

## Data Glacier Team: The Data Dynamos

Batch Code: LISUM04

### Internship Cohort: September to December 2021

#### 1. Team member's details:

Shari Arangi  
[Shariarangi504@gmail.com](mailto:Shariarangi504@gmail.com)  
Ph +254 0742164535  
We! Hub Victoria LTD, Kenya  
Data Science

Ateeb Aqil Aziz  
[Atteeb.aqil@gmail.com](mailto:Atteeb.aqil@gmail.com)  
Ph +92 3044693262  
Data Science

Jennifer Turley  
[Jennifer.turley@ucdconnect.ie](mailto:Jennifer.turley@ucdconnect.ie)  
Ph +353 83 065 7252  
University College Dublin, Ireland  
Data Science

Queen Echerenwa  
[Quin\\_codes@outlooks.com](mailto:Quin_codes@outlooks.com)  
+234 806 352 8847  
Data Science

#### 2. Problem Description

Data Science: Healthcare – Persistency of a Drug Project

This project utilizes data relating to treatments for improving bone density.

One of the challenges in medicine is to understand patient-specific factors that contribute to treatment outcomes. An important factor relates to patient persistence with prescribed medical therapy. The persistency of a drug for a patient in this data set refers to whether a given patient took the medication and continued taking it as prescribed. With patient persistency as the target variable and other variables of interest such as age, race, ethnicity, religion, gender along with medical comorbidities considered as dependent variables, this project seeks to determine whether and to what extent these other variables correlate with the desired positive patient persistency for the prescribed bone density treatment.

#### 3. Business Understanding

Pharmaceutical companies profit when their drugs are utilized by patients. To the extent patients fail to follow through by taking their medication as prescribed, the pharmaceutical company loses. The first loss comes directly from the patient not continuing to fill the prescription. However, as prescribing doctors may not always

be aware of a patient's non-persistence with a medication, there is a secondary risk that doctors could blame their patient's lack of improvement on the drug rather than on the patient's non-persistence taking it. The pharmaceutical company therefore has an interest in understanding why and when patients do not persist in treatment with their drug so this information can be shared with doctors. By identifying which factors put a patient at risk for non-persistence, doctors can then work proactively to educate similar patients to reduce non-persistence, leading to better patient outcomes and improved reputation and profit for the pharmaceutical company.

#### 4. Project Lifecycle

- a. Weeks 7/8 Data cleaning and feature engineering
- b. Weeks 9/10 Model development and model selection, from preliminary evaluation of results to selection of final model with explanation for the choice
- c. Weeks 11/12 Final model deployment and comprehensive reporting of results including PPT presentation

#### 5. Data intake report

Name: Drug Persistency Project

Report date: 6 November 2021

Internship Batch: Sept to December 2021

Version:<1.0>

Data intake by: Shari Arangi

Data intake reviewers: Queen Echerenwa, Jennifer Turley

Data storage location: <https://github.com/jenturley27/DataDynamosTeamProject>

##### Tabular data details:

<b>Total number of observations</b>	3424
<b>Total number of files</b>	1
<b>Total number of features</b>	69
<b>Base format of the file</b>	Xlsx
<b>Size of the data</b>	898 KB

##### note

The unique row id in the data is the **patient ID**

The column **Count\_Of\_Risks** and **Dexa\_Freq\_During\_Rx** display outliers

#### 6. Github repo link:

<https://github.com/jenturley27/DataDynamosTeamProject>