Assessment I: Statistical Interpretation and Exploratory Data Analysis

Analysis of the Human Development Index (HDI): A Data-Driven Exploration of Global and Regional Development Patterns.

Student Name
Isha Thapa

Student ID
2548377

Course Name:
5CS037- Concepts and Technologies of AI

Date of Submission
9th Jan, 2026

Table of Contents:

Introduction:

The Human Development Index (HDI) is an annual composite measure that evaluates countries' achievement in key dimensions such as Life Expectancy (LE), Gross National Income (GNI), and educational outcomes. In this dataset, HDI scores are available for multiple countries from 1990 to 2022. HDI is calculated as a geometric means of three components representing health, education and income.

In this project, we begin with Exploratory Data Analysis (EDA) to examine the short trends from 2020 to 2022, conduct regional comparisons between Middle East and South Asia to identify key disparities.

The dataset used is "Human_Development_Index_Dataset.csv" by *Lucas Yukiolmafuko.* Published by United Nation Development Programme (UNDP). For this report, we focus primarily on variables such as Human Development Index (HDI), Gross National Income (GNI), Gender Development Index (GDI) and Life Expectancy Index (LEI).

Problem-wise Analysis:

Problem 1A – Single Year HDI Exploration

Methods:
For this section, data for the year 2022 was filtered from original dataset. After necessary data wrangling, descriptive statistics were computed, including mean, median, mode, maximum, minimum. Countries were also grouped into HDI categories.

Key results:

- Lowest HDI: Somalia (0.38)
- Highest HDI: Switzerland (0.967)
- Highest GNI per capita: Liechtenstein (146,673.2415)

Visualization and tables:
HDI values were categorized into four standard levels:

- Low
- Medium
- High
- Very High

Interpretation and discussion:

The distribution of HDI in 2022 reveals significant inequality between Somalia (Sub-Saharan Africa) and Switzerland (Europe and East Asia). It reflects the strong socio-economic development and higher standards of living standards of “Very High” HDI performing countries. In contrast, “Low” HDI performing countries faces challenges related to limited resources, weak governance structures.

In addition to this, countries like Switzerland and Liechtenstein with high GNI per capita highlights how economic strength correlates with higher human development. However, HDI is also affected by other factors such as health and education, indicating income alone is not the sole driver of development.

Problem 1B – HDI Trend Analysis (2020–2022)

Methods:
Data from 2020 to 2022 was extracted and cleaned. Trends in HDI change were analyzed, average HDI values were computed across regions for the three years.

Key results:

- Andorra showed the greatest improvement in HDI from 2020 to 2022.
- Ukraine showed decline in HDI from 0.762 to 0.734.
- Sub-Saharan Africa had the lowest HDI
- Europe and Central Asia had the highest HDI

Visualization:

- Line graph: HDI Trend for Selected Countries (2020–2022),
- Bar graph: Average HDI by Region (2020–2022),
- Box chart: HDI Distribution Across Years (2020–2022),
- Scatter plot: HDI vs. GNI per Capita (2020–2022).

Interpretation and discussion:

From 2020 to 2022, most countries showed small but consistent improvement in HDI scores. This may reflect post-pandemic recovery in health (life expectancy), education and economic activities that were disrupted in 2020. However, country like Ukraine showed declining HDI due to geo-political conflicts, underscoring how external crises directly affects HDI.

At regional level, Europe and Central Asia maintain highest average HDI while Sub Saharan Africa remains lowest. This regional gap appears highlighting developmental disparities.

Problem 2 – Advanced HDI Exploration

Method:

- For this analysis, I filtered the original dataset to South Asian countries
- Performed data wrangling and calculated Composite Development Score
- Identified top 5 performers
- Detected outliers in HDI and GNI
- Calculated Pearson correlation coefficients between:
    - GNI & HDI,
    - LEI & HDI
- Examined development "gaps" to highlight inequalities

Key results:

- Top 5 Composite Performers: Maldives, Sri -Lanka, Bhutan, India, Bangladesh.
- Maldives is the outlier country in GNI
- LEI showed strong correlation with HDI
- Maldives showed largest development gaps
- Afghanistan showed smallest gaps

Visualizations and tables:

- Bar chart: Top 5 South Asian Countries by Composite Score (2022)
- Scatter Plot: GNI vs HDI with outliers highlighted
- Scatter Plots: HDI vs GDI and HDI vs LEI
- Scatter Plot: GNI-HDI gaps (top 3 vs bottom 3)

Interpretation and discussion:

In South Asia, the Composite Development Score indicates Maldives and Sri Lanka as leading developmental performers, largely driven by strong performance in income (gross income per capita) and health (life expectancy) relative to neighboring countries. Outlier analysis shows Maldives is an economic outlier, showing higher GNI levels compared to the region as whole.

Correlation tests show LEI has stronger positive relation with HDI than GNI, suggesting improvements in health sectors may cause more developmental gains. The gaps showed in Maldives and Afghanistan indicates internal disparities where Maldives showed more unbalanced distribution of development factors whereas Afghanistan showed overall lower development with low internal disparities.

Problem 3 – Comparative Regional Analysis: South Asia vs Middle East

Methods:

- Filtered the dataset from problem 1B into South Asia (SA) and Middle East (ME)
- performed data wrangling and computed:
    - Descriptive Statistics (mean, standard deviation)
    - Top/Bottom performers 2022 of each region (Middle East Vs South Asia)
    - Regional comparisons for GNI, LEI and GDI
    - Measure of disparity (Range, Coefficient of Variation)
    - Pearson correlations of each region between:
        - GDI and HDI
        - LEI and HDI
    - Identified outliers in HDI and GNI

Key Results:

- Middle East performs better overall, with higher mean HDI.
    - ME: 0.7889
    - SA: 0.6395
- Top Performers (2022):
    - ME: UAE, Israel, Bahrain
    - SA: Sri -Lanka, Maldives, Bhutan
- Bottom Performers (2022):
    - ME: Lebanon, Iraq, Yemen
    - SA: Nepal, Pakistan, Afghanistan
- GNI showed the greatest disparity between two regions
- South Asia: LEI more strongly correlated with HDI
- Middle East: GDI more strongly correlated with HDI
- South Asia showed fewer outliers than the Middle East

Visualization and tables:

- Bar chart: Top/Bottom HDI performers by region (2022)
- Bar chart: Regional Comparison of Development Metrics (2020–2022)
- Scatter Plot: HDI vs GDI and HDI vs LEI
- Scatter plot: GNI vs HDI with outliers highlighted

Interpretation and discussion:

Comparative analysis between South Asia and the Middle East demonstrates development divergence. Middle East exhibits higher average HDI indicating greater economic capacity. South Asia shows relatively low average HDI with consistent/unform performance among countries.

Interestingly, the relationship between development indicators differs across regions. In South, HDI is strongly influenced by Life Expectancy (LEI), indicating health outcomes play a central role in development. In Middle East, HDI correlates more strongly with Gender Development Index (GDI), gender disparities contribute significantly to overall development levels.

The greater number of outliers in the Middle East indicates internal inequality within region, smaller number in South Asia suggests a more consistent development.

Conclusion:

Summary:

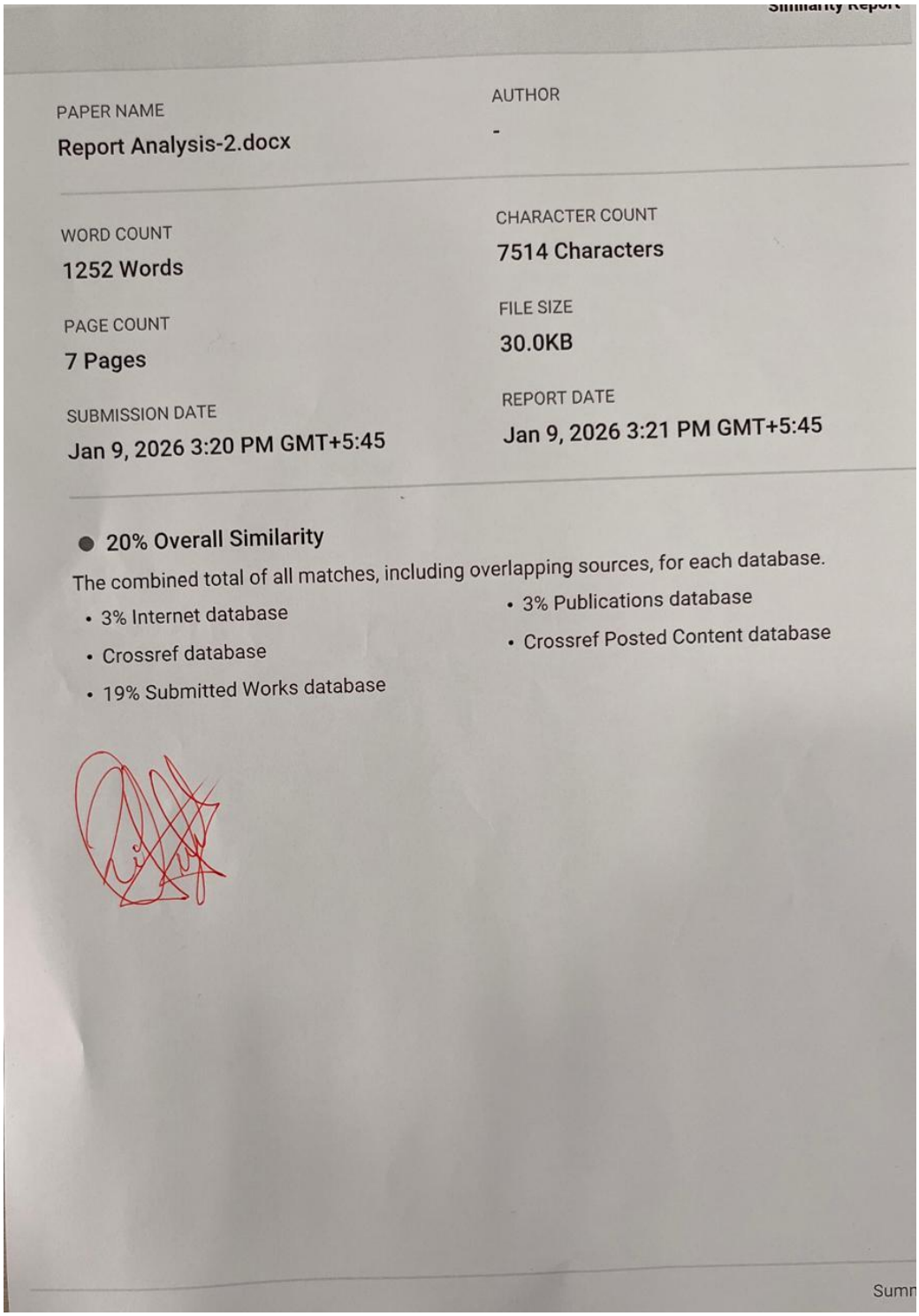
- This study analyzed Human Development Index (HDI) from 2020 to 2021, exploring regional disparities, relations between multiple development indicators.
- Improvements in health and income were found to contribute significantly to higher HDI performance across countries.
- This analysis also revealed that different development indicators (LEI, GDI) influence differently across regions.
- HDI can be affected by external shocks, as shown in case of Ukraine, where geo-political conflict contributed to decline in development.
- In addition to above, outlier patterns suggest that internal disparities and uneven development pathways can create significant variation within regions.

Limitation:

- The analysis use data from only 2020 – 2022, which provides short time window for observing development trends and long-term patterns.
- The period includes COVID – 19 and geo-political conflicts, which temporary affects HDI values.
- Missing data in dataset were imputed using mean or mode substitutions, which can reduce analytical accuracy and may mask real variation.
- Additionally, HDI is a composite score that summarizes key aspects of development, but other relevant factors such as education quality, economic diversification or healthcare capacity are not captured in this dataset.

Implications:

- Analyzing longer time series data to better identify structural trends.
- Including socio-economic variables such as education quality, poverty rates, etc. for more comprehensive analysis
- Using datasets with fewer missing values to increase reliability



Similarity Report

PAPER NAME
Report Analysis-2.docx

AUTHOR
-

WORD COUNT
1252 Words

CHARACTER COUNT
7514 Characters

PAGE COUNT
7 Pages

FILE SIZE
30.0KB

SUBMISSION DATE
Jan 9, 2026 3:20 PM GMT+5:45

REPORT DATE
Jan 9, 2026 3:21 PM GMT+5:45

● 20% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- Crossref database
- 19% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

Summ