PROJECT PROPOSAL

Analysis of Human Development Index (HDI)

UNIVERSITY OF CALGARY

Fall 2023: DATA 604- L01

Group #7:

Prashant Dhungana-30080130

Prateek Kaushik-30229287

Mariya Mathews-30192182

Arteen Rafiei-30043409

Nisha Pillai-30158934

INTRODUCTION

The chosen subject matter of this project revolves around the analysis of Human Development Index (HDI) and its relationship with various societal factors. The HDI is a summary measure of human development that was introduced in 1990 by the United Nations Development Programme (UNDP) [1]. The UNDP evaluates and ranks countries based on three fundamental dimensions of human development: longevity, education and standard of living. Longevity is measured by life expectancy at birth, reflecting the overall health and longevity of a nation's population. The education dimension is measured through two metrics: expected years of schooling for children entering schools and the mean years of schooling for the adult population. The standard of living is measured using the Gross National Income (GNI) per capita. The purpose of this project is to uncover insights into how different aspects of human development interact with one another and to understand how countries' development is affected by a combination of factors, including gender inequality and environmental pressures. The motivation of this project is to understand changing trends in HDI, investigate gender disparities and their impact on development, and assess the environmental pressures affecting human development. Through this exploration, we aim to generate valuable insights that can inform policymakers, researchers, and society at large.

DATASET-1

The dataset is extracted directly from the official website of the United Nations Development Programme (UNDP) [1]. The dataset is created from Human Development Reports by UNDP and falls under the Creative Commons Attribution 3.0 IGO License [2]. It is an open-sourced data by the UNDP for research, academic, and policy-making purposes.

The dataset mainly focuses on the dimensions that the UNDP uses to compute the Human Development Index (HDI) for 2021-2022. The dataset contains a total of seven tables that focus on different human development composite indices. For this project, we will primarily focus on five tables that are of interest to each team member which are:

Table 1: Human Development Index and its components

- This table focuses on the three dimensions used to compute the HDI for different countries and includes metrics such as life expectancy, schooling years, and gross national income per capita.

Table 2: Human Development Index trends, 1990-2021

- This table provides information on the evolution of HDI values over the past decades for various countries.

Table 4: Gender Development Index

- This table offers insights into gender-specific development indices for various countries that focuses on the three dimensions of HDI.

Table 5: Gender Inequality Index

- This table provides insights on gender disparities in human development metrics across various countries.

Table 7: Planetary pressures-adjusted Human Development Index

- This table has metrics such as carbon dioxide emissions and material footprint that focuses on the environment considerations into traditional HDI, reflecting a broad view of development across different countries.

The other two tables from our dataset that are not of our interest for this project are Table 3: Inequality-adjusted Human Development Index and Table 6: Multidimensional Poverty Index. The dataset is in Excel Workbook format and all the tables from the datasets have on average 228 rows each and contain various columns ranging from 9-15 depending on each table. The dataset contains mostly numerical values and the common column in all the tables is the country name that will be used to join various tables to answer some of our guiding questions which will be discussed in detail below.

DATASET-2

In order to answer our guiding questions, we will create an Excel table to categorize countries into their respective continents. This table will only contain two columns 'Country' and 'Continents' and will allow us to analyze and understand regional and demographic trends as required by our guiding questions.

METHODOLOGY

Our project will focus on data analysis and SQL queries to uncover patterns and relationships within the datasets. We plan to use Tableau for data visualization, Python for data wrangling, data cleaning, and feature engineering, as well as MySQL for database management and storing excel workbook data.

The table below illustrates the potential columns of interest and given dataset along with their table. We will be using 2 datasets, wherein Dataset-1: Human Development Index data, Dataset-2: Countries grouped in continents.

All the tables are joined using column "Country"

Guiding Question	Dataset and Table	Columns of Interest
1	**Dataset 1: Table 2** **Dataset 2: Table 1**	Dataset1: Country and HDI Value (1990-2021) Dataset 2: Country, Continent
2	**Dataset 1: Table 1,4** **Dataset 2: Table 1**	Dataset1:Country, HDI Rank(Table 1), Country, Mean years of Schooling (Male & Female), Estimated gross national income per capita (Male & Female) (Table 4) Dataset2: Country, Continent
3	**Dataset 1: Table 1,7** **Dataset 2: Table 1**	Dataset1:Country, Life expectancy at birth (Table 1),Carbon dioxide emissions per capita,Material footprint per capita(Table 7) Dataset2: Country, Continent

The key technologies and tools we will utilize to answer our questions include:

- **Python**: Python will serve as the primary programming language for data wrangling, and data cleaning for missing values from the dataset.
- **Tableau**: We will leverage Tableau to create visually appealing and informative data visualizations. These visualizations will help us present our findings in an easily understandable manner.
- **MySQL**: For efficient data management and storage, we will employ MySQL as our relational database management system. This system will help us organize the datasets and streamline the process of SQL queries.

GUIDING QUESTIONS

1) **How has the Human Development Index (HDI) changed over the past three decades across different countries and continents?**

For this inquiry, we plan to analyze Dataset-1 Table 2: Trends in Human Development Index from 1990 to 2021. Specifically, we will pay close attention to the columns labeled 'HDI Values across 1990-2021', 'Country', and 'Continents' from Dataset-2 Table 1 to join countries into their respective continents. Our primary focus is to determine whether the HDI has increased or decreased over the past three decades in various continents. The visualizations and analysis will provide us with an overview of the countries that have excelled, that have declined or those that have remained the same.

2) **Is there a significant wage gap between males and females with similar education levels across the respective levels of HDI in different continents?**

Here we will investigate Dataset-1 Table 4: Gender Development Index, in our analysis. This table will be examined with a focus on the columns 'Country,' 'Mean years of schooling,' and 'Estimated gross national income per capita' across both genders. 'Mean years of schooling' represents the average number of years of education received by people aged 25 and older in a given country for both genders. 'Estimated gross national income per capita' reflects the average income earned by the citizens of a country across both genders.

We will begin by examining the mean years of schooling for males and females in various countries. If the difference in the mean years of schooling is negligible (+/-0.5 years), i.e. men and women have almost the same number of years of schooling, we will then delve into the disparity in the Gross National Income between both genders and identify which countries show higher disparity. If we find countries exhibiting income disparity between men and women even after receiving the same level of education we will know that we have succeeded in answering the guiding question. Countries will be classified into different HDI tiers ('Very High HDI,' 'High HDI,' 'Medium HDI,' and 'Low HDI') based on the information derived from Dataset-1 Table 1.

We will join Tables 1 and 4 from Dataset-1 using the common 'Country' column. Our research focuses on examining the wage gap within a geographical framework by combining Dataset-1 Table 1, and Table 4 with Dataset-2 Table 1 which contains continent data.

The visualization will provide a worldwide perspective on countries with Very High HDI, displaying wage disparities continent by continent at a geographical level. The same visualization approach will be applied to the High HDI, Medium HDI, and Low HDI categories using choropleth maps and stacked bar charts.

3) **Are carbon dioxide emissions and material footprint a good indicator of life expectancy across different continents?**

Our analysis will investigate 'CO2 emissions per capita (production)' and 'material footprint per capita' as proxies for life expectancy within the global context. CO2 emissions per capita measures the total CO2 emissions produced by a country, divided by its population. Material footprint per capita quantifies the amount of raw materials extracted to meet a country's consumption needs.

By integrating these environmental indicators from Dataset-1 Table 7 (Planetary pressures-adjusted HDI) with life expectancy data from Table 1 (Human Development Index and its components), we aim to determine the correlation and predictive strength of these environmental factors with life expectancy across countries. They will then be grouped by continents from Dataset-2 Table 1 (Continents) to gain inferences across different continents. It is important to note that there are limitations to our data as life expectancy is a multifaceted attribute and cannot be explained solely off of environmental factors. Through the use of heat maps, scatter plots, and trend lines within our visualizations, we will present a clear illustration of how life expectancy relates to environmental indicators across continents.

SUCCESS CRITERIA

The project will be considered successful if it can provide clear visualizations that reveal trends and insights in response to the guiding questions and deliver a compelling narrative that accurately reflects the complexities and nuances of human development data across the world.

REFERENCES

[1] U. Nations, “Human Development Index,” United Nations. Available: https://hdr.undp.org/data-center/human-development-index. [Accessed: Nov. 06, 2023]

[2] U. Nations, “Terms of use,” United Nations. Available: https://hdr.undp.org/terms-use. [Accessed: Nov. 06, 2023]

[3] B. Herre, P. Arriagada, and M. Roser, “The Human Development Index and related indices: what they are and what we can learn from them,” Our World in Data, Nov. 2023, Available: https://ourworldindata.org/human-development-index. [Accessed: Nov. 06, 2023]

[4] “What Is the Human Development Index (HDI)?,” Investopedia. Available: https://www.investopedia.com/terms/h/human-development-index-hdi.asp. [Accessed: Nov. 06, 2023]

[5] G. Dongre, “What is Human Development Index?,” Medium, Jun. 12, 2021. Available: https://ganeshdongre.medium.com/what-is-human-development-index-9f2f991dfd88. [Accessed: Nov. 06, 2023]

[6] “Measuring development | Learn economics.” Available: https://www.learn-economics.co.uk/Measuring-development.html. [Accessed: Nov. 06, 2023]

[7] “Build a Basic View to Explore Your Data.” Available: https://help.tableau.com/current/pro/desktop/en-us/getstarted_buildmanual_ex1basic.htm. [Accessed: Nov. 06, 2023]

[8] “How to Automate Data Cleaning in Python?,” GeeksforGeeks, May 31, 2023. Available: https://www.geeksforgeeks.org/how-to-automate-data-cleaning-in-python/.[Accessed: Nov. 06, 2023]

[9] E. Ortiz-Ospina and M. Roser, “Economic Inequality by Gender,” Our World in Data, Mar. 2018, Available: https://ourworldindata.org/economic-inequality-by-gender [Accessed: Nov. 06, 2023]

Appendix

Project Timeline

S.No	Timeline	Tasks to Achieve
1	**Week 1-2**	Dataset Exploration & Understanding, Database Setup, Data Cleaning & Preprocessing using Python
2	**Week 3-4**	Storing the Data and SQL Queries for Data Retrieval
3	**Week 5**	Data Visualization using Tableau, Exploratory Data Analysis and Pattern Discovery
4	**Week 6**	Finalizing the Project, Generating Insights, and Documentation