Group Details:

Group Name	SoleDS
Name	Anju Paul
Email	anjupaul14@yahoo.co.in
Country	UK
College	Anglia Ruskin University
Specialization	Data Science

Problem Description:

To understand the persistency of a drug as per the prescription given by the physician is an important question faced by the pharmaceutical companies. The problem here is to build a classification model to understand the persistency (persistent or not) of a drug for the given dataset.

EDA Presentation & Recommendation:

Summary

- From the Exploratory Data Analysis done, we are able to find how the different features/variables affects drug persistency.
- Dropped the unwanted columns
- The categorical variables are converted to numerical variables
- A total of 6 principal components can contribute a 0.4692567637481701% of variance.

Recommendation

For the purpose of automating the process of drug persistency identification, the following machine learning models can be used:

- **Logistic regression** – It is a type of linear model that is used for binary classification. It predicts output which is a categorical dependent variable. Such predictions are like yes or no, A or B, etc.
- **Random Forest** – It is a type of Bagging Ensemble Learning Classification. It predicts by the averaging of a number of decision tree classifiers.
- Adaboost and XGBoost – It is a type of Boosting model. It converts weak learners to strong learners.