

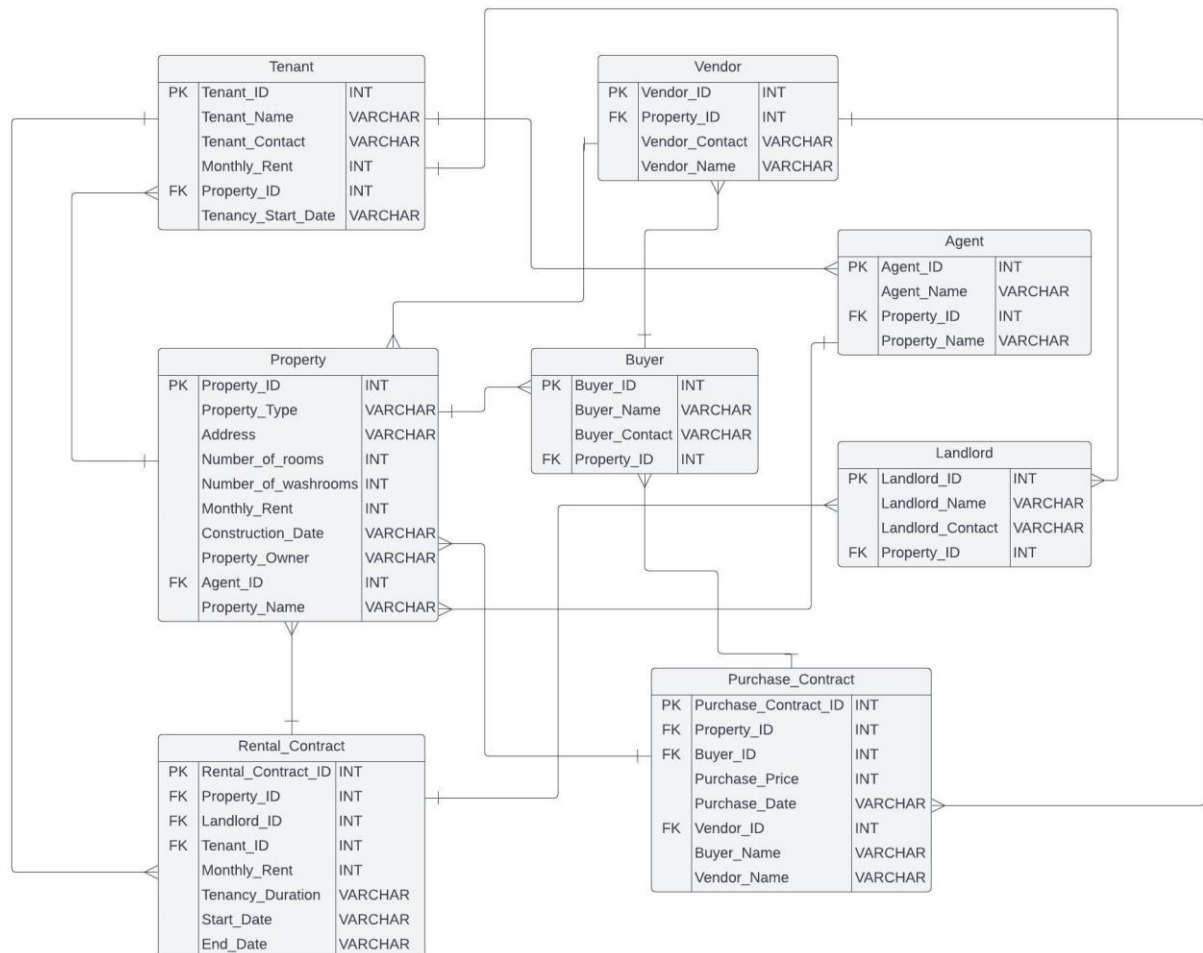
DATA ANALYTICS IN ACTION ASSIGNMENT

PART B

Word Count: 750

gxnj57 (001110537)

1) ER DIAGRAM – ESTATE AGENCY DATABASE



2-1) Table depicting the data cleaning steps for economic indicators:

Column Name	Situation of the Column	Cleaning action/steps	Justification/Explanation
Continent	Spelling mistakes in the continent names.	Used the pandas 'replace' method to correct the spelling mistakes. Replaced: <ol style="list-style-type: none"> 1. 'Afric' : 'Africa' 2. 'Asi' : 'Asia' 3. 'ASIA' : 'Asia' 4. 'Euro' : 'Europe' 5. 'EuRope' : 'Europe' 6. 'N America' : 'North America' 	Correcting spelling problems ensures categorical data integrity and correctness. This ensures accurate and representative continent analysis. Misclassifying data in misspelled categories can lead to incorrect analysis.
CO2kt	Missing/Null values	The missing values were substituted with the column's median.	The median can impute missing values in skewed distributions and outliers since it is resilient to outliers. The median preserves data dispersion since extreme numbers do not affect it like the mean. This method is advocated to preserve the dataset's central tendency. (Little, R. J. A., & Rubin, D.B, 2002).
HealthPC\$	Missing/Null values	The missing values were substituted with the column's median.	Similar to 'CO2kt', 'HealthPC\$' uses the median to impute missing data since it resists outliers and extreme values. This strategy prevents anomalous, extreme values from skewing the data's central trend, preserving it for analysis (Schafer, J. L, 1999).

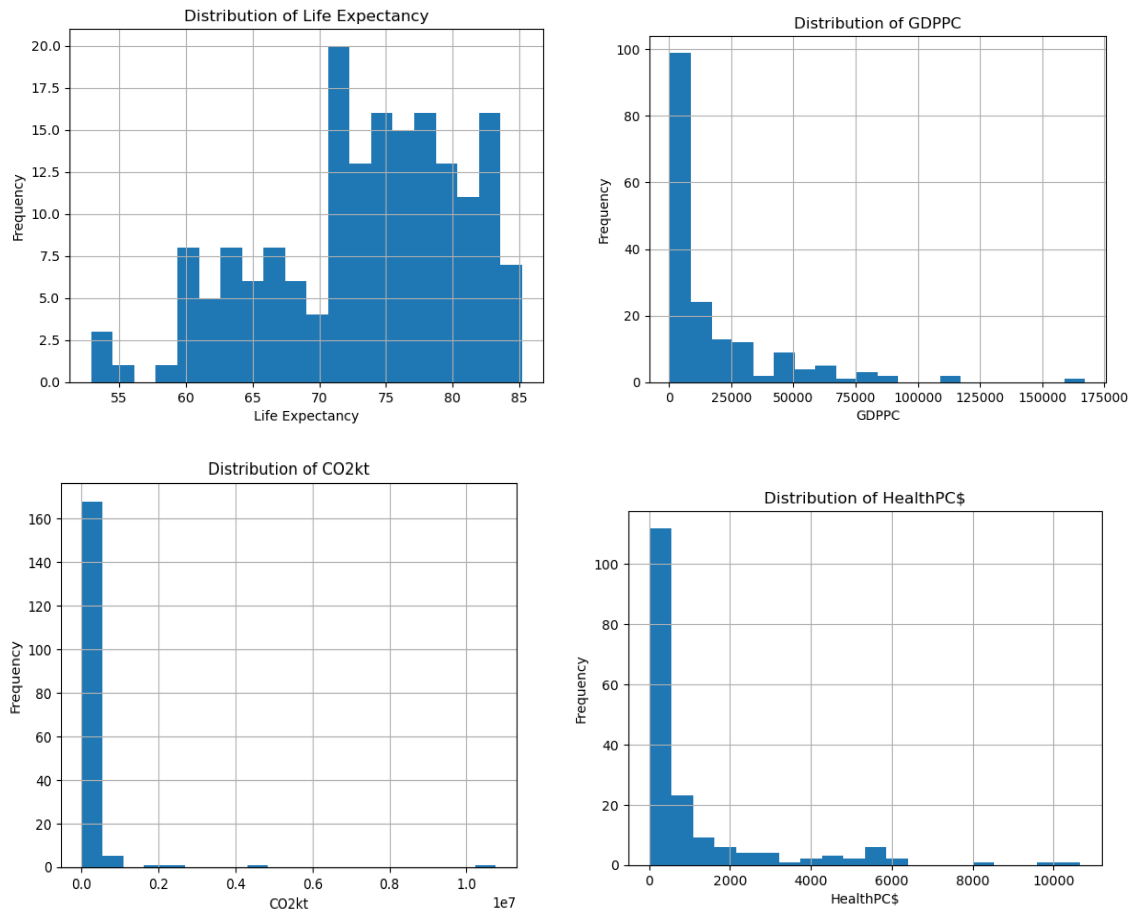


Figure 1: Distribution of columns across the dataset

From the above figures, it can be noted that the distributions are either slightly or highly skewed. This indicates the presence of outliers which affect the mean significantly. Hence, median is used for imputation in these circumstances to retain the data's structure and avoid bias from the mean in skewed distributions.

2-2) Effect of Pandemic on Economic Indicators:

1. GDP PER CAPITA

On comparison of GDPPC across the continents (2019 vs 2020) as seen from Figure 2, it can be noted that the GDPPC has decreased across all continents. The global pandemic's beginning in 2020 has far-reaching economic consequences. Figure 2 shows that GDPPC dropped across all continents, which is indicative of the pandemic's negative effects on economic activity and output. The mean GDPPC decreased from \$17,858.37 in 2019 to \$16,590.26 in 2020.

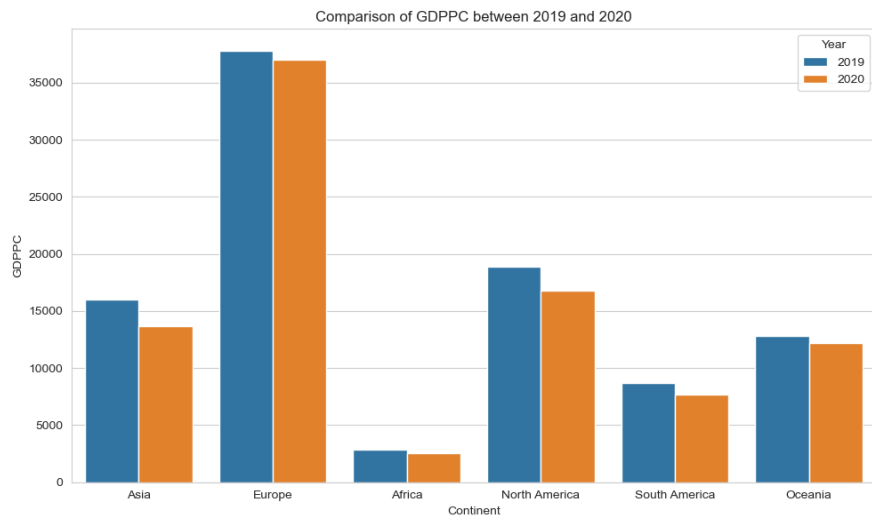


Figure 2: Comparison of GDPPC between 2019 and 2020

Asia has had the greatest loss in GDP when compared to other continents. Figure 3 shows that China has the biggest loss in GDPPC among Asian countries, with the value plummeting by more than 50% in just one year.

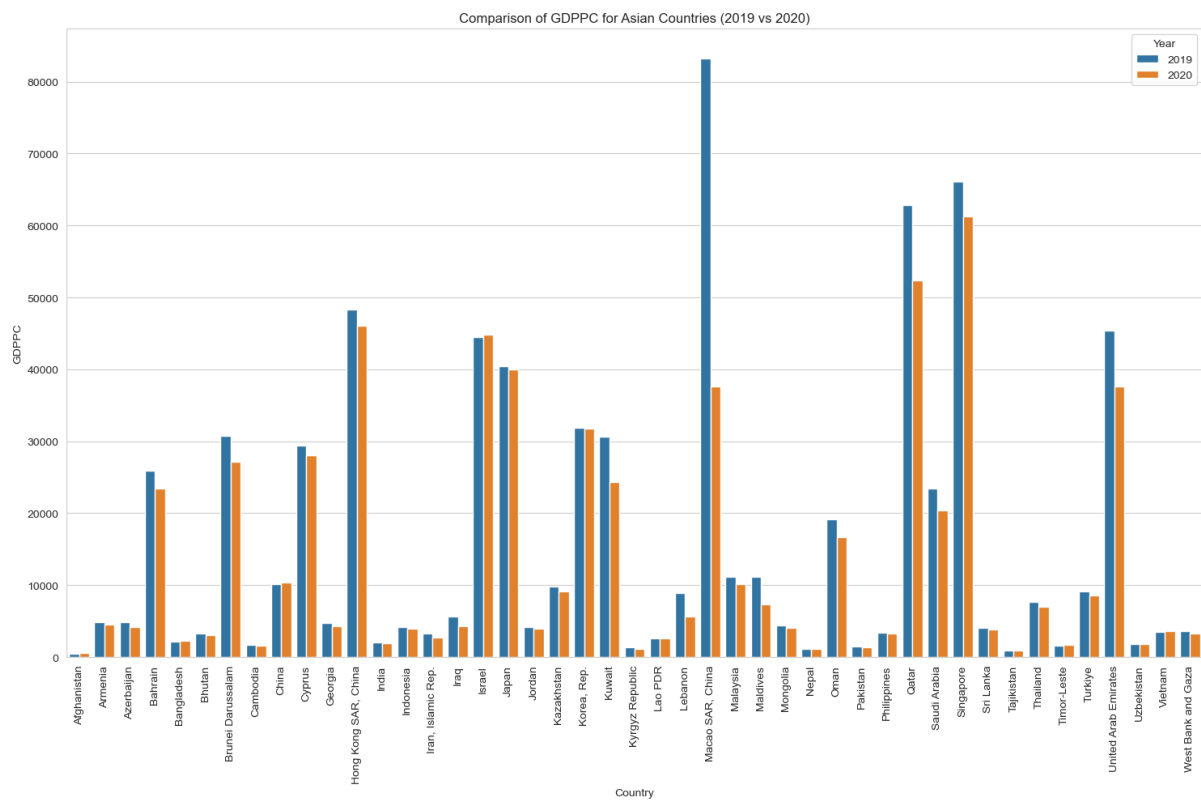


Figure 3: Comparison of GDPPC for Asian Countries (2019 vs 2020)

2. LIFE EXPECTANCY:

Figure 4 shows that the mean life expectancy has declined slightly from 73.35 years in 2019 to 72.67 years in 2020. This decrease can be linked to a number of variables, including the indirect and direct impacts on the public health.

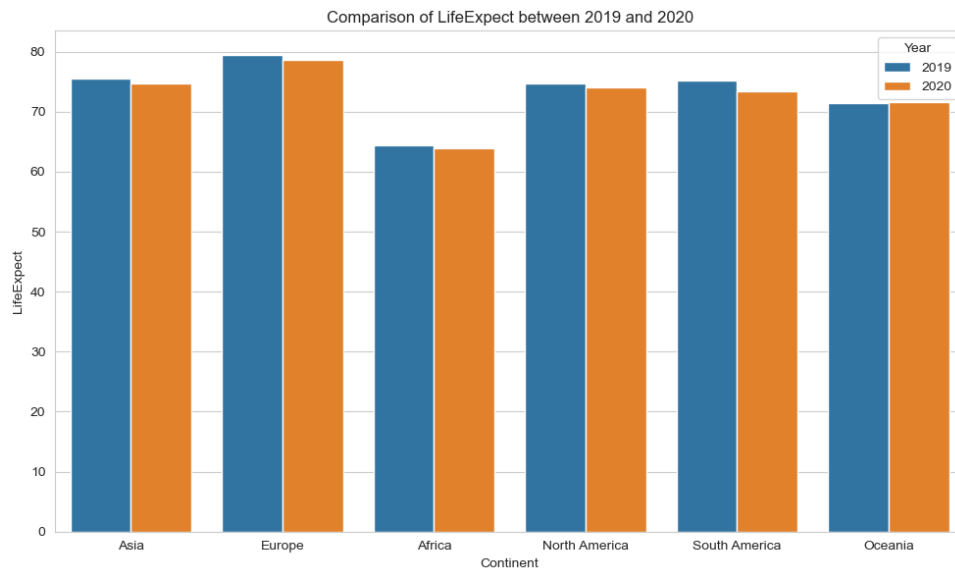


Figure 4: Comparison of Life Expectancy by Continent

3. HEALTHCARE EXPENSES PER CAPITA:

In 2019, average healthcare spending per capita was \$1112.16, and by 2020, it had risen to \$1167.17 as seen from Figure 5. This rising trend indicates a greater emphasis on health-related spending during the pandemic, possibly due to the need for additional healthcare capacity and research along with any necessary response measures.

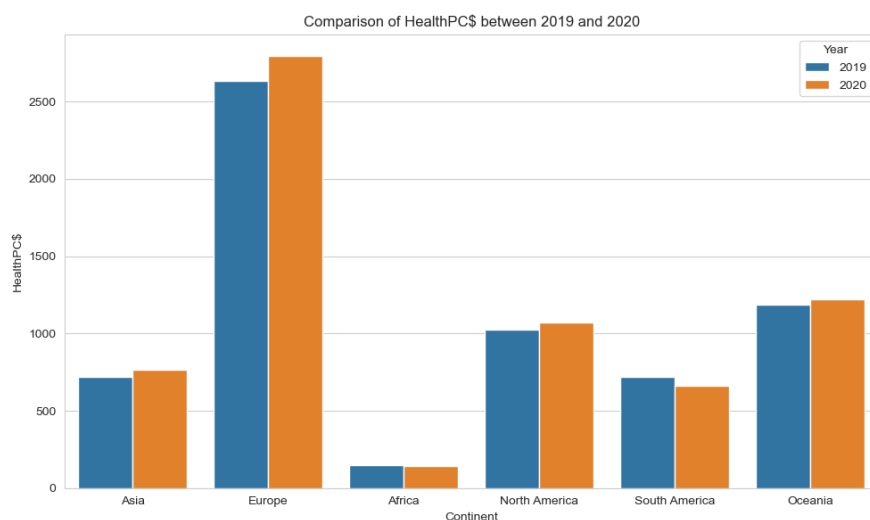


Figure 5: Health Expenditure Per Capita by Continent

4. MANUFACTURING EXPENSES PER CAPITA:

Figure 6 shows that the average manufacturing expenditure decreases from \$75,914.14 to \$73,541.69. The pandemic brought with it issues such as supply chain disruptions, personnel constraints, and changes in consumer behaviour, all of which are expected to have an impact on manufacturing sector performance. Since manufacturing spending is a large component of GDP, this reduction may lead to a contraction in overall economic production.

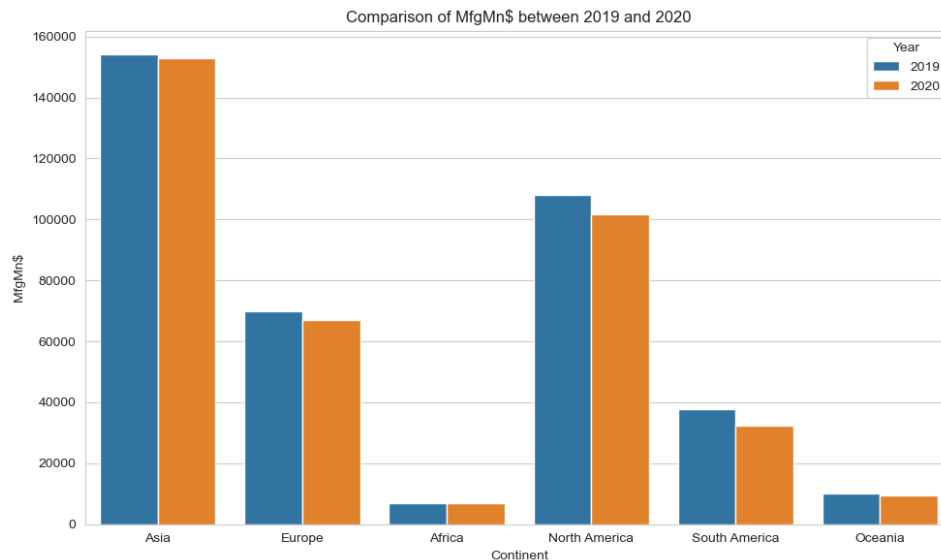


Figure 6: Manufacturing Expenses by Continent

5. AGRICULTURE EXPENSES PER CAPITA:

Figure 7 shows that the average agriculture expenditure increases from \$18,997.623 to \$20,072.67. The observed growth in agricultural spending can be ascribed to a variety of causes, including changes in agricultural regulations, technical developments, and increased demand for agricultural goods. Asia's significant increase in agricultural costs demonstrates resilience and stresses the importance of the agriculture sector in sustaining economic stability and development during the pandemic.

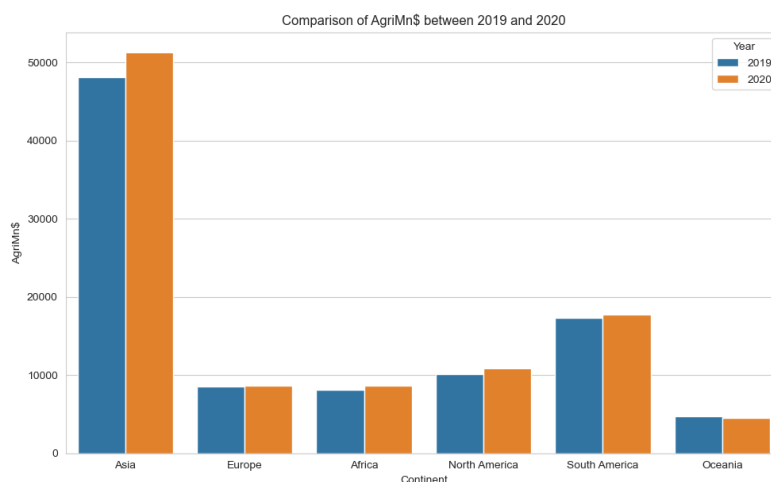


Figure 7: Agriculture Expenses by Continent

5. CO2 EMISSIONS PER CAPITA:

The observed decrease in average CO2 emissions per capita, as shown in Figure 8, from 186,390.25 units in 2019 to 178,424.17 units in 2020, indicates a significant environmental impact of the worldwide pandemic on carbon emissions. The decrease in CO2 emissions, which is directly related to economic activity, implies a temporary halt in industry and transportation, emphasizing the complex relationship between environmental indicators and economic dynamics.

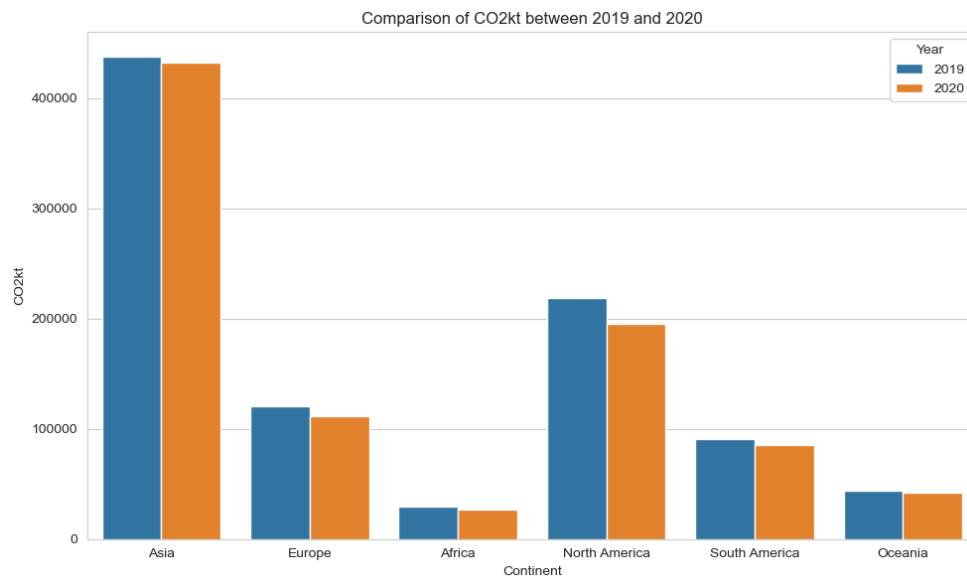


Figure 8: CO2 Emissions by Continent

References:

- 1] Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*.
- 2] Schafer, J. L. (1999). *Analysis of Incomplete Multivariate Data*.