



Education vs Unemployment in South Africa: A Data Analysis

Module: NDTA 631 – Data Analysis and Visualization



Institution: Sol Plaatje University

Group Members: Boitelo Katleho Motsoto 202213685

Lebohang Rametsi 202300826

Nobuhle Thembalami Sibiya 202317888

Khensani Precious Maluleke 202306725

Vanessa Skosana 202226439

Zusiphe Xelelo 202301876

Mihlali Fana 202342128

Introduction

South Africa has a special economic and social problem: unemployment is consistently high even with significant investments in education. A major research topic is brought up by this paradox: do lower unemployment rates result from higher educational enrollment?

Two datasets from the World Bank Open Data platform school enrollment percentages and unemployment rates (by age and gender) were used in this project to examine this subject. We sought to clean, filter, and analyze the data in a way that reveals long-term trends by fusing Excel features (conditional formatting and visualization) with programming tools (Python, Pandas, NumPy, SQLite, and Matplotlib).

The results shed light on the relationship between education and work while also highlighting more serious structural issues with South Africa's labor system.

Dataset Details

The data for this analysis was sourced from **World Bank Open Data**. We selected two datasets:

1. **School Enrollment (% of school-age population)** – This dataset captures how many students are enrolled compared to the official school-age population.
2. **Unemployment Data** – Youth unemployment, as well as breakdowns by gender (Female, Male, and Total).

Dataset overview:

- Country: South Africa (filtered from global datasets).
- Years covered: ~1990–2022.
- Variables:
 - Female_Unemployment (%)
 - Male_Unemployment (%)
 - Total_Unemployment (%)
 - Education_Enrollment (%)

Both datasets were merged on **Country** and **Year** (TIME_PERIOD). This created a combined dataset for analysis.

https://data360.worldbank.org/en/indicator/WB_HCP_UNE_2EAP_MF_Y

https://data360.worldbank.org/en/indicator/WB_WDI_SE_SEC_ENRR

merged dataset - https://solplaatjeuniversity-my.sharepoint.com/:x/g/personal/202213685_spu_ac_za/EafELdJ74RJMrcYJBkZf0EBilyQ

[TA2IS8hkg2XO_9e1gQ?e=oDQw9Z&nav=MTVfe0Y0MEE0REQxLTVEMzUtNDMzM1BODIELThGRDM4Njg3MUJDNH0](https://data.worldbank.org/indicator/SL.UV.CD?locations=SA)

Data Preparation

Raw datasets often contain inconsistencies such as missing values, extra countries, or misaligned labels. The following cleaning steps were performed in Python (Jupyter Notebook):

1. **Filtering:** Only South Africa was retained, since the study focused exclusively on this country.
2. **Column renaming:** Indicators such as “*Youth Unemployment, Female (% of female labor force ages 15–24)*” were shortened to *Female_Unemployment*.
3. **Missing values:** Rows with no unemployment or enrollment data were dropped.
4. **Merging:** The two datasets were joined on **Year**.

After cleaning, the dataset contained consistent yearly records for South Africa, ready for numerical and visual analysis.

Numerical Analysis

Using **Pandas** and **NumPy**, we calculated several metrics:

- **Averages:**
 - Average Female Unemployment: ~44.34%
 - Average Male Unemployment: ~36.79%
 - Average Education_Enrollment: ~92%
 - Average Total:~40.1%
- **Gender Gap:**
 - Female unemployment was consistently higher than male unemployment by **8–10 percentage points**.
 - This gap persisted across decades, highlighting ongoing gender disparities in the labor market.
- **Trends:**
 - Education enrollment steadily increased, especially after the early 2000s.
 - Unemployment remained high, showing little correlation with education.

Data Visualisation

Several charts were created using **Matplotlib** and **Seaborn** to illustrate patterns:

1. Line Chart – Unemployment over Year

- Shows unemployment rising after 2010, peaking close to 50%.

2. Line Chart – Education Enrollment over Time

- Steady improvement in enrollment, surpassing 100% in recent years.

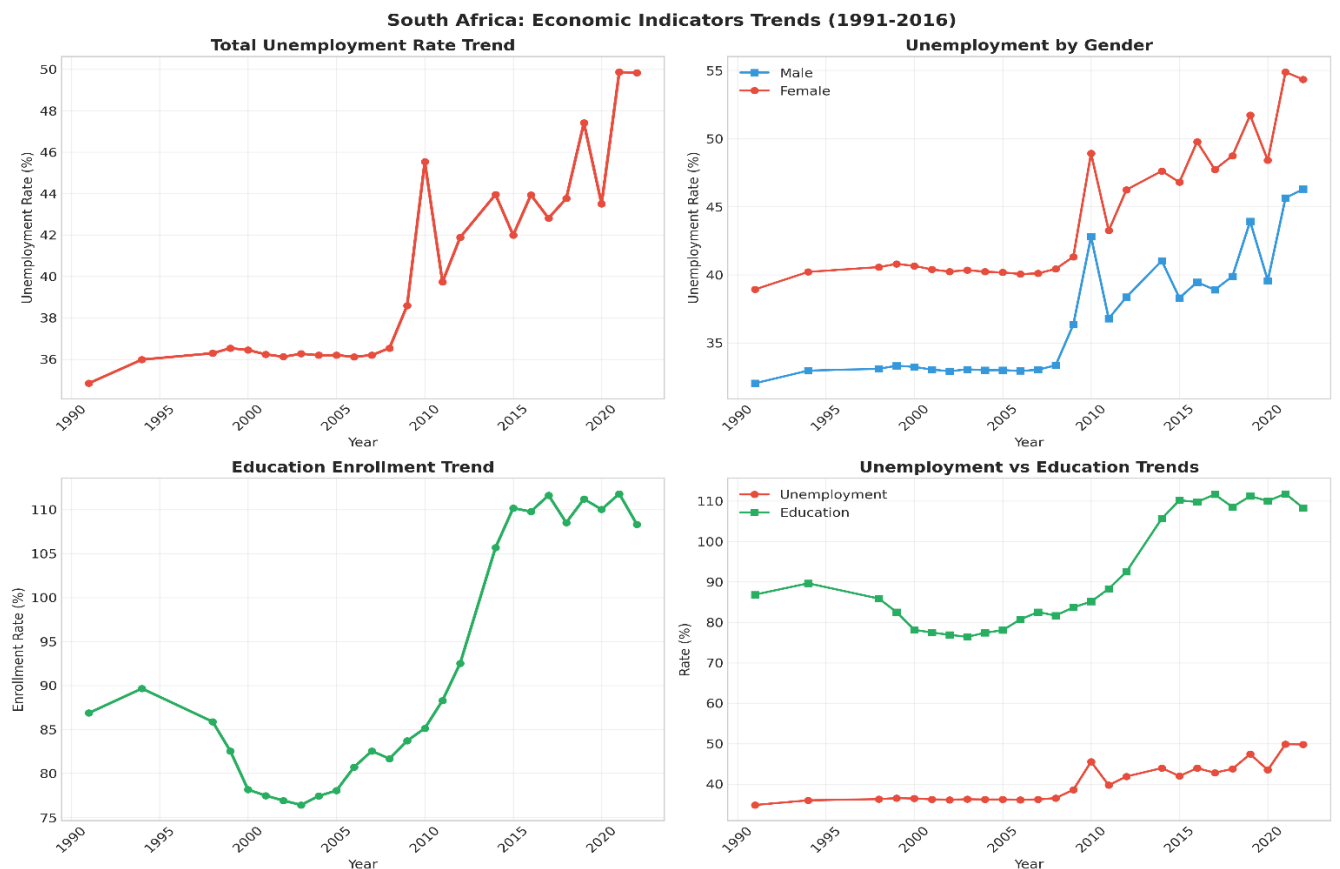
3. Comparison Plot (Enrollment vs Unemployment)

- Demonstrates that higher enrollment does not correlate with lower unemployment.

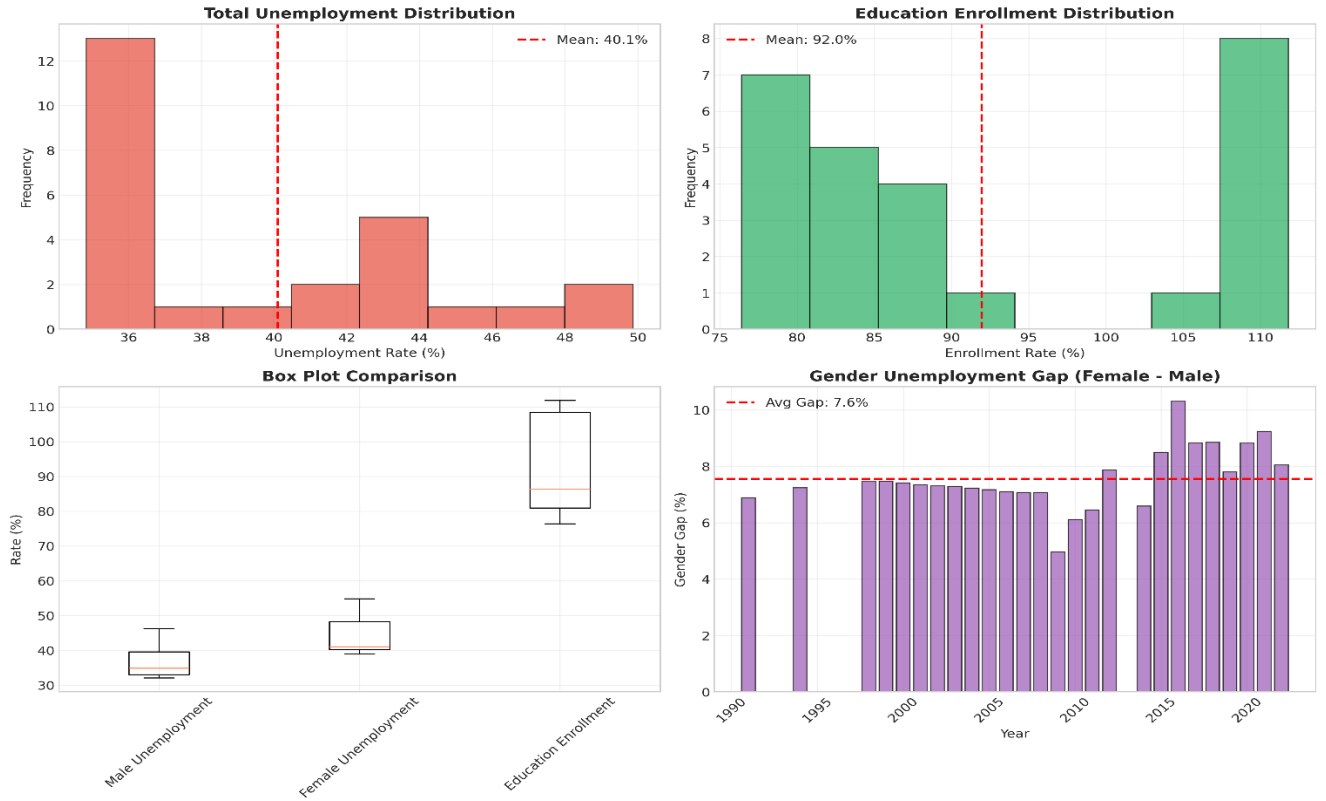
4. Gender Trends

- Female unemployment consistently above male unemployment.

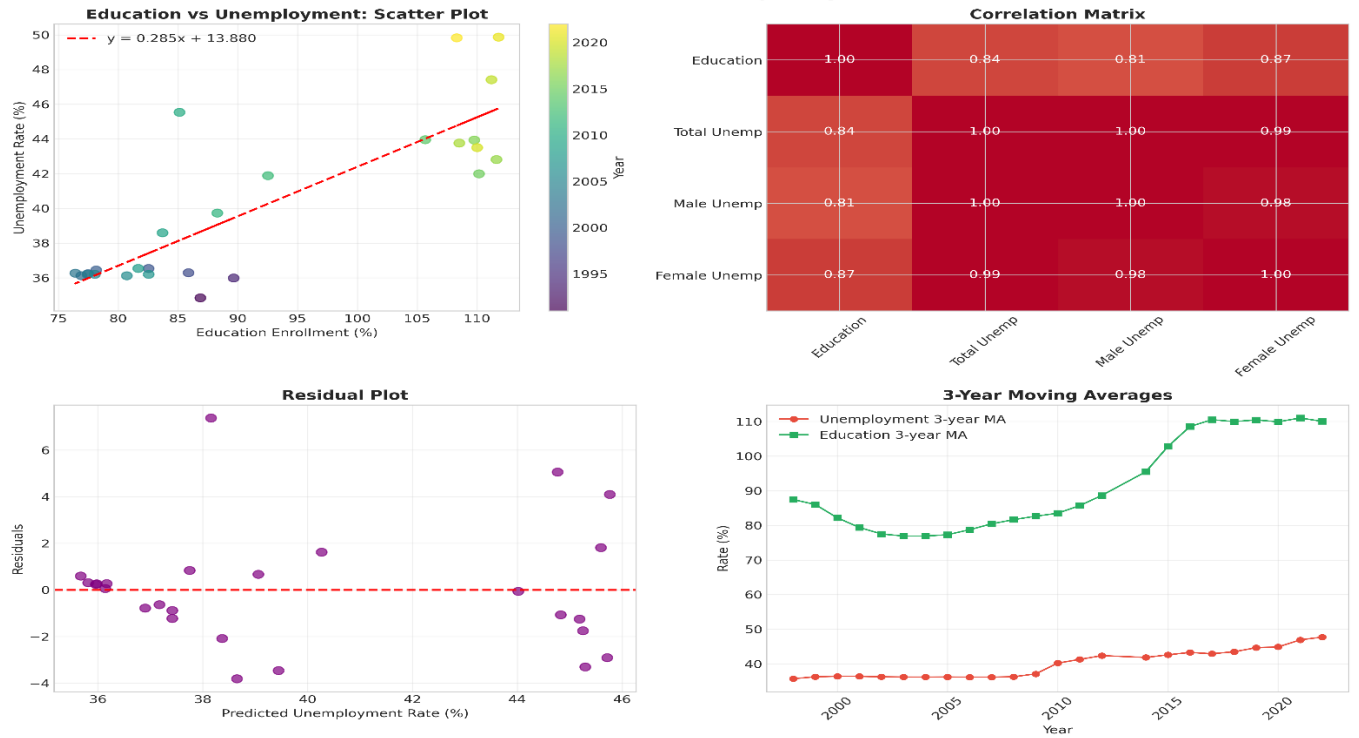
These visualisations make the trends easier to understand than raw numbers alone.

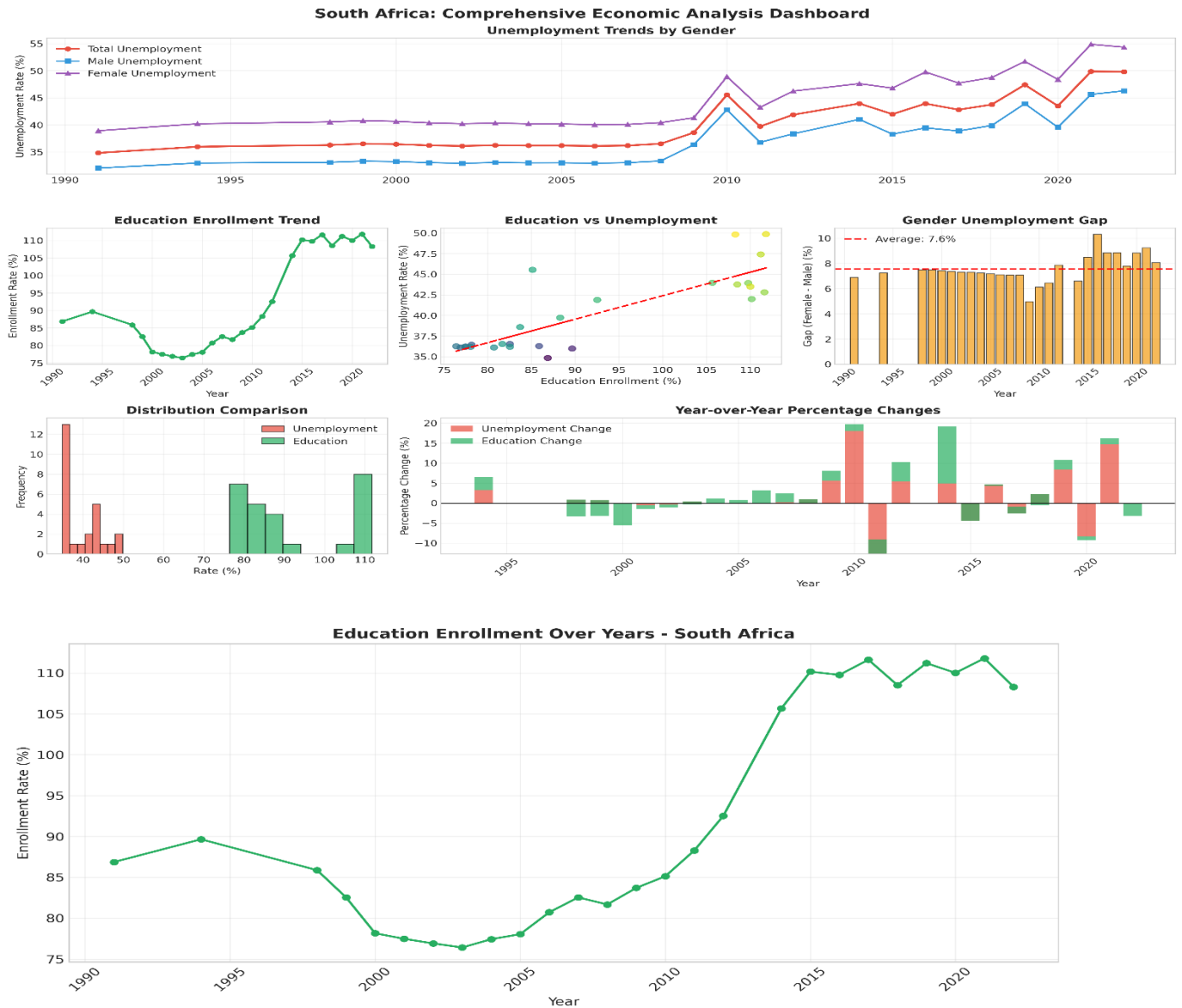


South Africa: Distribution Analysis



South Africa: Relationship Analysis





Database Integration

An **SQLite database** was created to manage the cleaned dataset. Three tables were included:

- school_enrollment
- youth_unemployment
- education_vs_youth_jobs (merged table)

We also tested **update and delete operations**, ensuring that records could be modified safely. Finally, the database was reloaded into Pandas to continue analysis in Python.

south_africa_analysis.db X

south_africa_analysis.db

Rows: 25 Filter 25 rows... Upgrade to PRO

	Country	Year	Female_Un...	Male_Un...	Total_Un...	Education_Enr...
1	South Africa	1994	40.215	32.973	35.984	89.6537322998047
2	South Africa	1998	40.568	33.104	36.289	85.8771896362305
3	South Africa	1999	40.798	33.33	36.534	82.5389862060547
4	South Africa	2000	40.657	33.243	36.444	78.1558532714844
5	South Africa	2001	40.394	33.047	36.236	77.4757308959961
6	South Africa	2002	40.234	32.932	36.12	76.9122772216797
7	South Africa	2003	40.347	33.063	36.263	76.4124908447266
8	South Africa	2004	40.227	33.001	36.194	77.4297485351562
9	South Africa	2005	40.182	33.012	36.198	78.0467529296875
10	South Africa	2006	40.053	32.947	36.117	80.7229995727539
11	South Africa	2007	40.106	33.035	36.198	82.5455474853516
12	South Africa	2008	40.433	33.364	36.536	81.6763763427734
13	South Africa	2009	41.327	36.353	38.59	83.7058792114258
14	South Africa	2010	48.911	42.799	45.534	85.1267929077148
15	South Africa	2011	43.247	36.79	39.729	88.2877883911133
16	South Africa	2012	46.24	38.367	41.884	92.5222702026367
17	South Africa	2014	47.615	41.028	43.954	105.673553466797
18	South Africa	2015	46.798	38.304	41.99	110.168930053711
19	South Africa	2016	49.767	39.463	43.933	109.774078369141
20	South Africa	2017	47.729	38.898	42.811	111.6274033
21	South Africa	2018	48.739	39.894	43.76	108.5205688
22	South Africa	2019	51.714	43.919	47.411	111.2068024
23	South Africa	2020	48.399	39.574	43	110.0174789
24	South Africa	2021	54.879	45.644	49.866	111.8019638
25	South Africa	2022	54.334	46.285	49.826	108.3046188

Python & Excel Data Analysis

Beyond Python, Excel was used to apply **conditional formatting** for further insights:

- **High Unemployment:** Any year with Total_Unemployment > 40% was highlighted.
- **Low Enrollment:** Enrollment below 80% was highlighted in yellow.
- **Data Bars:** Applied to unemployment to visualize severity.

F1 : X ✓ fx Education_Enrollment

	A	B	C	D	E	F	G
	Country	TIME_PERIOD	Female_Unemployment	Male_Unemployment	Total_Unemployment	Education_Enrollment	
1	South Africa	1994	40.215	32.973	35.984	89.6537323	
2	South Africa	1998	40.568	33.104	36.289	85.87718964	
3	South Africa	1999	40.798	33.33	36.534	82.53898621	
4	South Africa	2000	40.657	33.243	36.444	78.15585327	
5	South Africa	2001	40.394	33.047	36.236	77.4757309	
6	South Africa	2002	40.234	32.932	36.12	76.91227722	
7	South Africa	2003	40.347	33.063	36.263	76.41249084	
8	South Africa	2004	40.227	33.001	36.194	77.42974854	
9	South Africa	2005	40.182	33.012	36.198	78.04675293	
10	South Africa	2006	40.053	32.947	36.117	80.72299957	
11	South Africa	2007	40.106	33.035	36.198	82.54554749	
12	South Africa	2008	40.433	33.364	36.536	81.67637634	
13	South Africa	2009	41.327	36.353	38.59	83.70587921	
14	South Africa	2010	48.911	42.799	45.534	85.12679291	
15	South Africa	2011	43.247	36.79	39.729	88.28778839	
16	South Africa	2012	46.24	38.367	41.884	92.5222702	
17	South Africa	2014	47.615	41.028	43.954	105.6735535	
18	South Africa	2015	46.798	38.304	41.99	110.1689301	
19	South Africa	2016	49.767	39.463	43.933	109.7740784	
20	South Africa	2017	47.729	38.898	42.811	111.6274033	
21	South Africa	2018	48.739	39.894	43.76	108.5205688	
22	South Africa	2019	51.714	43.919	47.411	111.2068024	
23	South Africa	2020	48.399	39.574	43	110.0174789	
24	South Africa	2021	54.879	45.644	49.866	111.8019638	
25	South Africa	2022	54.334	46.285	49.826	108.3046188	

SA_Data

Ready Accessibility: Good to go

Findings from Excel:

According to the Excel conditional formatting, South Africa's unemployment rate has been extremely high for a long time, it frequently exceeds the 40% crisis level and even approaches 50% after 2010. Enrollment in education also increased gradually, from less than 80% in the late 1990s to more than 100% in the most recent years. Better access to education alone won't address the unemployment situation, though, as the rise in education hasn't resulted in lower unemployment, which suggests deeper economic problems.

Conclusion

Three main conclusions emerged from the analysis:

1. High jobless Persists: For many years, particularly after 2010, South Africa's jobless rate has continuously remained above 40%.
2. Education Enrollment Improved: In recent years, enrollment has risen significantly and surpassed 100%.
3. No Direct Link: Unemployment has not decreased in spite of increased access to education. This suggests that there are systemic problems with the economy, such as a lack of jobs, a mismatch in skills, or poor job creation.

The necessity for focused interventions is further supported by the gender analysis, which also showed that women experience greater unemployment rates than males.

References

-World Bank Open Data: <https://data.worldbank.org>

1. Lam, D., Leibbrandt, M. and Mlatsheni, C., 2009. Education and youth unemployment in South Africa. In Labor markets and economic development (pp. 90-109). Routledge.
<https://www.taylorfrancis.com/chapters/edit/10.4324/9780203881699-6/education-youth-unemployment-south-africa-david-lam-murray-leibbrandt-cecil-mlatsheni>
2. Reuters, 2025. South Africa's unemployment rate rises to 32.9% in first quarter 2025. [online] Available at: <https://www.reuters.com/world/africa/south-africas-unemployment-rate-rises-329-first-quarter-2025-05-13>
3. Statistics South Africa (Stats SA), 2025. Quarterly Labour Force Survey, Q2 2025: Media Release. [online] Available at: <https://www.statssa.gov.za/publications/P0211/Media%20Release%20QLFS%20Q2%202025.pdf>

4. Statistics South Africa (Stats SA), 2021. Labour Market Dynamics and Education Report. [online] Available at: <https://www.statssa.gov.za/publications/Report-03-01-81/PresentationUpdatedFinal.pdf>

GITHUB REPOSITORY LINK

<https://github.com/Leboh-Tech20/NDTA631-GROUP-ASSIGNMENT.git>