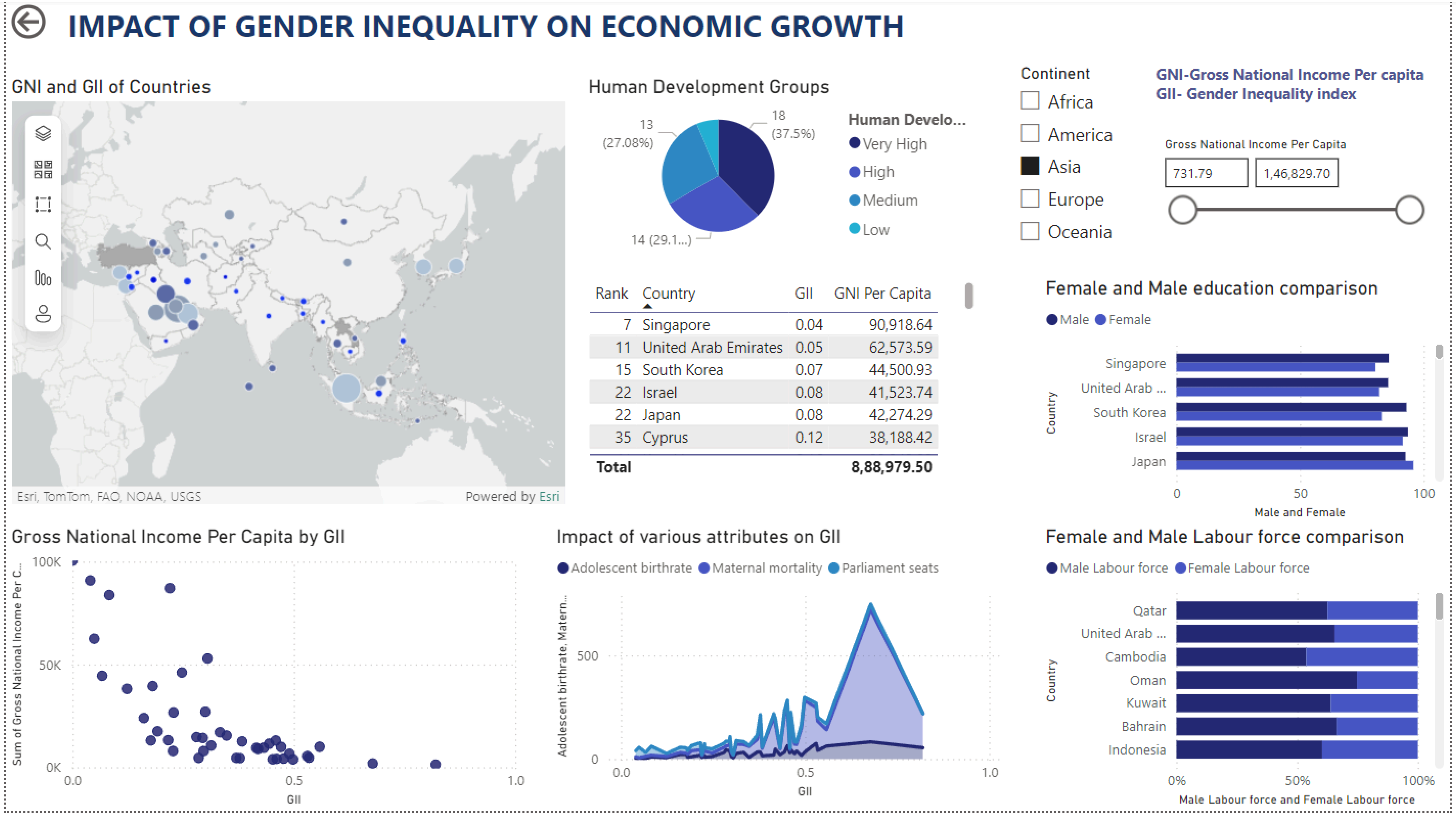
```python
# Plotting the results
plt.scatter(y_test, y_pred)
plt.xlabel('Actual GNI_Per_Capita')
plt.ylabel('Predicted GNI_Per_Capita')
plt.title('Actual vs Predicted GNI Per Capita')
plt.show()
```
✓ 0.1s



-DASHBOARD SCREENSHOTS-



-DASHBOARD SCREENSHOTS-



HYPOTHESIS AND SOLUTION

Formulated Hypothesis:

Hypothesis: Higher gender inequality negatively impacts economic growth, as measured by GNI per capita.

Implementation of policies to promote women employment and reducing gender gaps in labor force and education can have a great impact on economic performance. By analyzing data on gender disparities in education, employment, and political representation across various regions, we quantified how these inequalities hinder Economic Growth.

Proposed Solution

To address the impact of gender inequality on economic growth, we propose a multifaceted solution that includes policy reforms, educational programs, and economic incentives designed to improve gender equality. Key components include:

1. Policy Reforms: Implement laws and policies that promote gender equality in education, workforce participation, and political representation.
2. Educational Programs: Increase access to secondary and tertiary education for girls and women, with a focus on STEM fields.
3. Economic Incentives: Provide financial incentives for companies that demonstrate gender equality in their workforce and leadership positions.


THANK YOU