

School Name: Blossom Academy

Program : Data Analytics

Cohort : Kigali-Rwanda

Step-by-Step Capstone Project Assignment

End-to-End Data Pipeline & Analytics Solution

Tools Used

- **KoboToolbox:** Data collection
- **VS Code & Python:** ETL (Extract, Transform, Load)
- **Github:** Storing code from Python (Vcode and Project)
- **PostgreSQL:** Data storage and analytics
- **Power BI:** Power Query, data modeling, and visualization

Project Overview

Students will design and implement an **end-to-end data pipeline** where data is:

Collected Kobotool box → Processed with Python (vs code) → Stored in PostgreSQL → Fetched into Power BI → Visualized for insights.

You may choose **any real-world scenario and collect data**(survey, operational, geospatial, etc.). The emphasis is on a **clean, reproducible, and professional ETL workflow.**

Step-by-Step Workflow

1. Problem Definition

- Define the problem, objectives, and expected insights
- Identify intended users or stakeholders

Deliverable: 1-page project overview

2. Data Collection (KoboToolbox)

- Design a validated KoboToolbox form
- Test with sample submissions

Deliverable: KoboToolbox form link or XLSForm

3. ETL with Python (VS Code & Github)

- Extract data from KoboToolbox (API/CSV)
- Transform and clean data using Python (vscode)

- Load data into PostgreSQL staging tables

Deliverable: Python ETL script + proof of data in PostgreSQL

4. PostgreSQL Data Modeling

- Design clean and analytical tables or views
- Apply naming standards and indexing

Deliverable: SQL schema & table creation scripts

5. Data Transformation & Analytics (PostgreSQL) (if needed)

- Clean and transform data using SQL
- Create derived fields and KPIs
- Answer at least **three analytical questions**

Deliverable: SQL transformation & analytics queries

6. Power BI: Data Fetching & Visualization

- Connect **Power BI to PostgreSQL**
- Use **Power Query** for final shaping
- Build an interactive dashboard

Dashboard Requirements:

- Minimum 7 visuals (one Page)
- Filters/slicers
- KPI

Deliverable: Power BI (.pbix) file + screenshots

⌚ Total: 1 Week

This assignment follows a **real-world ETL → Analytics → Visualization workflow**, aligned with industry practices demonstrated in end-to-end data pipeline tutorials.

CheckPoint:

After completing the data visualization, invite colleagues or other participants to fill out the KoboToolbox survey. Then refresh the Power BI dashboard. If you can see the updates reflected in the KPIs or visualizations, congratulations!!!! Your project is complete and ready for submission.

Prepared By:

Theoneste NDAYISENGA & Jeannine MUKASHEMA