**Executive Summary**

This project focuses on the development of a country through various actors covered in Social Progress Index. Social Progress Index contains various categories through which we can judge a countries growth. Categories covered in Social Progress Index are as follows:

1. Nutrition and Basic Medical Care
2. Water and Sanitation
3. Shelter
4. Personal Safety
5. Access to basic knowledge
6. Access to Information
7. Health and Wellness
8. Environmental Quality
9. Personal Rights
10. People Freedom Choice
11. Tolerance and Inclusion
12. Access to advanced Education

As you can see, Social Progress Index covers all the areas with detailed information and authenticities. Our goal is to make sense out from this data and we do that by providing different functions to our sample data. Descriptive Analysis, Regression statistics and Linear Regression showed remarkable results on our sample data. The summary of data was well calculated by descriptive analysis. Regression statistics should always be less than 0.5 and it was less than 0.5 in almost all the cases. Last but not the least Linear Regression is the best tool for future prediction and analysis and it results were somewhat accurate in both the cases. These are just few mathematical ways through which we make our data meaningful.

1. **Introduction**

Sustainable development of a country is the organizing principle for meeting the goals of its citizen while sustaining the resources. In order to sustain the resources of development, we make use of an index called as Social Progress Index. Social progress index tells us about the society, its requirements and its consumptions. This index acts as a building block for a better idea about the society and helps us in sustaining the quality of life. For this project, we used the social progress index from 2017. Our data has 50 variables with 12 different categories of 182 countries. These categories are further divided into different parameters to cover the whole society in different ways. Our aim is to work on a sample of this huge dataset and make meaning out from it. To get the meaning out from a data includes different analysis like descriptive analysis, linear regression and correlating that gives us the idea for a measures that should be taken for a productive future.

1. **Descriptive Analysis**

Descriptive Analysis is an abstract of a given data. The data is usually in such huge numbers that we work on the averages of these dataset. The descriptive analysis provides us with a summary that includes the following parameters:-

1. Mean
2. Standard Error
3. Median
4. Mode
5. Standard Deviation
6. Sample variance
7. Kurtosis
8. Skewness
9. Range
10. Minimum
11. Maximum
12. Sum
13. Count

These 13 parameters are very essential to define a dataset and can be used in the analysis of the dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| Access to piped water |  | Levels of violent crime |  |
| *Parameters* | *Values* | *Parameters* | *Values* |
|  |  |  |  |
| Mean | 61.3 | Mean | 2.8 |
| Standard Error | 4.3 | Standard Error | 0.1 |
| Median | 73.6 | Median | 3.0 |
| Mode | 100.0 | Mode | 3.0 |
| Standard Deviation | 36.8 | Standard Deviation | 1.1 |
| Sample Variance | 1352.2 | Sample Variance | 1.2 |
| Kurtosis | -1.5 | Kurtosis | -0.4 |
| Skewness | -0.4 | Skewness | 0.3 |
| Range | 97.6 | Range | 4.0 |
| Minimum | 2.4 | Minimum | 1.0 |
| Maximum | 100.0 | Maximum | 5.0 |
| Sum | 4414.9 | Sum | 201.5 |
| Count | 72.0 | Count | 72.0 |
| Confidence Level(95.0%) | 8.6 | Confidence Level(95.0%) | 0.3 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Press Freedom Index* |  | *Wastewater Treatment* |  |
| *Parameters* | *Values* | *Parameters* | *Values* |
|  |  |  |  |
| Mean | 32.7 | Mean | 23.2 |
| Standard Error | 1.6 | Standard Error | 3.8 |
| Median | 30.1 | Median | 5.3 |
| Mode | #N/A | Mode | 0.0 |
| Standard Deviation | 13.9 | Standard Deviation | 32.4 |
| Sample Variance | 193.3 | Sample Variance | 1046.7 |
| Kurtosis | 0.4 | Kurtosis | 0.1 |
| Skewness | 0.6 | Skewness | 1.3 |
| Range | 63.0 | Range | 99.3 |
| Minimum | 8.6 | Minimum | 0.0 |
| Maximum | 71.6 | Maximum | 99.3 |
| Sum | 2356.2 | Sum | 1668.4 |
| Count | 72.0 | Count | 72.0 |
| Confidence Level(95.0%) | 3.3 | Confidence Level(95.0%) | 7.6 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Satisfied demand for contraception* |  | *Inequality in the attainment of education* |  |
| *Parameters* | *Values* | *Parameters* | *Values* |
|  |  |  |  |
| Mean | 64.6 | Mean | 0.201165574 |
| Standard Error | 2.3 | Standard Error | 0.017714541 |
| Median | 70.8 | Median | 0.164075 |
| Mode | 83.1 | Mode | #N/A |
| Standard Deviation | 19.5 | Standard Deviation | 0.150312863 |
| Sample Variance | 380.2 | Sample Variance | 0.022593957 |
| Kurtosis | -0.8 | Kurtosis | -1.27876081 |
| Skewness | -0.7 | Skewness | 0.398188147 |
| Range | 71.2 | Range | 0.4823155 |
| Minimum | 19.4 | Minimum | 0.0139045 |
| Maximum | 90.6 | Maximum | 0.49622 |
| Sum | 4647.9 | Sum | 14.4839213 |
| Count | 72.0 | Count | 72 |
| Confidence Level(95.0%) | 4.6 | Confidence Level(95.0%) | 0.035321791 |

1. **Confidence Interval**

Confidence Interval with 95%

Water and Sanitation

|  |  |  |  |
| --- | --- | --- | --- |
| **95% Confidence Interval** |  | **Lower** | **Higher** |
| Mean | 61.31807933 | 52.67712355 | 69.9590351 |
| Confidence Level(95.0%) | 8.640955774 |  |  |

Access to Advanced Education

|  |  |  |  |
| --- | --- | --- | --- |
| **95% Confidence Level** |  | **Lower** | **Higher** |
| Mean | 0.201165574 | 0.165843783 | 0.236487365 |
| Confidence Level(95.0%) | 0.035321791 |  |  |

**4) Hypothesis**

1. **Access to Advanced Education is higher among European countries than African countries**

|  |  |  |
| --- | --- | --- |
| ***Column1*** | ***Europe*** | ***Africa*** |
| **Mean** | **0.050522** | **0.350539** |
| **Variance** | **0.001045** | **0.009459** |
| **Observations** | **20** | **20** |
| **Hypothesized Mean Difference** | **0** |  |
| **df** | **23** |  |
| **t Stat** | **-13.09133** |  |
| **P(T<=t) one-tail** | **1.91E-12** |  |
| **t Critical one-tail** | **1.713872** |  |
| **P(T<=t) two-tail** | **3.82E-12** |  |
| **t Critical two-tail** | **2.068658** |  |

1. **The difference in terms of Personal Safety between Asian and American countries**

|  |  |  |
| --- | --- | --- |
| ***Parameters*** | ***Asian*** | ***American*** |
| **Mean** | **2.705882353** | **3.428571429** |
| **Variance** | **1.376838235** | **1.340659341** |
| **Observations** | **17** | **14** |
| **Hypothesized Mean Difference** | **0** |  |
| **df** | **28** |  |
| **t Stat** | **-1.718974706** |  |
| **P(T<=t) one-tail** | **0.048328996** |  |
| **t Critical one-tail** | **1.701130934** |  |
| **P(T<=t) two-tail** | **0.096657992** |  |
| **t Critical two-tail** | **2.048407142** |  |

1. **difference in terms of Environmental Quality between European and American countries**

|  |  |  |
| --- | --- | --- |
| ***Parameters*** | ***Europe*** | ***American*** |
| **Mean** | **0.050521525** | **3.428571429** |
| **Variance** | **0.001045332** | **1.340659341** |
| **Observations** | **20** | **14** |
| **Hypothesized Mean Difference** | **0** |  |
| **df** | **13** |  |
| **t Stat** | **-10.9132065** |  |
| **P(T<=t) one-tail** | **3.23432E-08** |  |
| **t Critical one-tail** | **1.770933396** |  |
| **P(T<=t) two-tail** | **6.46864E-08** |  |
| **t Critical two-tail** | **2.160368656** |  |

1. **Correlation and Regression**

**Regression: Environmental Quality variable vs. Water Sanitation variable**

|  |  |
| --- | --- |
| **SUMMARY OUTPUT** |  |
| ***Regression Statistics*** |  |
| **Multiple R** | 0.100898195 |
| **R Square** | 0.010180446 |
| **Adjusted R Square** | -0.003959834 |
| **Standard Error** | 13.93133177 |
| **Observations** | 72 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ***df*** | ***SS*** | ***MS*** | ***F*** | ***Significance F*** |
| **Regression** | **1** | **139.7** | **139.7** | **0.7** | **0.4** |
| **Residual** | **70** | **13585.7** | **194.1** |  |  |
| **Total** | **71** | **13725.5** |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **37.37** | **5.71** | **6.54** | **0.00** | **25.97** | **48.77** | **25.97** | **48.77** |
| **Satisfied demand for contraception** | **-0.07** | **0.08** | **-0.85** | **0.40** | **-0.24** | **0.10** | **-0.24** | **0.10** |

**Correlation**

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Correlation | -0.1008982 |

**Regression: Access to Information and communication vs. Personal freedom and choice**

|  |  |
| --- | --- |
| ***Regression Statistics*** |  |
| Multiple R | 0.65 |
| R Square | 0.43 |
| Adjusted R Square | 0.42 |
| Standard Error | 28.01 |
| Observations | 72 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Parameters*** | ***df*** | ***SS*** | ***MS*** | ***F*** | ***Significance F*** |
| Regression | 1 | 41081.98 | 41081.98 | 52.36 | 0.00 |
| Residual | 70 | 54921.89 | 784.60 |  |  |
| Total | 71 | 96003.88 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Parameters*** | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** |
| Intercept | 44.09 | 4.07 | 10.83 | 0.00 | 35.97 | 52.21 | 35.97 |
| Wastewater treatment | 0.74 | 0.10 | 7.24 | 0.00 | 0.54 | 0.95 | 0.54 |

**Correlation**

|  |  |
| --- | --- |
| **Parameter** | **Values** |
| C**orrelation** | **0.65415598** |

1. **Conclusions and limitation**

From freedom to wellness, the Social Progress Index covers almost all the aspects of a society and helps us take better measure for future. For this particular dataset, the descriptive analysis is a really good way of having the gist of the whole data. Averages are the building blocks of any analysis and descriptive analysis helps us with just a click. Further, we randomly selected different variables from categories 2 and 12 and calculated the confidence interval. This interval tells us about 95% true meaning of a population. Hypothesis testing also helped us in a better understanding and the plausibility of our hypothesis. After calculated the given data, we made a linear regression model with line that best fit our data. Linear regression is probably the best way in predictive analysis. The regression should always be calculated below 0.5 and it came out less than 0.5 in both the cases which is a good observation. With this, we also calculated the correlations between the variables that gave us a different idea about the variables and its relations.

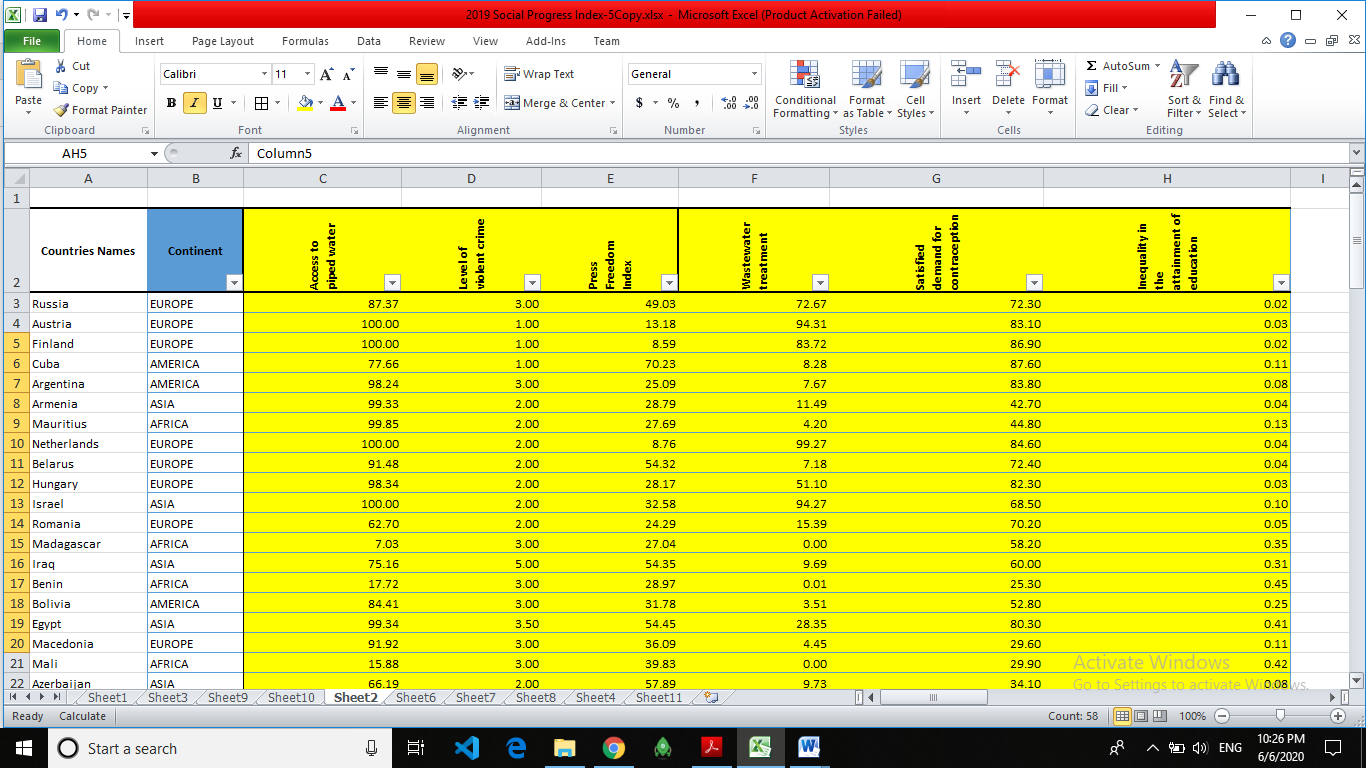
Working with such diverse variables and vast categories about a county was very intuitive and interesting but the only drawback I felt was that the selection of the data. With such huge amounts of data we tend to randomly select our sample which may lead in loss in accuracy. Rest the results were quite interesting to find.

Reference list

**Appendices**

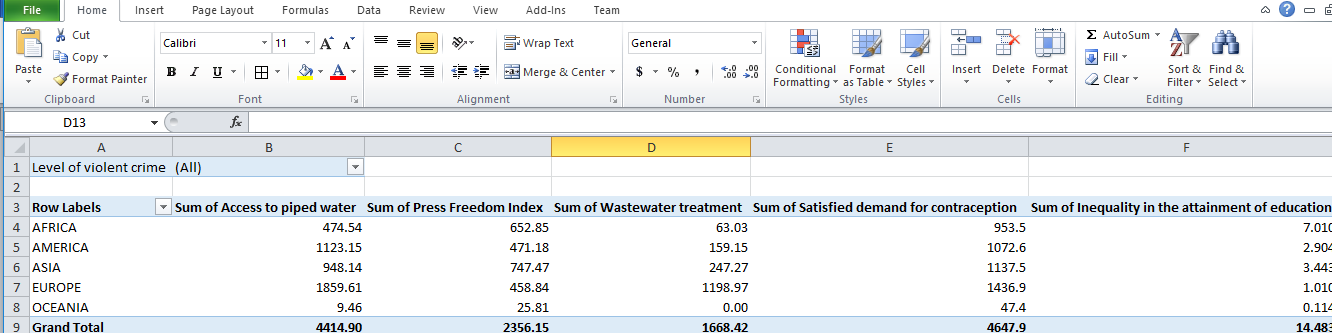
1. **Random data**

Random selection of data.



Bar representation of data

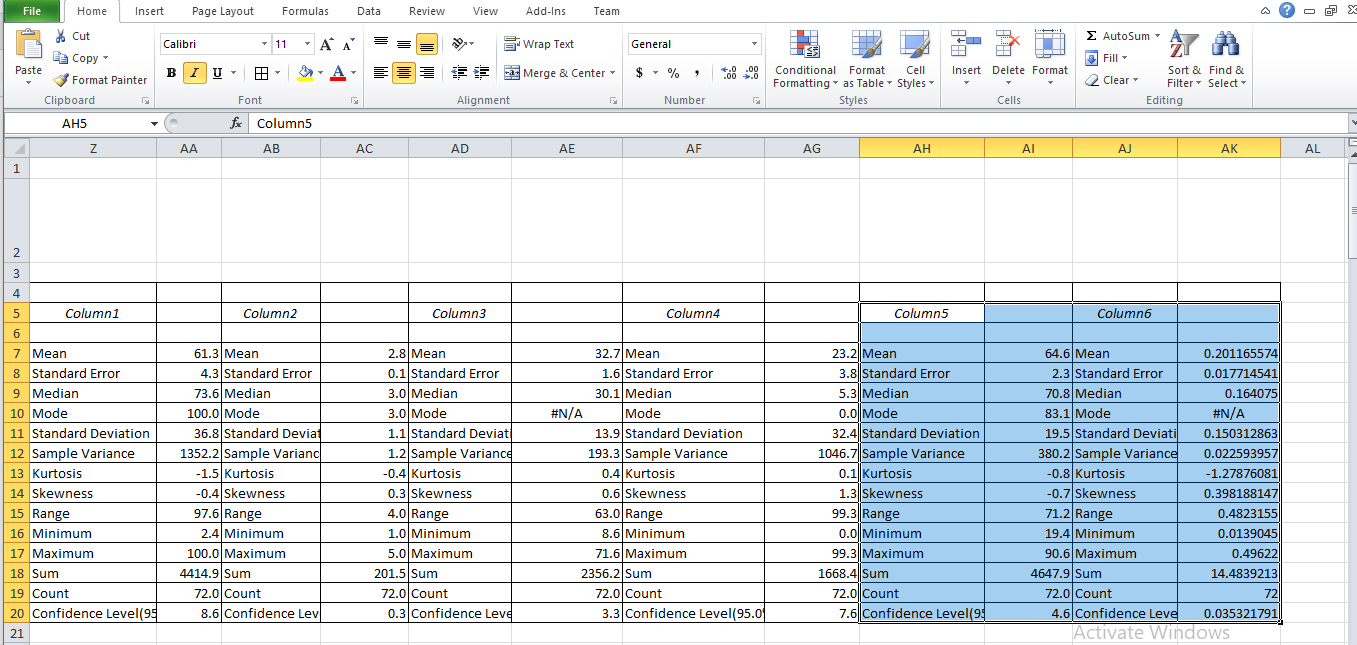
**Pivot table of selected random data**

****

**Bar representation using pivot tables**

**Pie chart representation using pivot tables**

**Descriptive analysis done in section 2**

****

**Data used for Hypothesis in section 4**

1. **Access to Advanced Education is higher among European countries than African countries**

|  |  |
| --- | --- |
| Europe | Africa |
| 0.02 | 0.13 |
| 0.03 | 0.35 |
| 0.02 | 0.45 |
| 0.04 | 0.42 |
| 0.04 | 0.37 |
| 0.03 | 0.42 |
| 0.05 | 0.26 |
| 0.11 | 0.41 |
| 0.02 | 0.50 |
| 0.05 | 0.28 |
| 0.12 | 0.24 |
| 0.06 | 0.22 |
| 0.05 | 0.35 |
| 0.03 | 0.46 |
| 0.02 | 0.22 |
| 0.07 | 0.46 |
| 0.03 | 0.35 |
| 0.08 | 0.35 |
| 0.12 | 0.40 |
| 0.01 | 0.39 |

1. **The difference in terms of Personal Safety between Asian and American countries**

|  |  |
| --- | --- |
| Asian | American |
| 2.00 | 1.00 |
| 2.00 | 3.00 |
| 5.00 | 3.00 |
| 3.50 | 2.00 |
| 2.00 | 5.00 |
| 1.00 | 3.00 |
| 5.00 | 4.00 |
| 3.00 | 5.00 |
| 1.00 | 3.00 |
| 2.00 | 4.00 |
| 3.00 | 3.00 |
| 3.00 | 4.00 |
| 3.00 | 5.00 |
| 2.00 | 3.00 |
| 4.00 |  |
| 2.50 |  |
| 2.00 |  |

1. **The difference in terms of Environmental Quality between European and American countries**

|  |  |
| --- | --- |
| Europe | American |
| 0.02 | 1.00 |
| 0.03 | 3.00 |
| 0.02 | 3.00 |
| 0.04 | 2.00 |
| 0.04 | 5.00 |
| 0.03 | 3.00 |
| 0.05 | 4.00 |
| 0.11 | 5.00 |
| 0.02 | 3.00 |
| 0.05 | 4.00 |
| 0.12 | 3.00 |
| 0.06 | 4.00 |
| 0.05 | 5.00 |
| 0.03 | 3.00 |
| 0.02 |  |
| 0.07 |  |
| 0.03 |  |
| 0.08 |  |
| 0.12 |  |
| 0.01 |  |

**Data used for Correlation and Regression for section 5.**

|  |  |  |
| --- | --- | --- |
| **Continent** | **Access to piped water** | **Wastewater treatment** |
| EUROPE | 87.37 | 72.67 |
| EUROPE | 100.00 | 94.31 |
| EUROPE | 100.00 | 83.72 |
| AMERICA | 77.66 | 8.28 |
| AMERICA | 98.24 | 7.67 |
| ASIA | 99.33 | 11.49 |
| AFRICA | 99.85 | 4.20 |
| EUROPE | 100.00 | 99.27 |
| EUROPE | 91.48 | 7.18 |
| EUROPE | 98.34 | 51.10 |
| ASIA | 100.00 | 94.27 |
| EUROPE | 62.70 | 15.39 |
| AFRICA | 7.03 | 0.00 |
| ASIA | 75.16 | 9.69 |
| AFRICA | 17.72 | 0.01 |
| AMERICA | 84.41 | 3.51 |
| ASIA | 99.34 | 28.35 |
| EUROPE | 91.92 | 4.45 |
| AFRICA | 15.88 | 0.00 |
| ASIA | 66.19 | 9.73 |
| EUROPE | 96.84 | 71.86 |
| ASIA | 58.19 | 0.00 |
| AFRICA | 6.99 | 0.00 |
| AMERICA | 97.49 | 3.89 |
| AMERICA | 84.63 | 6.80 |
| ASIA | 12.17 | 0.00 |
| AFRICA | 6.42 | 0.00 |
| AMERICA | 72.13 | 40.42 |
| EUROPE | 100.00 | 94.52 |
| EUROPE | 100.00 | 77.62 |
| AFRICA | 21.69 | 3.20 |
| AFRICA | 32.63 | 0.00 |
| EUROPE | 98.22 | 57.18 |
| EUROPE | 97.88 | 51.11 |
| AFRICA | 5.36 | 0.00 |
| AFRICA | 12.59 | 1.44 |
| AFRICA | 22.27 | 0.30 |
| AFRICA | 15.75 | 4.20 |
| ASIA | 24.35 | 3.30 |
| AFRICA | 8.72 | 0.00 |
| AFRICA | 63.85 | 5.40 |
| ASIA | 98.45 | 56.53 |
| ASIA | 28.37 | 0.00 |
| AMERICA | 87.85 | 12.24 |
| ASIA | 47.37 | 0.00 |
| AFRICA | 24.77 | 0.18 |
| ASIA | 56.57 | 11.52 |
| AMERICA | 92.33 | 39.72 |
| ASIA | 58.19 | 4.20 |
| ASIA | 61.37 | 15.60 |
| EUROPE | 98.97 | 59.78 |
| EUROPE | 100.00 | 95.30 |
| AFRICA | 2.42 | 0.00 |
| EUROPE | 53.93 | 14.00 |
| AFRICA | 82.01 | 44.10 |
| AMERICA | 66.32 | 0.00 |
| ASIA | 43.13 | 2.58 |
| EUROPE | 100.00 | 91.60 |
| AMERICA | 93.70 | 17.56 |
| AFRICA | 14.84 | 0.00 |
| EUROPE | 100.00 | 89.78 |
| OCEANIA | 9.46 | 0.00 |
| AMERICA | 9.80 | 0.00 |
| AFRICA | 5.75 | 0.00 |
| ASIA | 8.16 | 0.00 |
| AMERICA | 82.94 | 0.74 |
| EUROPE | 81.97 | 5.14 |
| EUROPE | 100.00 | 63.01 |
| AFRICA | 7.99 | 0.00 |
| ASIA | 11.78 | 0.00 |
| AMERICA | 83.52 | 5.00 |
| AMERICA | 92.13 | 13.32 |

|  |  |  |
| --- | --- | --- |
| **Continent** | **Satisfied demand for contraception** | **Press Freedom Index** |
| EUROPE | 72.30 | 49.03 |
| EUROPE | 83.10 | 13.18 |
| EUROPE | 86.90 | 8.59 |
| AMERICA | 87.60 | 70.23 |
| AMERICA | 83.80 | 25.09 |
| ASIA | 42.70 | 28.79 |
| AFRICA | 44.80 | 27.69 |
| EUROPE | 84.60 | 8.76 |
| EUROPE | 72.40 | 54.32 |
| EUROPE | 82.30 | 28.17 |
| ASIA | 68.50 | 32.58 |
| EUROPE | 70.20 | 24.29 |
| AFRICA | 58.20 | 27.04 |
| ASIA | 60.00 | 54.35 |
| AFRICA | 25.30 | 28.97 |
| AMERICA | 52.80 | 31.78 |
| ASIA | 80.30 | 54.45 |
| EUROPE | 29.60 | 36.09 |
| AFRICA | 29.90 | 39.83 |
| ASIA | 34.10 | 57.89 |
| EUROPE | 76.60 | 14.31 |
| ASIA | 83.10 | 30.73 |
| AFRICA | 52.80 | 54.10 |
| AMERICA | 89.90 | 11.10 |
| AMERICA | 66.20 | 38.03 |
| ASIA | 44.90 | 37.75 |
| AFRICA | 19.40 | 40.59 |
| AMERICA | 83.60 | 27.90 |
| EUROPE | 80.30 | 19.92 |
| EUROPE | 59.30 | 30.35 |
| AFRICA | 76.80 | 31.16 |
| AFRICA | 29.80 | 24.03 |
| EUROPE | 63.10 | 23.89 |
| EUROPE | 79.50 | 12.40 |
| AFRICA | 37.80 | 29.94 |
| AFRICA | 54.90 | 28.65 |
| AFRICA | 76.10 | 28.78 |
| AFRICA | 65.90 | 35.08 |
| ASIA | 71.10 | 27.61 |
| AFRICA | 42.90 | 24.62 |
| AFRICA | 76.80 | 42.64 |
| ASIA | 70.40 | 28.67 |
| ASIA | 67.10 | 71.58 |
| AMERICA | 82.80 | 44.11 |
| ASIA | 82.30 | 61.15 |
| AFRICA | 36.20 | 35.84 |
| ASIA | 90.60 | 44.53 |
| AMERICA | 83.20 | 49.33 |
| ASIA | 66.30 | 30.16 |
| ASIA | 75.60 | 54.55 |
| EUROPE | 78.00 | 22.26 |
| EUROPE | 80.60 | 14.80 |
| AFRICA | 38.70 | 30.71 |
| EUROPE | 62.00 | 28.83 |
| AFRICA | 73.00 | 31.60 |
| AMERICA | 87.30 | 28.82 |
| ASIA | 54.70 | 44.66 |
| EUROPE | 80.40 | 8.89 |
| AMERICA | 86.80 | 32.62 |
| AFRICA | 31.60 | 39.89 |
| EUROPE | 86.00 | 14.18 |
| OCEANIA | 47.40 | 25.81 |
| AMERICA | 48.70 | 24.66 |
| AFRICA | 40.40 | 29.03 |
| ASIA | 71.70 | 42.08 |
| AMERICA | 81.60 | 33.63 |
| EUROPE | 26.90 | 29.92 |
| EUROPE | 82.80 | 16.66 |
| AFRICA | 42.20 | 22.66 |
| ASIA | 74.10 | 45.94 |
| AMERICA | 64.40 | 23.29 |
| AMERICA | 73.90 | 30.59 |