



Data Collection and Preprocessing Phase

| Data Conccion and | 1 reprocessing r hase |
|-------------------|--|
| | |
| Date | 28 June 2024 |
| | |
| Team ID | 739785 |
| 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.

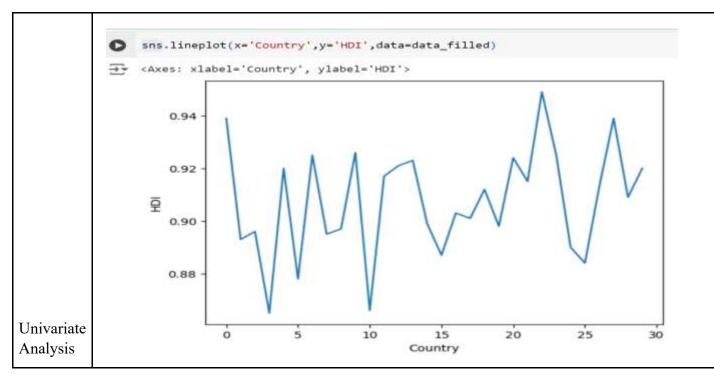




| Section | Desc | riptio |)n | | | | | | | | | | | | | | |
|---------|--------------------|--|--|--|--|--|---|--|-------------------------------------|---|--|--|---|--|--|---|--|
| | | riptiv | e Ana | alysis | S:- | LIFE . | hair | Gross sytemal | ONI per capits | Change in MEI rank | Antraga amual MII | Confficient of house | Inequality is life something. | Inequality- adjusted | Inequality | Inequality- stjusted | Despulity |
| | | Unramed: 8 | tri | HCC Name | 100 | espectancy | years of schooling | income (GMI) per capita | minus HOI rank | 2810-2015 | greath 1598-1888 | irequitty | (N) 2018- 2015 | | eturation(X) | aducation index | (10) |
| | cour | | 1# 185.000000 | 160 Narii 180 000000 | 1000000 381 | | | | minus HOL | | growth | | DO 2018- | espectancy | ******(X) | | |
| | | e 195.000000 | | | | espectancy | schooling | per capita | mines HOI rask | 2910-2015 | 20045 2000-2000 144,000000 0.773611 | irequiity | (N) 2618- 2615 185,000000 17,155135 | espectancy Endex | | index | (N) 154300000 34,115584 |
| | cour | e 195,000000 e 17,000000 | 195.000000 | 182 000000 | 186 000000 | 193.000000 | \$69000000 18E 000000 | per capita 103,000000 | 158 000000 2 200042 17 300425 | 2818-2615 188.000000 4.021277 5.442040 | 97045 1999-1999 144.000000 0.773611 0.760704 | | (N) 2628- 2625 185,000000 17,155135 12,107865 | 155,000000 0 190741 | 160 000000 30 378125 14 621657 | 118 (00000) | (10) 154 300000 24 115584 10 511501 |
| | coor | e 195,000000 e 87,000000 i 56,435804 | 195.000000 98.000000 96.425804 1.000000 | 188.000000 94.319148 54.425001 1.000000 | 188 000000 0 688335 0 185152 0 353000 | 195,000000 71,267568 8,328011 43,900000 | 18E000000 8:371277 3:068029 1:400000 | per capita 163,060000 17239-562591 16052-861648 254,060000 | 12000000 179.000000 179.0000000 | 2918-2919 188.000000 40.03377 5.440940 35.000000 | 979-1000 144.00000 0.773611 0.700704 -1.55000 | 15120000 30.68934 10.212353 5.30000 | (K) 2618- 2625 185,000000 17,155135 12,197885 2,600000 | 6990151 6990151 6 990741 6 350000 | 160 000000 30 378125 14 631467 1 450000 | 158 00000 0:527722 0:224298 0:134000 | (N) 154 800000 24 15584 10 51501 8 900000 |
| D. | COOR | e 195,000000 e 87,000000 i 56,435804 | 165.000000 08.000000 165.435804 | 182 000000 84 319148 54 425501 | 15E 000000 0 408935 0 185452 | 199.000000 71.297368 8.328311 43.90000 85.550000 | 166 000000 8:371277 3:068008 | per capital 103 000000 17239 502591 19052 801048 | 158 000000 2 200042 17 300425 | 2918-2915 188.000000 -0.021277 5.440940 -25.000000 -2.000000 | growth 1598-1600 144.00000 0.773611 0.700704 -1.55000 0.425000 | 151 200000 30 669934 10 210353 5.30000 10 250000 | (K) 2618- 2615 185,000000 17,155135 12,107885 2,600000 6,700000 | 489454 456 080000 5 860151 5 190741 5 25000 5 524000 | 160 000000 20 278125 14 631467 1 400000 6 525000 | 168 000000 0:527722 0:224236 | (N) 154,000000 24,115584 10,511501 8,900000 78,300000 |
| Data | coor mas std | e 195 000000 87 000000 96 435804 0 100000 | 195.000000 98.000000 96.425804 1.000000 | 188.000000 94.319148 54.425001 1.000000 | 188 000000 0 688335 0 185152 0 353000 | 195,000000 71,267568 8,328011 43,900000 | 18E000000 8:371277 3:068029 1:400000 | per capita 163,060000 17239-562591 16052-861648 254,060000 | 12000000 179.000000 179.0000000 | 2918-2919 188.000000 40.03377 5.440940 35.000000 | 979-1000 144.00000 0.773611 0.700704 -1.55000 | 15120000 30.68934 10.212353 5.30000 | (K) 2618- 2625 185,000000 17,155135 12,197885 2,600000 | 6990151 6990151 6 990741 6 350000 | 160 000000 30 378125 14 631467 1 450000 | 158 00000 0:527722 0:224298 0:134000 | (N) 154 800000 24 15584 10 51501 8 900000 |

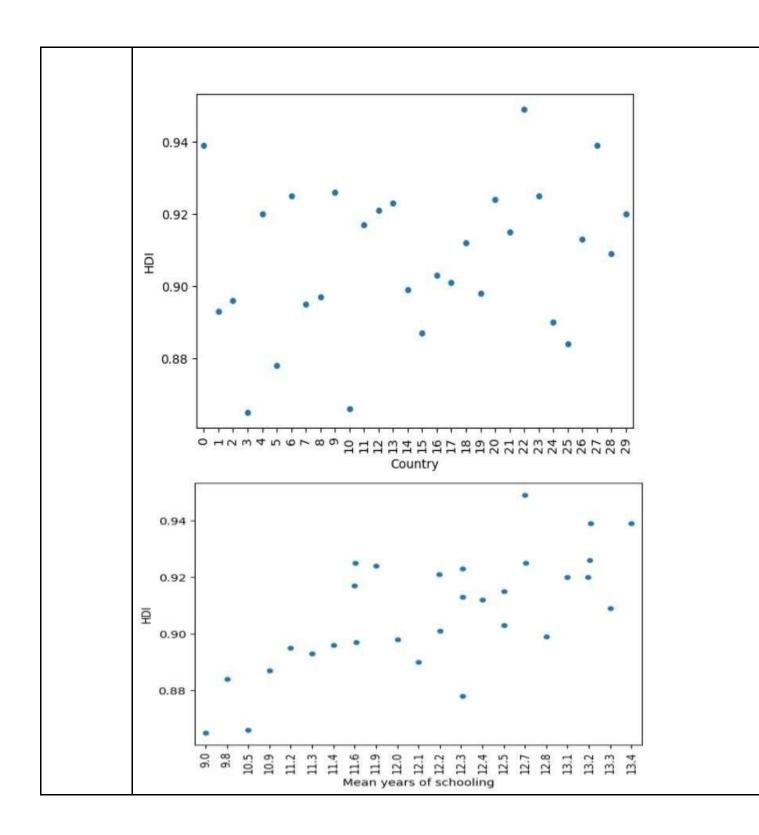






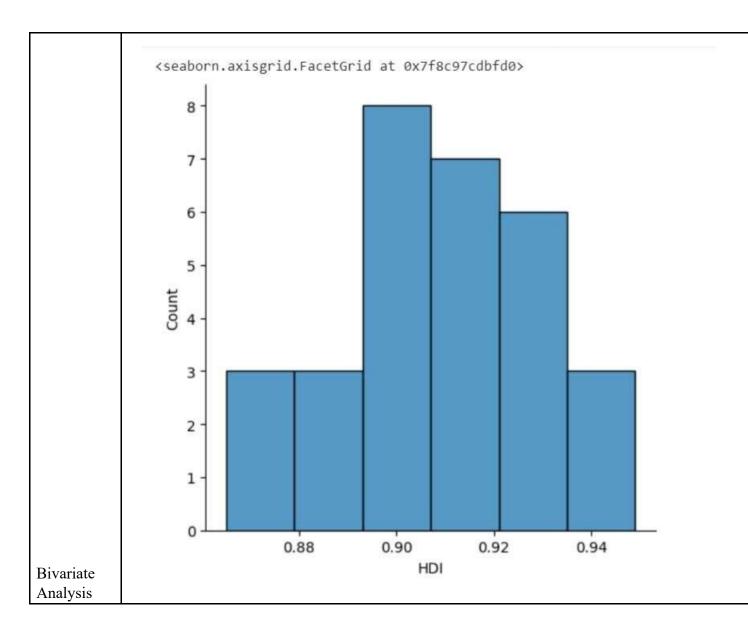






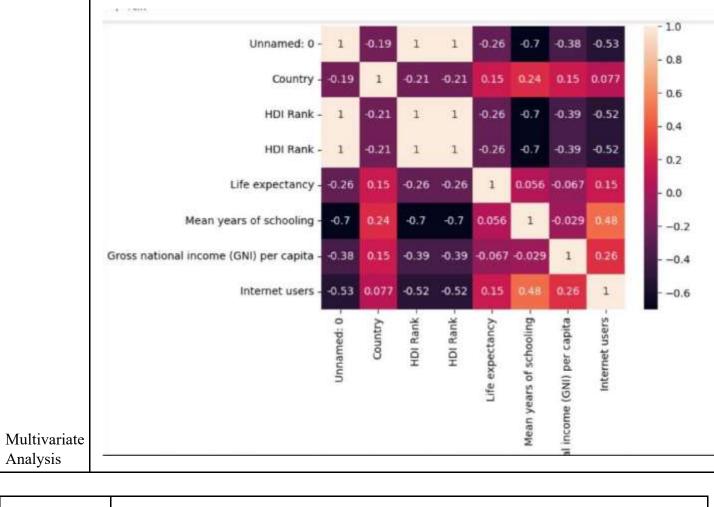












| Outliers and Anomalies | - |
|---------------------------|-----------------------|
| Data Preproces | sing Code Screenshots |





| | 747 | data | , tead() | | | | | | | | | | | | | | | | | |
|------------------------|---|--|--|--|---|---|--|---|--|---|---|---|--|---|---|---|--|--|--|--|
| | | | 11000// | | | | | | | | | | | | | | | | | |
| | 7 | | Urramed : | | Country | MDI Rank | | Life expectancy | | of Incom | capita capita renk capita | | 11 97 | icient human exp | | nequality- adjusted life expectancy index | Ehequalit 1 education(X | educat | ted in income | |
| | | 0 | - 4 | э т. | Nova | 1.0 | 0.049 | 81.3 | 12 | 2.7 67614 | 0 5.0 | 6.0 | 2 | 5.4 | 3.3 | 0.018 | 2 | 4 0 | 894 104 | |
| | | 1 | - 51 | 2 | Austrati | 2.0 | 0.009 | 82.5 | 33 | 3.2 42822 | 0 19.0 | 1.0 | - | 8.0 | 4.3 | 0.921 | 1 | 9 0 | 821. 17. | |
| | | 2 | - 2 | 2 3 | Switzertoni | 2.0 | 0.609 | 83.1 | 13 | 3.4 56964 | 0 7.0 | 0.0 | | 6.4 | 3.8 | 0.654 | 5 | 7 0 | 840 15.7 | |
| | | 3 | (3 | 1 + | Germany | 4.0 | 0.626 | 81:1 | 13 | 3.2 45000 | 0 13.0 | 0,0 | - | 7.0 | 3.7 | 0,905 | 2 | 6 0 | 881 14.1 | |
| | | 4 | | 4 3: | Dennari | 5.0 | 0.105 | 80.4 | 12 | 2.7 44519 | 0 13.0 | 2.0 | 2 | 7.0 | 3.8 | 0.894 | 3 | 0 0 | 890 143 | |
| | | 5 109 | es ≈ 82 cc | plums | | | | | | | | | | | | | | | | |
| | | 4 | | | | | | | | | | | | | | | | | | |
| pading Data | | | | | | | | | | | | | | | | | | | | |
| | | | | | ercoled (FEE) | | | | | | | | | | | | | | | |
| | 1 | no. | a_filler | Liber | beta_Filled.e | Hert.), 259 | /Lane-tree | 11- | | | | | | | | | | | | |
| | 1 | Đ | insted: | Is the | Heli yotu | HES MAN | LLPs peritoricy s | Name not years of integling | (ART) cons per cons | er Charge he je HOT ek rank sa 3658- st 3655 | Coefficia of him bromals | m enymitten | e espectar | nt Irequals to cy mhuntano | to moreover | ed to income | Deline | Income requelity in (Quintale regia) 2018-2028 DE | Deceme quality (False (utio) sp-pers | |
| | | | 0 | 1 | 22 10 / | 100 | 817 | | 70 kg - 5 | | 3.4 | E 1300 | 0 89400 | 80 2 | 40 100 | 01 10.400 | 0.000 | 3.00000 | 2 9000000 | |
| | | | - 1 | 2 | 1 20 1 | | 40.9 | | 0023 19 | | 0.0 | | | | NO 1103 | | | | 1.60000 | |
| | | 1 | 3 | 1 | 9 40 1 | | 811 | | 6000 TI | | 64 | | | | 700 H943 800 H860 | | | | 1.20000E 1.10000E | |
| | | 4 | . 6 | 1 | 0 00 0 | | 10.4 | 127 4 | | | 7.0 | | | | 100 1100 | | | | F.000000 | |
| | | 4 | 5 | 4 | 21 10 1 | | 30.2 | | 11122 .0 | | | | | | 1313 | | | | 1,34503 | |
| | | | 1.0 | 10 | 20 10 1 | | 811 | | | | - 68 | 17000 | | | | | | | 1.300006 | |
| | | , | + | 4 | 12 88 9 | | | 12.2 4 | O/1907 175 | 0 11 | 3.7 | E 17000 | 0 89000 | 10 31 | 100 1100 | 00 76,300 | 0.75800 | 5.3000001 | | |
| | | 7 | 1 0 | | 13 80 1 | | 817 | 122 3 | | 10 10 | 87 | | | | 500 8.860 500 8.664 | | | | 1.000000 | |
| andling | | 7 | | 1 11 | 4 100 | 1801 | 10.1 10.1 | 127 3 121 4 | 17960 26 12920 12 | 10 78 10 14 | 67 | E 47000 | 0 89570 0 89520 | 00 21 00 31 | 100 1 000 100 1 100 | 00 17.400 | 0.7900 67500 | 6.00000 5.60000 | 1,300000 | |
| | | | | 7 | 28 70 0 | HIJA | 817 | 11.8 : 4 | 41260 8 | 10 21 | 44 | | 0 8940 | 00 45 | 200 1 1014 | 0.00 | 0.020 | #3000E | 1.000000 | |
| | *1 | n elies Sectify agents | T 0 0 0 0 11 11 11 11 11 11 11 11 11 11 1 | 17 12 centing cell order + detail | 17 80 1 4 100 1 28 100 1 11 120 1 | 0.000 0.000 0.000 0.000 discoler | 10.7 0.3 76.2 96.3 | 12.2 3 | 1760 36 1250 12 1260 13 1260 13 | 10 14 10 14 10 38 | 87 | E 6 1000 | 0 8:9570 0 8:9520 0 8:9980 | 00 21 00 31 00 54 | 100 1 100 | 00 11.700 00 17.400 00 27.000 | 0 74600 6 75606 0 94010 | 6-300000 5-800000 9-900000 | 200 × 71 71 | |
| | # 17 187 # A MA | tt en rillea (destrict agentral agentral | TO THE TOTAL | 17 12 coming col (crit coling to chick and | 17 NO 1 4 NO 1 28 NO 1 11 NO 1 1 NO 1 1 NO 1 | 0.000 000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0. | NIT 0.3 Pt2 913 dec[*ejer degeliek] | 122 3 132 4 132 5 136 5 | THE STATE OF THE S | 10 78 10 18 10 38 10 28 | 97 97 98 | E 2 #0000 E 4 70000 E 2 #0000 | 0 89730 0 89730 0 89800 0 89800 | 00 21 00 31 00 54 | 100 1 000 100 1 100 100 1 100 | 00 11.700 00 17.400 00 27.000 | 6 75000 6 75000 6 95000 6 75000 | 6.00000 5.40000 9.40000 5.44000 | 1.30000 2.00000 1.340833 | |
| | # 7 199 # A 198 # 3 8 19 | tt en rillea (destrict agentral agentral | of the state of th | III 12 conting col crim colony colo | 17 SO I A 100 | 0.000 000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0. | SET SEE SEE SEE SEE SEE SEE SEE SEE SEE | 122 3 121 4 122 5 126 5 126 5 | principal and a second and a se | 10 78 19 15 19 28 | 57 CR 68 | E 28000 E 47000 E 5,0000 E 2,0000 E 2,0000 | 0 89730 0 89730 0 89800 0 89800 | 00 21 00 31 00 54 00 6 | to 1 | 00 11 70 00 00 00 00 00 00 00 00 00 00 00 00 | 0 PRIOR C 75000 C 7500 | # DONORS | 1,30000 2,00000 1,240633 1,240633 1016 1016 1016 1016 1016 1016 1016 | |
| | 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | the section of the se | T S S S S S S S S S S S S S S S S S S S | III 12 consising to distribute the least to describe the least t | 12 00 1 2 100 1 2 100 1 | ONE OF THE PROPERTY OF THE PRO | ELT CLS TREE SELECTION OF THE PROPERTY OF SELECTION OF SE | 122 3 121 4 132 5 130 5 130 5 | ONI Capeto Charles on the Capeto Charles on | 10 78 12 13 14 15 16 16 16 16 16 16 16 16 16 16 16 16 16 | 57 QX 60 60 FFacaet of Summ in equality | e 2 8000 to 4 7000 to 6 9000 to 2 80000 to 2 800000 to | o service appoints of the control of | 00 21 00 34 00 54 00 64 5 6 6 6 6 6 6 7 6 7 6 7 7 8 7 8 8 8 8 8 8 | 200 5.000 200 2.000 200 3.000 201 2.000 201 201 2.000 201 201 2.000 201 201 2.000 201 2.000 201 2.000 201 2.000 201 2.000 201 2.000 | 00 11 1000 00 17 400 00 27 000 48 96 104 00 | 0 7000 0 7500 0 6000 0 7500 0 7500 1 3000 1 3000 | # SOCORE S AND S A | 1.30000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.000000 2.000000 2.000000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.00000 2.000000 2.000000 2.000000 2.00000000 | |
| andling issing Data | # 7 199 # A 198 # 3 8 19 | the section of the se | of the state of th | 17 12 12 12 12 12 13 14 15 16 16 17 16 16 17 16 17 17 17 17 17 17 17 17 17 17 17 17 17 | 17 SO I A SOO I A SO | OSP OSD | ELT CLL TR2 ELT TR2 EL | 122 3 121 4 122 5 128 5 | 001 001 001 001 001 001 001 001 001 001 | 10 78 19 15 19 28 | SA CA | 0 2 8000 0 4 7000 0 5 0000 0 2 8000 0 3 355 3 3 355 3 3 355 3 3 355 3 3 355 3 3 355 3 3 355 | 0 89700 0 89800 0 89800 0 89800 1 10 | Brangelity (1) (N) (N) (N) (N) (N) (N) (N) (N) (N) (N | tomassiniy significant behaviors | 00 11 100 00 17 400 00 27 000 48 96 104 96 104 00 000 000 000 000 000 000 000 000 00 | 0.7000 0.7500 0.6000 0.75008 | ESCOCIO SANOCIO SAN | 1.30000 2.00000 1.3000 1.30000 1.30000 1.30000 1.30000 1.30000 1.30000 1.30000 1.3000 1.30000 | |





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