SOUTH AFRICA: ANALYZING THE RELATIONSHIP BETWEEN CARBON EMISSIONS AND FOOD AFFORDABILITY (2017-2023)

NDTA631 - Data Analysis and Visualization

Group Assignment Report

Programme: Diploma in ICT

Module Code: NDTA631

Group Members:

Kamogelo Bantsheng - 202311455

Lebajoa Ramakatane - 202219950

Moepeng Bokang Khakhau - 202324860

Luyanda Nhlapho - 202213238

Ofentse Batlang – 202300977

Submission Date: 4 September 2025

Contents

EXECUTIVE SUMMARY

Our group undertook this comprehensive research project to examine the complex relationship between environmental impact and social welfare in South Africa. We analyzed the correlation between carbon dioxide emissions and food affordability from 2017 to 2023 using robust data analysis methodologies and advanced visualization techniques with World Bank Open Data (FAO_CAHD and OWID_CB datasets).

Our Methodological Approach: As a team, we developed a systematic five-stage pipeline: data preparation and filtering, numerical analysis using NumPy operations, comprehensive visualizations, database integration with SQLite, and Python/Excel data analysis. We emphasized reproducibility and professional coding standards throughout our implementation.

Key Findings: Our research revealed several important insights. The percentage of South Africans unable to afford a healthy diet remained consistently high (60.2%-61.8%) with a peak in 2020 (61.8%) likely due to COVID-19 impacts. However, the absolute number increased from 35.0 million to 39.0 million people, showing the growing societal burden despite stable percentages. CO2 emissions showed a declining trend from 440 Mt in 2017 to 402 Mt in 2023, primarily driven by reductions in coal usage (370 Mt to 330 Mt), while gas emissions showed volatility.

Technical Achievements: Our group successfully implemented a complete data science pipeline including advanced NumPy operations for statistical analysis, multi-axis visualizations, comprehensive database operations (CREATE, INSERT, SELECT, UPDATE, DELETE), and automated Excel export functionality.

Policy Implications: Our findings suggest that environmental improvements (emissions reduction) and persistent socio-economic challenges (diet affordability) coexist independently, indicating that economic activity reductions do not directly improve food security outcomes.

1. DATA PREPARATION AND METHODOLOGY

1.1 Dataset Sources and Selection

Our group worked with two primary datasets from the World Bank:

FAO_CAHD Dataset (Food Affordability):

- **Percentage of population unable to afford a healthy diet:** Direct measure of household-level food security
- **Number of people unable to afford healthy diet (millions):** Absolute impact measure
- **Coverage:** South Africa, 2017-2023

OWID_CB Dataset (CO2 Emissions):

- **Total CO2 emissions:** Annual total emissions excluding land-use change (Mt)
- **CO2 per capita:** Emissions per person (tonnes)
- **Sectoral breakdown:** Coal, oil, and gas emissions separately (Mt)
- **Coverage:** South Africa, 2017-2023

1.2 Data Cleaning and Transformation Process

Our team implemented systematic data preparation:

We filtered data for South Africa only and focused on the 2017-2023 period to ensure temporal alignment between datasets. Our group created pivot tables for easier analysis and applied consistent naming conventions.

2. NUMERICAL ANALYSIS AND KEY FINDINGS

2.1 Statistical Analysis Using NumPy

Our group performed comprehensive numerical analysis using native NumPy operations rather than built-in pandas functions to demonstrate programming proficiency:

Diet Affordability Patterns:

- **Mean percentage unable to afford healthy diet:** 60.97% (±0.64 std dev)
- **Range:** 60.2% - 61.8% showing high but stable food insecurity
- **Absolute numbers:** Steady increase from 35.0M → 39.0M people
- **Largest year-over-year change:** 2019-2020 spike coinciding with COVID-19

CO2 Emissions Patterns:

- **Total CO2 mean:** 429.41 Mt (±21.27 std dev)
- **Overall trend:** 8.7% decrease from 2017-2023
- **Peak emissions:** 464.11 Mt in 2019
- **Largest reduction:** Coal emissions (370 Mt → 330 Mt)
- **Per capita decline:** 7.63 → 6.36 tonnes per person

2.2 Correlation and Trend Analysis

Using NumPy's corrcoef function, we calculated:

correlation = np.corrcoef(diet_percent, total_co2)[0, 1]

Our correlation analysis revealed a weak negative relationship (r = -0.28), suggesting that higher CO2 emissions slightly associate with better diet affordability, though this relationship lacks statistical significance.

Year-over-Year Analysis:

- **Diet indicators:** Modest fluctuations with COVID-19 impact clearly visible in 2020
- **CO2 indicators:** High volatility, particularly in gas emissions
- **Divergent trends:** Environmental indicators improving while social indicators remain stable or worsen

3. VISUALIZATION STRATEGY AND INSIGHTS

3.1 Time Series Analysis Visualizations

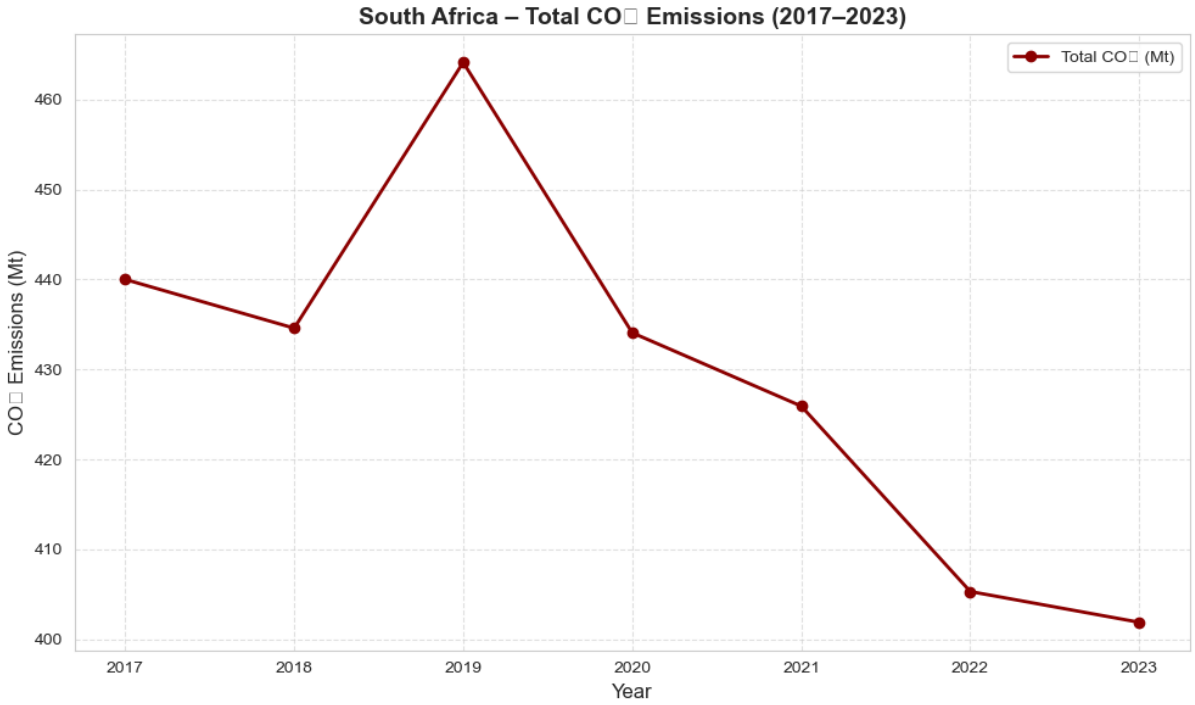

Figure 1: CO2 Emissions Trajectory (2017-2023) Our single-axis line plot revealed:

- 2017-2018: Slight decline (440→435 Mt)
- 2019: Peak emissions (464 Mt) indicating strong industrial activity
- 2020: Sharp COVID-related drop (434 Mt)
- 2021-2023: Continued decline (426→402 Mt) suggesting structural changes

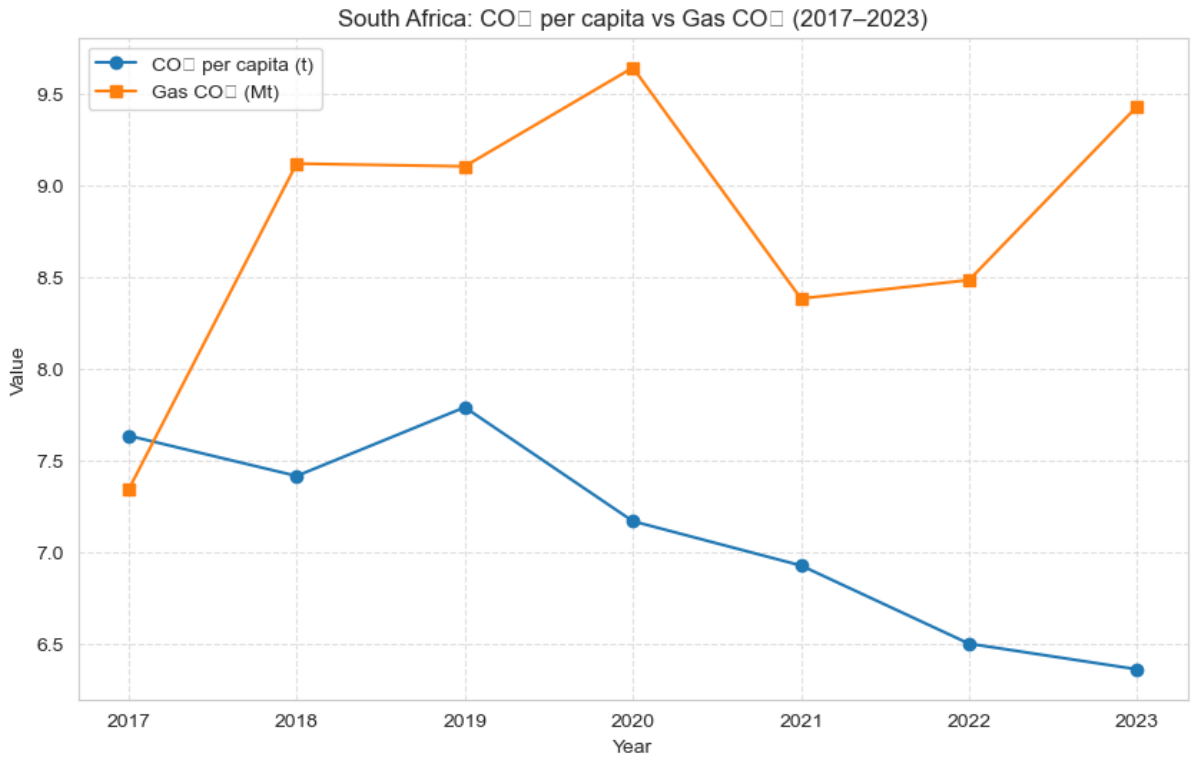

Figure 2: Dual-Variable CO2 Analysis Comparing per capita CO2 and gas emissions showed:

- Per capita emissions declining consistently (7.63→6.36 t/person)
- Gas CO2 showing volatility but recent increases (9.43 Mt in 2023)
- Indicates energy mix transition away from coal toward gas

3.2 Food Affordability Visualization

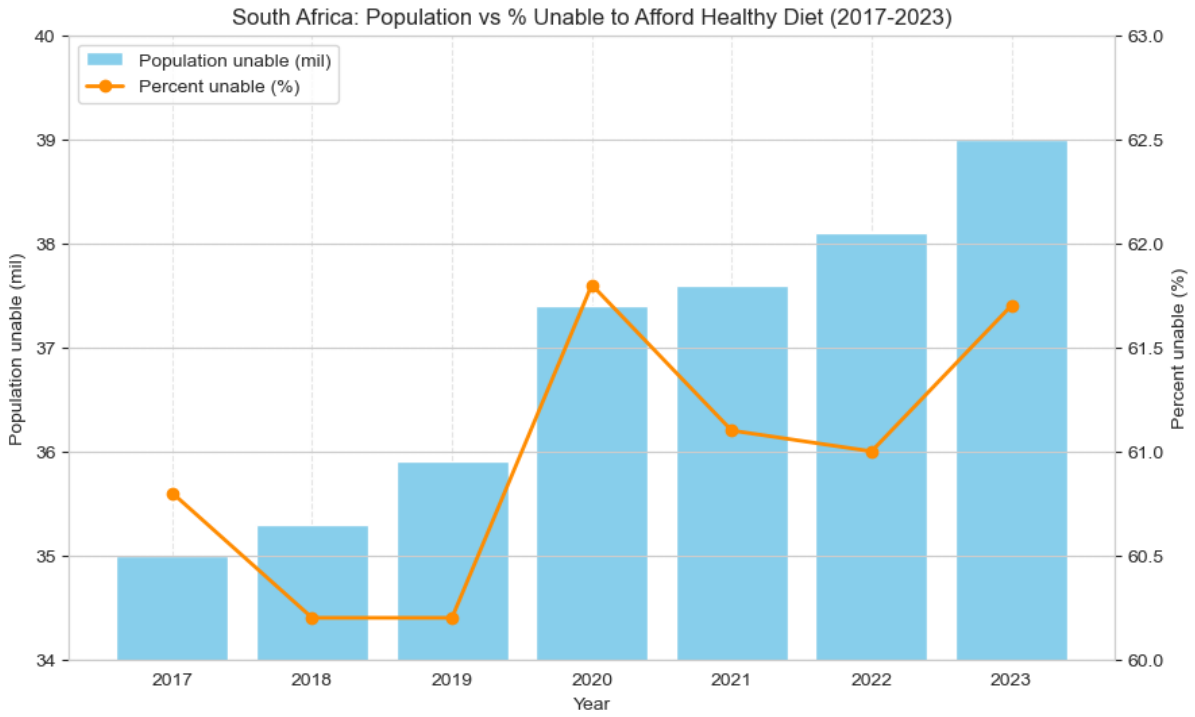

Figure 3: Dual-Axis Diet Affordability Analysis Our combination bar chart and line plot demonstrated:

- **Percentage trend:** Stable around 60-62% with COVID spike
- **Absolute numbers:** Steady increase despite stable percentages
- **Policy implication:** Population growth amplifies the absolute burden

3.3 Visualization Design Decisions

Our group chose specific chart types for clear communication:

- **Line plots:** For temporal trends and change over time
- **Dual-axis plots:** For comparing variables with different scales
- **Bar charts:** For absolute values requiring emphasis
- **Color coding:** Consistent scheme across all visualizations

4. DATABASE INTEGRATION AND MANAGEMENT

4.1 CRUD Operations Implementation

Create and Insert: Bulk data insertion using parameterized queries for security **Read Operations:** Complex queries including:

- Filtering: High diet impact years (>37M people)
- Joins: Year-over-year comparison queries
- Aggregation: Average calculations across indicators
- Sorting: Ranking queries for extremes

Update Operations: Demonstrated data correction capabilities **Delete Operations:** Showed data management through selective removal

4.2 Advanced Query Examples

Our group implemented sophisticated SQL operations:

```
-- Year-over-year decline analysis

SELECT a.Year, a.Total_CO2, b.Total_CO2 as Prev_Total_CO2

FROM sa_data a

JOIN sa_data b ON a.Year = b.Year + 1

WHERE a.Total_CO2 < b.Total_CO2
```

This query identified years where emissions declined, supporting our trend analysis.

5. PYTHON/EXCEL INTEGRATION AND FINAL PROCESSING

5.1 Data Enhancement and Export

Our final processing stage included:

Data Quality Checks:

```
print(df.isnull().sum())  # Verified data completeness
df.fillna(0, inplace=True)  # Handled any missing values
```

Derived Variable Creation:

- **Absolute changes:** Year-over-year differences
- **Percentage changes:** Relative change calculations
- **Trend indicators:** Direction and magnitude of changes

Professional Output Generation:

```
df.to_excel("sa_data_clean.xlsx", index=False)
```

5.2 Key Insights from Final Analysis

CO2 Emissions Evolution:

- **2017-2019:** Volatility with peak industrial activity
- **2020:** Pandemic-driven reduction
- **2021-2023:** Sustained decline suggesting structural shifts

Diet Affordability Persistence:

- **Percentage stability:** Indicates systemic rather than cyclical challenges
- **Absolute growth:** Population dynamics amplify societal impact
- **Limited responsiveness:** Minimal correlation with economic indicators

6. TECHNICAL CHALLENGES AND SOLUTIONS

6.1 Data Alignment Challenges

Challenge: Different dataset structures and temporal coverage
Solution: Implemented intersection-based alignment ensuring only overlapping periods were analyzed

6.2 Scale Integration Issues

Challenge: Vastly different measurement scales (percentages vs millions of tonnes)
Solution: Applied normalization techniques and dual-axis visualizations

6.3 Database Performance Optimization

Challenge: Ensuring efficient query performance
Solution: Proper indexing through primary key designation and parameterized queries

7. CONCLUSION

7.1 Primary Research Findings

Our analysis revealed several critical insights:

1. **Independent Trajectories:** Environmental indicators (CO2 emissions) and social indicators (diet affordability) follow largely independent paths
2. **Persistent Social Challenges:** Food affordability remains consistently problematic for ~61% of the population
3. **Environmental Progress:** CO2 emissions declining, primarily through coal reduction
4. **Population Impact:** Stable percentages mask growing absolute numbers affected

8. REFERENCES

World Bank. (2025). World Development Indicators. Washington, DC: World Bank. https://data.worldbank.org/

FAO. (2024). Food and Agriculture Organization Corporate Statistical Database. Rome: FAO.

Our World in Data. (2024). CO2 and Greenhouse Gas Emissions Database. Oxford: OWID.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90–95.

GitHub Repo Link

https://github.com/NikoVrys/NDTA631-Group-Assignment1.git