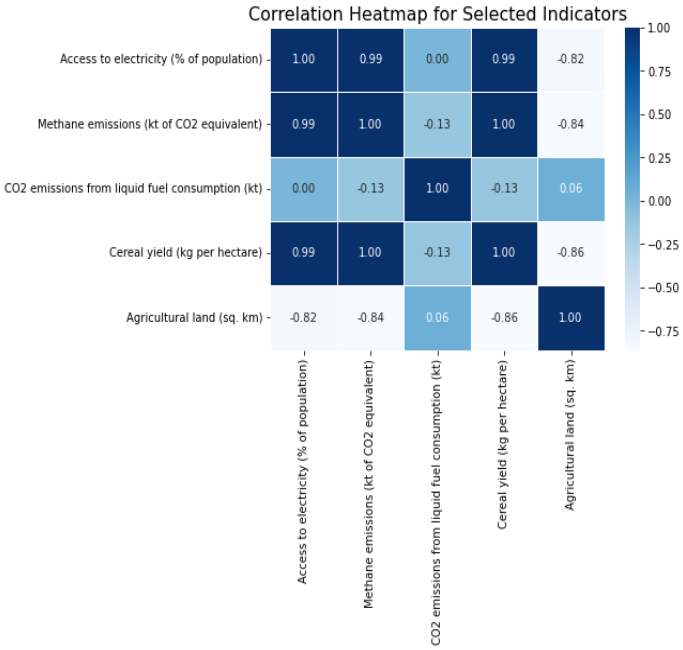
# APPLIED DATA SCIENCE – 1

## ASSIGNMENT 2: STATISTICS AND TRENDS

**TITLE:** A Comprehensive Exploration of Trends in World Bank Indicators for Environmental, Agricultural, and Energy Metrics.

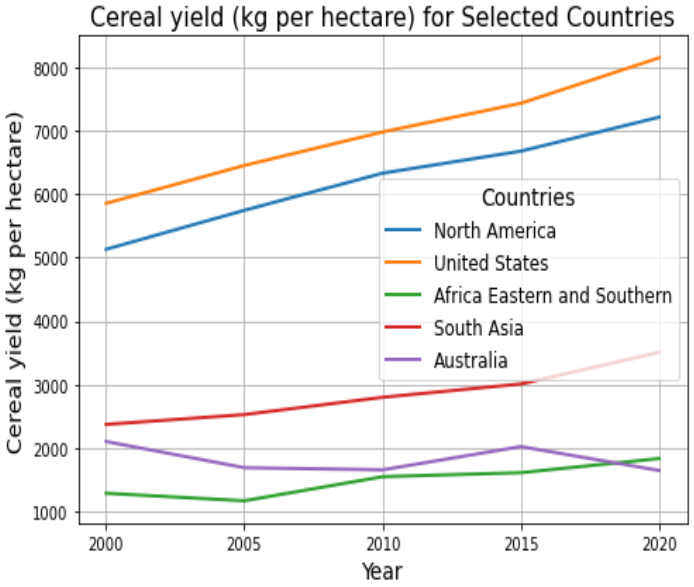
**ABSTRACT:** The report utilises World Bank data for an exploratory data analysis (EDA) focusing on a diverse set of indicators, including environmental, agricultural, and energy-related metrics. The analysis showcases correlations between indicators, revealing that countries with high electricity access, lower methane and CO2 emissions, efficient cereal yields, and balanced agricultural land use are likely to exhibit a more sustainable and developed socio-economic and environmental profile. From the conclusions provided, lower values in these indicators may highlight areas for improvement, suggesting the need for targeted interventions.

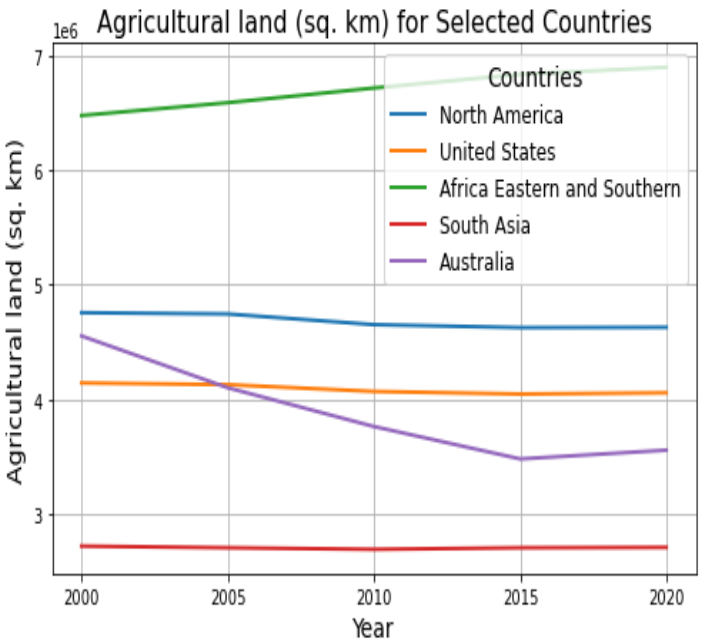
**NAME: HARIKRISHNAN MARIMUTHU**
**STUDENT ID: 22076986**
**DATA REFERENCE LINK:** https://data.worldbank.org/topic/climate-change
**GITHUB LINK:** https://github.com/Harikrishnan03/ADS1_Assignment2




The correlation heatmap shows that there is a significant negative connection with agricultural land and a strong positive correlation (0.99) between electricity access and methane emissions. There are Moderate positive correlations (0.13, 0.25) appear between methane and liquid fuel emissions, cereal yield, and agricultural land. Correlation should not be equated with causality The observed positive correlation between electricity access and methane emissions may stem from fossil fuel usage. Furthermore, the moderate negative correlation (-0.50) between CO2 emissions from liquid fuel consumption and cereal yield prompts investigation into environmental impacts on crop productivity. The heatmap provides valuable insights, signalling the need for more in-depth statistical exploration.

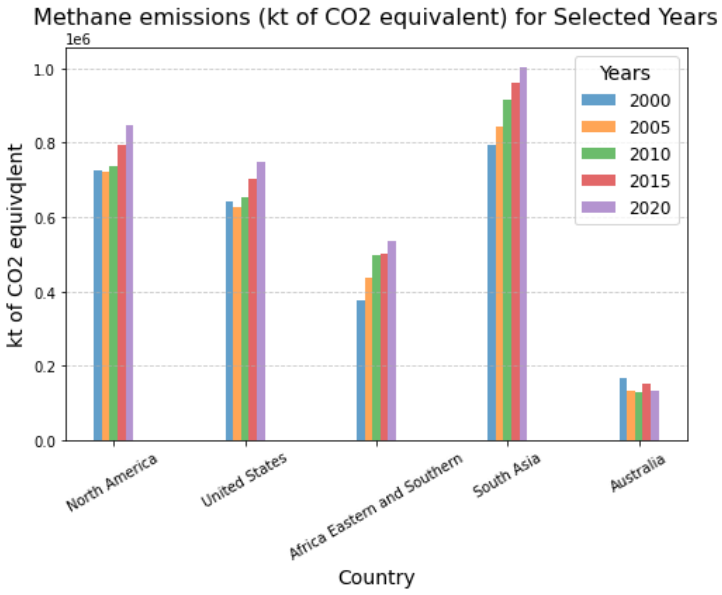
The plot shows cereal yield trends (2000–2020), with the United States leading with a 1.5% annual increase, followed by North America (1.2%), South Asia (0.9%), and Africa Eastern and Southern (0.6%). Climate, technology, and infrastructural variations are the main causes of disparities, with the United States experiencing advantage conditions. Increased agricultural investment, increased financial conditions, and advances in technology have led to higher cereal yields in every country. Maintaining this pattern is critical for achieving the world's food needs, but long-term agricultural sustainability depends on addressing issues like climate change, water scarcity, and degradation of land.

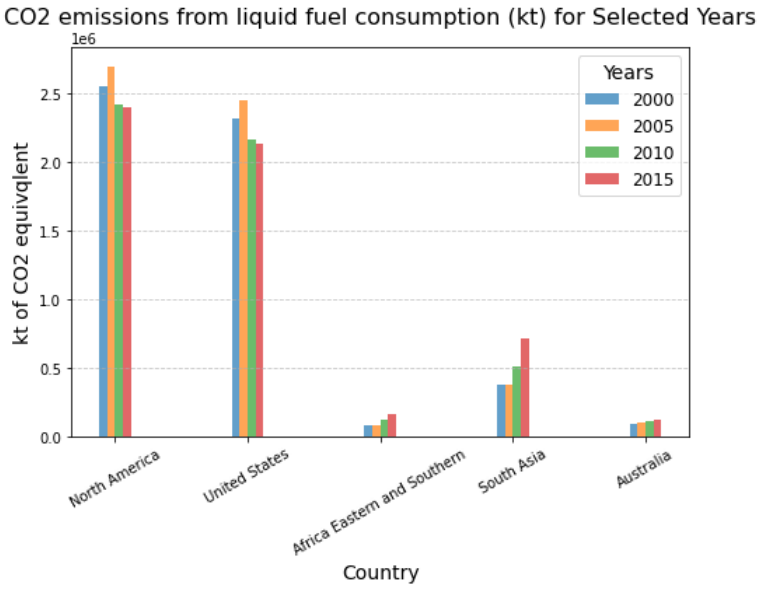



The plot depicts the global land-cover trend from 2000 to 2020, with an increase from 4.9 to 5.1 billion hectares. Africa Eastern and Southern lead with 1.7 billion hectares, followed by North America, the United States, Australia, and South Asia. Economic development, population growth, and supportive government policies contribute to increased agricultural land in these regions. While expanding agricultural land meets global food needs, it raises environmental concerns such as water pollution and deforestation. Recognizing limited land availability, adopting sustainable practices, and enhancing productivity are imperative for meeting global food demand without compromising crucial ecosystems like wetlands and forests.

The plot shows the dynamics of land cover for each region between 2000 and 2020. In North America and the United States, land cover remained relatively stable, with gradual increases observed. Meanwhile, Africa Eastern and Southern witnessed significant growth from 375,632 sq. km. in 2000 to 535,800 sq. km. in 2020. South Asia consistently expanded, reaching 1,004,819 sq. km in 2020. Australia's land cover fluctuated, experiencing a dip in 2010 and a subsequent rise, totaling 131,484 sq. km in 2020. These trends may stem from factors like economic development, population growth, and supportive government policies, underscoring the need for comprehensive environmental and agricultural planning.




The plot shows land cover dynamics (in square kilometres) across various regions from 2000 to 2015. The patterns in North America and the US were unstable, peaking in 2005. Africa Eastern and Southern exhibited overall growth, notably increasing from 77,461.56 sq. km in 2000 to 122,216.89 sq. km in 2010. Consistent growth saw South Asia reach 714,580.96 sq. km in 2015. Australia had a positive trend, which was especially apparent between 2000 and 2010. These patterns highlight the significance of comprehensive environmental and land management strategies by showing the dynamic nature of changes in land cover.

The distribution across five bins is represented by the histogram values for "access to electricity (% of population)" in South Asia. Regions with comparatively poorer access are probably included in the first bin (57.89%–66.73%), while significantly improved access is included in the second bin (66.73%–75.11%). While the fourth bin (86.78%–96.19%) indicates wide availability, the third bin (75.11%–86.78%) indicates continued growth in access. Near-universal access is indicated by the fifth bin, which is 96.19% and above. This distribution helps identify locations for focused interventions and development projects by offering a nuanced picture of power accessibility.