



Academic Year	Module	Assessment Number	Assessment Type
2025	5CS037– Concepts & Technologies of AI	1	Report

Assignment - I

Statistical Interpretation and Exploratory Data Analysis : Human Development Index (HDI) Analysis Report

Student Id : 2501453

Student Name : Aditri Prajapati

Section : L5CG11

Module Leader : Siman Giri

Tutor : Robin Tuladhar

Submitted on : 10-01-2026

Table of Contents

1. Introduction

1.1 Overview of Human Development Index

1.2 Objectives of the Analysis

1.3 Scope and Limitations

2. Methodology

2.1 Data Sources and Description

2.2 Analytical Tools and Techniques

2.3 Data Processing Pipeline

3. Problem 1A: Single Year HDI Exploration (2022)

3.1 Methods and Approach

3.2 Key Results and Findings

3.3 Visualizations and Analysis

3.4 Interpretation and Discussion

4. Problem 1B: HDI Trend Analysis (2020-2022)

4.1 Methods and Approach

4.2 Key Results and Findings

4.3 Visualizations and Analysis

4.4 Interpretation and Discussion

5. Problem 2: Advanced HDI Exploration - South Asia Focus

5.2 Key Results and Findings

5.3 Visualizations and Analysis

5.4 Interpretation and Discussion

6. Problem 3: Comparative Regional Analysis

6.1 Methods and Approach

6.2 Key Results and Findings

6.3 Visualizations and Analysis

6.4 Interpretation and Discussion

7. Conclusion and Recommendations

7.1 Summary of Key Findings

7.2 Policy Implications

7.3 Limitations of the Study

7.4 Future Research Directions

8. References

9. Appendix

1. Introduction

1.1 Overview of Human Development Index

The United Nations Development Programme (UNDP) created the Human Development Index (HDI) in 1990, a composite indicator, to gauge a nation's general degree of human development.

In contrast to purely economic indicators like Gross Domestic Product(GDP), HDI takes a multifaceted approach by combining health, education and income dimensions. Life expectancy at birth is a measure of health; expected and mean years of schooling are a measure of education; and Gross National Income(GNI) per capita is a measure of standard of living. HDI offers a more comprehensive and people-centered evaluation of development outcomes by incorporating these dimensions.

The HDI score ranges between 0 and 1, where higher values indicate higher levels of human development. Countries are grouped into four development categories: Very High, High, Medium, and Low Human Development. This classification allows meaningful comparison across countries and regions while highlighting persistent inequalities in development outcomes.

1.2 Objectives of the Analysis

The main goal of this analysis is to use HDI data to investigate regional and global trends in human development.

In particular, the study aims to:

- Examine HDI levels across countries for a single year (2022)
- Analyze short-term HDI trends between 2020 and 2022
- Conduct an in-depth regional analysis focusing on South Asia
- Develop a composite development score using selected HDI-related indicators
- Compare development patterns across different global regions

1.3 Scope and Limitations

This analysis covers all countries for which HDI data is available from 1990-2022, with particular focus on 2020-2022 for trend analysis and South Asia/Middle East for regional comparisons.

This study utilizes secondary data derived from internationally recognized development datasets. Despite the dataset's comprehensiveness, the analysis is constrained by the variables' availability and the chosen time frame. In addition, proxy measures and normalization techniques are used because some indicators lack official index values, which may not precisely replicate UNDP's official HDI computation methodology.

2. Methodology

2.1. Data Sources and Description

The dataset used in this analysis contains country-level HDI values and associated indicators such as life expectancy, GNI per capita, and regional classifications. The data spans multiple years, allowing both cross-sectional and temporal analysis. The year 2022 is treated as the most recent reference point for single-year and comparative assessments.

Unnamed: 0	iso3	country	year	hdi	life_expectancy	pop_millions	hdi_f	hdi_m	life_expec_f	life_expec_m	expec_yr_school	expec_yr_school_f	expec_yr_school_m	mean_yr_school	
0	1	AFG	Afghanistan	1990	0.284	45.967	10.694796	NaN	NaN	48.397	43.709	2.936460	2.117230	4.532768	0.871962
1	2	AFG	Afghanistan	1991	0.292	46.663	10.745167	NaN	NaN	49.144	44.353	3.228456	2.246242	4.768261	0.915267
2	3	AFG	Afghanistan	1992	0.299	47.596	12.057433	NaN	NaN	50.320	45.070	3.520452	2.383115	5.015989	0.958573
3	4	AFG	Afghanistan	1993	0.307	51.466	14.003760	NaN	NaN	52.739	50.216	3.812448	2.528328	5.276587	1.001878
4	5	AFG	Afghanistan	1994	0.300	51.495	15.455554	NaN	NaN	53.544	49.531	4.104445	2.682390	5.550723	1.045184

Table 1 : Dataset Overview

2.2 Analytical Tools and Techniques

The analysis was conducted using Python, primarily leveraging libraries such as Pandas for data manipulation, NumPy for numerical operations, and Matplotlib/Seaborn for data visualization. Statistical techniques include descriptive statistics, min–max normalization, ranking, interquartile range (IQR) analysis, and correlation analysis.

2.3 Data Processing Pipeline

The data processing workflow involved importing the dataset, inspecting data types, handling missing values, and filtering relevant subsets for each problem. Duplicate records were removed where identified, and missing values in critical variables were either excluded or justified based on analytical relevance. Missing values in key variables such as HDI were removed to maintain analytical validity, while other missing indicators were documented rather than imputed to avoid bias. With the help of feature engineering techniques, derived variables such as Composite Score and GNI-HDI Gap was created. The cleaned and processed data were then used for statistical analysis and visualization, followed by interpretation of results. This systematic pipeline ensured data quality and analytical reliability.

3. Problem 1A: Single Year HDI Exploration (2022)

3.1 Methods and Approach

In problem 1A, the data was filtered based on observations from 2022. Missing HDI data was filtered out for countries that did not have this data. Descriptive statistics were used to understand the data distribution for HDI, and countries were grouped based on the HDI criteria set by UNDP. Top-performing countries were further segmented based on GNI per capita.

Data Cleaning Decisions:

- Missing HDI values were dropped as HDI is the primary variable of interest
- Missing gender indicators were noted but not imputed due to potential bias
- Numeric columns were converted to appropriate data types

For facilitating a comparative analysis of development achievements, a new variable named HDI Category was developed for the year 2022. For determining these categories, four official groups of human development, as classified by UNDP, which include Low, Medium, High, and Very High, were used based on the values of HDI. All countries have been categorized based on their respective HDI values. Classification correctness was verified by inspecting category counts and reviewing sample country assignments.

3.2 Key Results and Findings

The 2022 dataset contains HDI values for 187 countries. The global mean HDI is 0.72, with a median of 0.73, indicating a slight left-skew in the distribution. While a majority of countries fall into the High and Very High HDI categories, a non-negligible proportion remains in Medium and Low development groups.

```
BASIC STATISTICAL ANALYSIS - HDI 2022
Mean HDI: 0.7229
Median HDI: 0.7395
Standard Deviation: 0.1530
HDI Range: [0.3800, 0.9670]
```

Fig 1 : HDI Statistics for 2022

After applying the HDI classification rules, a new column titled **“HDI Category”** was successfully added to the 2022 dataset. Each country was assigned to one of the four UNDP development groups based solely on its HDI value.

To verify the correctness of the classification:

- The distribution of countries across HDI categories was examined.
- Sample rows containing country names, HDI values, and assigned categories were reviewed manually.
- The dataset structure was inspected to confirm the presence of the new column.

These verification steps ensured that all countries were classified accurately and that no records were left uncategorized.

```
=== Adding HDI Category Column ===
=== HDI Category Distribution ===
HDI Category
Very High    71
High         54
Medium       44
Low          35
Name: count, dtype: int64
```

Fig 2 : HDI Category Distribution

3.3 Visualizations and Analysis

Country with Highest HDI:

Country: Switzerland

HDI: 0.9670

Country with Lowest HDI:

Country: Somalia

HDI: 0.3800

Top 5 Countries by HDI:

	country	hdi
5609	Switzerland	0.967
4322	Norway	0.966
2540	Iceland	0.959
2474	Hong Kong, China (SAR)	0.956
5576	Sweden	0.952

Bottom 5 Countries by HDI:

	country	hdi
5345	Somalia	0.380
5411	South Sudan	0.381
1088	Central African Republic	0.387
1121	Chad	0.394
4223	Niger	0.394

Fig 3 : Countries with HDI Comparison

Filtering Countries with HDI > 0.800

Number of countries with HDI > 0.800: 71

Top 10 Countries by GNI per capita (HDI > 0.800)

country	hdi	gross_inc_percap
Liechtenstein	0.942	146673.24150
Qatar	0.875	95944.37754
Singapore	0.949	88761.14559
Ireland	0.950	87467.51391
Luxembourg	0.927	78554.23640
United Arab Emirates	0.937	74103.71494
Switzerland	0.967	69432.78669
Norway	0.966	69189.76165
United States	0.927	65564.93798
Hong Kong, China (SAR)	0.956	62485.50516

Fig 4 : Top 10 countries by GNI per capita among those with Very High HDI

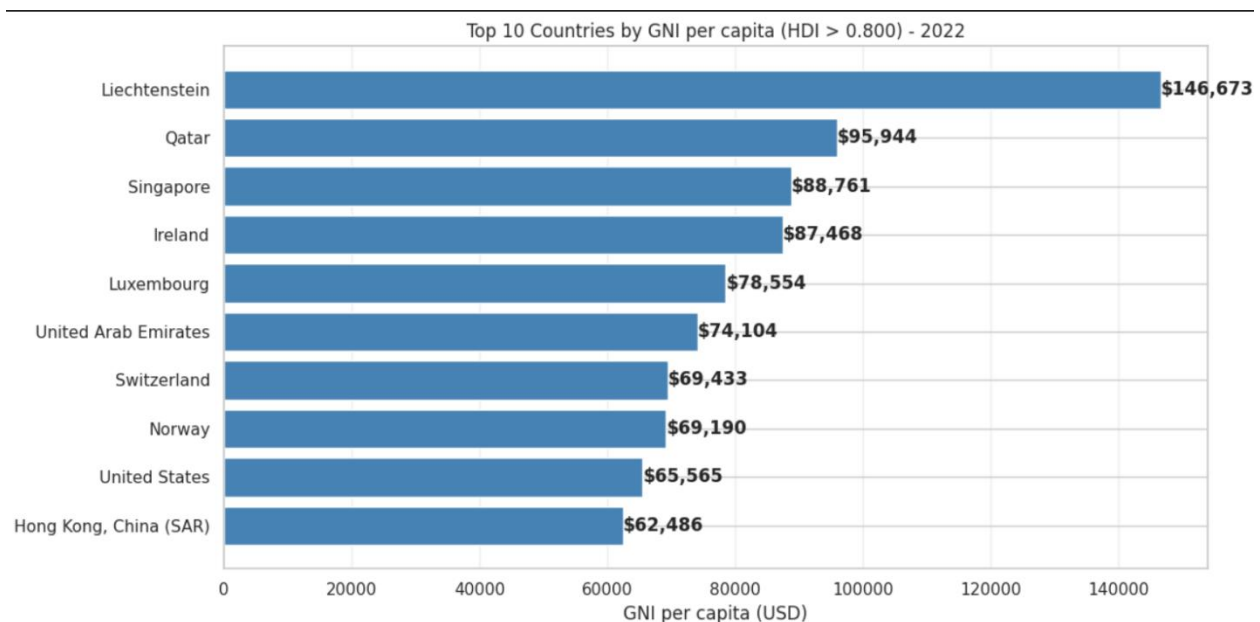


Fig 5 : Bar graph of Top 10 countries by GNI per capita among those with Very High HDI

=== Sample from each HDI Category ===

Very High HDI Countries (sample):

country	hdi	HDI Category
Andorra	0.884	Very High
Antigua and Barbuda	0.826	Very High
Argentina	0.849	Very High

High HDI Countries (sample):

country	hdi	HDI Category
Albania	0.789	High
Algeria	0.745	High
Armenia	0.786	High

Medium HDI Countries (sample):

country	hdi	HDI Category
...		
country	hdi	HDI Category
Afghanistan	0.462	Low
Benin	0.504	Low
Burkina Faso	0.438	Low

Fig 6: Sample from each HDI category

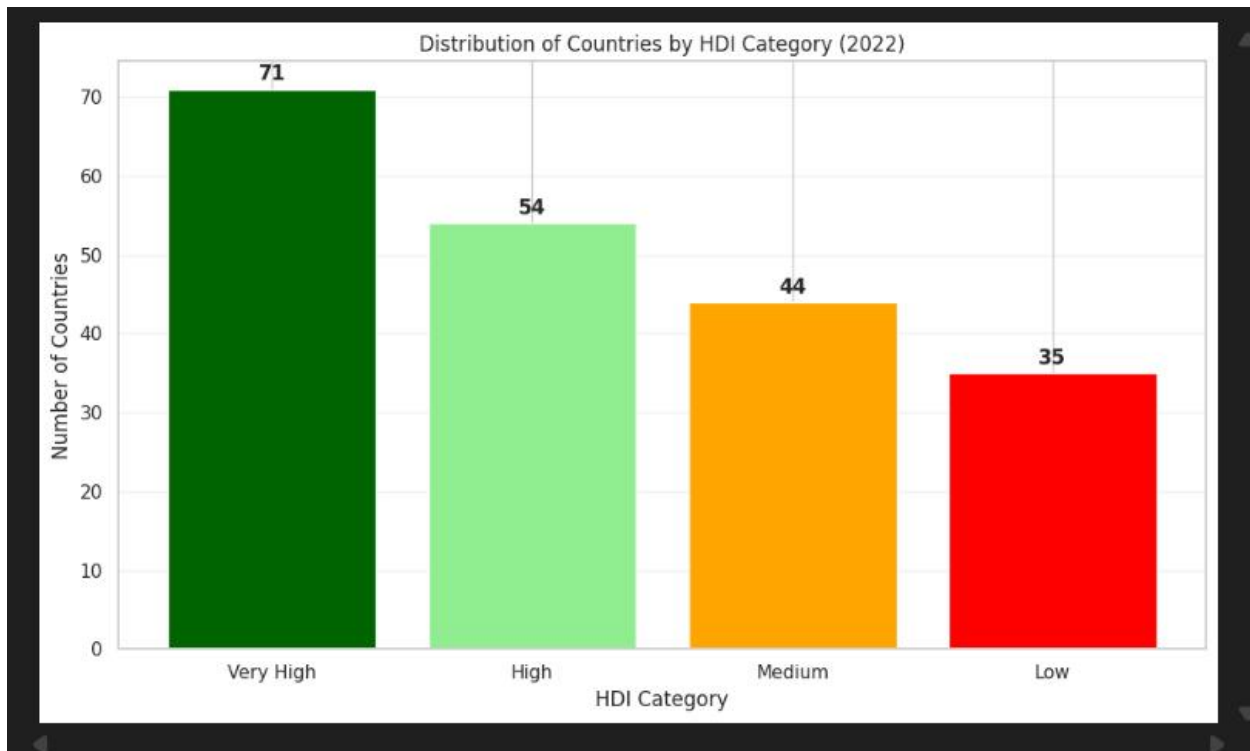


Fig 7: Bar graph of Distribution of Countries by HDI category

3.4 Interpretation and Discussion

The HDI category classification reveals clear global development inequalities. Countries classified under Very High Human Development typically demonstrate strong performance in health, education, and income indicators. In contrast, countries in the Low and Medium categories often face challenges related to limited healthcare access, lower educational attainment, and reduced economic capacity.

By converting continuous HDI values into categorical development groups, the analysis becomes more interpretable for policy discussion and comparative studies. This classification also facilitates subsequent regional and trend-based analyses conducted in later sections of the report.

Finally, the updated dataset containing the newly added HDI Category column was saved as "HDI category added.csv", as required by the assignment, and included in the final submission.

4. Problem 1B: HDI Trend Analysis (2020–2022)

4.1 Methods and Approach

This section examines HDI trends during the period 2020–2022, capturing the impact of the COVID-19 pandemic. Only countries with complete HDI data for all three years were retained. Year-over-year changes were calculated, and regional averages were compared to assess recovery patterns.

4.2 Key Results and Findings

The global mean HDI declined in 2020 but showed gradual recovery in 2021 and 2022. While most countries experienced modest improvements by 2022, recovery was uneven across regions.

```
...  
Countries with HDI increase 2020-2021: 81  
Countries with HDI increase 2021-2022: 168  
Average HDI change 2020-2021: -0.0009  
Average HDI change 2021-2022: 0.0045
```

Fig 8: Global HDI trends

Countries with HDI data for all three years: 203

=== Question 1: Which countries show the greatest improvement in HDI from 2020 to 2022? ===

Top 10 countries with greatest HDI improvement:

Andorra: +0.041
Azerbaijan: +0.038
Ecuador: +0.031
Maldives: +0.025
Mexico: +0.024
San Marino: +0.023
Tajikistan: +0.023
Bahamas: +0.022
Kuwait: +0.021
Türkiye: +0.020

=== Question 2: Did any countries experience a decline in HDI? Provide possible reasons. ===

Yes, 54 countries experienced HDI decline:

Top 5 countries with largest HDI decline:

Timor-Leste: -0.067
Ukraine: -0.028
Afghanistan: -0.026
Namibia: -0.024
Lebanon: -0.019

Fig 9 :Countries with the largest HDI improvements and declines

4.3 Visualizations and Analysis

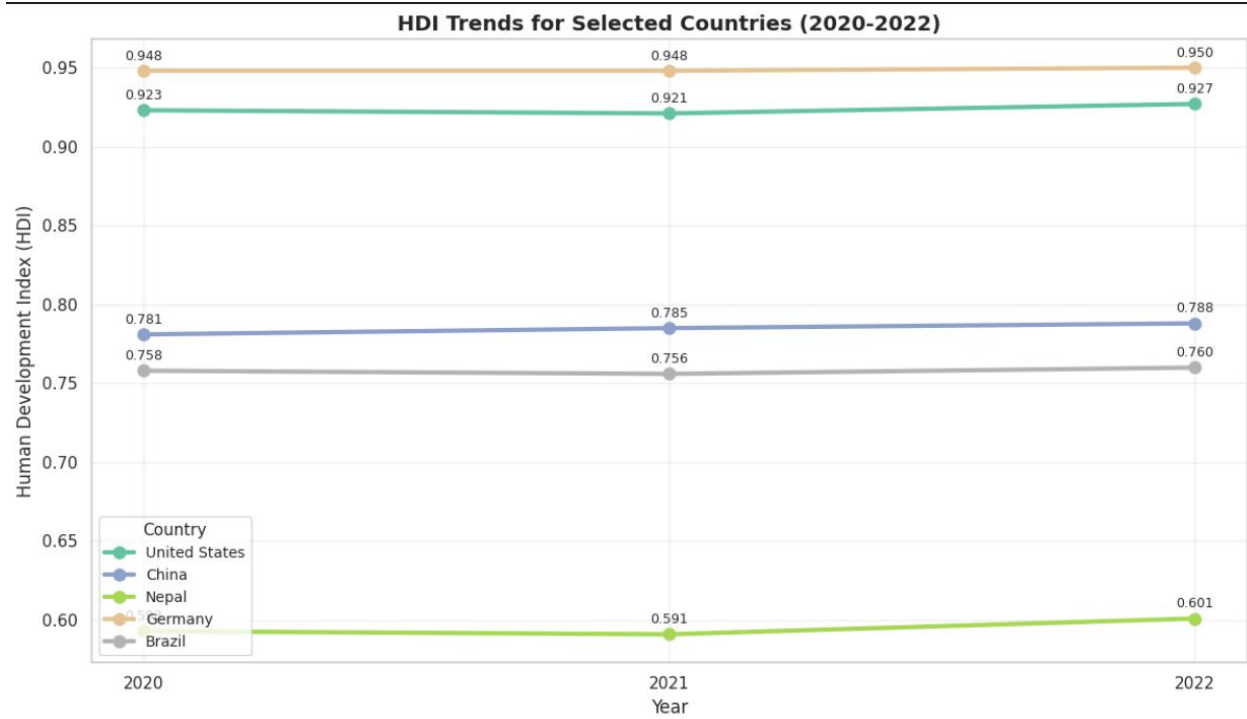


Fig 10 : Illustrating HDI trends for selected countries

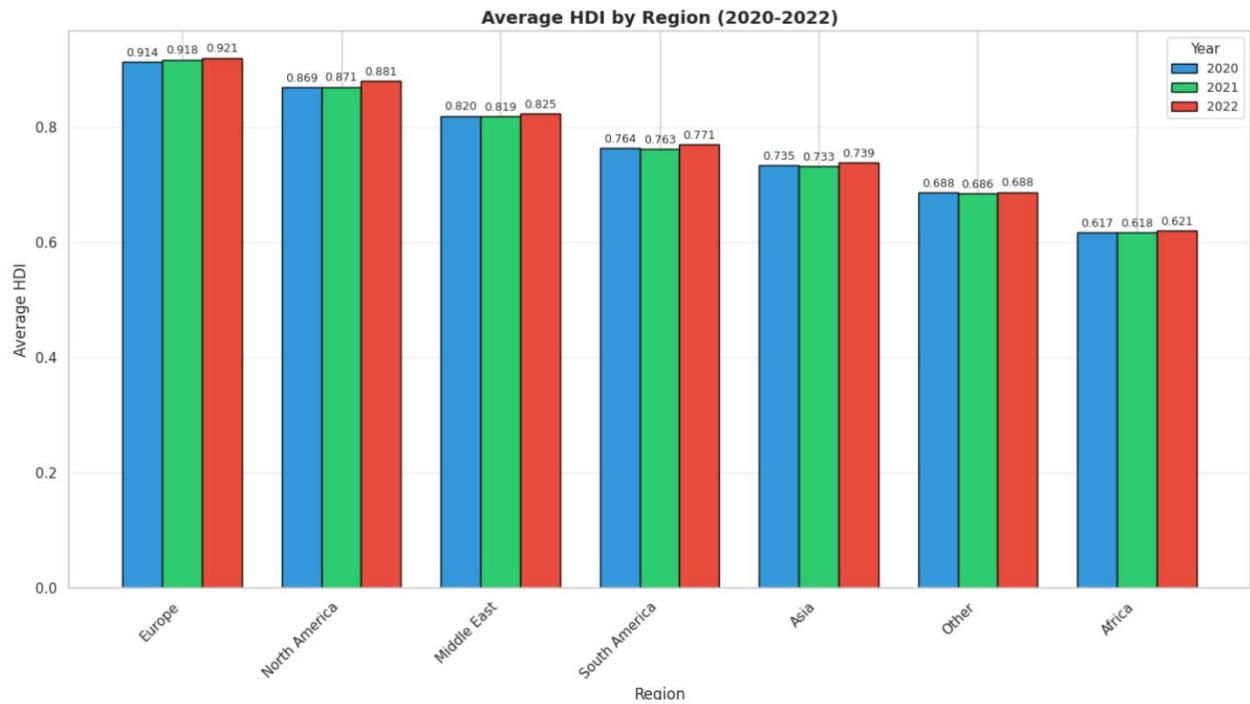


Fig 11: Showing regional average HDI values

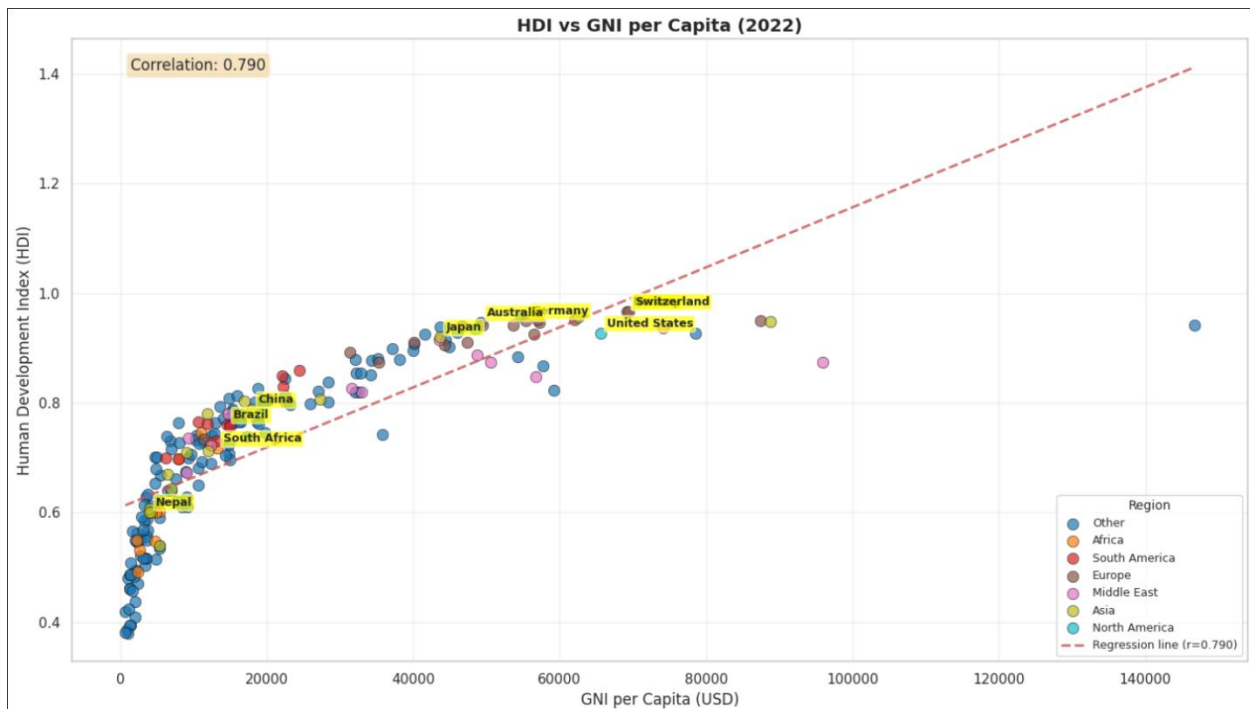


Fig 12: Illustrating HDI vs GNI per Capita(2022)

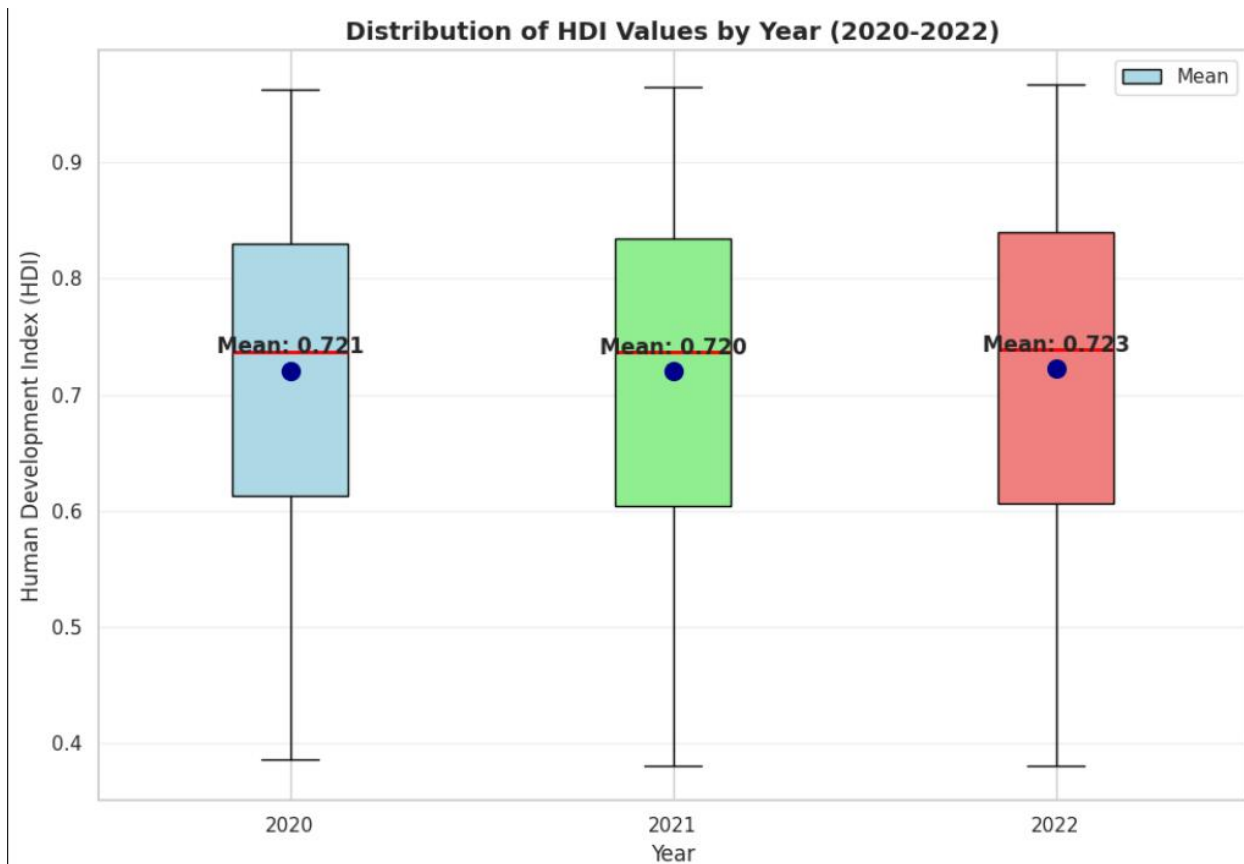


Fig 13: Presenting HDI distributions using box plots

4.4 Interpretation and Discussion

The pandemic caused the first recorded global decline in HDI since its inception. Recovery trajectories varied based on healthcare capacity, digital infrastructure, and economic diversification. Countries with resilient health systems and robust digital access recovered more quickly, highlighting the importance of systemic preparedness for future shocks.

5. Problem 2: Advanced HDI Exploration – South Asia Focus

5.1 Methods and Approach

South Asian countries were analysed in detail using a composite development score based on life expectancy and GNI per capita. These indicators were normalised using min–max scaling to ensure comparability. Additional analyses included outlier detection using IQR, percentile-based gap analysis, and correlation analysis between HDI and its components.

5.2 Key Results and Findings

South Asia exhibits wide variation in development outcomes, ranging from High HDI countries such as Sri Lanka to Low HDI countries such as Afghanistan. Differences between HDI rankings and composite score rankings highlight varying strengths across health, income, and education dimensions.

=== South Asian Countries Ranked by Composite Score (2022) ===				
country	Composite_Score	hdi	life_expectancy	gross_inc_percap
Maldives	0.600000	0.762	80.839	18846.792190
Sri Lanka	0.410342	0.780	76.610	11899.498470
Bhutan	0.315326	0.681	72.229	10624.873880
Bangladesh	0.269390	0.670	73.698	6511.122178
India	0.177463	0.644	67.744	6950.526798
Nepal	0.173122	0.601	70.484	4025.554685
Pakistan	0.128527	0.540	66.431	5374.270423
Afghanistan	0.000000	0.462	62.879	1335.205733

Fig 14: South Asian HDI rankings and composite score comparisons

5.3 Visualizations and Analysis

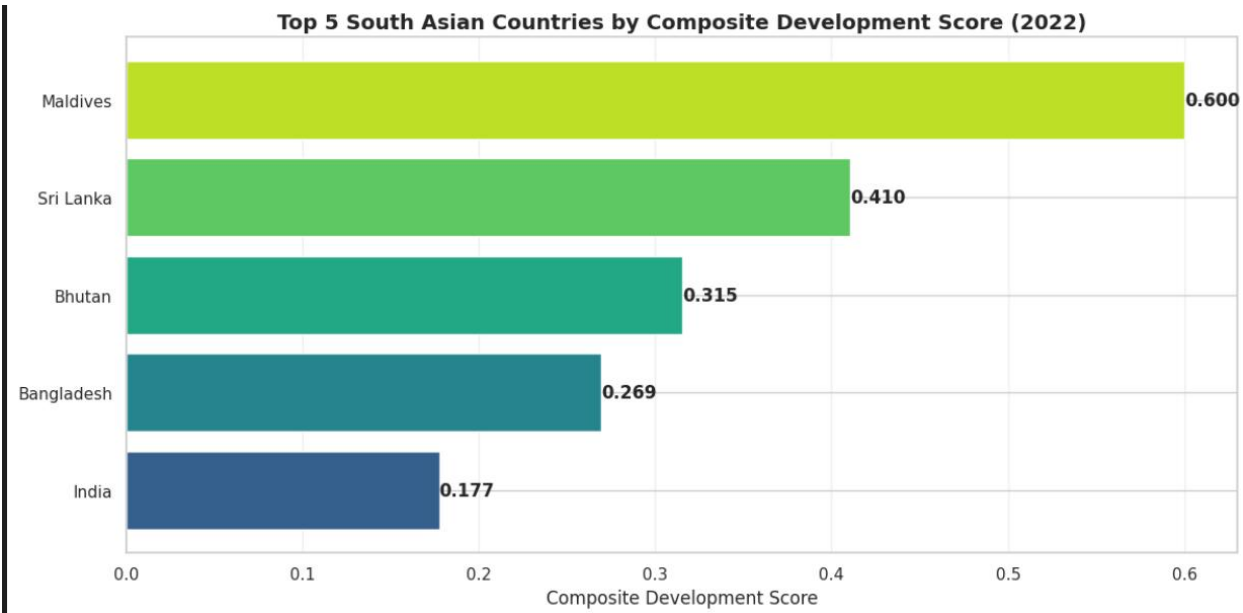


Fig 15: Showing composite score rankings

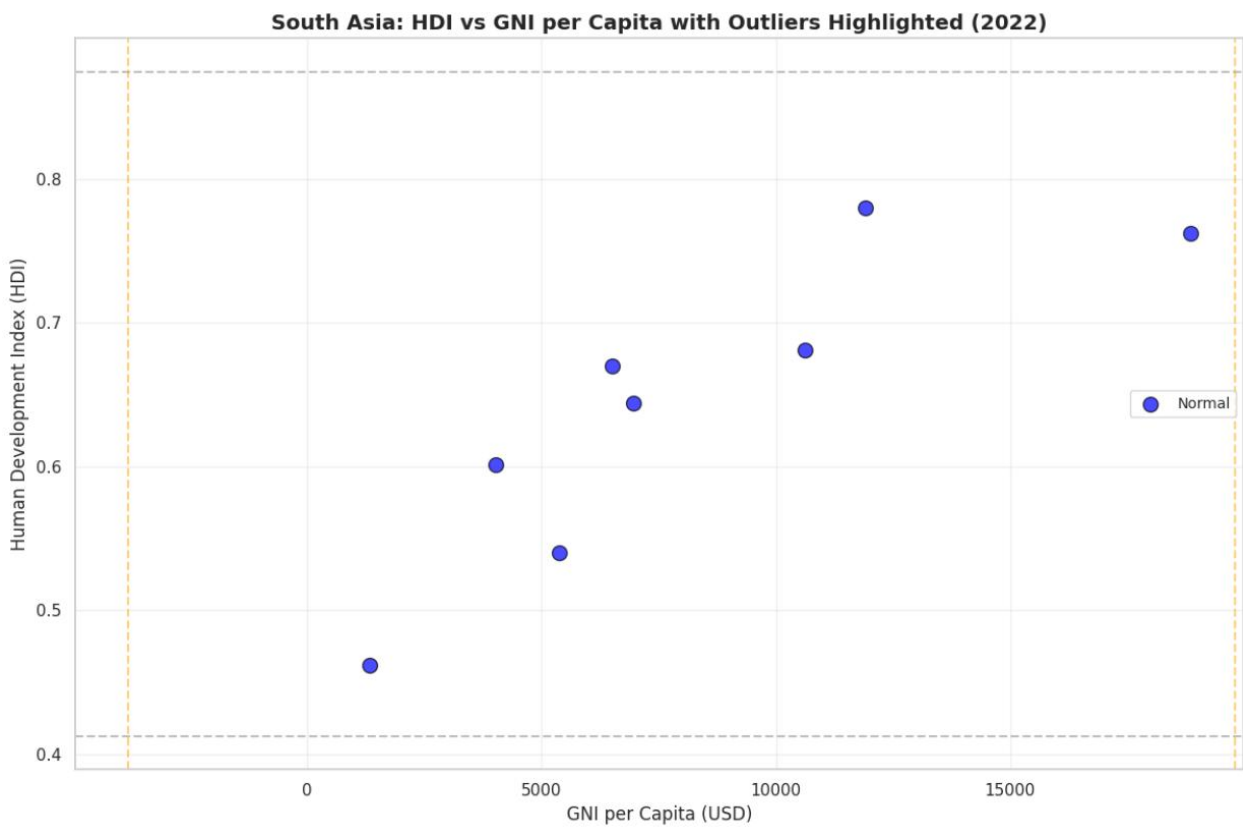


Fig 16: Displaying HDI versus GNI per capita with outliers highlighted

5.4 Interpretation and Discussion

The analysis reveals that some South Asian countries achieve relatively strong human development outcomes despite limited income levels, indicating efficient social policy implementation. Conversely, income-rich countries do not always translate economic resources into equivalent development gains.

6. Problem 3: Comparative Regional Analysis

6.1 Methods and Approach

A comparative analysis was conducted between South Asia and the Middle East using descriptive statistics, correlation analysis, and visual comparisons. Both regions were evaluated using identical metrics to ensure consistency.

6.2 Key Results and Findings

The Middle East outperforms South Asia in terms of average HDI, life expectancy, and income levels. However, internal disparities exist in both regions, particularly in conflict-affected countries.

=== Sample Data - South Asia (2020-2022) ===			
	year	country	hdi
0	2020	Afghanistan	0.488
1	2021	Afghanistan	0.473
2	2022	Afghanistan	0.462
39	2020	Bangladesh	0.657
40	2021	Bangladesh	0.662
=== Sample Data - Middle East (2020-2022) ===			
	year	country	hdi
36	2020	Bahrain	0.884
37	2021	Bahrain	0.884
38	2022	Bahrain	0.888
240	2020	Iraq	0.661
241	2021	Iraq	0.667

Fig 17: Table comparing regional HDI statistics

6.3 Visualizations and Analysis

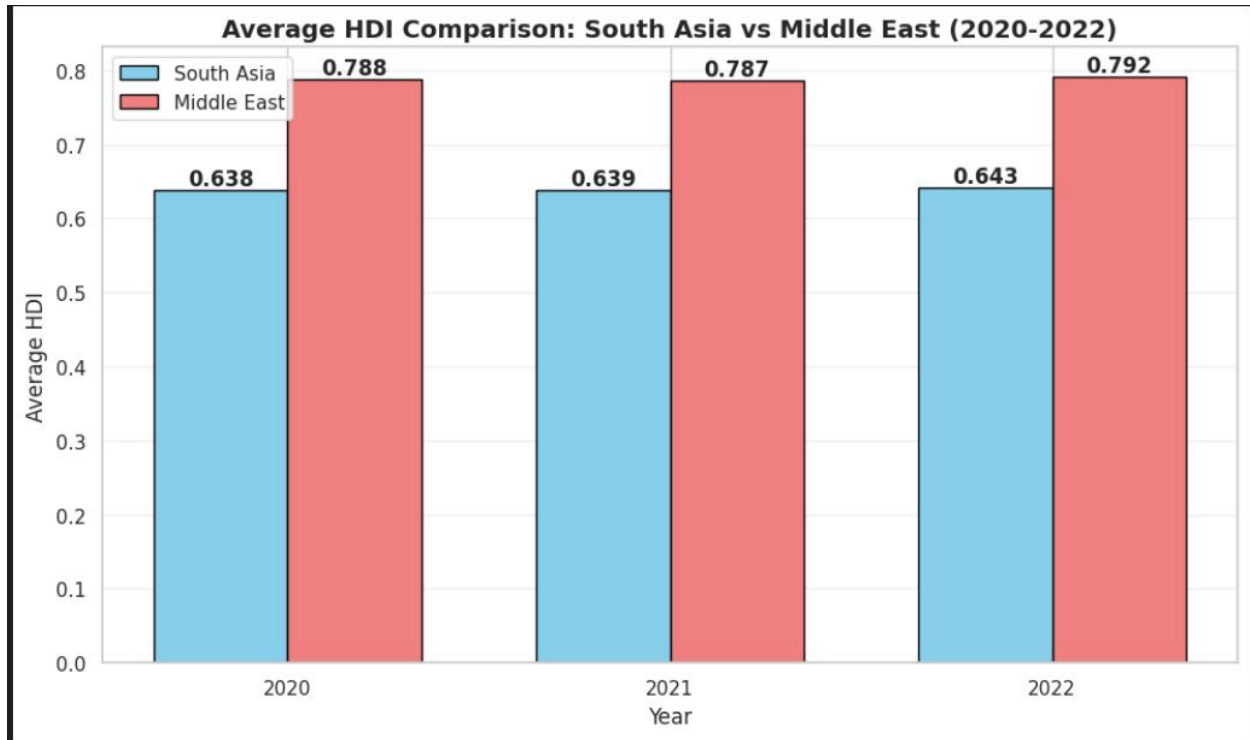


Fig 18: Showing regional HDI comparisons

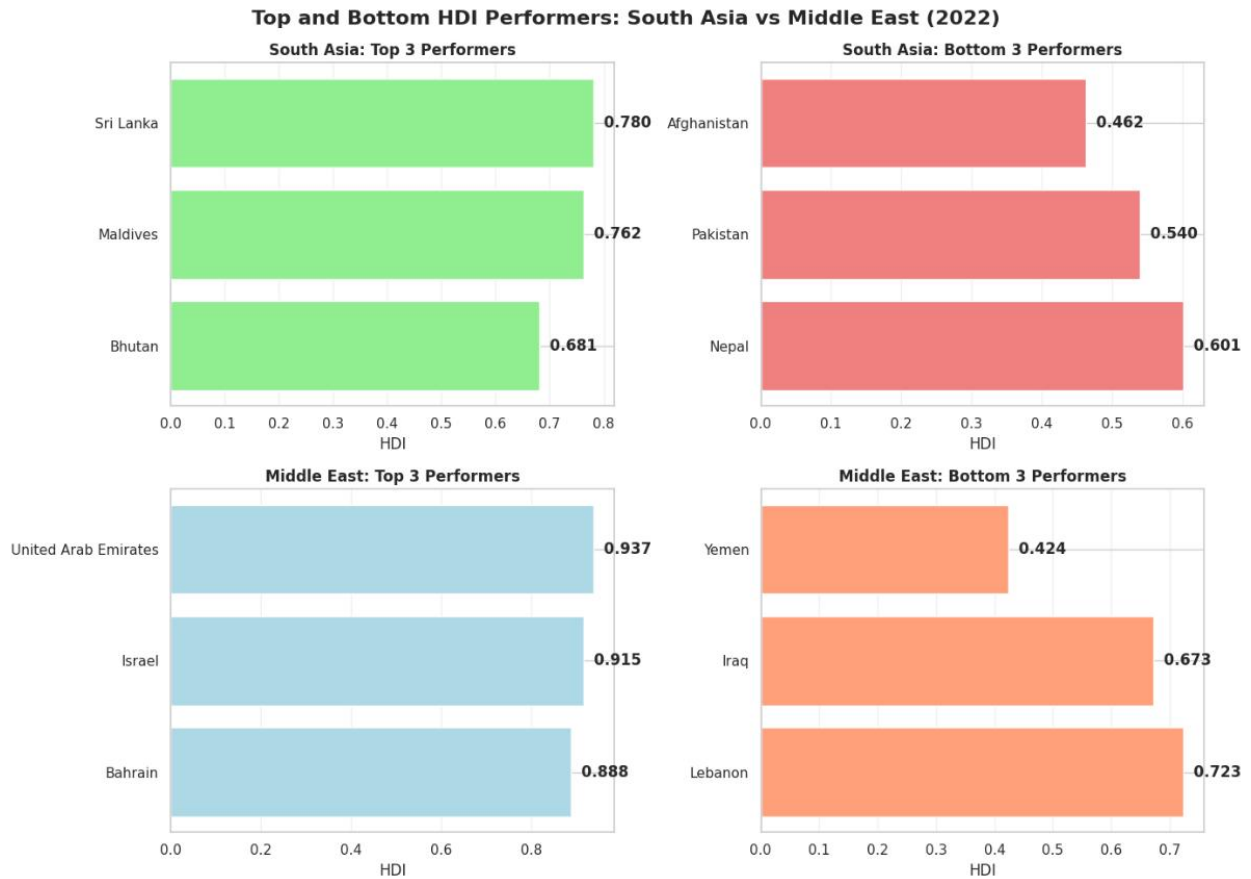


Fig 19: Comparing top and bottom performers

6.4 Interpretation and Discussion

Despite significant income advantages, the Middle East does not exhibit proportionally higher HDI gains compared to South Asia. This suggests that development efficiency and governance play a crucial role alongside economic resources.

7. Conclusion and Recommendations

7.1 Summary of Key Findings

This study analyzed Human Development Index (HDI) data from 1990–2022 to examine global and regional development patterns. The results show steady long-term global improvement in human development, with most countries now classified as High or Very High HDI. However, significant inequalities persist across and within regions. The COVID-19 period (2020–2022) caused temporary setbacks in HDI, particularly affecting health and income dimensions, though most countries demonstrated gradual recovery by 2022. The South Asia analysis highlighted wide internal disparities, while comparative analysis showed the Middle East outperforming South Asia in income but not proportionally in overall human development outcomes.

7.2 Policy Implications

The findings suggest that economic growth alone is insufficient to achieve high human development. Policymakers should prioritize investments in healthcare, education, and gender equality to ensure balanced development. Countries that perform well despite limited income levels demonstrate the importance of efficient social policies. For conflict-affected nations, restoring political stability and basic services is essential before sustainable development gains can occur.

7.3 Limitations of the Study

This analysis is subject to several limitations. First, missing or incomplete data for certain countries, particularly those affected by conflict, may reduce accuracy. Second, the use of country-level averages masks sub-national inequalities. Third, the Composite Score used in South Asia focuses only on health and income, excluding education, which may oversimplify development measurement.

7.4 Future Research Directions

Future studies could extend this work by analysing longer-term HDI trends, incorporating sub-national or inequality-adjusted HDI measures, and applying machine learning techniques to predict development trajectories. Including environmental sustainability indicators would also provide a more holistic view of human development in the context of global challenges such as climate change.

8. References

(UNDP, 2023)

(Yukiolmafuko, 2024)

References

UNDP, U. N. D. P., 2023. *Human Development Report 2023/2024*, s.l.: s.n.

Yukiolmafuko, L., 2024. Human_Development_Index_Dataset.csv [Modified dataset for academic use].

9. Appendix

Appendix A: Complete Code Availability

The complete Jupyter notebook with all analysis code is available at:

https://github.com/AditriPrajapati1/AI_Assignment-I

Appendix B: Generated Datasets

The following datasets were generated during the analysis and are included in the submission folder and GitHub repository:

1.HDI category added.csv

2.HDI problem1B.csv

3.HDI SouthAsia.csv

4.HDI SouthAsia 2020 2022.csv

5.HDI MiddleEast 2020 2022.csv

Appendix C: Plagiarism Report

The similarity report was generated using Canvas plagiarism detection. The overall similarity index is within the acceptable academic threshold.

Similarity Report	
PAPER NAME	AUTHOR
2501453_AditriPrajapati.pdf	-
WORD COUNT	CHARACTER COUNT
2203 Words	12887 Characters
PAGE COUNT	FILE SIZE
25 Pages	1.0MB
SUBMISSION DATE	REPORT DATE
Jan 9, 2026 10:49 AM GMT+5:45	Jan 9, 2026 10:49 AM GMT+5:45
● 19% Overall Similarity	
The combined total of all matches, including overlapping sources, for each database.	
<ul style="list-style-type: none">• 6% Internet database• 2% Publications database• Crossref database• Crossref Posted Content database• 17% Submitted Works database	



Academic Year	Module	Assessment Number	Assessment Type
2025	CS037- Concepts & Technologies of AI	1	Report

Share



Assignment - I

Page 1 of 25

Top sources

All Sources



19%

Overall Similarity

- 1

University of Wolverhampton on 2026-01-07
SUBMITTED WORKS

2%
- 2

University of Wolverhampton on 2026-01-08
SUBMITTED WORKS

2%
- 3

University of Iowa on 2024-12-18
SUBMITTED WORKS

2%
- 4

University of Wolverhampton on 2026-01-08
SUBMITTED WORKS

1%
- 5

University of Wolverhampton on 2026-01-08
SUBMITTED WORKS

1%
- 6

University of Wolverhampton on 2026-01-07
SUBMITTED WORKS

1%
- 7

gupea.ub.gu.se
INTERNET

<1%

>>

Similarity Report

PAPER NAME

2501453_AditriPrajapati.pdf

AUTHOR

-

WORD COUNT

2203 Words

CHARACTER COUNT

12887 Characters

PAGE COUNT

25 Pages

FILE SIZE

1.0MB

SUBMISSION DATE

Jan 9, 2026 10:49 AM GMT+5:45

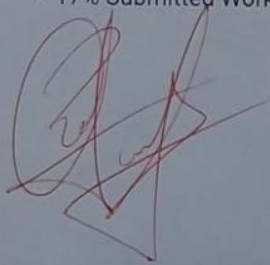
REPORT DATE

Jan 9, 2026 10:49 AM GMT+5:45

● 19% Overall Similarity

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

- 6% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 17% Submitted Works database



Tutor: Robin Tuladhar