Team:

- Group Name: Titans
- Members:
    - Name: Cynthia Onyia
    - Email: cyn4jun@gmail.com
    - Country: United Kingdom
    - College/Company: University of Wolverhampton
    - Specialization: Data Science
-
    - Name: Christopher Avbenake
    - Email: kristoville@gmail.com,
    - Country: United Kingdom,
    - College/Company: Retail Assets Solution,
    - Specialization: Data Science

Problem Description:

ABC Pharma Company approached a data analytics firm to understand the persistency of drugs with physician prescriptions. The objective is to identify key factors affecting drug persistency through data analysis using machine learning.

Business Understanding:

In understanding the persistency of drugs prescribed by physician will help ABC pharma company to optimize their marketing strategies, improve patient's adherence, and potentially increase drug sales. This project will aim to build a predictive model that can identify the likelihood of a patient adhering to prescribed medication based on clinical, demographic, and treatment factors.

Project Lifecycle and Deadline:

1. Problem Understanding
2. Data Understanding
3. Data Cleaning and Feature Engineering
4. Model Development
5. Model Selection
6. Model Evaluation
7. Reporting and Documentation
8. Deployment

Total Duration**: 30 to 40 Days**

Data Intake Report:

- Source: Provided Healthcare dataset
- Data Type: Structured Data
- Format: Excel (.xlsx)
- Key Variables:
    - Patient ID
    - Persistency_Flag (Target Variable)
    - Age, Gender, Race, Ethnicity, Region
    - Physician Specialty, T-Score, Risk Segment
    - Dexa Scan Frequency, Fragility Fracture Recency
    - Glucocorticoid Usage, Injectable Experience
    - Risk Factors, Comorbidity, Concomitancy, Adherence

All of the above incudes data exploration, cleaning, and feature engineering.

Data Intake Report

Name: **Healthcare - Persistency of a drug**
Report date: **Proposed date of 30th June 2025**
Internship Batch: **LISUM44**
Version:<1.0>
Data intake by: **Cynthia Onyia**
Data intake reviewer: **Christopher Avbenake**
Data storage location: **https://github.com/MelCyn/Healthcare--Drug-Persistency**
Tabular data details:

Total number of observations	3425
Total number of files	1
Total number of features	69
Base format of the file	xlsx
Size of the data	899KB ie 0.899 ~ 0.9MB

Proposed Approach:
Since this is the beginning of the project, the approach of validation and assumptions will be noted during the course of the project.

NB: This report just gives the details of the raw data.