Gender Development Index Project

Team Atreides

Samwel Omondi, Laura Abuko

Github link

Notebook link

Tableau link

presentation

- ## Business Understanding

Introduction

The Gender Development Index (GDI) measures gender gaps in human development achievements by accounting for disparities between women and men in three basic dimensions of human development—health, knowledge, and living standards using the same component indicators as in the Human Development Index (HDI). The GDI is the ratio of the HDIs calculated separately for females and males using the same methodology as in the HDI. It is a direct measure of the gender gap showing the female HDI as a percentage of the male HDI.

Problem statement

United Nations launched its sustainable development agenda in 2015, reflecting the growing understanding by Member States that a development model that is sustainable for this and future generations offers the best path forward for reducing poverty and improving the lives of people everywhere.Gender disparity a variable measured by gender development index is still a problem globally with the claim of the index being higher presumed average of 0.943. Team atreides has been contracted to figure out how we can predict future GDI index. As the Analytics team we will help provide an indication of the global status in terms of gender development while providing insights into some of the key dimensional indexes that might drive gender development outcomes.We will be performing hypothesis testing to try to prove or disprove the claim.

Metric for Success

Gender Development Index Project

To determine the gender development index globally for our analysis using various dimension index we will need to obtain the following as our measure of success:

- Find and deal with outliers, anomalies, and missing data within the dataset.
- Comprehensive EDA with aim of univariate, bivariate and multivariate analysis
- Choose the best predictive model with an accuracy score of 70% and above
- Perform hypothesis testing to prove and disprove the stated claim

2. Data Understanding

Data used for analysis was obtained from an open data source kaggle;link.
The provided dataset is an aggregation, by country GDI, HDI, and life expectancy of men and women in 2019. The data has 190 rows and 14 columns
The following shows the information on the dataset:

```
HDI Rank                    189 non-null    float64
 1   Country                    189 non-null    object
 2   GDI_Value                  190 non-null    object
 3   GDI_Group                  190 non-null    object
 4   HDI_Female                 190 non-null    object
 5   HDI_Male                   190 non-null    object
 6   Lif_Expec_Female           190 non-null    object
 7   Lif_Excep_Male             190 non-null    object
 8   Excep_Yrs_Schooling_Female 190 non-null    object
 9   Excep_Yrs_Schooling_Male   190 non-null    object
 10  Mean_Yrs_Schooling_Female  190 non-null    object
 11  Mean_Yrs_Schooling_Male    190 non-null    object
 12  GNI_PC_Female              190 non-null    object
 13  GNI_PC_Male
```

3. Data preparation

Experimental Design

The following steps outline the experimental design;

- Defining the research question
- Finding and dealing with outliers, anomalies, and missing data within the dataset
- Plot appropriate univariate and bivariate summaries recording our observations
- Implementing our solution by performing hypothesis testing
- Challenging the solution by providing insights on how to make improvements

Deriving new attributes

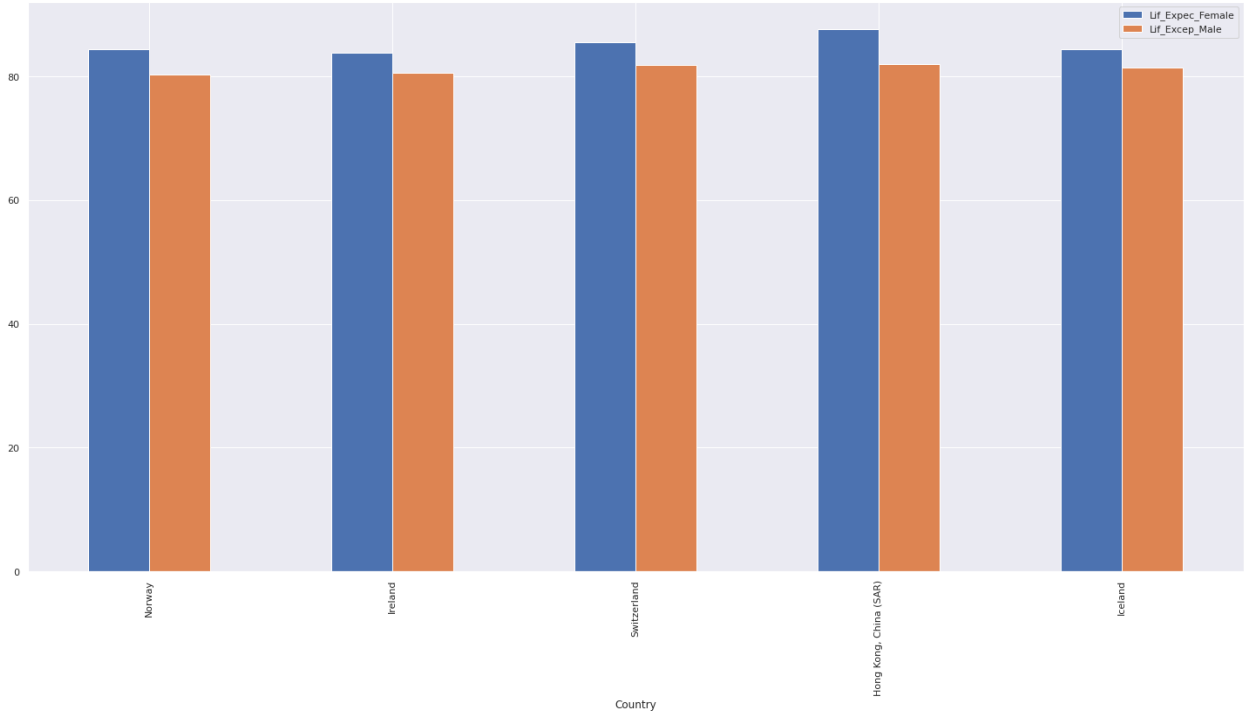
Gender Development Index Project

1. Which countries have the largest and lowest GDI?
2. What are the average years of education a female receives in the top GDI countries vs the least ranked countries?
3. What are the average years of education a male receives in the top vs bottom GDI countries?
4. What is the average gross national income per capita a female receives in the top vs the bottom GDI countries?
5. What is the average gross national income per capita a male receives in the top vs bottom GDI countries?

4. Data Analysis

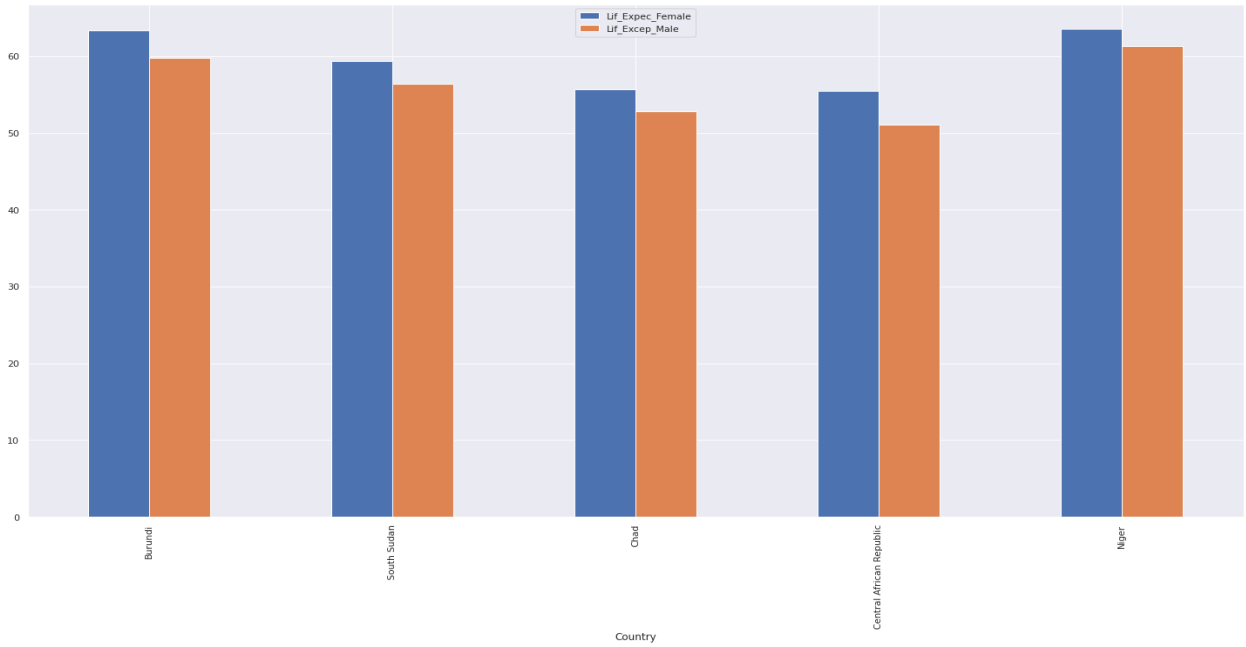
Life Expectancy

The diagram below shows Maximum life expectancy for top countries is 80years with HOng kong having the highest life expectancy.

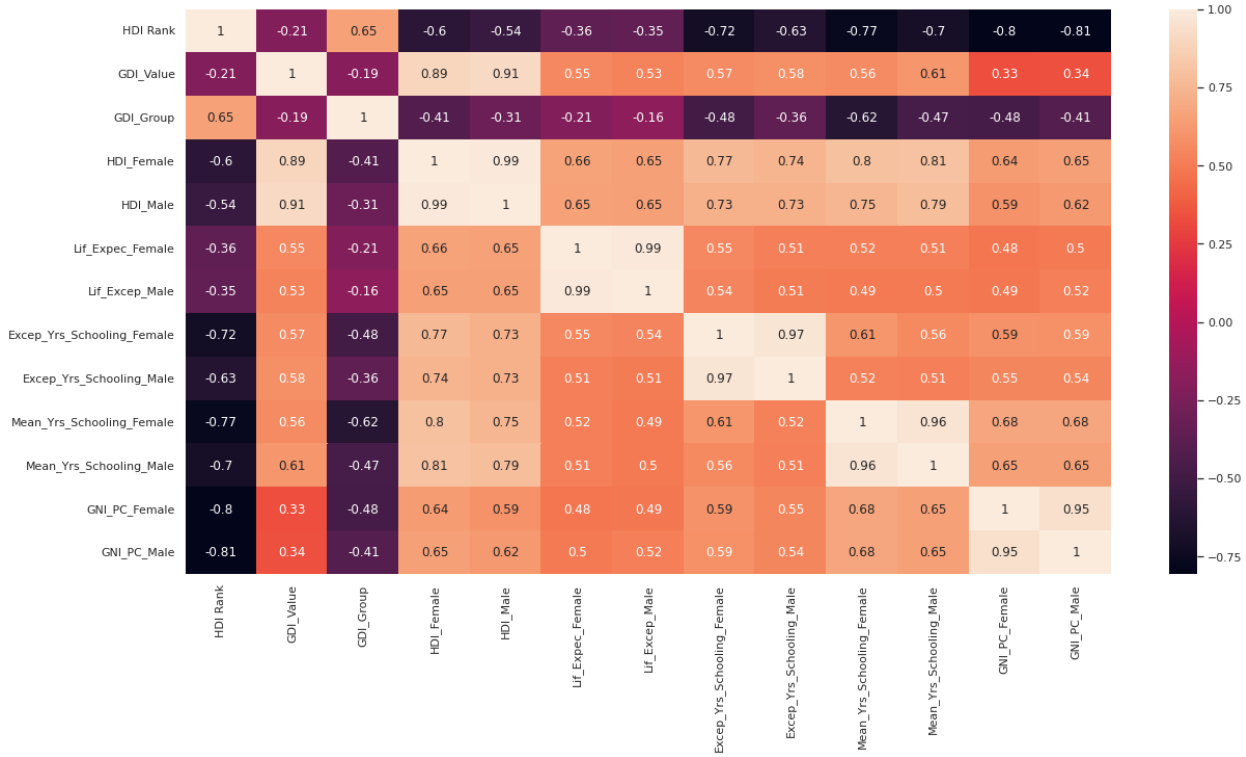


The maximum life expectancy for Bottom countries in the GDI is 60years with Chad and Central African Republic having the lowest life expectancy

Gender Development Index Project



Data correlation



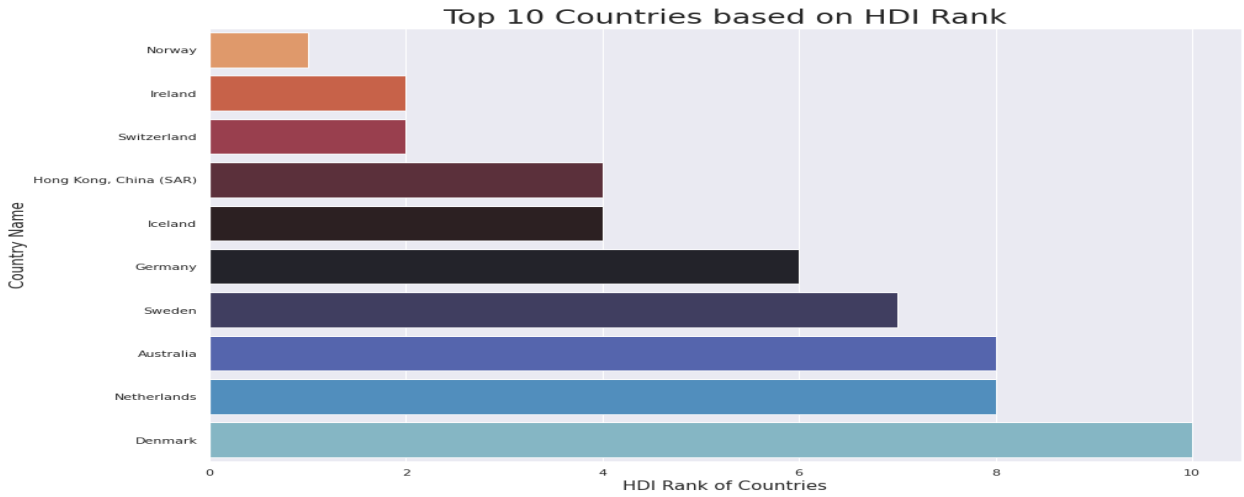
- There is strong correlation(0.99) between HDI_male and HDI female
- There is also a strong correlation between GDI value and HDI for each gender (0.89 corr in female and 0.91 in male).

Gender Development Index Project

- There is also a strong correlation between gross national income in male vs female (0.95)

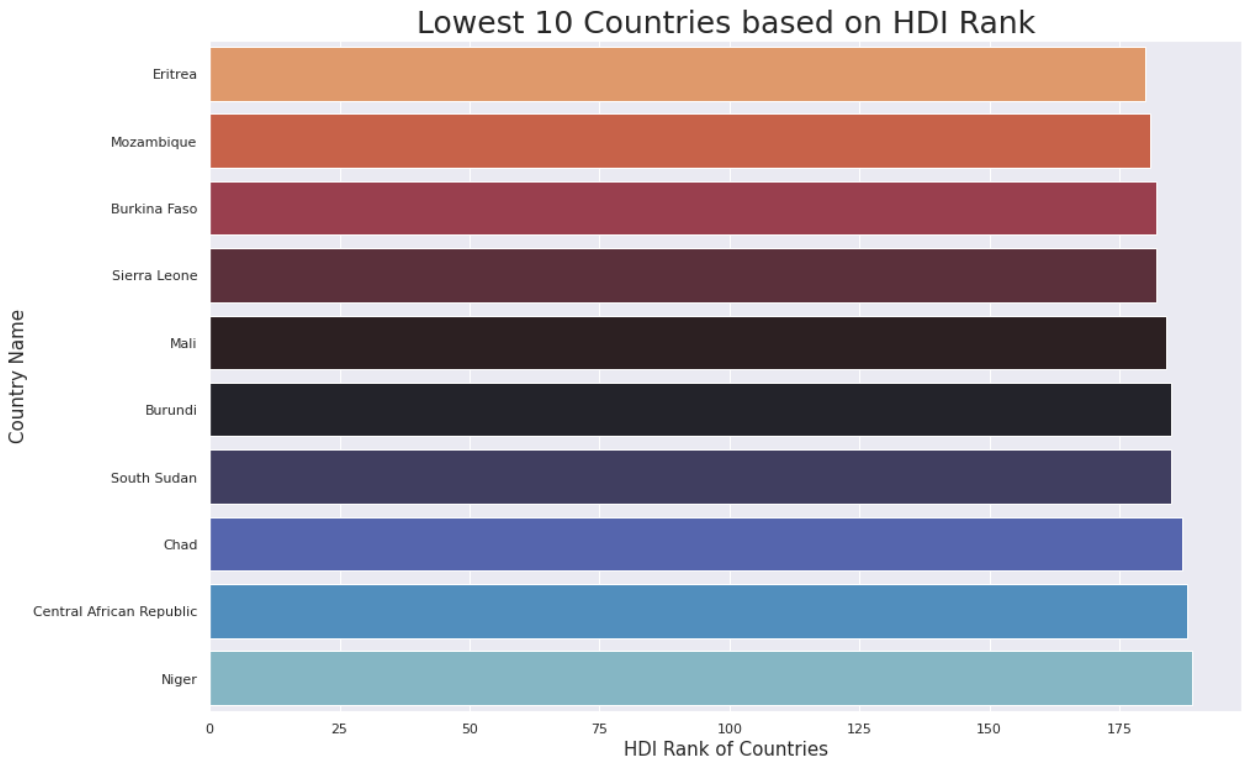
Countries in the GDI

Norway had the highest Human development index in 2019 being closely followed by both ireland and switzerland



Niger had the lowest Human development index ,followed by the central African Republic then chad.

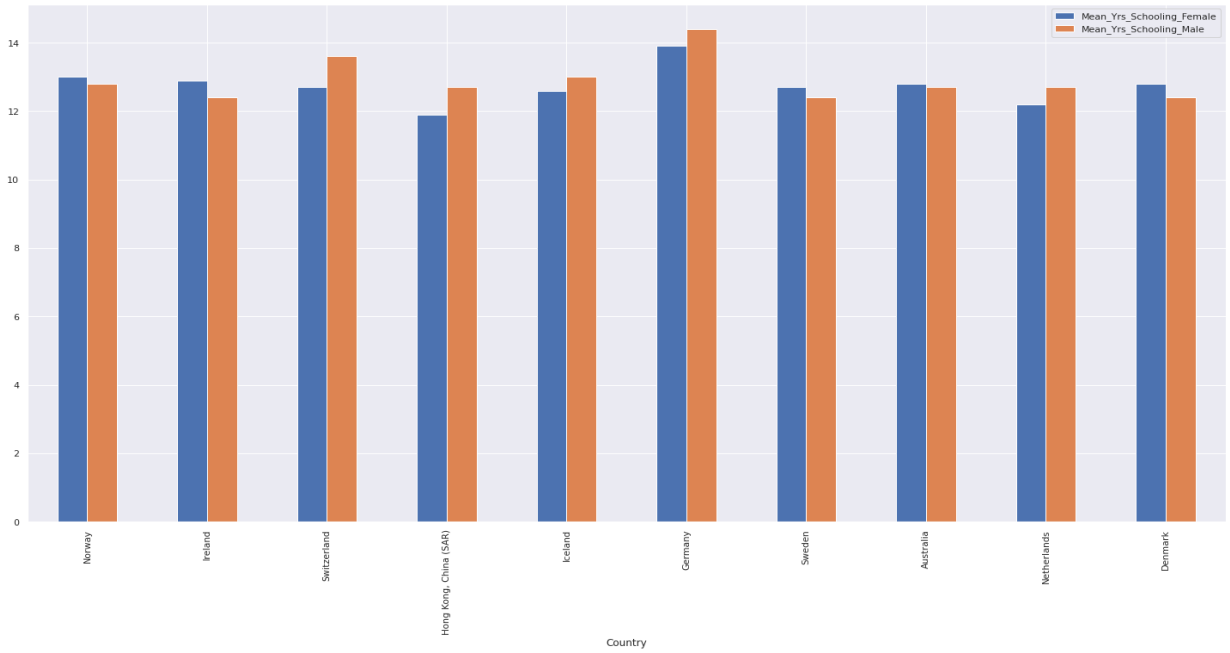
Gender Development Index Project

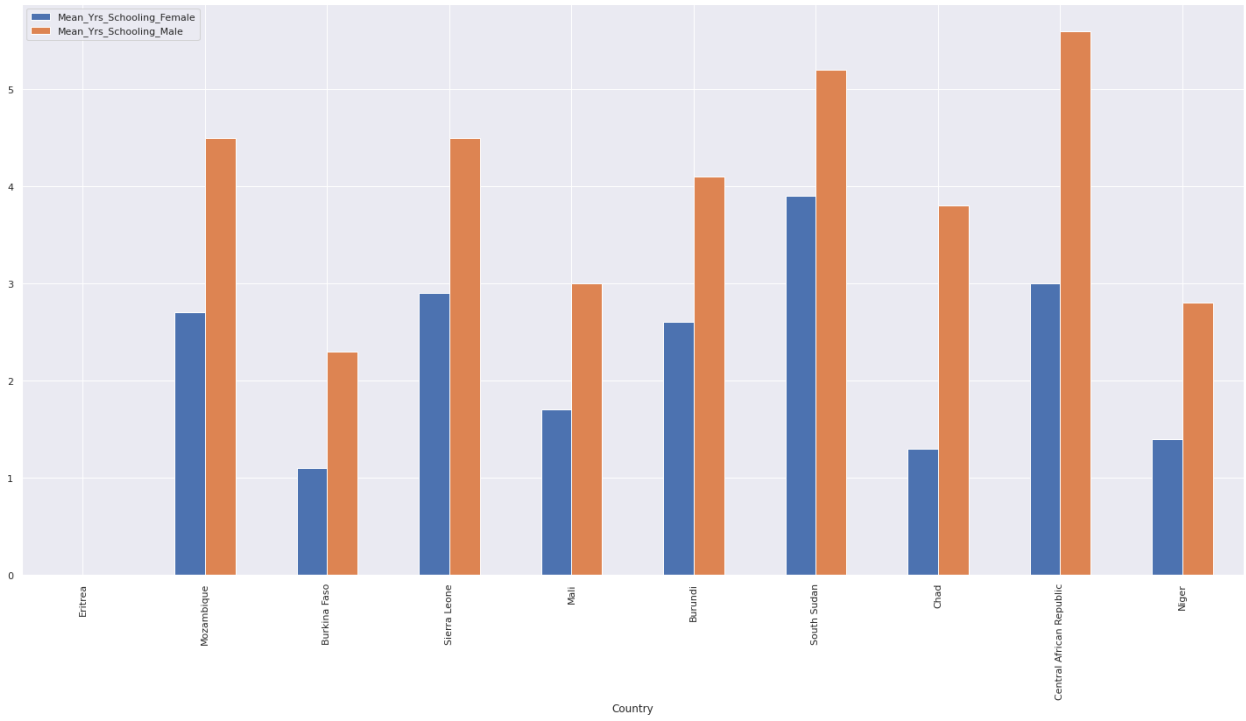


Education

- The average years for which a female receives education in a top HDI ranked country is 12-14, whereas it is a maximum of 3 years in a country ranked bottom 10 on the HDI index.
- The average years for which a male receives education in a top HDI ranked country is 12-14.4, whereas it is a maximum of 5.6 years in a country ranked bottom 10 on the HDI index.

Gender Development Index Project



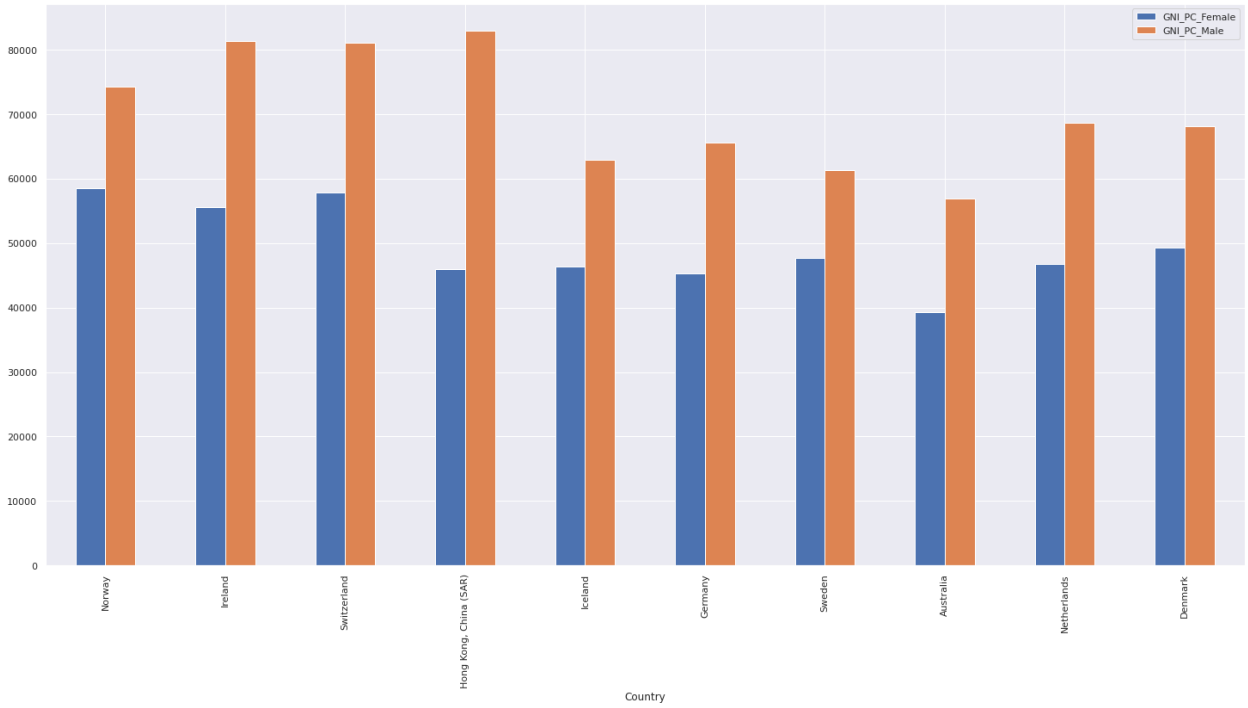


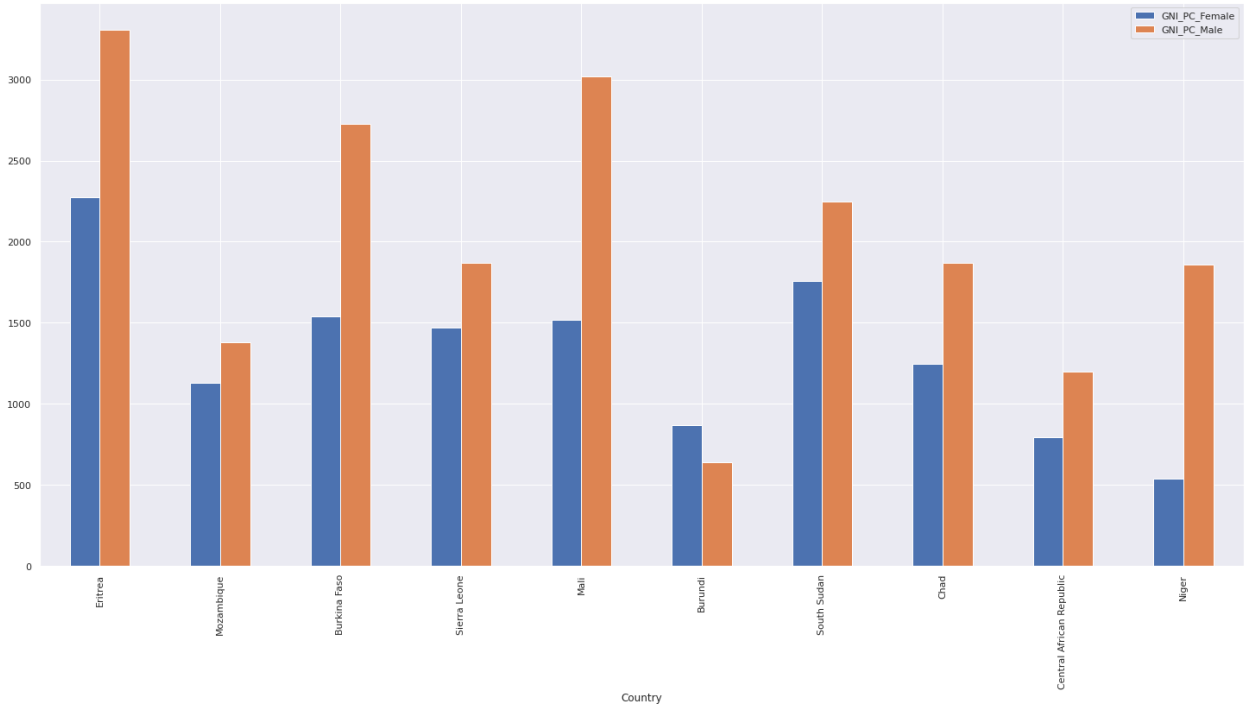
Gross national income

- The average gross national income per capita a female receives in the top 10 HDI ranked countries is between 40k - 55k, whereas it is between 500 - 2k in countries ranked bottom 10 on the HDI index.

Gender Development Index Project

- The average gross national income per capita a Male receives in the top 10 HDI ranked countries is between 55k - 80k, whereas it is between 600 - 3k in countries ranked bottom 10 on the HDI index.





5. Hypothesis Testing

Step 1: Specify Null Hypothesis

Gender Development Index Project

The Average global gender development index equals 0.943

Step 2: Specify Alternative Hypothesis
The Average global gender development index is higher than 0.943

Step 3: Set the Significance Level (a)
Our significance level will be 0.005

Step 4: Defining Sample Size

Yamane's formula sample size
n = N/(1+N(e)2

Where:
n = sample size
e = precision
N = populationsize

n = 189/ ((1 + 189(0.05*0.05))
n = 128

From this, we arrived at a sample of **128**

Step 5: Calculate the Test Statistic and Corresponding P-Value
We used Z-test since our sample satisfied all the assumptions

```
0.7525794837339217
accept null hypothesis
```

We See our p-value is 0.752 which indicates we have not enough evidence to reject the null in favor of the alternative and therefore we accept the null.

Step 6: Confidence Interval

```
Confidence interval:
(0.7655511229714637, 0.8765895020285356)
```

Our population mean is 0.83 and the Confidence Interval of our sample is (0.76, 0.88) This observation tells us that indeed our sample is accurate to test our hypothesis and prove our null hypothesis to be true.

6. Discussion and Conclusion

Gender Development Index Project

1. Norway had the highest Human development index in 2019 being closely followed by both ireland and switzerland
2. Niger had the lowest Human development index ,followed by central African Republic then chad.(Due to the effect that economies can often suffer from the "snowball" effect, the fact that two-thirds of the nation live below the poverty line of making less than $1 a day, Niger's low HDI is a direct result from poor healthcare, a meager economy, and a lack of centralized, nationalized infrastructure.)
3. The average years for which a female receives education in a top HDI ranked country is 12-14, whereas it is a maximum of 3 years in a country ranked bottom 10 on the HDI index!
4. The average years for which a male receives education in a top HDI ranked country is 12-14.4, whereas it is a maximum of 5.6 years in a country ranked bottom 10 on the HDI index!
5. The average gross national income per capita a female receives in the top 10 HDI ranked countries is between 40k - 55k, whereas it is between 500 - 2k in countries ranked bottom 10 on the HDI index!
6. The average gross national income per capita a Male receives in the top 10 HDI ranked countries is between 55k - 80k, whereas it is between 600 - 3k in countries ranked bottom 10 on the HDI index!(Female are still struggling to earn as much as men even in low income countries)
7. Using LDA algorithm we are able to make a prediction of a countries HDI group with 73% accuracy
8. Using Z test to test our hypothesis we are able to confidently say we have no enough evidence to reject the null in favor of the alternative and therefore we accept the null(Null Hypothesis -the Average global gender development index equals 0.943: Alternative Hypothesis - the Average global gender development index is higher than 0.943)

7. Recommendations

- The position of women, as well as gender norms, may differ between ethnic, religious, educational and other groups living in different areas of a country. Hence there is a need for an index that reveals the gender differences in human development within countries on a global scale.

Gender Development Index Project

- It’s important to avoid old and sometimes well-ingrained stereotypes such as ‘blue for boys’ and ‘pink for girls’, or ‘tough boys’ and ‘caring’ girls, boys don't cry etc
- Eliminating the wage gap and hiring stereotypes could also benefit the world economy in many other ways.
- Gender gap is affected by other aspects in the spectrum other than just the three dimensions discussed here. Focusing also on these areas would have a ripple effect in bridging the overall gender gap. some of these areas include politics(Develop a global partnership for development), Eradicate extreme poverty and hunger, Reduce child mortality, Improve maternal health