

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 28 June 2024 |
| Team ID | 739785 |
| Project Title | A Comprehensive Measure Of Well-Being:A Human Development Using Machine Learning |
| Maximum Marks | 6 Marks |

Data Exploration and Preprocessing Report

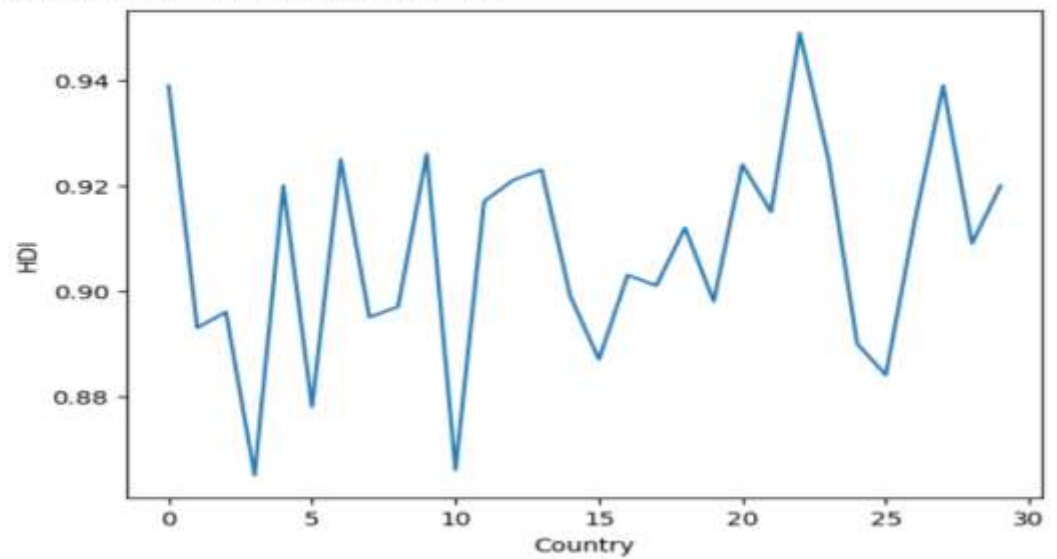
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

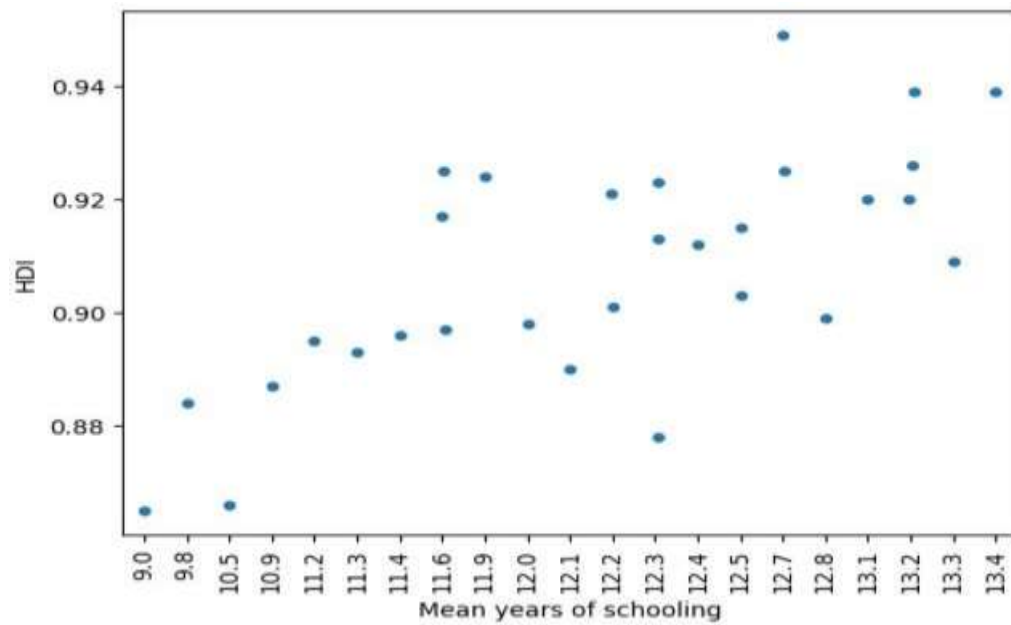
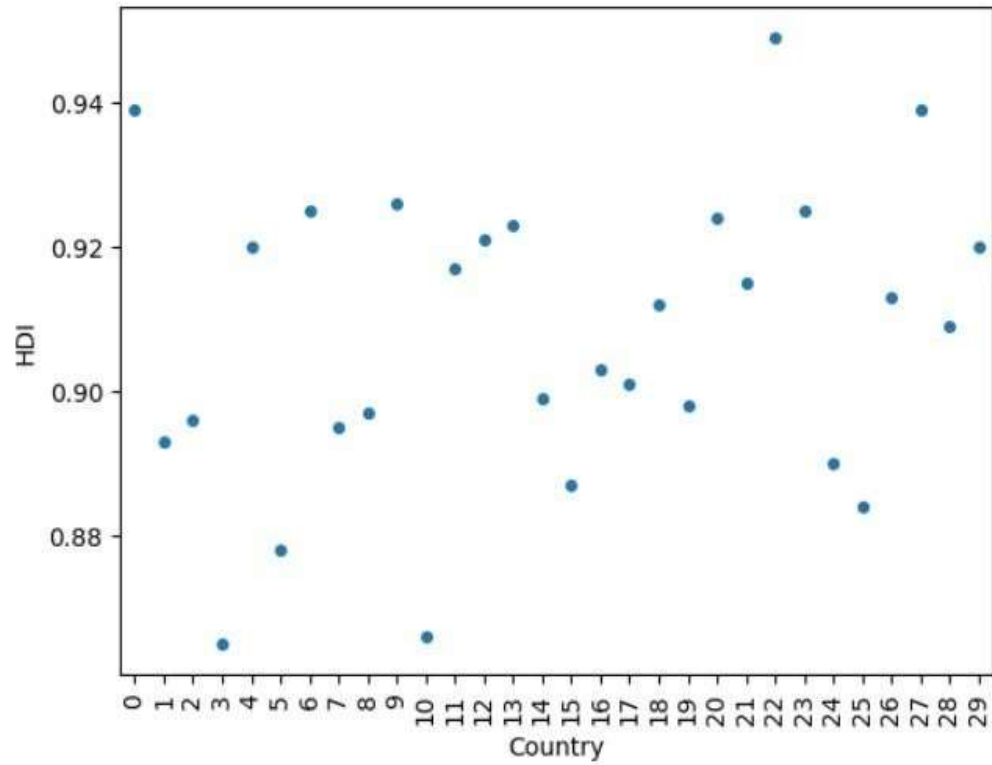
| Section | Description | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|--|------------|------------|------------|------------|-----------------|-------------------------|--|--|-------------------------------|-------------------------------------|-------------------------------------|---|---|---|-------------------------------------|-------------------------------------|--------------------------|-------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------|-----------|-----------|-----------|----------|-----------|----------|--------------|----------|-----------|----------|-----------|-----------|----------|-----------|----------|-----------|-----|-----------|-----------|-----------|----------|----------|----------|--------------|-----------|----------|----------|-----------|-----------|----------|-----------|----------|-----------|-----|----------|----------|----------|----------|-----------|----------|------------|------------|------------|-----------|----------|----------|----------|----------|----------|----------|-----|-----------|-----------|-----------|----------|-----------|----------|-------------|-----------|-----------|----------|-----------|----------|----------|----------|----------|-----------|-----|-----------|-----------|-----------|----------|-----------|----------|--------------|----------|----------|----------|-----------|-----------|----------|-----------|----------|-----------|-----|------------|------------|------------|----------|-----------|-----------|--------------|-----------|----------|----------|-----------|-----------|----------|-----------|----------|-----------|-----|------------|------------|------------|----------|-----------|-----------|---------------|-----------|-----------|----------|-----------|-----------|----------|-----------|----------|-----------|
| Data Overview | <div>Descriptive Analysis:-</div> <div><pre>data.describe()</pre><table><thead><tr><th></th><th>Unnamed: 0</th><th>Id</th><th>HDI Rank</th><th>HDI</th><th>Life expectancy</th><th>Mean years of schooling</th><th>Gross national income (GNI) per capita</th><th>GNI per capita minus HDI rank</th><th>Change in HDI rank 2010-2015</th><th>Average annual HDI growth 1990-2000</th><th>Coefficient of human inequality</th><th>Inequality in life expectancy (N) 2010-2015</th><th>Inequality-adjusted life expectancy index</th><th>Inequality in education(N)</th><th>Inequality-adjusted education index</th><th>Inequality in income (N)</th></tr></thead><tbody><tr><td>count</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td><td>195.000000</td></tr><tr><td>mean</td><td>87.900000</td><td>98.000000</td><td>94.319148</td><td>0.668835</td><td>71.267368</td><td>9.371277</td><td>17239.502591</td><td>2.239362</td><td>-0.021277</td><td>0.773611</td><td>30.669294</td><td>17.155135</td><td>0.868151</td><td>20.278125</td><td>0.527722</td><td>24.115584</td></tr><tr><td>std</td><td>50.425804</td><td>98.425804</td><td>54.425521</td><td>0.155152</td><td>8.328811</td><td>3.068028</td><td>19052.809848</td><td>17.309420</td><td>5.443946</td><td>0.702704</td><td>10.219353</td><td>12.197885</td><td>0.190741</td><td>14.631457</td><td>0.224238</td><td>10.511501</td></tr><tr><td>min</td><td>0.000000</td><td>1.000000</td><td>1.000000</td><td>0.352000</td><td>48.900000</td><td>1.400000</td><td>294.000000</td><td>-79.000000</td><td>-35.000000</td><td>-1.550000</td><td>5.300000</td><td>2.600000</td><td>0.253000</td><td>1.400000</td><td>0.134000</td><td>8.900000</td></tr><tr><td>25%</td><td>48.500000</td><td>46.500000</td><td>47.750000</td><td>0.577750</td><td>65.550000</td><td>6.075000</td><td>3527.000000</td><td>-7.900000</td><td>-2.000000</td><td>0.425000</td><td>10.860000</td><td>6.700000</td><td>0.524000</td><td>6.575000</td><td>0.324000</td><td>16.300000</td></tr><tr><td>50%</td><td>87.900000</td><td>98.000000</td><td>94.500000</td><td>0.729500</td><td>73.300000</td><td>8.650000</td><td>10361.000000</td><td>4.900000</td><td>0.000000</td><td>0.760000</td><td>26.200000</td><td>13.600000</td><td>0.707000</td><td>17.490000</td><td>0.527000</td><td>21.300000</td></tr><tr><td>75%</td><td>145.500000</td><td>146.500000</td><td>139.750000</td><td>0.811250</td><td>77.900000</td><td>11.125000</td><td>23384.000000</td><td>13.000000</td><td>2.000000</td><td>1.150000</td><td>28.500000</td><td>20.100000</td><td>0.813000</td><td>34.525000</td><td>0.732250</td><td>29.475000</td></tr><tr><td>max</td><td>194.900000</td><td>195.000000</td><td>195.000000</td><td>0.948000</td><td>84.200000</td><td>13.400000</td><td>129615.000000</td><td>40.000000</td><td>15.000000</td><td>3.630000</td><td>44.800000</td><td>40.200000</td><td>0.959000</td><td>48.300000</td><td>0.921000</td><td>58.400000</td></tr></tbody></table></div> | | Unnamed: 0 | Id | HDI Rank | HDI | Life expectancy | Mean years of schooling | Gross national income (GNI) per capita | GNI per capita minus HDI rank | Change in HDI rank 2010-2015 | Average annual HDI growth 1990-2000 | Coefficient of human inequality | Inequality in life expectancy (N) 2010-2015 | Inequality-adjusted life expectancy index | Inequality in education(N) | Inequality-adjusted education index | Inequality in income (N) | count | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | mean | 87.900000 | 98.000000 | 94.319148 | 0.668835 | 71.267368 | 9.371277 | 17239.502591 | 2.239362 | -0.021277 | 0.773611 | 30.669294 | 17.155135 | 0.868151 | 20.278125 | 0.527722 | 24.115584 | std | 50.425804 | 98.425804 | 54.425521 | 0.155152 | 8.328811 | 3.068028 | 19052.809848 | 17.309420 | 5.443946 | 0.702704 | 10.219353 | 12.197885 | 0.190741 | 14.631457 | 0.224238 | 10.511501 | min | 0.000000 | 1.000000 | 1.000000 | 0.352000 | 48.900000 | 1.400000 | 294.000000 | -79.000000 | -35.000000 | -1.550000 | 5.300000 | 2.600000 | 0.253000 | 1.400000 | 0.134000 | 8.900000 | 25% | 48.500000 | 46.500000 | 47.750000 | 0.577750 | 65.550000 | 6.075000 | 3527.000000 | -7.900000 | -2.000000 | 0.425000 | 10.860000 | 6.700000 | 0.524000 | 6.575000 | 0.324000 | 16.300000 | 50% | 87.900000 | 98.000000 | 94.500000 | 0.729500 | 73.300000 | 8.650000 | 10361.000000 | 4.900000 | 0.000000 | 0.760000 | 26.200000 | 13.600000 | 0.707000 | 17.490000 | 0.527000 | 21.300000 | 75% | 145.500000 | 146.500000 | 139.750000 | 0.811250 | 77.900000 | 11.125000 | 23384.000000 | 13.000000 | 2.000000 | 1.150000 | 28.500000 | 20.100000 | 0.813000 | 34.525000 | 0.732250 | 29.475000 | max | 194.900000 | 195.000000 | 195.000000 | 0.948000 | 84.200000 | 13.400000 | 129615.000000 | 40.000000 | 15.000000 | 3.630000 | 44.800000 | 40.200000 | 0.959000 | 48.300000 | 0.921000 | 58.400000 |
| | | Unnamed: 0 | Id | HDI Rank | HDI | Life expectancy | Mean years of schooling | Gross national income (GNI) per capita | GNI per capita minus HDI rank | Change in HDI rank 2010-2015 | Average annual HDI growth 1990-2000 | Coefficient of human inequality | Inequality in life expectancy (N) 2010-2015 | Inequality-adjusted life expectancy index | Inequality in education(N) | Inequality-adjusted education index | Inequality in income (N) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | count | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | mean | 87.900000 | 98.000000 | 94.319148 | 0.668835 | 71.267368 | 9.371277 | 17239.502591 | 2.239362 | -0.021277 | 0.773611 | 30.669294 | 17.155135 | 0.868151 | 20.278125 | 0.527722 | 24.115584 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | std | 50.425804 | 98.425804 | 54.425521 | 0.155152 | 8.328811 | 3.068028 | 19052.809848 | 17.309420 | 5.443946 | 0.702704 | 10.219353 | 12.197885 | 0.190741 | 14.631457 | 0.224238 | 10.511501 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | min | 0.000000 | 1.000000 | 1.000000 | 0.352000 | 48.900000 | 1.400000 | 294.000000 | -79.000000 | -35.000000 | -1.550000 | 5.300000 | 2.600000 | 0.253000 | 1.400000 | 0.134000 | 8.900000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 25% | 48.500000 | 46.500000 | 47.750000 | 0.577750 | 65.550000 | 6.075000 | 3527.000000 | -7.900000 | -2.000000 | 0.425000 | 10.860000 | 6.700000 | 0.524000 | 6.575000 | 0.324000 | 16.300000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 50% | 87.900000 | 98.000000 | 94.500000 | 0.729500 | 73.300000 | 8.650000 | 10361.000000 | 4.900000 | 0.000000 | 0.760000 | 26.200000 | 13.600000 | 0.707000 | 17.490000 | 0.527000 | 21.300000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 75% | 145.500000 | 146.500000 | 139.750000 | 0.811250 | 77.900000 | 11.125000 | 23384.000000 | 13.000000 | 2.000000 | 1.150000 | 28.500000 | 20.100000 | 0.813000 | 34.525000 | 0.732250 | 29.475000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | max | 194.900000 | 195.000000 | 195.000000 | 0.948000 | 84.200000 | 13.400000 | 129615.000000 | 40.000000 | 15.000000 | 3.630000 | 44.800000 | 40.200000 | 0.959000 | 48.300000 | 0.921000 | 58.400000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Univariate
Analysis

```
sns.lineplot(x='Country',y='HDI',data=data_filled)
```

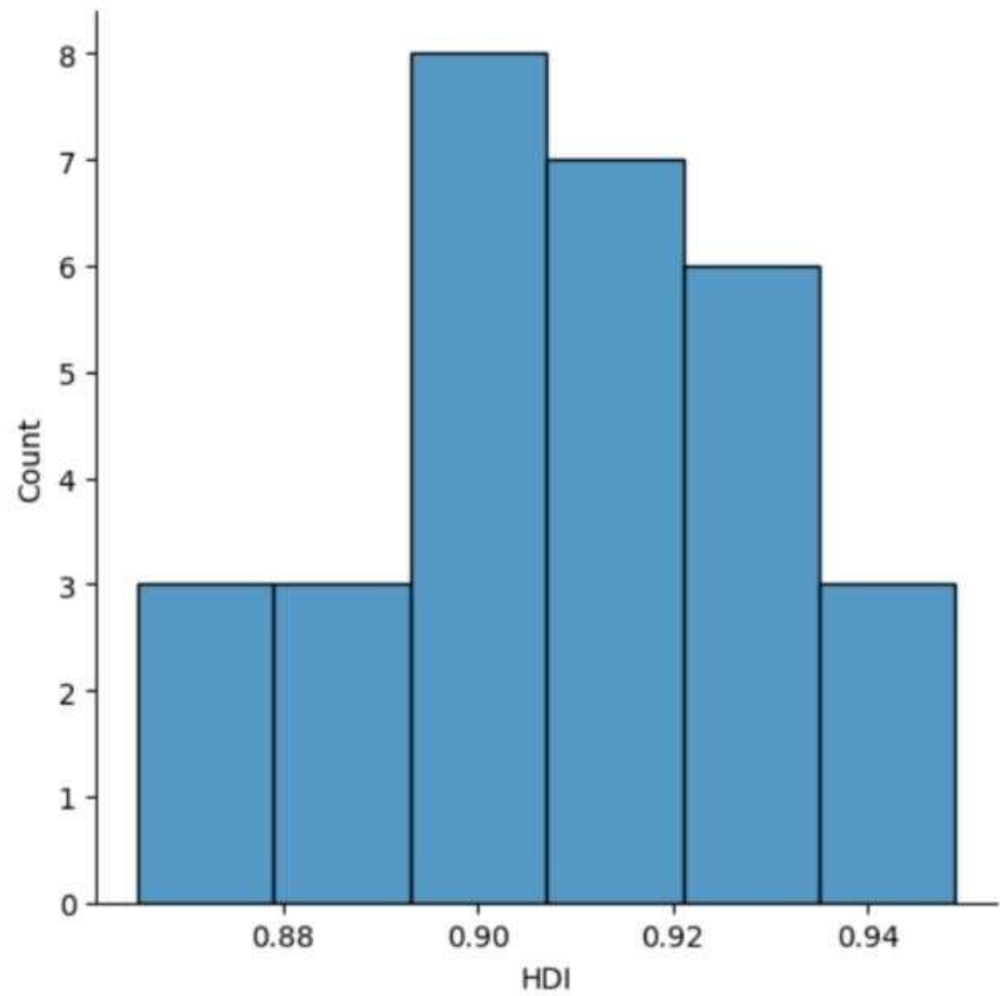
```
<Axes: xlabel='Country', ylabel='HDI'>
```



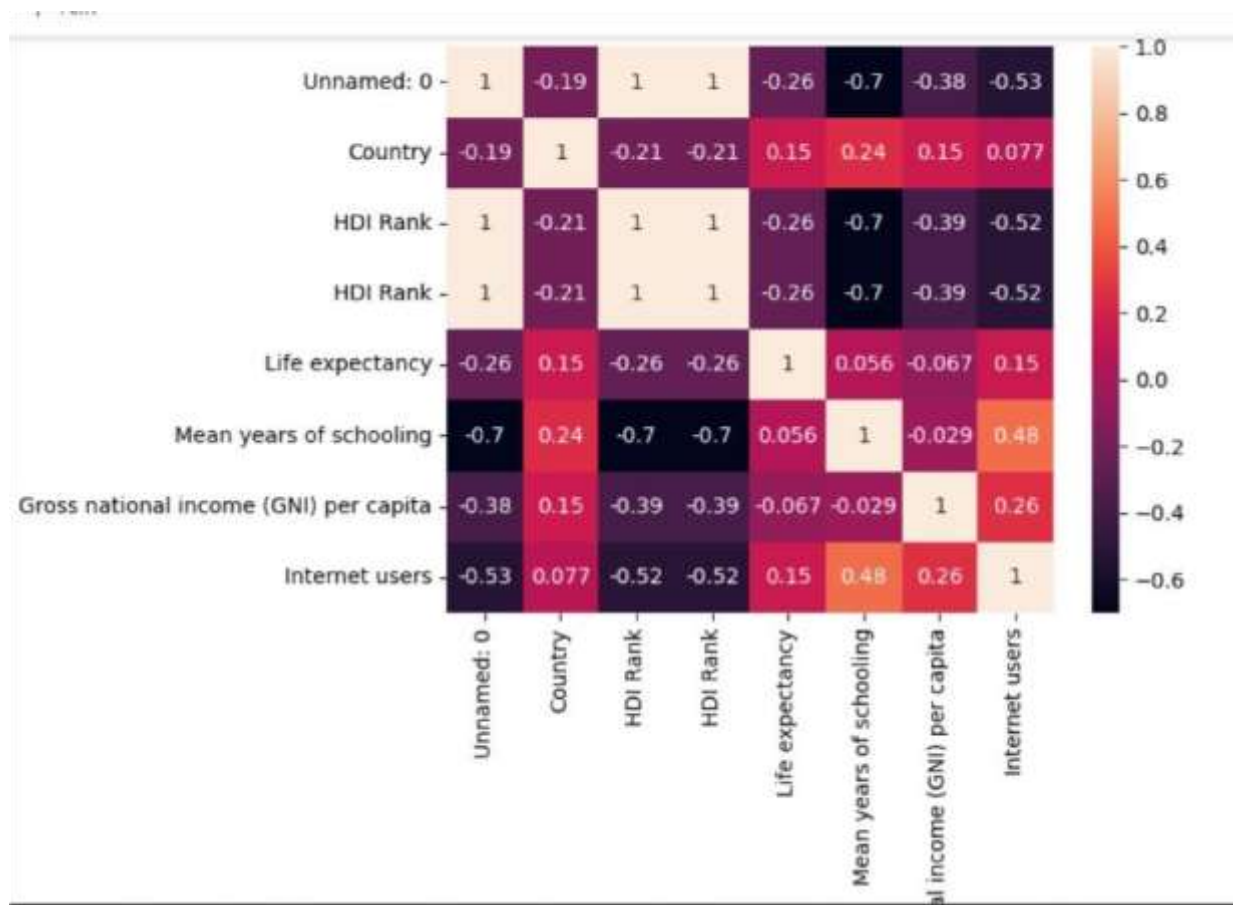


Bivariate
Analysis

<seaborn.axisgrid.FacetGrid at 0x7f8c97cdbfd0>



Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data

```
[4] data=pd.read_csv("../content/HDI.csv")
```

```
[5] data.head()
```

| Unnamed: 0 | Id | Country | HDI Rank | HDI | Life expectancy | Mean years of schooling | Gross national income (GNI) per capita | GNI per capita rank | Change in HDI rank | Coefficient of human inequality | Inequality in life expectancy (%) 2010-2015 | Inequality-adjusted life expectancy index | Inequality in education(X) | Inequality-adjusted education index | Inequality in income (%) | Inequality-adjusted income index | | |
|------------|----|---------|-------------|-----|-----------------|-------------------------|--|---------------------|--------------------|---------------------------------|---|---|----------------------------|-------------------------------------|--------------------------|----------------------------------|------|---|
| 0 | 0 | 1 | Norway | 1.0 | 0.949 | 81.7 | 12.7 | 67014.0 | 5.0 | 0.0 | — | 5.4 | 3.3 | 0.918 | 2.4 | 0.894 | 10.4 | 0 |
| 1 | 1 | 2 | Australia | 2.0 | 0.939 | 82.5 | 13.2 | 42822.0 | 19.0 | 1.0 | — | 8.0 | 4.3 | 0.921 | 1.9 | 0.921 | 17.7 | 0 |
| 2 | 2 | 3 | Switzerland | 2.0 | 0.939 | 83.1 | 13.4 | 56364.0 | 7.0 | 0.0 | — | 8.4 | 3.8 | 0.934 | 5.7 | 0.940 | 15.7 | 0 |
| 3 | 3 | 4 | Germany | 4.0 | 0.926 | 81.1 | 13.2 | 40000.0 | 13.0 | 0.0 | — | 7.0 | 3.7 | 0.905 | 2.6 | 0.881 | 14.8 | 0 |
| 4 | 4 | 5 | Denmark | 5.0 | 0.925 | 80.4 | 12.7 | 44519.0 | 13.0 | 2.0 | — | 7.0 | 3.8 | 0.894 | 3.0 | 0.896 | 14.3 | 0 |

5 rows x 18 columns

5 rows x 18 columns

Handling Missing Data

```
[18] data_filled = data_encoded.fillna(data_encoded.mean())
```

```
data_filled.fillna(data_filled.mean(), inplace=True)
```

| Unnamed: 0 | Id | Country | HDI Rank | HDI | Life expectancy | Mean years of schooling | Gross national income (GNI) per capita | GNI per capita rank | Change in HDI rank | Coefficient of human inequality | Inequality in life expectancy (%) 2010-2015 | Inequality-adjusted life expectancy index | Inequality in education (%) | Inequality-adjusted education index | Inequality in income (%) | Inequality-adjusted income index | Income inequality (Gini ratio) 2010-2015 | Income inequality (Palme ratio) 2010-2015 |
|------------|----|---------|----------|------|-----------------|-------------------------|--|---------------------|--------------------|---------------------------------|---|---|-----------------------------|-------------------------------------|--------------------------|----------------------------------|--|---|
| 0 | 0 | 1 | 22 | 1.0 | 0.949 | 81.7 | 12.7 | 67014.0 | 5.0 | 0.0 | 5.40 | 1.00000 | 0.91800 | 2.40 | 0.8940 | 10.40 | 0.8920 | 3.80000 |
| 1 | 1 | 2 | 8 | 2.0 | 0.939 | 82.5 | 13.2 | 43822.0 | 19.0 | 1.0 | 8.00 | 4.00000 | 0.92100 | 1.90 | 0.9210 | 17.70 | 0.7530 | 6.00000 |
| 2 | 2 | 3 | 27 | 2.0 | 0.939 | 83.1 | 13.4 | 56364.0 | 7.0 | 0.0 | 8.40 | 1.00000 | 0.93400 | 5.70 | 0.9430 | 15.70 | 0.8860 | 4.90000 |
| 3 | 3 | 4 | 9 | 4.0 | 0.926 | 81.1 | 13.2 | 40000.0 | 13.0 | 0.0 | 7.00 | 0.70000 | 0.90500 | 2.60 | 0.8810 | 14.80 | 0.7070 | 4.90000 |
| 4 | 4 | 5 | 8 | 5.0 | 0.925 | 80.4 | 12.7 | 44519.0 | 13.0 | 2.0 | 7.00 | 1.00000 | 0.89400 | 3.00 | 0.8960 | 14.30 | 0.7990 | 4.00000 |
| 5 | 5 | 6 | 21 | 3.0 | 0.925 | 82.2 | 11.6 | 39152.0 | 10 | 0.0 | 6.88 | 0.00000 | 0.94300 | 6.124 | 0.9194 | 16.184 | 0.79008 | 5.46523 |
| 6 | 6 | 7 | 28 | 7.0 | 0.904 | 81.7 | 11.8 | 40326.0 | 8.0 | 2.0 | 6.88 | 1.70000 | 0.91400 | 4.200 | 0.8940 | 12.430 | 0.81230 | 4.30000 |
| 7 | 7 | 8 | 12 | 8.0 | 0.923 | 81.9 | 12.3 | 43796.0 | 11.0 | 1.0 | 7.70 | 0.70000 | 0.90500 | 3.900 | 0.8830 | 16.300 | 0.79960 | 5.30000 |
| 8 | 8 | 9 | 12 | 9.0 | 0.921 | 82.7 | 12.2 | 37080.0 | 20.0 | 7.0 | 5.70 | 2.80000 | 0.92700 | 2.300 | 0.8940 | 11.700 | 0.79900 | 4.00000 |
| 9 | 9 | 10 | 4 | 10.0 | 0.920 | 82.2 | 13.1 | 42582.0 | 12.0 | 1.0 | 6.70 | 4.70000 | 0.91300 | 3.900 | 0.8900 | 17.400 | 0.75000 | 5.00000 |
| 10 | 10 | 11 | 28 | 10.0 | 0.920 | 79.2 | 13.2 | 50140.0 | 1.0 | 3.0 | 12.80 | 6.10000 | 0.89800 | 5.800 | 0.8930 | 27.000 | 0.64030 | 9.10000 |
| 11 | 11 | 12 | 11 | 12.0 | 0.917 | 84.2 | 11.6 | 54295.0 | 2.0 | 2.0 | 6.88 | 2.80000 | 0.93900 | 6.124 | 0.9194 | 16.184 | 0.79008 | 5.46523 |

Data Transformation

```
# Import pandas as pd
from sklearn.preprocessing import OneHotEncoder

# Identify categorical columns
categorical_columns = data.select_dtypes(include=["object"]).columns

# Apply One-Hot Encoding to categorical columns
data_encoded = pd.get_dummies(data, columns=categorical_columns, drop_first=True)

# Display the first few rows of the encoded dataset
data_encoded.head()
```

| Unnamed: 0 | Id | Country | HDI Rank | HDI | Life expectancy | Mean years of schooling | Gross national income (GNI) per capita | GNI per capita rank | Change in HDI rank | Coefficient of human inequality | Inequality in life expectancy | | Inequality in education | | Inequality in income | | Income inequality (Palme ratio) 2010-2015 | Income inequality (Palme ratio) 2010-2015 | |
|------------|----|---------|----------|-----|-----------------|-------------------------|--|---------------------|--------------------|---------------------------------|---|---|-----------------------------|-------------------------------------|--------------------------|----------------------------------|---|---|-----|
| | | | | | | | | | | | Inequality in life expectancy (%) 2010-2015 | Inequality-adjusted life expectancy index | Inequality in education (%) | Inequality-adjusted education index | Inequality in income (%) | Inequality-adjusted income index | | | |
| 0 | 0 | 1 | 22 | 1.0 | 0.949 | 81.7 | 12.7 | 67014.0 | 5.0 | 0.0 | 5.4 | 3.3 | 0.918 | 2.4 | 0.894 | 10.4 | 0.892 | 3.8 | 2.9 |
| 1 | 1 | 2 | 8 | 2.0 | 0.939 | 82.5 | 13.2 | 42822.0 | 19.0 | 1.0 | 8.0 | 4.3 | 0.921 | 1.9 | 0.921 | 17.7 | 0.753 | 6.0 | 1.8 |
| 2 | 2 | 3 | 27 | 2.0 | 0.939 | 83.1 | 13.4 | 56364.0 | 7.0 | 0.0 | 8.4 | 3.8 | 0.934 | 5.7 | 0.940 | 15.7 | 0.886 | 4.9 | 1.2 |
| 3 | 3 | 4 | 9 | 4.0 | 0.926 | 81.1 | 13.2 | 40000.0 | 13.0 | 0.0 | 7.0 | 3.7 | 0.905 | 2.6 | 0.881 | 14.8 | 0.707 | 4.9 | 1.8 |
| 4 | 4 | 5 | 8 | 5.0 | 0.925 | 80.4 | 12.7 | 44519.0 | 13.0 | 2.0 | 7.0 | 3.8 | 0.894 | 3.0 | 0.896 | 14.3 | 0.796 | 4.0 | 1.0 |

5 rows x 20 columns

5 rows x 19 columns

| | |
|---------------------|---|
| Feature Engineering | Attached the codes in final submission. |
| Save Processed Data | - |