

# Mthobisi Khoza

0681027155 | [khozamthobisi7@gmail.com](mailto:khozamthobisi7@gmail.com) | [linkedin.com/in/mthobisi-khoza](https://www.linkedin.com/in/mthobisi-khoza) | [github.com/BeAHappyMoron](https://github.com/BeAHappyMoron)

## EDUCATION

---

### University of the Free State

*Bachelor of Science in Computer Science, Minor in Management*

Qwaqwa, Phuthaditjhaba

February 2022 – November 2024

## EXPERIENCE

---

### Data Analyst Trainee

*Auditor-General South Africa*

April 2025 – Present

*Lynnwood, PTA*

- \* Utilized Power BI and SQL Server to design, refresh, and optimize interactive dashboards to enable financial and operational audit.
- \* Applied Apache Spark (PySpark) and Python for large-scale data processing, predictive modeling, and machine learning-based insights.
- \* Optimized SQL queries and developed ETL pipelines to ensure efficient data transformation, loading, and high-performance Power BI reporting.

### Business Calculations (Statistics) Tutor

*University of the Free State*

June 2024 – October 2024

*Qwaqwa, Phuthaditjhaba*

- \* Facilitated Business Calculations tutorials for undergraduate students, simplifying complex statistical concepts and improving learner comprehension.
- \* Designed and delivered engaging practical exercises and summaries aligned with course outcomes to reinforce statistical methods and formulas.
- \* Provided one-on-one academic support to students, including those registered with the Center for Universal Access and Disability Support (CUADS), fostering inclusive learning.

## PROJECTS

---

### Korema Farms 3D Website | *JavaScript, React.js, Three.js, Tailwind CSS*

February 2025 – Present

- \* Developed a full-stack web application using with Flask serving a REST API with React as the frontend
- \* Implemented GitHub OAuth to get data from user's repositories
- \* Visualized GitHub data to show collaboration
- \* Used Celery and Redis for asynchronous tasks

### PySpark SQL Server ETL | *Python, SQL Server, Apache Spark (PySpark), JDBC, Git*

May 2025 – Present

- \* Developed a PySpark ETL pipeline that connects to a SQL Server database using JDBC to extract data from dynamic source tables.
- \* Applied data cleaning techniques including trimming, lowercasing, filling missing values by type, and dropping duplicates for consistency and quality.
- \* Normalized string columns to proper case and dynamically handled schema fields to ensure flexible processing of various tables.
- \* Automated clean data export back to SQL Server under a new destination table using overwrite mode for downstream analytics.

## TECHNICAL SKILLS

---

**Languages:** Java, Python, C#, SQL Server, Apache Spark (PySpark), JavaScript, HTML/CSS, Tailwind CSS, Android Development (Java)

**Frameworks:** React, Three.js, . NET

**Developer Tools:** Git, Visual Studio, Webstorm, VS Code, Visual Studio, PyCharm, Jupyter

**Libraries:** Pandas, NumPy, Matplotlib, Seaborn, JDBC