Group details

Group Name: TBC

Team member:

Name	Email	Country	College	Specialization
Yuchi Chen	ychen306@ur.rochester.edu	United States	University of Rochester	Data Science
Bolutife Akinlawon	bolu.akinlawon@gmail.com	United Kingdom	UWE, Bristol	Data Science
Alexis Collier	colliera75@gmail.com	United States	Emory University Bootcamp	Data Science
Han-Fu Lin	hanfu.lin@mail.utoronto.ca	Canada	University of Totonto	Data science

Github Repo link: [Healthcare Persistency of a drug](#)

Problem description

A pharmaceutical company conducts a large number of clinical trials in order to study the durability of a new drug. These trials record a large number of different attributes of experimental subjects and the results of the experiment by means of control variables. The company wants to use the data to understand what properties affect the drug's durability.

Data understanding

Type of Data: 2 numeric columns + 66 categorical columns

There are not any missing values for all columns in the dataset.

Outlier detection: Though some columns are highly skewed, such as Dexa_Freq_During_Rx, there is no value that should be considered an outlier.

To deal with the numeric column Dexa_Freq_During_Rx, consider doing nothing or normalization using log transformation or capping, or binning it into ranges manually. After the number of Dexa scans reaches a threshold, higher values will

not make a big difference, so transformation can be applied to this right-skewed feature.

Other skewed features:

'Gender',

'Dexa_Freq_During_Rx',

'Adherent_Flag',

'Risk_Type_1_Insulin_Dependent_Diabetes',

'Risk_Osteogenesis_Imperfecta',

'Risk_Rheumatoid_Arthritis',

'Risk_Untreated_Chronic_Hyperthyroidism',

'Risk_Untreated_Chronic_Hypogonadism',

'Risk_Untreated_Early_Menopause',

'Risk_Patient_Parent_Fractured_Their_Hip',

'Risk_Chronic_Liver_Disease',

'Risk_Low_Calcium_Intake',

'Risk_Poor_Health_Frailty',

'Risk_Excessive_Thinness',

'Risk_Hysterectomy_Oophorectomy',

'Risk_Estrogen_Deficiency',

'Risk_Immobilization',

'Risk_Recurring_Falls'