# Assignment5

August 3, 2024

## 1 Name:Krupanidhi Jena

## 2 Reg NO:2141010032

## 3 Sec:16(p)

## 4 B.tech (CSE)

### 4.1 Perform data preprocessing on a dataset (e.g., Titanic dataset) including cleaning, handling missing values, transformation, normalization, encoding,and feature engineering for predictive modeling.

## 5 importing libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[3]:
```
df=pd.read_csv('Per Capita GDP of All Countries 1970 to 2022.csv')
df.head(30)
```

[3]:

| | Sr.No | Country | 1970 | 1971 | \ |
|---|---|---|---|---|---|
| 0 | 1 | Afghanistan | 162.642176 | 166.224832 | |
| 1 | 2 | Albania | 974.554499 | 975.216820 | |
| 2 | 3 | Algeria | 373.670080 | 380.125666 | |
| 3 | 4 | Andorra | 5008.901637 | 5305.716707 | |
| 4 | 5 | Angola | 631.514939 | 648.910896 | |
| 5 | 6 | Anguilla | 621.101316 | 725.213850 | |
| 6 | 7 | Antigua and Barbuda | 524.913170 | 635.428324 | |
| 7 | 8 | Argentina | 1425.357894 | 1678.388787 | |
| 8 | 9 | Aruba | 3343.551506 | 3539.154720 | |
| 9 | 10 | Australia | 3590.037728 | 3884.030700 | |
| 10 | 11 | Austria | 2059.268192 | 2381.629933 | |
| 11 | 12 | Bahamas | 4259.305052 | 4286.281941 | |
| 12 | 13 | Bahrain | 1896.176458 | 1969.702274 | |
| 13 | 14 | Bangladesh | 91.732366 | 81.575518 | |
| 14 | 15 | Barbados | 903.012578 | 1011.028545 | |

| | | | | |
|---|---|---|---|---|
| 15 | 16 | Belgium | 2773.481236 | 3085.109750 |
| 16 | 17 | Belize | 230.906758 | 287.472565 |
| 17 | 18 | Benin | 106.519449 | 110.668874 |
| 18 | 19 | Bermuda | 5104.598512 | 5726.867894 |
| 19 | 20 | Bhutan | 206.680778 | 214.769378 |
| 20 | 21 | Bolivia (Plurinational State of) | 220.272271 | 233.407375 |
| 21 | 22 | Botswana | 112.786843 | 145.847758 |
| 22 | 23 | Brazil | 367.125782 | 400.134371 |
| 23 | 24 | British Virgin Islands | 1212.165327 | 1021.866809 |
| 24 | 25 | Brunei Darussalam | 1686.777121 | 1789.110415 |
| 25 | 26 | Bulgaria | 1048.590520 | 1029.050237 |
| 26 | 27 | Burkina Faso | 90.208795 | 90.534198 |
| 27 | 28 | Burundi | 70.168159 | 70.494594 |
| 28 | 29 | Cabo Verde | 253.916821 | 280.349382 |
| 29 | 30 | Cambodia | 114.439733 | 114.442138 |

| | 1972 | 1973 | 1974 | 1975 | 1976 | \ |
|---|---|---|---|---|---|---|
| 0 | 141.365323 | 149.744069 | 181.598776 | 194.669036 | 205.674097 | |
| 1 | 976.521166 | 978.617771 | 981.265874 | 984.901966 | 989.749878 | |
| 2 | 496.991248 | 624.193795 | 875.023948 | 989.243356 | 1075.725904 | |
| 3 | 6285.301526 | 7824.531307 | 9078.009005 | 10076.494834 | 9817.312977 | |
| 4 | 644.659613 | 762.730111 | 827.378464 | 589.868220 | 547.946934 | |
| 5 | 850.887123 | 957.808007 | 1049.881766 | 1140.561267 | 1086.473476 | |
| 6 | 871.867830 | 1215.586135 | 1300.717871 | 1333.904761 | 1109.562071 | |
| 7 | 1556.526579 | 1724.205313 | 1907.094439 | 1996.493745 | 2080.525773 | |
| 8 | 3725.350305 | 3890.231590 | 4052.355158 | 4220.081349 | 4410.324290 | |
| 9 | 4466.183034 | 6349.496844 | 7458.406447 | 7857.260452 | 8374.850181 | |
| 10 | 2924.947592 | 3891.702402 | 4631.855170 | 5286.897178 | 5679.783584 | |
| 11 | 4493.308784 | 5024.039305 | 5962.291182 | 5454.223469 | 5885.411989 | |
| 12 | 2367.250664 | 3196.301037 | 4150.198216 | 4154.484554 | 5681.200623 | |
| 13 | 89.573360 | 95.156689 | 110.858841 | 113.449744 | 123.688300 | |
| 14 | 1151.329867 | 1397.942497 | 1663.330643 | 1945.368774 | 2100.621610 | |
| 15 | 3835.929146 | 4907.873258 | 5747.637943 | 6726.770164 | 7276.188739 | |
| 16 | 367.842221 | 589.632882 | 775.941108 | 884.028604 | 682.906881 | |
| 17 | 133.354183 | 159.761063 | 167.784664 | 199.472350 | 207.426900 | |
| 18 | 6361.395010 | 7257.860841 | 8390.519081 | 9233.392266 | 10316.059272 | |
| 19 | 219.656318 | 239.095644 | 272.404723 | 246.316138 | 242.802694 | |
| 20 | 266.735184 | 265.051936 | 428.605157 | 467.805263 | 524.111294 | |
| 21 | 183.040570 | 257.793130 | 340.548656 | 376.534984 | 387.465855 | |
| 22 | 502.649167 | 694.649640 | 869.604041 | 998.737047 | 1153.288405 | |
| 23 | 1056.718684 | 1182.915253 | 1396.465109 | 1513.845653 | 1627.158899 | |
| 24 | 2364.014359 | 3649.148266 | 8738.105417 | 9190.990734 | 10827.978164 | |
| 25 | 1196.232688 | 1417.550265 | 1531.647164 | 1343.908397 | 1418.145113 | |
| 26 | 113.660892 | 129.249127 | 139.749595 | 162.986762 | 153.009253 | |
| 27 | 68.829523 | 85.537148 | 94.637389 | 110.655636 | 119.380471 | |
| 28 | 313.553639 | 383.896232 | 381.439736 | 440.132704 | 407.185546 | |
| 29 | 111.730841 | 94.327387 | 96.574950 | 106.966181 | 120.382292 | |

| | 1977 | … | 2013 | 2014 | 2015 | \ |
|---|---|---|---|---|---|---|
| 0 | 232.778698 | … | 631.060444 | 596.048898 | 554.004007 | |
| 1 | 994.937066 | … | 4425.430822 | 4586.606188 | 3950.390255 | |
| 2 | 1225.090692 | … | 5519.777576 | 5516.229431 | 4197.417286 | |
| 3 | 10384.507961 | … | 44735.004774 | 45619.782548 | 38869.870777 | |
| 4 | 578.443953 | … | 5228.511165 | 5371.216648 | 4130.930080 | |
| 5 | 1210.224756 | … | 20569.480211 | 22047.422883 | 22756.280768 | |
| 6 | 1260.564830 | … | 13350.149137 | 14004.811212 | 14861.882707 | |
| 7 | 2304.812606 | … | 14469.003850 | 13241.393385 | 14908.621297 | |
| 8 | 4624.873192 | … | 26514.863550 | 26940.258722 | 28419.243101 | |
| 9 | 8207.700253 | … | 66820.508917 | 62402.632438 | 52321.003543 | |
| 10 | 6812.731674 | … | 50718.407330 | 51719.250909 | 44179.468449 | |
| 11 | 6117.865558 | … | 27404.162248 | 28554.857734 | 29821.354971 | |
| 12 | 6798.708509 | … | 25790.856791 | 25464.800667 | 22795.370758 | |
| 13 | 127.333275 | … | 1130.874771 | 1282.106878 | 1449.562923 | |
| 14 | 2376.961899 | … | 16923.590920 | 16964.752264 | 17023.694365 | |
| 15 | 8464.645137 | … | 46981.053794 | 47838.536915 | 41086.169118 | |
| 16 | 860.294383 | … | 5865.470901 | 6031.548403 | 6098.600079 | |
| 17 | 221.598241 | … | 1213.950476 | 1249.840095 | 1040.863784 | |
| 18 | 11184.709840 | … | 101813.371965 | 101258.622050 | 105404.472317 | |
| 19 | 259.909829 | … | 2409.430145 | 2589.893727 | 2746.377297 | |
| 20 | 515.093216 | … | 2853.797161 | 3022.462884 | 2975.648811 | |
| 21 | 474.406397 | … | 6436.601866 | 6844.033074 | 5869.729556 | |
| 22 | 1311.006139 | … | 12258.564849 | 12071.404655 | 8783.214413 | |
| 23 | 1658.406126 | … | 39499.947657 | 41504.608056 | 43647.074848 | |
| 24 | 12759.537324 | … | 43949.635599 | 41035.760222 | 30681.450167 | |
| 25 | 1444.620074 | … | 7511.506523 | 7742.783575 | 6945.432877 | |
| 26 | 180.646058 | … | 762.087122 | 766.350661 | 631.648325 | |
| 27 | 138.319814 | … | 241.549503 | 257.822931 | 262.001758 | |
| 28 | 379.053210 | … | 3758.638926 | 3735.706612 | 3167.796310 | |
| 29 | 113.917657 | … | 1015.220881 | 1098.074538 | 1170.742816 | |

| | 2016 | 2017 | 2018 | 2019 | 2020 | \ |
|---|---|---|---|---|---|---|
| 0 | 526.164709 | 534.020133 | 513.981256 | 512.910125 | 512.749930 | |
| 1 | 4117.006970 | 4521.752217 | 5268.116608 | 5359.239078 | 5288.989573 | |
| 2 | 3967.200546 | 4134.936055 | 4171.795392 | 4021.983602 | 3354.157303 | |
| 3 | 39904.138181 | 40632.206416 | 42904.822583 | 41328.634079 | 37207.175636 | |
| 4 | 3468.520268 | 4042.681382 | 3240.862207 | 2569.629757 | 1639.951023 | |
| 5 | 21619.849507 | 18718.318718 | 21240.732202 | 24495.015171 | 16582.659014 | |
| 6 | 15862.651663 | 16110.312400 | 17514.355864 | 18187.779712 | 15303.278112 | |
| 7 | 12767.455354 | 14609.783370 | 11816.649537 | 10006.693111 | 8565.153126 | |
| 8 | 28449.728927 | 29329.092343 | 30918.483584 | 31902.757333 | 24008.127822 | |
| 9 | 54054.216565 | 57489.804356 | 58327.726542 | 54407.644621 | 56001.553064 | |
| 10 | 45277.769467 | 47429.535841 | 51466.603137 | 50067.472961 | 48839.268984 | |
| 11 | 30124.376154 | 31303.530973 | 32400.956691 | 33564.695734 | 23998.168238 | |
| 12 | 22867.275453 | 24349.836847 | 25415.750065 | 25869.206182 | 23433.491494 | |

| | | | | | |
|---|---|---|---|---|---|
| 13 | 1651.719965 | 1785.836613 | 1931.478758 | 2110.916696 | 2230.583106 |
| 14 | 17381.365087 | 17881.563253 | 18271.252253 | 19063.102291 | 16882.501523 |
| 15 | 42038.215343 | 44162.256256 | 47455.523320 | 46554.245138 | 45517.826647 |
| 16 | 6099.957102 | 6049.391948 | 6001.024262 | 6134.215233 | 5185.158070 |
| 17 | 1049.820303 | 1095.274133 | 1194.437603 | 1170.965879 | 1240.733104 |
| 18 | 109426.650599 | 112539.075464 | 113690.519262 | 116333.291544 | 107559.619559 |
| 19 | 2988.978262 | 3427.173682 | 3389.777295 | 3564.617808 | 3181.340129 |
| 20 | 3013.502708 | 3280.008213 | 3471.006951 | 3472.380831 | 3068.812555 |
| 21 | 6411.553469 | 6705.339802 | 6947.818053 | 6679.198025 | 5863.185478 |
| 22 | 8680.735421 | 9896.717511 | 9121.020092 | 8883.903092 | 6923.700197 |
| 23 | 47270.587444 | 43569.527611 | 45719.465963 | 49448.872917 | 46222.905209 |
| 24 | 26761.567232 | 28186.946531 | 31240.444448 | 30748.320153 | 27179.344445 |
| 25 | 7442.667846 | 8237.834760 | 9319.715187 | 9767.173368 | 10058.979650 |
| 26 | 665.786066 | 711.184631 | 779.202768 | 772.542487 | 840.250143 |
| 27 | 268.733291 | 295.620936 | 288.626394 | 283.669870 | 282.431058 |
| 28 | 3312.702076 | 3534.343583 | 3860.454811 | 3928.309096 | 3223.735602 |
| 29 | 1281.105971 | 1400.899265 | 1533.315985 | 1671.914393 | 1577.911739 |

| | 2021 | 2022 |
|---|---|---|
| 0 | 378.082892 | 344.649112 |
| 1 | 6281.046775 | 6655.258391 |
| 2 | 3700.320562 | 4273.922183 |
| 3 | 42065.970756 | 42293.297081 |
| 4 | 2026.030952 | 3183.699796 |
| 5 | 19106.060927 | 28561.285003 |
| 6 | 16740.348196 | 18880.028883 |
| 7 | 10775.999836 | 13867.913268 |
| 8 | 29127.759386 | 33300.838821 |
| 9 | 67616.301208 | 67866.792484 |
| 10 | 53720.125274 | 52608.760637 |
| 11 | 28260.367048 | 31458.354292 |
| 12 | 26860.074554 | 30146.509534 |
| 13 | 2449.024978 | 2527.521270 |
| 14 | 17507.467994 | 20238.784242 |
| 15 | 51737.760269 | 49986.832591 |
| 16 | 6060.968224 | 6984.216959 |
| 17 | 1360.911474 | 1302.850067 |
| 18 | 113524.811093 | 117568.241306 |
| 19 | 3560.197894 | 3704.344581 |
| 20 | 3345.023002 | 3600.121635 |
| 21 | 7238.796096 | 7737.654680 |
| 22 | 7696.785130 | 8917.673509 |
| 23 | 46915.895058 | 46996.749944 |
| 24 | 31448.913608 | 37152.466238 |
| 25 | 12204.368627 | 13301.923710 |
| 26 | 904.474930 | 845.760531 |
| 27 | 298.170349 | 312.874378 |

```
28    3557.837775    3902.589049
29    1625.235021    1759.608024

[30 rows x 55 columns]
```

[5]:
```
df.shape
```

[5]:
```
(187, 55)
```

[104]:
```
df.isnull()
```

[104]:
```
      Sr.No  Country   1970   1971   1972   1973   1974   1975   1976   1977  \
0    False    False  False  False  False  False  False  False  False  False
1    False    False  False  False  False  False  False  False  False  False
2    False    False  False  False  False  False  False  False  False  False
3    False    False  False  False  False  False  False  False  False  False
4    False    False  False  False  False  False  False  False  False  False
..     …        …    …      …      …      …      …      …      …      …
182  False    False  False  False  False  False  False  False  False  False
183  False    False  False  False  False  False  False  False  False  False
184  False    False  False  False  False  False  False  False  False  False
185  False    False  False  False  False  False  False  False  False  False
186  False    False  False  False  False  False  False  False  False  False

     …   2013   2014   2015   2016   2017   2018   2019   2020   2021   2022
0    …  False  False  False  False  False  False  False  False  False  False
1    …  False  False  False  False  False  False  False  False  False  False
2    …  False  False  False  False  False  False  False  False  False  False
3    …  False  False  False  False  False  False  False  False  False  False
4    …  False  False  False  False  False  False  False  False  False  False
..   …    …    …    …    …    …    …    …    …    …    …
182  …   True   True   True   True   True   True   True   True   True   True
183  …   True   True   True   True   True   True   True   True   True   True
184  …   True   True   True   True   True   True   True   True   True   True
185  …  False  False  False  False  False  False  False  False  False  False
186  …  False  False  False  False  False  False  False  False  False  False

[187 rows x 55 columns]
```

## 6 finding and cleaning missing values (Preprocessing)

[9]:
```
null_values = df.isnull().sum()
print(null_values)
```

```
Sr.No      0
Country    0
1970       0
```

| | |
|---|---|
| 1971 | 0 |
| 1972 | 0 |
| 1973 | 0 |
| 1974 | 0 |
| 1975 | 0 |
| 1976 | 0 |
| 1977 | 0 |
| 1978 | 0 |
| 1979 | 0 |
| 1980 | 0 |
| 1981 | 0 |
| 1982 | 0 |
| 1983 | 0 |
| 1984 | 0 |
| 1985 | 0 |
| 1986 | 0 |
| 1987 | 0 |
| 1988 | 0 |
| 1989 | 0 |
| 1990 | 0 |
| 1991 | 5 |
| 1992 | 5 |
| 1993 | 5 |
| 1994 | 6 |
| 1995 | 6 |
| 1996 | 6 |
| 1997 | 6 |
| 1998 | 6 |
| 1999 | 6 |
| 2000 | 6 |
| 2001 | 6 |
| 2002 | 6 |
| 2003 | 6 |
| 2004 | 6 |
| 2005 | 6 |
| 2006 | 6 |
| 2007 | 6 |
| 2008 | 6 |
| 2009 | 6 |
| 2010 | 6 |
| 2011 | 7 |
| 2012 | 7 |
| 2013 | 8 |
| 2014 | 8 |
| 2015 | 8 |
| 2016 | 8 |
| 2017 | 8 |
| 2018 | 8 |

```
2019       8
2020       8
2021       8
2022       8
dtype: int64
```

[11]:
```
data_types = df.dtypes
print(data_types)
```

```
Sr.No        int64
Country     object
1970       float64
1971       float64
1972       float64
1973       float64
1974       float64
1975       float64
1976       float64
1977       float64
1978       float64
1979       float64
1980       float64
1981       float64
1982       float64
1983       float64
1984       float64
1985       float64
1986       float64
1987       float64
1988       float64
1989       float64
1990       float64
1991       float64
1992       float64
1993       float64
1994       float64
1995       float64
1996       float64
1997       float64
1998       float64
1999       float64
2000       float64
2001       float64
2002       float64
2003       float64
2004       float64
2005       float64
2006       float64
```

```
2007       float64
2008       float64
2009       float64
2010       float64
2011       float64
2012       float64
2013       float64
2014       float64
2015       float64
2016       float64
2017       float64
2018       float64
2019       float64
2020       float64
2021       float64
2022       float64
dtype: object
```

[13]:
```
stats = df.describe()
print(stats)
```

```
           Sr.No          1970          1971          1972          1973  \
count  187.000000    187.000000    187.000000    187.000000    187.000000
mean    94.000000   1077.965621   1193.635635   1387.392333   1737.889426
std     54.126395   1476.554395   1645.255727   1955.808883   2513.633214
min      1.000000     44.521365     47.939054     50.492965     56.515341
25%     47.500000    218.262278    230.269381    264.588670    313.333994
50%     94.000000    456.141164    476.565736    560.263804    692.681596
75%    140.500000   1212.608199   1243.107653   1412.715352   1914.779634
max    187.000000  11706.408105  12907.451868  15647.128859  20160.375873

                1974          1975          1976          1977          1978  \
count    187.000000    187.000000    187.000000    187.000000    187.000000
mean    2193.118218   2409.693036   2599.626234   2867.841320   3235.847755
std     3325.517167   3782.597478   4151.170437   4574.309050   5006.278328
min       62.343096     71.074239     75.712074     77.457973     82.355693
25%      380.446157    403.195817    410.993977    475.383223    540.537250
50%      869.604041    889.056832    900.504248   1033.105677   1173.577242
75%     2404.689149   2740.750028   2886.698838   3271.799736   3518.304483
max    23922.839175  27444.369312  31691.670308  35646.372841  36521.055340

       …           2013          2014          2015          2016  \
count  …     179.000000    179.000000    179.000000    179.000000
mean   …   19115.279408  19389.855328  17510.426356  17654.065018
std    …   28707.970468  29297.366217  26149.778473  26398.826760
min    …     241.549503    257.822931    262.001758    268.733291
25%    …    2083.375879   2087.905584   2127.484188   2126.149482
50%    …    6954.774166   6959.839547   6228.426299   6163.861179
```

```
75%    …   24026.889514   24752.837664   23136.104683   23782.244659
max    …  185001.796240  195511.394093  170270.150743  174294.012870

                2017           2018           2019           2020  \
count     179.000000     179.000000     179.000000     179.000000
mean    18470.677082   19556.597351   19378.379540   17893.779980
std     27265.450226   29029.398865   28746.558990   27266.242929
min       295.620936     288.626394     283.669870     282.431058
25%      2120.321972    2351.493347    2366.391819    2200.828660
50%      6705.339802    6947.818053    6955.881770    6327.598962
75%     25004.516355   26123.860763   26832.177371   22836.749167
max    173611.784301  193968.166388  199382.781207  182537.304601

                2021           2022
count     179.000000     179.000000
mean    20454.420382   21213.009641
std     32313.750589   32925.369671
min       298.170349     312.874378
25%      2435.555445    2465.901217
50%      6635.463923    6984.216959
75%     25846.571854   29958.822758
max    235132.745557  240535.043309

[8 rows x 54 columns]
```

## 7 cleaning and transformation

[15]:
```
df_filled = df.fillna(df['1991'].mean())
df_filled.head(30)
```

[15]:

| | Sr.No | Country | 1970 | 1971 | \ |
|---|---|---|---|---|---|
| 0 | 1 | Afghanistan | 162.642176 | 166.224832 | |
| 1 | 2 | Albania | 974.554499 | 975.216820 | |
| 2 | 3 | Algeria | 373.670080 | 380.125666 | |
| 3 | 4 | Andorra | 5008.901637 | 5305.716707 | |
| 4 | 5 | Angola | 631.514939 | 648.910896 | |
| 5 | 6 | Anguilla | 621.101316 | 725.213850 | |
| 6 | 7 | Antigua and Barbuda | 524.913170 | 635.428324 | |
| 7 | 8 | Argentina | 1425.357894 | 1678.388787 | |
| 8 | 9 | Aruba | 3343.551506 | 3539.154720 | |
| 9 | 10 | Australia | 3590.037728 | 3884.030700 | |
| 10 | 11 | Austria | 2059.268192 | 2381.629933 | |
| 11 | 12 | Bahamas | 4259.305052 | 4286.281941 | |
| 12 | 13 | Bahrain | 1896.176458 | 1969.702274 | |
| 13 | 14 | Bangladesh | 91.732366 | 81.575518 | |
| 14 | 15 | Barbados | 903.012578 | 1011.028545 | |
| 15 | 16 | Belgium | 2773.481236 | 3085.109750 | |

| | | | | | |
|---|---|---|---|---|---|
| 16 | 17 | Belize | 230.906758 | 287.472565 | |
| 17 | 18 | Benin | 106.519449 | 110.668874 | |
| 18 | 19 | Bermuda | 5104.598512 | 5726.867894 | |
| 19 | 20 | Bhutan | 206.680778 | 214.769378 | |
| 20 | 21 | Bolivia (Plurinational State of) | 220.272271 | 233.407375 | |
| 21 | 22 | Botswana | 112.786843 | 145.847758 | |
| 22 | 23 | Brazil | 367.125782 | 400.134371 | |
| 23 | 24 | British Virgin Islands | 1212.165327 | 1021.866809 | |
| 24 | 25 | Brunei Darussalam | 1686.777121 | 1789.110415 | |
| 25 | 26 | Bulgaria | 1048.590520 | 1029.050237 | |
| 26 | 27 | Burkina Faso | 90.208795 | 90.534198 | |
| 27 | 28 | Burundi | 70.168159 | 70.494594 | |
| 28 | 29 | Cabo Verde | 253.916821 | 280.349382 | |
| 29 | 30 | Cambodia | 114.439733 | 114.442138 | |

| | 1972 | 1973 | 1974 | 1975 | 1976 | \ |
|---|---|---|---|---|---|---|
| 0 | 141.365323 | 149.744069 | 181.598776 | 194.669036 | 205.674097 | |
| 1 | 976.521166 | 978.617771 | 981.265874 | 984.901966 | 989.749878 | |
| 2 | 496.991248 | 624.193795 | 875.023948 | 989.243356 | 1075.725904 | |
| 3 | 6285.301526 | 7824.531307 | 9078.009005 | 10076.494834 | 9817.312977 | |
| 4 | 644.659613 | 762.730111 | 827.378464 | 589.868220 | 547.946934 | |
| 5 | 850.887123 | 957.808007 | 1049.881766 | 1140.561267 | 1086.473476 | |
| 6 | 871.867830 | 1215.586135 | 1300.717871 | 1333.904761 | 1109.562071 | |
| 7 | 1556.526579 | 1724.205313 | 1907.094439 | 1996.493745 | 2080.525773 | |
| 8 | 3725.350305 | 3890.231590 | 4052.355158 | 4220.081349 | 4410.324290 | |
| 9 | 4466.183034 | 6349.496844 | 7458.406447 | 7857.260452 | 8374.850181 | |
| 10 | 2924.947592 | 3891.702402 | 4631.855170 | 5286.897178 | 5679.783584 | |
| 11 | 4493.308784 | 5024.039305 | 5962.291182 | 5454.223469 | 5885.411989 | |
| 12 | 2367.250664 | 3196.301037 | 4150.198216 | 4154.484554 | 5681.200623 | |
| 13 | 89.573360 | 95.156689 | 110.858841 | 113.449744 | 123.688300 | |
| 14 | 1151.329867 | 1397.942497 | 1663.330643 | 1945.368774 | 2100.621610 | |
| 15 | 3835.929146 | 4907.873258 | 5747.637943 | 6726.770164 | 7276.188739 | |
| 16 | 367.842221 | 589.632882 | 775.941108 | 884.028604 | 682.906881 | |
| 17 | 133.354183 | 159.761063 | 167.784664 | 199.472350 | 207.426900 | |
| 18 | 6361.395010 | 7257.860841 | 8390.519081 | 9233.392266 | 10316.059272 | |
| 19 | 219.656318 | 239.095644 | 272.404723 | 246.316138 | 242.802694 | |
| 20 | 266.735184 | 265.051936 | 428.605157 | 467.805263 | 524.111294 | |
| 21 | 183.040570 | 257.793130 | 340.548656 | 376.534984 | 387.465855 | |
| 22 | 502.649167 | 694.649640 | 869.604041 | 998.737047 | 1153.288405 | |
| 23 | 1056.718684 | 1182.915253 | 1396.465109 | 1513.845653 | 1627.158899 | |
| 24 | 2364.014359 | 3649.148266 | 8738.105417 | 9190.990734 | 10827.978164 | |
| 25 | 1196.232688 | 1417.550265 | 1531.647164 | 1343.908397 | 1418.145113 | |
| 26 | 113.660892 | 129.249127 | 139.749595 | 162.986762 | 153.009253 | |
| 27 | 68.829523 | 85.537148 | 94.637389 | 110.655636 | 119.380471 | |
| 28 | 313.553639 | 383.896232 | 381.439736 | 440.132704 | 407.185546 | |
| 29 | 111.730841 | 94.327387 | 96.574950 | 106.966181 | 120.382292 | |

| | 1977 | … | 2013 | 2014 | 2015 | \ |
|---|---|---|---|---|---|---|
| 0 | 232.778698 | … | 631.060444 | 596.048898 | 554.004007 | |
| 1 | 994.937066 | … | 4425.430822 | 4586.606188 | 3950.390255 | |
| 2 | 1225.090692 | … | 5519.777576 | 5516.229431 | 4197.417286 | |
| 3 | 10384.507961 | … | 44735.004774 | 45619.782548 | 38869.870777 | |
| 4 | 578.443953 | … | 5228.511165 | 5371.216648 | 4130.930080 | |
| 5 | 1210.224756 | … | 20569.480211 | 22047.422883 | 22756.280768 | |
| 6 | 1260.564830 | … | 13350.149137 | 14004.811212 | 14861.882707 | |
| 7 | 2304.812606 | … | 14469.003850 | 13241.393385 | 14908.621297 | |
| 8 | 4624.873192 | … | 26514.863550 | 26940.258722 | 28419.243101 | |
| 9 | 8207.700253 | … | 66820.508917 | 62402.632438 | 52321.003543 | |
| 10 | 6812.731674 | … | 50718.407330 | 51719.250909 | 44179.468449 | |
| 11 | 6117.865558 | … | 27404.162248 | 28554.857734 | 29821.354971 | |
| 12 | 6798.708509 | … | 25790.856791 | 25464.800667 | 22795.370758 | |
| 13 | 127.333275 | … | 1130.874771 | 1282.106878 | 1449.562923 | |
| 14 | 2376.961899 | … | 16923.590920 | 16964.752264 | 17023.694365 | |
| 15 | 8464.645137 | … | 46981.053794 | 47838.536915 | 41086.169118 | |
| 16 | 860.294383 | … | 5865.470901 | 6031.548403 | 6098.600079 | |
| 17 | 221.598241 | … | 1213.950476 | 1249.840095 | 1040.863784 | |
| 18 | 11184.709840 | … | 101813.371965 | 101258.622050 | 105404.472317 | |
| 19 | 259.909829 | … | 2409.430145 | 2589.893727 | 2746.377297 | |
| 20 | 515.093216 | … | 2853.797161 | 3022.462884 | 2975.648811 | |
| 21 | 474.406397 | … | 6436.601866 | 6844.033074 | 5869.729556 | |
| 22 | 1311.006139 | … | 12258.564849 | 12071.404655 | 8783.214413 | |
| 23 | 1658.406126 | … | 39499.947657 | 41504.608056 | 43647.074848 | |
| 24 | 12759.537324 | … | 43949.635599 | 41035.760222 | 30681.450167 | |
| 25 | 1444.620074 | … | 7511.506523 | 7742.783575 | 6945.432877 | |
| 26 | 180.646058 | … | 762.087122 | 766.350661 | 631.648325 | |
| 27 | 138.319814 | … | 241.549503 | 257.822931 | 262.001758 | |
| 28 | 379.053210 | … | 3758.638926 | 3735.706612 | 3167.796310 | |
| 29 | 113.917657 | … | 1015.220881 | 1098.074538 | 1170.742816 | |

| | 2016 | 2017 | 2018 | 2019 | 2020 | \ |
|---|---|---|---|---|---|---|
| 0 | 526.164709 | 534.020133 | 513.981256 | 512.910125 | 512.749930 | |
| 1 | 4117.006970 | 4521.752217 | 5268.116608 | 5359.239078 | 5288.989573 | |
| 2 | 3967.200546 | 4134.936055 | 4171.795392 | 4021.983602 | 3354.157303 | |
| 3 | 39904.138181 | 40632.206416 | 42904.822583 | 41328.634079 | 37207.175636 | |
| 4 | 3468.520268 | 4042.681382 | 3240.862207 | 2569.629757 | 1639.951023 | |
| 5 | 21619.849507 | 18718.318718 | 21240.732202 | 24495.015171 | 16582.659014 | |
| 6 | 15862.651663 | 16110.312400 | 17514.355864 | 18187.779712 | 15303.278112 | |
| 7 | 12767.455354 | 14609.783370 | 11816.649537 | 10006.693111 | 8565.153126 | |
| 8 | 28449.728927 | 29329.092343 | 30918.483584 | 31902.757333 | 24008.127822 | |
| 9 | 54054.216565 | 57489.804356 | 58327.726542 | 54407.644621 | 56001.553064 | |
| 10 | 45277.769467 | 47429.535841 | 51466.603137 | 50067.472961 | 48839.268984 | |
| 11 | 30124.376154 | 31303.530973 | 32400.956691 | 33564.695734 | 23998.168238 | |
| 12 | 22867.275453 | 24349.836847 | 25415.750065 | 25869.206182 | 23433.491494 | |
| 13 | 1651.719965 | 1785.836613 | 1931.478758 | 2110.916696 | 2230.583106 | |

| | | | | | |
|---|---|---|---|---|---|
| 14 | 17381.365087 | 17881.563253 | 18271.252253 | 19063.102291 | 16882.501523 |
| 15 | 42038.215343 | 44162.256256 | 47455.523320 | 46554.245138 | 45517.826647 |
| 16 | 6099.957102 | 6049.391948 | 6001.024262 | 6134.215233 | 5185.158070 |
| 17 | 1049.820303 | 1095.274133 | 1194.437603 | 1170.965879 | 1240.733104 |
| 18 | 109426.650599 | 112539.075464 | 113690.519262 | 116333.291544 | 107559.619559 |
| 19 | 2988.978262 | 3427.173682 | 3389.777295 | 3564.617808 | 3181.340129 |
| 20 | 3013.502708 | 3280.008213 | 3471.006951 | 3472.380831 | 3068.812555 |
| 21 | 6411.553469 | 6705.339802 | 6947.818053 | 6679.198025 | 5863.185478 |
| 22 | 8680.735421 | 9896.717511 | 9121.020092 | 8883.903092 | 6923.700197 |
| 23 | 47270.587444 | 43569.527611 | 45719.465963 | 49448.872917 | 46222.905209 |
| 24 | 26761.567232 | 28186.946531 | 31240.444448 | 30748.320153 | 27179.344445 |
| 25 | 7442.667846 | 8237.834760 | 9319.715187 | 9767.173368 | 10058.979650 |
| 26 | 665.786066 | 711.184631 | 779.202768 | 772.542487 | 840.250143 |
| 27 | 268.733291 | 295.620936 | 288.626394 | 283.669870 | 282.431058 |
| 28 | 3312.702076 | 3534.343583 | 3860.454811 | 3928.309096 | 3223.735602 |
| 29 | 1281.105971 | 1400.899265 | 1533.315985 | 1671.914393 | 1577.911739 |

| | 2021 | 2022 |
|---|---|---|
| 0 | 378.082892 | 344.649112 |
| 1 | 6281.046775 | 6655.258391 |
| 2 | 3700.320562 | 4273.922183 |
| 3 | 42065.970756 | 42293.297081 |
| 4 | 2026.030952 | 3183.699796 |
| 5 | 19106.060927 | 28561.285003 |
| 6 | 16740.348196 | 18880.028883 |
| 7 | 10775.999836 | 13867.913268 |
| 8 | 29127.759386 | 33300.838821 |
| 9 | 67616.301208 | 67866.792484 |
| 10 | 53720.125274 | 52608.760637 |
| 11 | 28260.367048 | 31458.354292 |
| 12 | 26860.074554 | 30146.509534 |
| 13 | 2449.024978 | 2527.521270 |
| 14 | 17507.467994 | 20238.784242 |
| 15 | 51737.760269 | 49986.832591 |
| 16 | 6060.968224 | 6984.216959 |
| 17 | 1360.911474 | 1302.850067 |
| 18 | 113524.811093 | 117568.241306 |
| 19 | 3560.197894 | 3704.344581 |
| 20 | 3345.023002 | 3600.121635 |
| 21 | 7238.796096 | 7737.654680 |
| 22 | 7696.785130 | 8917.673509 |
| 23 | 46915.895058 | 46996.749944 |
| 24 | 31448.913608 | 37152.466238 |
| 25 | 12204.368627 | 13301.923710 |
| 26 | 904.474930 | 845.760531 |
| 27 | 298.170349 | 312.874378 |
| 28 | 3557.837775 | 3902.589049 |

```
29    1625.235021    1759.608024

[30 rows x 55 columns]
```

## 8 filling null value with their mean values

[17]:
```
null_values = df_filled.isnull().sum()
print(null_values)
```

```
Sr.No      0
Country    0
1970       0
1971       0
1972       0
1973       0
1974       0
1975       0
1976       0
1977       0
1978       0
1979       0
1980       0
1981       0
1982       0
1983       0
1984       0
1985       0
1986       0
1987       0
1988       0
1989       0
1990       0
1991       0
1992       0
1993       0
1994       0
1995       0
1996       0
1997       0
1998       0
1999       0
2000       0
2001       0
2002       0
2003       0
2004       0
2005       0
```

```
2006       0
2007       0
2008       0
2009       0
2010       0
2011       0
2012       0
2013       0
2014       0
2015       0
2016       0
2017       0
2018       0
2019       0
2020       0
2021       0
2022       0
dtype: int64
```

[19]:
```
df['1991']
```

[19]:
```
0        310.846728
1        483.676684
2       1781.785125
3      25265.665752
4       1323.873293
             ...
182             NaN
183             NaN
184             NaN
185      428.503428
186     1219.184684
Name: 1991, Length: 187, dtype: float64
```

[21]:
```
df_filled['1991']
```

[21]:
```
0        310.846728
1        483.676684
2       1781.785125
3      25265.665752
4       1323.873293
             ...
182     7287.373880
183     7287.373880
184     7287.373880
185      428.503428
186     1219.184684
Name: 1991, Length: 187, dtype: float64
```

## 9 splitting into data sets

[77]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(df_filled.iloc[:,0:2],df_filled.
 ↪iloc[:,1],test_size=0.2)
x_train1,x_test1,y_train1,y_test1=train_test_split(df_filled.iloc[:,3:
 ↪],df_filled.iloc[:,1],test_size=0.2)
```

[71]:
```
x_train
```

[71]:
```
     Sr.No  Country
145    146      145
42      43       42
131    132      131
132    133      132
101    102      101
..     …       …
36      37       36
99     100       99
70      71       70
153    154      153
127    128      127

[149 rows x 2 columns]
```

[73]:
```
y_train
```

[73]:
```
145    145
42      42
131    131
132    132
101    101
      …
36      36
99      99
70      70
153    153
127    127
Name: Country, Length: 149, dtype: int64
```

[29]:
```
y_test
```

[29]:
```
88                              Japan
141                            Rwanda
6                 Antigua and Barbuda
147                      Saudi Arabia
58                           Eswatini
```

```
131                     Papua New Guinea
119                                Nepal
133                                 Peru
85                                Israel
82           Iran (Islamic Republic of)
13                              Bangladesh
66                                 Gambia
145                             San Marino
157      St. Vincent and the Grenadines
169                               Türkiye
154                          South Africa
121                         New Caledonia
106                       Marshall Islands
20     Bolivia (Plurinational State of)
108                             Mauritius
182         Yemen Arab Republic (Former)
186                              Zimbabwe
12                                Bahrain
96                                Liberia
111                                Monaco
123                              Nicaragua
70                               Greenland
128                               Pakistan
135                                 Poland
155                                  Spain
179                                Vanuatu
9                                Australia
37                  China, Hong Kong SAR
73                                 Guinea
112                               Mongolia
140                               Romania
136                              Portugal
130                                Panama
Name: Country, dtype: object
```

## 10 Normalization

[79]:
```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(x_train1)
x_train_scaled=scaler.transform(x_train1)
x_test_scaled=scaler.transform(x_test1)
```

[85]:
```
x_train_scaled
```

[85]:
```
array([[0.24829608, 0.2532637 , 0.26601518, …, 0.32059066, 0.29280592,
        0.31190155],
       [0.02821228, 0.02647178, 0.02267136, …, 0.00787131, 0.0076163 ,
        0.0082323 ],
       [0.0170519 , 0.01574796, 0.01544994, …, 0.01531366, 0.01354501,
        0.014813  ],
       …,
       [0.00826243, 0.00693105, 0.00613807, …, 0.01980255, 0.01784015,
        0.01957137],
       [0.17617682, 0.16830133, 0.15700323, …, 0.1727311 , 0.15337393,
        0.14291763],
       [0.32564137, 0.35579295, 0.47751522, …, 0.20424841, 0.18723486,
        0.22217096]])
```

[87]:
```
x_test_scaled
```

[87]:
```
array([[0.00818264, 0.00785798, 0.00864206, …, 0.00594753, 0.00510911,
        0.00476748],
       [0.0582819 , 0.05959812, 0.06425626, …, 0.20340032, 0.19052576,
        0.14255445],
       [0.01576   , 0.01323765, 0.01175881, …, 0.01936628, 0.01825834,
        0.019912  ],
       …,
       [0.00053693, 0.00040381, 0.00080719, …, 0.00186506, 0.00189233,
        0.00258819],
       [0.00658172, 0.00737042, 0.00917204, …, 0.02980799, 0.02922503,
        0.03077977],
       [0.20156512, 0.19024702, 0.16700832, …, 0.21853097, 0.19728772,
        0.18907223]])
```

## 11 label edcoding and feature engineering for predictive modeling

[31]:
```
from sklearn.preprocessing import LabelEncoder
le= LabelEncoder()
```

[35]:
```
le.classes_
```

[35]:
```
array(['Afghanistan', 'Albania', 'Algeria', 'Andorra', 'Angola',
       'Anguilla', 'Argentina', 'Aruba', 'Austria', 'Bahamas', 'Barbados',
       'Belgium', 'Belize', 'Benin', 'Bermuda', 'Bhutan', 'Botswana',
       'Brazil', 'British Virgin Islands', 'Brunei Darussalam',
       'Bulgaria', 'Burkina Faso', 'Burundi', 'Cabo Verde', 'Cambodia',
       'Cameroon', 'Canada', 'Cayman Islands', 'Central African Republic',
       'Chad', 'Chile', 'China', 'China, Macao SAR', 'Colombia',
       'Comoros', 'Congo', 'Cook Islands', 'Costa Rica', 'Cuba', 'Cyprus',
       'Czechoslovakia (Former)', "Côte d'Ivoire", 'D.P.R. of Korea',
       'D.R. of the Congo', 'Denmark', 'Djibouti', 'Dominica',
```

```
       'Dominican Republic', 'Ecuador', 'Egypt', 'El Salvador',
       'Equatorial Guinea', 'Ethiopia (Former)', 'Fiji', 'Finland',
       'Former Netherlands Antilles', 'France', 'French Polynesia',
       'Gabon', 'Germany', 'Ghana', 'Greece', 'Grenada', 'Guatemala',
       'Guinea-Bissau', 'Guyana', 'Haiti', 'Honduras', 'Hungary',
       'Iceland', 'India', 'Indonesia', 'Iraq', 'Ireland', 'Italy',
       'Jamaica', 'Jordan', 'Kenya', 'Kiribati', 'Kuwait',
       "Lao People's DR", 'Lebanon', 'Lesotho', 'Libya', 'Liechtenstein',
       'Luxembourg', 'Madagascar', 'Malawi', 'Malaysia', 'Maldives',
       'Mali', 'Malta', 'Mauritania', 'Mexico', 'Micronesia (FS of)',
       'Montserrat', 'Morocco', 'Mozambique', 'Myanmar', 'Namibia',
       'Nauru', 'Netherlands', 'New Zealand', 'Niger', 'Nigeria',
       'Norway', 'Oman', 'Palau', 'Paraguay', 'Philippines',
       'Puerto Rico', 'Qatar', 'Republic of Korea',
       'Saint Kitts and Nevis', 'Saint Lucia', 'Samoa',
       'Sao Tome and Principe', 'Senegal', 'Seychelles', 'Sierra Leone',
       'Singapore', 'Solomon Islands', 'Somalia', 'Sri Lanka',
       'State of Palestine', 'Sudan (Former)', 'Suriname', 'Sweden',
       'Switzerland', 'Syrian Arab Republic', 'Thailand', 'Togo', 'Tonga',
       'Trinidad and Tobago', 'Tunisia', 'Turks and Caicos Islands',
       'Tuvalu', 'U.R. of Tanzania: Mainland', 'USSR (Former)', 'Uganda',
       'United Arab Emirates', 'United Kingdom', 'United States',
       'Uruguay', 'Venezuela (Bolivarian Republic of)', 'Viet Nam',
       'Yemen Democratic (Former)', 'Yugoslavia (Former)', 'Zambia'],
      dtype=object)
```

[33]:
```
le.fit(y_train)
```

[33]:
```
LabelEncoder()
```

[51]:
```
df_filled["Country"]=le.fit_transform(df_filled["Country"])
df_filled
```

[51]:

| | Sr.No | Country | 1970 | 1971 | 1972 | 1973 | \ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 162.642176 | 166.224832 | 141.365323 | 149.744069 | |
| 1 | 2 | 1 | 974.554499 | 975.216820 | 976.521166 | 978.617771 | |
| 2 | 3 | 2 | 373.670080 | 380.125666 | 496.991248 | 624.193795 | |
| 3 | 4 | 3 | 5008.901637 | 5305.716707 | 6285.301526 | 7824.531307 | |
| 4 | 5 | 4 | 631.514939 | 648.910896 | 644.659613 | 762.730111 | |
| .. | … | … | … | … | … | … | |
| 182 | 183 | 182 | 76.222959 | 87.998562 | 124.898913 | 110.398671 | |
| 183 | 184 | 183 | 94.929351 | 85.003136 | 92.397759 | 73.556033 | |
| 184 | 185 | 184 | 704.150562 | 757.937976 | 783.579953 | 1011.118968 | |
| 185 | 186 | 185 | 360.651807 | 357.771353 | 409.332165 | 504.983508 | |
| 186 | 187 | 186 | 449.078443 | 503.731409 | 603.030370 | 718.359564 | |

| | 1974 | 1975 | 1976 | 1977 | … | 2013 | \ |
|---|---|---|---|---|---|---|---|

```
0     181.598776    194.669036   205.674097    232.778698  …    631.060444
1     981.265874    984.901966   989.749878    994.937066  …   4425.430822
2     875.023948    989.243356  1075.725904   1225.090692  …   5519.777576
3    9078.009005  10076.494834  9817.312977  10384.507961  …  44735.004774
4     827.378464    589.868220   547.946934    578.443953  …   5228.511165
..           …             …            …             …  …           …
182   130.344863    154.173397   163.609346    181.710840  …   7287.373880
183    76.071783     75.559923    95.488404    114.550976  …   7287.373880
184  1385.814039   1537.956907  1719.654598   2071.555089  …   7287.373880
185   604.411205    539.065320   553.914220    528.811611  …   1745.014154
186   836.465455    889.056832   851.930852    838.458026  …   1408.367809

              2014          2015          2016          2017          2018  \
0      596.048898    554.004007    526.164709    534.020133    513.981256
1     4586.606188   3950.390255   4117.006970   4521.752217   5268.116608
2     5516.229431   4197.417286   3967.200546   4134.936055   4171.795392
3    45619.782548  38869.870777  39904.138181  40632.206416  42904.822583
4     5371.216648   4130.930080   3468.520268   4042.681382   3240.862207
..            …             …             …             …             …
182   7287.373880   7287.373880   7287.373880   7287.373880   7287.373880
183   7287.373880   7287.373880   7287.373880   7287.373880   7287.373880
184   7287.373880   7287.373880   7287.373880   7287.373880   7287.373880
185   1696.117261   1283.982258   1227.976126   1497.004453   1475.226627
186   1407.034293   1410.329175   1421.787790   1494.170144   1570.886029

              2019          2020          2021          2022
0      512.910125    512.749930    378.082892    344.649112
1     5359.239078   5288.989573   6281.046775   6655.258391
2     4021.983602   3354.157303   3700.320562   4273.922183
3    41328.634079  37207.175636  42065.970756  42293.297081
4     2569.629757   1639.951023   2026.030952   3183.699796
..            …             …             …             …
182   7287.373880   7287.373880   7287.373880   7287.373880
183   7287.373880   7287.373880   7287.373880   7287.373880
184   7287.373880   7287.373880   7287.373880   7287.373880
185   1268.212284    954.356128   1135.078255   1455.546876
186   1471.513983   1382.591426   1507.994790   1618.733070

[187 rows x 55 columns]
```

[41]:
```
y_train=le.fit_transform(y_train)
y_test=le.fit_transform(y_test)
```

[43]:
```
y_train
```

[43]:
```
array([ 32,  89,  92, 107,  13,  71,  55,   8, 143,  38, 116,  82,  95,
        51, 109,  98,  60,  64,   3,  22, 113, 129,  47,   7,  67, 139,
```

```
 93,  41,  94, 148,   0, 108, 115,  42,  16,  40, 134,  29,  65,
137, 144, 138,  30,  18,  19,  86,  66,  73,  45,   2,  50, 133,
101, 141,  69,  99,  88, 131,  53,  84, 135, 132,  97,  37,  44,
 91,  27,   4, 118,  77,  72, 112,  57, 140,  15, 147,  62, 120,
 70,  43, 128,  83, 127, 114,  31,  48,  63,  14,  96, 136,  80,
124,  85,  76,  21, 103, 123,  75,  28,  78, 100,  20, 117, 104,
 52,  46,  68,  49,  87,  56,   1,  34,  39, 126,  36,  61, 142,
 11, 106,  17,  58,  24,   6, 111,   5, 145,  12,  25, 102,  23,
  9, 119, 105, 125, 130, 146,  81,  33, 122, 121,  10,  79,  54,
 35,  90,  74,  59, 110,  26], dtype=int64)
```

[45]:
```
y_test
```

[45]:
```
array([12, 28,  0, 30,  6, 23, 18, 24, 11, 10,  3,  7, 29, 33, 34, 31, 19,
       14,  4, 15, 36, 37,  2, 13, 16, 20,  8, 21, 25, 32, 35,  1,  5,  9,
       17, 27, 26, 22], dtype=int64)
```

[ ]: