**TEAM MEMBERS DETAILS**

GROUP NAME: Health Research

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Email | Country | College/Company | Specialisation |
| Kelvin Mpofu | mpofukelvintafadzwa@gmail.com | South Africa | n/a | Data science |
| Purity Nyagweth | purityeverniter@gmail.com | Kenya | n/a | Data Science |
| Reshma Jayapalan | reshma.jayapalan@gmail.com | UAE | n/a | Data Science |
| Alhanouf Alghamdi | hanouf.haz@gmail.com | Saudi Arabia | n/a | Data Science |

**Problem Description**

ABC pharma company has a challenge in understanding drug persistency as per physician prescription and to solve this problem it wants to automate the process of identification.

Objective is to build a classification model for drug persistency identification.

This will automate the process of identifying drug persistency for ABC pharma company thus helping the company to understand drug persistency as per physician prescription.

**Data Understanding**

Summary of the data received in excel sheet:



Data shape – 3424 rows x 69 columns

Categorical variables – 67

Numerical variables - 2

Problems with the data:

Number of missing values – 0

Outliers – There are outliers in the numerical columns; Dex\_Freq\_During\_Rx and Count\_Of\_Risk

Skewed data – The numerical columns have positively skewed data.