

In [1]: # Data Exploration and Preprocessing#Handling Missing Values#Handling Outliers#Handling Skewness#Categorical Data Encoding

In [2]: #EXPLORATORY DATA ANALYSIS - PREPROCESSINGimport pandas as pdimport numpy as npimport matplotlib.pyplot as pltimport seaborn as snsimport matplotlib.pyplot as pltimport warningswarnings.filterwarnings('ignore')

In [3]: #LOADING THE DATA SETWBD = pd.read\_csv("WORLD BANK DATA.csv")

Out[3]:

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...	2013	2014	2015	2016
0	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	56682.0	57475.0	58178.0	58782.0	...	102880.0	103594.0	104257.0	104874.0
1	Africa Eastern and Southern	AFE	Population, total	SP.POP.TOTL	130692579.0	134169237.0	137835590.0	141630546.0	145605995.0	149742351.0	...	567892149.0	583651101.0	600008424.0	616377605.0
2	Afghanistan	AFG	Population, total	SP.POP.TOTL	8622466.0	8790140.0	8969047.0	9157465.0	9355514.0	9565147.0	...	31541209.0	32716210.0	33753499.0	34636207.0
3	Africa Western and Central	AFW	Population, total	SP.POP.TOTL	97256290.0	99314028.0	101445032.0	103667517.0	105959979.0	108336203.0	...	387204553.0	397855507.0	408690375.0	419778384.0
4	Angola	AGO	Population, total	SP.POP.TOTL	5357195.0	5441333.0	5521400.0	5599827.0	5673199.0	5736582.0	...	26147002.0	27128337.0	28127721.0	29154746.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
261	Kosovo	KXK	Population, total	SP.POP.TOTL	947000.0	966000.0	994000.0	1022000.0	1050000.0	1078000.0	...	1818117.0	1812771.0	1788196.0	1777557.0
262	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	5753386.0	5860197.0	5973803.0	6097298.0	...	26984002.0	27753304.0	28516545.0	29274002.0
263	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16898464.0	17503133.0	18042215.0	18063097.0	19187194.0	...	53873616.0	54729551.0	55876504.0	56422274.0
264	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	3323427.0	3431381.0	3542764.0	3658024.0	...	15234976.0	15737793.0	16248230.0	16767761.0
265	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	4049778.0	4177931.0	4310332.0	4447149.0	...	13555422.0	1385753.0	14154937.0	14452704.0

266 rows x 67 columns

In [4]: # DATA PREPROCESSINGWBD = WBD.dropna()

Out[4]:

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...	2013	2014	2015	2016
0	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	56682.0	57475.0	58178.0	58782.0	...	102880.0	103594.0	104257.0	104874.0
1	Africa Eastern and Southern	AFE	Population, total	SP.POP.TOTL	130692579.0	134169237.0	137835590.0	141630546.0	145605995.0	149742351.0	...	567892149.0	583651101.0	600008424.0	616377605.0
2	Afghanistan	AFG	Population, total	SP.POP.TOTL	8622466.0	8790140.0	8969047.0	9157465.0	9355514.0	9565147.0	...	31541209.0	32716210.0	33753499.0	34636207.0
3	Africa Western and Central	AFW	Population, total	SP.POP.TOTL	97256290.0	99314028.0	101445032.0	103667517.0	105959979.0	108336203.0	...	387204553.0	397855507.0	408690375.0	419778384.0
4	Angola	AGO	Population, total	SP.POP.TOTL	5357195.0	5441333.0	5521400.0	5599827.0	5673199.0	5736582.0	...	26147002.0	27128337.0	28127721.0	29154746.0

5 rows x 67 columns

In [5]: WBD.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 67 columns):
#   Column      Non-Null Count  Dtype
---  --
0   Country Name 266 non-null    object
1   Country Code 266 non-null    object
2   Indicator Name 266 non-null   object
3   Indicator Code 266 non-null   object
4   1960         264 non-null    float64
5   1961         264 non-null    float64
6   1962         264 non-null    float64
7   1963         264 non-null    float64
8   1964         264 non-null    float64
9   1965         264 non-null    float64
10  1966         264 non-null    float64
11  1967         264 non-null    float64
12  1968         264 non-null    float64
13  1969         264 non-null    float64
14  1970         264 non-null    float64
15  1971         264 non-null    float64
16  1972         264 non-null    float64
17  1973         264 non-null    float64
18  1974         264 non-null    float64
19  1975         264 non-null    float64
20  1976         264 non-null    float64
21  1977         264 non-null    float64
22  1978         264 non-null    float64
23  1979         264 non-null    float64
24  1980         264 non-null    float64
25  1981         264 non-null    float64
26  1982         264 non-null    float64
27  1983         264 non-null    float64
28  1984         264 non-null    float64
29  1985         264 non-null    float64
30  1986         264 non-null    float64
31  1987         264 non-null    float64
32  1988         264 non-null    float64
33  1989         264 non-null    float64
34  1990         265 non-null    float64
35  1991         265 non-null    float64
36  1992         265 non-null    float64
37  1993         265 non-null    float64
38  1994         265 non-null    float64
39  1995         265 non-null    float64
40  1996         265 non-null    float64
41  1997         265 non-null    float64
42  1998         265 non-null    float64
43  1999         265 non-null    float64
44  2000         265 non-null    float64
45  2001         265 non-null    float64
46  2002         265 non-null    float64
47  2003         265 non-null    float64
48  2004         265 non-null    float64
49  2005         265 non-null    float64
50  2006         265 non-null    float64
51  2007         265 non-null    float64
52  2008         265 non-null    float64
53  2009         265 non-null    float64
54  2010         265 non-null    float64
55  2011         265 non-null    float64
56  2012         265 non-null    float64
57  2013         265 non-null    float64
58  2014         265 non-null    float64
59  2015         265 non-null    float64
60  2016         265 non-null    float64
61  2017         265 non-null    float64
62  2018         265 non-null    float64
63  2019         265 non-null    float64
64  2020         265 non-null    float64
65  2021         265 non-null    float64
66  2022         265 non-null    float64
dtypes: float64(63), object(4)
memory usage: 139.4+ KB
```

In [6]: # HOW MANY ROW AND COLUMNSWBD.shape

Out[6]: (266, 67)

In [7]: WBD.dtypes

Out[7]:

```
Country Name    object
Country Code    object
Indicator Name   object
Indicator Code   object
1960            float64
2019            float64
2019            float64
2019            float64
2022            float64
Length: 67, dtype: object
```

In [8]: # STATISTICAL ANALYSISWBD.describe(include='all').T

Out[8]:

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Country Name	266	266	San Marino	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Country Code	266	266		1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Indicator Name	266	1	Population, total	266	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Indicator Code	266	1	SP.POP.TOTL	266	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1960	264.0	NaN	NaN	NaN	117271174.712121	369543889.752433	2646.0	513221.25	3757485.5	267006062.75	3031472434.0
...	...	...	...	...	...	...	...	...	...	...	...
2018	265.0	NaN	NaN	NaN	312027628.377358	974687963.06586	10865.0	1797085.0	10395329.0	60421760.0	7661776612.0
2019	265.0	NaN	NaN	NaN	315711047.727733	985168976.435331	10956.0	1788878.0	10447666.0	59872579.0	7742682218.0
2020	265.0	NaN	NaN	NaN	31293645.637736	995229442.160067	11069.0	1790133.0	10606227.0	61704518.0	7820963775.0
2021	265.0	NaN	NaN	NaN	322518038.833962	1004211289.611352	11204.0	1786038.0	10505772.0	63888334.0	7888161297.0
2022	265.0	NaN	NaN	NaN	325483941.14717	1012173614.238201	11312.0	1761885.0	10526073.0	65497748.0	7951149546.0

67 rows x 11 columns

In [9]: WBD.mean().value\_counts()

Out[9]:

```
2.598113e+08    1
1.718262e+08    1
2.776655e+08    1
1.851494e+08    1
2.423883e+08    1
2.069199e+08    1
2.269322e+08    1
2.388908e+08    1
1.237232e+08    1
1.172712e+08    1
Length: 63, dtype: int64
```

In [10]: WBD.mode().value\_counts()

Out[10]:

Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	2013	2014	2015	2016	2017	2018	2019	20	20	
1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042
2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067
2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092
2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117
2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142
2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167
2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192
2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217
2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242
2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267
2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292
2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317
2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342
2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367
2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392
2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417
2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	2434	2435	2436	2437	2438	2439	2440	2441	2442
2443	2444	2445	2446	2447	2448	2449	2450	2451	2452	2453	2454	2455	2456	2457	2458	2459	2460	2461	2462	2463	2464	2465	2466	2467
2468	2469	2470	2471	2472	2473	2474	2475	2476	2477	2478	2479	2480	2481	2482	2483	2484	2485	2486	2487	2488	2489	2490	2491	2492
2493	2494	2495	2496	2497	2498	2499	2500	2501	2502	2503	2504	2505	2506	2507	2508	2509	2510	2511	2512	2513	2514	2515	2516	2517
2518	2519	2520	2521	2522	2523	2524	2525	2526	2527	2528	2529	2530	2531	2532	2533	2534	2535	2536	2537	2538	2539	2540	2541	2542
2543	2544	2545	2546	2547	2548	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558	2559	2560	2561	2562	2563	2564	2565		