**Ntokozo Matsipa's Project Overview**

**Risk Analysis Leave**

In this project, I took a lead role alongside a colleague to address a critical issue of leave underpayments and overpayments at the bank. We identified and validated discrepancies in employees' payments by analyzing historical data, including data that had been previously lost due to system maintenance. Our efforts provided the business with the assurance and validation needed to correct these financial inaccuracies. This project was executed using Python in Visual Studio Code, utilizing various libraries for data analysis.

**Consolidated View of People Risk in Human Capital**

For this initiative, we were tasked with creating a comprehensive business report for the Executive Committee (EXCO), covering all areas within the bank. I collaborated with three colleagues and focused on the data aspect of the report. My responsibilities included data collection, analysis, and visualization, which were then presented in a clear and accessible format using PowerPoint.

**Infobyte Presentation**

During a one-month online internship, I completed three tasks that demonstrated my capabilities in data analysis, machine learning, and data visualization. These tasks were encapsulated in three projects: CAR\_SALES\_PRICE, FLOWER\_CLASS\_PREDICTOR, and INDIA\_UNEMPLOYMENT. Each project showcased my skills and thorough understanding of these areas.

**CA Controls Target Definition**

In this project, we aimed to integrate various business areas into a unified controls target definition and subsequently create Hadoop tables. We utilized Hadoop for data processing and Excel for data management to ensure accurate and efficient control definitions.

**StatsSmart**

I collaborated with a team on a data science project called StatsSmart, which resulted in an innovative app that mimics the capabilities of a data scientist. This app allows users to upload a data file (such as a CSV), clean the data, visualize it, and gain insights through a chat interface connected to ChatGPT. Additionally, the app supports building machine learning models, selecting optimal features, identifying outliers through scatter plots, and more. This project was developed using Visual Studio Code, Python, and various extensions. You can explore the project further [here](https://github.com/Tokyo-The-skater/StatSmart_App).