



<b>ASSESSMENT:</b>	GROUP ASSIGNMENT	<b>PROGRAMME:</b>	Diploma in ICT
<b>MODULE CODE:</b>	NDTA 631	<b>MARKS:</b>	100
<b>MODULE DESCRIPTION:</b>	Data Analysis And Visualization	<b>DATE:</b>	4 August 2025
<b>EXAMINER(S):</b>	Melvin Kisten	<b>DUE DATE:</b>	4 September 2025 <b>(NO EXTENSIONS)</b>
<b>GROUP SIZE:</b>	6 to 7 students		

### **INSTRUCTIONS:**

- Answer ALL questions.
- Indicate your **full names and student number** on your **answer sheet, code, etc.**
- **Copying from a fellow student or any source will be treated as plagiarism.**

#### **Submission Requirements**

- **Report (7 to 9 pages) DOCX and PDF**
- **Video demo (Optional but recommended)**
- Create dedicated GitHub repository, include:
  - Comprehensive codebase or Jupyter Notebook or Python scripts
  - Execution instructions
  - Environment configuration
  - Maintain clean, commented code
  - Use version control (**git**)
  - Implement error handling

## Group Assignment - Data Analysis and Visualisation:

### Dataset(s)

Each group must pick at least **two** different **South African** datasets from World Bank Open Data:

<https://data360.worldbank.org/en/search?country=ZAF>

Your **two South African** datasets must work together to tell one clear story.

No two groups may use the same datasets.

Before starting, check what datasets other groups have chosen so you do not repeat the same datasets.

**Else you will lose 15% of your final group mark!**

### Make a complete data analysis and visualisation pipeline:

- Load, clean, and prepare a min of 2 real datasets
- Use Python (Pandas, NumPy, Matplotlib, Seaborn) or Excel, etc
- Build and query a database
- Create clear visualisations
- Write an actionable report

### Deliverables:

1. **Data Preparation (15 marks)**
  - Load data and clean it
  - Handle missing values
  - Generate descriptive stats and insights
2. **Numerical Analysis (20 marks)**
  - Use NumPy for calculations
  - Reshape arrays and perform operations
  - Explain findings
3. **Visualisation (20 marks)**
  - Create plots (bar, scatter, box, histograms, etc)
  - Explain trends and patterns
  - Use colour and labels clearly
4. **Database Integration (20 marks)**
  - Build and query a database

- Update and delete records safely
- Load database data into Pandas

**5. Python/Excel Data Analysis (15 marks)**

- Clean and transform data
- Apply conditional formatting
- Create charts and summarise findings

**6. Report & Demo (10 marks)**

- Write a 7–9 page report
- Include dataset details, methods, visualisations, and conclusions
- Present with clear demo

**Marking Rubric (100 Marks)**

Criteria	Excellent	Good	Poor
<b>Q1: Data Preparation (15)</b>	13–15	8–12	1–7
<b>Q2: NumPy Analysis (20)</b>	17–20	10–16	1–9
<b>Q3: Visualisation (20)</b>	17–20	10–16	1–9
<b>Q4: Database (20)</b>	17–20	10–16	1–9
<b>Q5: Python/Excel Analysis (15)</b>	13–15	8–12	1–7
<b>Q6: Report &amp; Demo (10)</b>	8–10	5–7	1–4

----- All the best -----