



Data Collection and Preprocessing Phase

11111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	reprocessing raise
Date	28 June 2024
Team ID	739773
Project Title	A Comprehensive Measure Of Well-Being:A Human Development Using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.





Section	D	esc	riptio	on														
		esc:	riptiv	e An	alysi	s:-												
	£		Unnamed:	Id	HDI Rank	HDI	Life expectancy	Mean years of schooling	Gross national income (GNI) per capita	GNI per capita rank minus HDI rank	Change in HDI rank 2010-2015	Average annual HDI growth 1990-2000	 Coefficient of human inequality	Inequality in life expectancy (%) 2010- 2015	Inequality- adjusted life expectancy index	Inequality in education(%)	Inequality- adjusted education index	Inequality in income (%)
	=	count	195.000000	195.000000	188.000000	188.000000	190.000000	188.000000	193.000000	188.000000	188.000000	144.000000	151.000000	185.000000	185.000000	160.000000	158.000000	154.000000
		mean	97.000000	98.000000	94.319149	0.698835	71.267368	8.371277	17239.502591	2.239362	-0.021277	0.773611	20.609934	17.155135	0.666151	20.278125	0.527722	24.115584
		std	56.435804 0.000000	1.000000	54.435501 1.000000	0.155153 0.352000	8.328811 48.900000	3.098028 1.400000	19052.801649 294.000000	17.309425 -79.000000	5.440940 -35.000000	0.762704	10.210353 5.300000	12.197865 2.600000	0.190741	14.631437	0.224236 0.134000	10.511501 8.900000
		25%	48.500000	49.500000	47.750000	0.577750	65.550000	6.075000	3527.000000	-7.000000	-2.000000	0.425000	10.950000	6.700000	0.524000	6.575000	0.324000	16.300000
		50%	97.000000	98.000000	94.500000	0.728500	73.300000	8.650000	10383.000000	4.000000	0.000000	0.760000	20.200000	13.600000	0.707000	17.450000	0.527000	21.300000
D .					139.750000	0.811250	77.000000	11.125000	23394.000000	13.000000	2.000000	1.150000	29.500000	26.100000	0.813000	34.525000	0.732250	29.475000
Data		max	194.000000	195.000000	188.000000	0.949000	84.200000	13.400000	129916.000000	48.000000	15.000000	3.630000	44.800000	46.200000	0.959000	48.300000	0.921000	56.400000
Overview																		

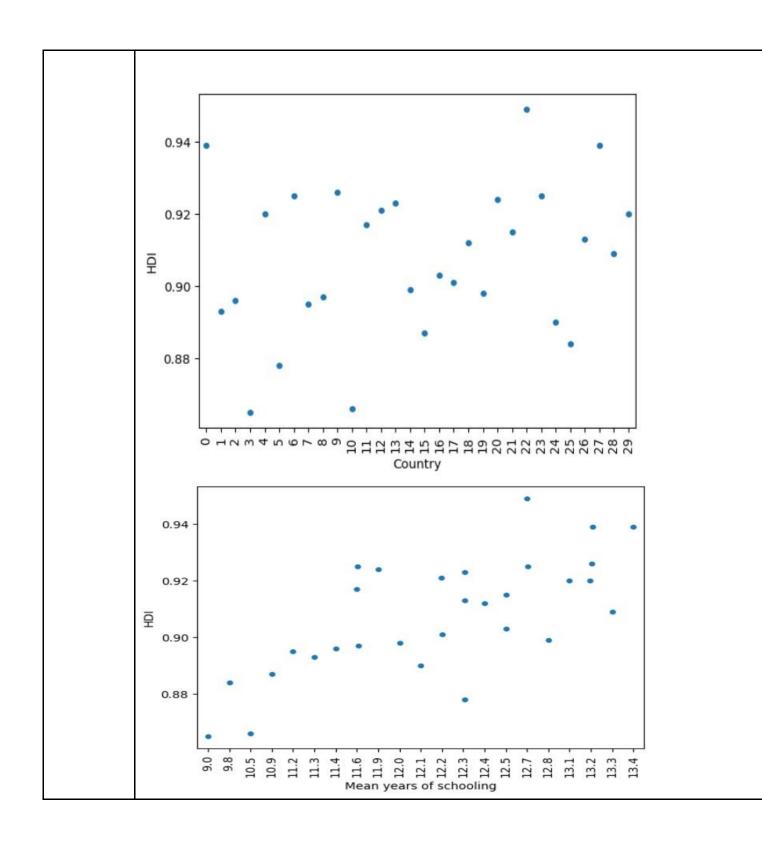




Univariate
Analysis

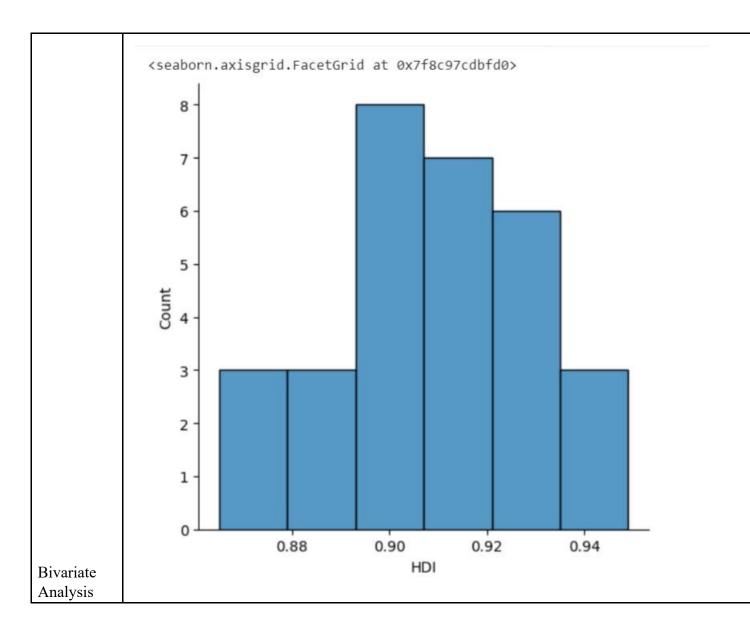






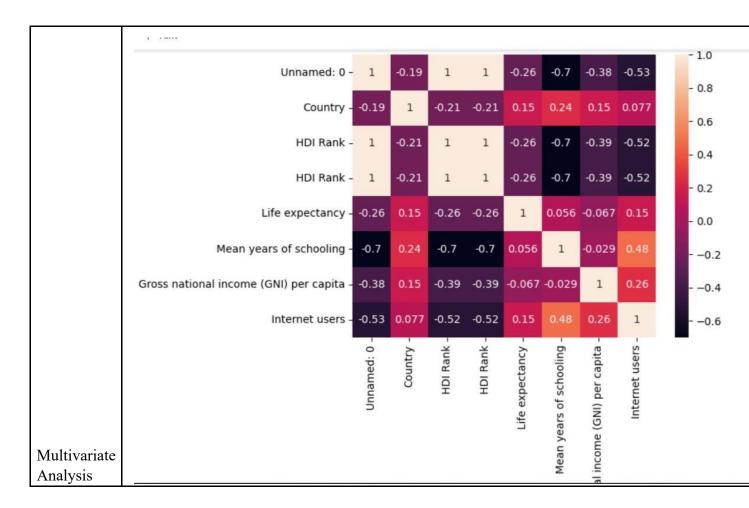












Outliers and Anomalies	-
Data Preproces	ssing Code Screenshots





Deciding Data																									
	I	[4]	data=p	od.rea	ad_c	sv(" <u>/</u>	content/	/HDI.cs	<u>v</u> ")																
Community Comm	Ţ	[5]	data.h	nead()																				
1 2 Augusta 20 Color E3 11 12 Augusta 20 Color E3 11 11 20 Augusta 12 Color E3 11 11 12 Augusta 12 Color E3 11 12 Augusta 12 Color E3 11 12 Augusta 12 Augusta E3 Color E3 E3 E3 E3 E3 E3 E3 E	:	⊕	Un			d	Country		HDI		ancv	years of	nation inco (GN	mal ome WI) oer	per of capita i rank minus HDI	in HDI rank 2010-	. of h	ient :	in life ectancy) 2010- e	adjusted life expectancy		in e	djusted ducation	in income	
2 2 3 September 20 Color Septemb			0		0	1	Norway	1.0	0.949		81.7	12.7	6761	4.0	5.0	0.0	***	5.4	3.3	0.918		2.4	0.894	10.4	0
## A 5 Gentlay 4.0 G.000 81.1 13.2 4500.0 13.0 G.0 7.0 3.7 G.005 2.8 G.008 14.8 Source of the control of th																									
## 4 5 Demail 50 0.005 00.4 127 440100 13.0 2.0 7.0 3.8 0.004 3.0 0.006 14.3 ## Committee of Committe																	***								
Color Colo					8 .																				
			5 rows	× 82 (colum																				
Control Cont			4																						
	Loading Data																								
Handling C		0	data_fi data_fi	illed.f	illna	data_f	illed.mea	n(),inpla	Life	Mean years of	national incom (GNI	s per C l capita i e rank) minus	n HDI rank 2010-		of human o	in life expectancy (%) 2010-	adjusted life expectancy	in	adjusted education	in income	adjusted income	inequality (Quintile ratio)	inequality (Palma ratio)		
## April Data 1		()	0	1	22	10 09	49	81.7	12.7		a rank			5.400			2 400	0.89400	10.400	0.88200				
Handling 3			1	1	2	0	2.0 0.9	39	82.5	13.2	42822	0 19.0	1.0		8.000	4.300000	0.921000	1.900	0.92100	17.700	0.75300	6.000000	1.400000		
Handling 4																									
Handling The content of the property of the									7777	12.7	44519.0	0 13.0	2.0		7.000			3.000	0.89600	14.300	0.78900	4.500000	1.000000		
Handling 8																									
## Display the first for rows of the encoded dataset data, encoded-head() 1		8																							
Missing Data 11 11 12 11 12 0 0 1917 842 116 542850 20 30 - 8888 2,80000 0,869000 6124 0,81848 16184 0,70008 5,45833 1245833 14 11 12 11 12 0 0 1917 842 116 542850 20 30 - 8888 2,80000 0,869000 6124 0,81848 16184 0,70008 5,45833 1245833 15 import pands as pd from sklears, preprocessing import OneHottRooder # Edentify categorical columns categorical_columns data_encoded = pd.get_dumies(data1, columnscategorical_columns, drop_first=True) # Oisplay the first fear rows of the encoded dataset data_encoded = pd.get_dumies(data1, columnscategorical_columns, drop_first=True) # Oisplay the first fear rows of the encoded dataset data_encoded = pd.get_dumies(data1, columnscategorical_columns, drop_first=True) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of the encoded dataset data_encoded.pead(columns) # Oisplay the first fear rows of th	Handling																								
import pandss as pd from sklearm.preprocessing import OneNotEncoder # Identify categorical columns = categorical_columns = data_select_dtypes(includew['object']).columns																									
# Identify categorical columns categorical columns datal.escoded = pd.get_dnamies(data), columns: # Apply One-Hoth Encoding to categorical columns data_escoded = pd.get_dnamies(data), columns: # Display the first few rows of the encoded dataset data_escoded.head() Unnamed: Id. Country HOI	rissing Data		,	10	16	-0	12.0 0.9	11	04.2	11.0	J4203.0	-2.U	9.0		0.000	2.00000	0.909000	0.124	0.01048	10.104	0.70006	0.443033	1.240033		
Unnamed Horizontal Part			from sk # Ident categor # Apply data_en # Displ	clearn. tify car rical_c One-H ncoded	preprotegor olumn lot En	ical co s = dat coding get_dur t few r	olumns tal.select to catego mmies(data	t_dtypes(orical co o1, colum	include= lumns ns=categ	orical_co		#rop_first=T	rve)												
Data 1 1 2 0 20 0939 825 132 428220 190 1.0 80 43 0921 19 0921 177 0753 60 14 2 2 3 27 20 0939 83.1 13.4 563640 70 00 84 38 0934 57 0840 157 0806 49 12 3 3 4 9 40 0926 81.1 132 450000 130 00 70 37 0905 26 0891 148 0787 46 1.1 4 4 5 6 50 0925 80.4 127 445190 130 2.0 70 38 0894 30 0896 14.3 0789 4.5 1.0		Ž.	Unna	amed:	Id C	ountry		HDI expe	ctancy	years of	nation inco (GN	al capita me rank II) minus er HDI	in HDI rank . 2010-	200	of human	in life expectancy (%) 2010-	adjusted life expectancy	Inequali education(in educati	ed in incom	y adjuste	inequali d (Quinti	ity inequal ile (Pa io) rat	ity lma io) co	
Data 2 2 3 27 20 0939 83.1 13.4 563640 7.0 0.0 84 3.8 0.934 5.7 0.840 15.7 0.806 4.9 1.2 3 3 4 9 4.0 0.926 81.1 13.2 45000.0 13.0 0.0 7.0 3.7 0.905 2.6 0.891 14.8 0.787 4.6 1.1 4 4 5 6 50 0.925 80.4 12.7 44519.0 13.0 2.0 7.0 3.8 0.894 3.0 0.896 14.3 0.789 4.5 1.0																									
Data 4 4 5 6 50 0925 80.4 12.7 44519.0 13.0 2.0 7.0 3.8 0.894 3.0 0.896 14.3 0.789 4.5 1.0																									
Section 100 cells mare	Data																								
Fransformation 1	Transformation																								





Feature Engineering	Attached the codes in final submission.
Save Processed Data	-