Final Project (Individual)

Team Primero

Date: 25th July 2021

LISUM01

Name: Reeha Khan
Email: khanreeha22@gmail.com
Country: Pakistan
College: National University of Sciences and Technology
Specialization: Data Science

Problem description

ABC is a pharmaceutical company that wants to understand the persistency of a drug as per the physician's prescription for a patient. The company has approached an Analytics company to automate this process of identification. This Analytics company has assigned the task to Team Primero to come up with a solution to automate the persistency of a drug for the client ABC.

Data Set

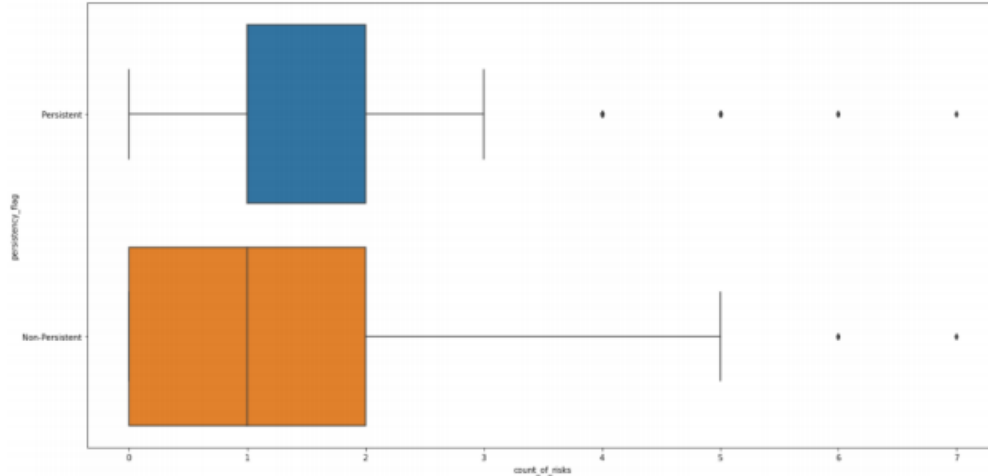| Bucket	Variable	Variable Description
Unique Row Id	Patient ID	Unique ID of each patient
Target Variable	Persistency_Flag	Flag indicating if a patient was persistent or not
Demographics	Age	Age of the patient during their therapy
	Race	Race of the patient from the patient table
	Region	Region of the patient from the patient table
	Ethnicity	Ethnicity of the patient from the patient table
	Gender	Gender of the patient from the patient table
	IDN Indicator	Flag indicating patients mapped to IDN
Provider Attributes	NTM - Physician Specialty	Specialty of the HCP that prescribed the NTM Rx
Clinical Factors	NTM - T-Score	T Score of the patient at the time of the NTM Rx (within 2 years prior from rxdate)
	Change in T Score	Change in Tscore before starting with any therapy and after receiving therapy (Worsened, Remained Same, Improved, Unknown)
	NTM - Risk Segment	Risk Segment of the patient at the time of the NTM Rx (within 2 years days prior from rxdate)
	Change in Risk Segment	Change in Risk Segment before starting with any therapy and after receiving therapy (Worsened, Remained Same, Improved, Unknown)
	NTM - Multiple Risk Factors	Flag indicating if patient falls under multiple risk category (having more than 1 risk) at the time of the NTM Rx (within 365 days prior from rxdate)
	NTM - Dexa Scan Frequency	Number of DEXA scans taken prior to the first NTM Rx date (within 365 days prior from rxdate)
	NTM - Dexa Scan Recency	Flag indicating the presence of Dexa Scan before the NTM Rx (within 2 years prior from rxdate or between their first Rx and Switched Rx; whichever is smaller and applicable)
	Dexa During Therapy	Flag indicating if the patient had a Dexa Scan during their first continuous therapy
	NTM - Fragility Fracture Recency	Flag indicating if the patient had a recent fragility fracture (within 365 days prior from rxdate)
	Fragility Fracture During Therapy	Flag indicating if the patient had fragility fracture during their first continuous therapy
	NTM - Glucocorticoid Recency	Flag indicating usage of Glucocorticoids (>=7,5mg strength) in the one year look-back from the first NTM Rx
	Glucocorticoid Usage During Therapy	Flag indicating if the patient had a Glucocorticoid usage during the first continuous therapy
	NTM - Injectable Experience	Flag indicating any injectable drug usage in the recent 12 months before the NTM OP Rx
Disease/Treatment Factor	NTM - Risk Factors	Risk Factors that the patient is falling into. For chronic Risk Factors complete lookback to be applied and for non-chronic Risk Factors, one year lookback from the date of first OP Rx
	NTM - Comorbidity	Comorbidities are divided into two main categories - Acute and chronic, based on the ICD codes. For chronic disease we are taking complete look back from the first Rx date of NTM therapy and for acute diseases, time period before the NTM OP Rx with one year lookback has been applied
	NTM - Concomitancy	Concomitant drugs recorded prior to starting with a therapy(within 365 days prior from first rxdate)
	Adherence	Adherence for the therapies

Data Types

The dimension of the dataset is (3424, 69). The features have the following datatypes. ("object" types mean categorical columns):

```
Ptid                                                              object
Persistency_Flag                                                 object
Gender                                                           object
Race                                                             object
Ethnicity                                                        object
Region                                                           object
Age_Bucket                                                       object
Ntm_Speciality                                                   object
Ntm_Specialist_Flag                                              object
Ntm_Speciality_Bucket                                            object
Gluco_Record_Prior_Ntm                                           object
Gluco_Record_During_Rx                                           object
Dexa_Freq_During_Rx                                              int64
Dexa_During_Rx                                                   object
Frag_Frac_Prior_Ntm                                              object
Frag_Frac_During_Rx                                              object
Risk_Segment_Prior_Ntm                                           object
Tscore_Bucket_Prior_Ntm                                          object
Risk_Segment_During_Rx                                           object
Tscore_Bucket_During_Rx                                          object
Change_T_Score                                                   object
Change_Risk_Segment                                              object
Adherent_Flag                                                    object
Idn_Indicator                                                    object
Injectable_Experience_During_Rx                                  object
Comorb_Encounter_For_Screening_For_Malignant_Neoplasms           object
Comorb_Encounter_For_Immunization                                object
Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx  object
Comorb_Vitamin_D_Deficiency                                      object
Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified             object
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx object
Comorb_Long_Term_Current_Drug_Therapy                            object
Comorb_Dorsalgia                                                 object
Comorb_Personal_History_Of_Other_Diseases_And_Conditions         object
Comorb_Other_Disorders_Of_Bone_Density_And_Structure             object
Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias  object
Comorb_Osteoporosis_without_current_pathological_fracture        object
Comorb_Personal_history_of_malignant_neoplasm                    object
Comorb_Gastro_esophageal_reflux_disease                          object
Concom_Cholesterol_And_Triglyceride_Regulating_Preparations      object
Concom_Narcotics                                                 object
Concom_Systemic_Corticosteroids_Plain                            object
Concom_Anti_Depressants_And_Mood_Stabilisers                     object
Concom_Fluoroquinolones                                          object
Concom_Cephalosporins                                            object
Concom_Macrolides_And_Similar_Types                              object
Concom_Broad_Spectrum_Penicillins                                object
Concom_Anaesthetics_General                                      object
Concom_Viral_Vaccines                                            object
Risk_Type_1_Insulin_Dependent_Diabetes                           object
Risk_Osteogenesis_Imperfecta                                     object
Risk_Rheumatoid_Arthritis                                        object
Risk_Untreated_Chronic_Hyperthyroidism                           object
Risk_Untreated_Chronic_Hypogonadism                              object
Risk_Untreated_Early_Menopause                                   object
Risk_Patient_Parent_Fractured_Their_Hip                          object
Risk_Smoking_Tobacco                                             object
Risk_Chronic_Malnutrition_Or_Malabsorption                       object
Risk_Chronic_Liver_Disease                                       object
Risk_Family_History_Of_Osteoporosis                              object
Risk_Low_Calcium_Intake                                          object
Risk_Vitamin_D_Insufficiency                                     object
Risk_Poor_Health_Frailty                                         object
Risk_Excessive_Thinness                                          object
Risk_Hysterectomy_Oophorectomy                                   object
Risk_Estrogen_Deficiency                                         object
Risk_Immobilization                                              object
Risk_Recurring_Falls                                             object
Count_Of_Risks                                                   int64
```
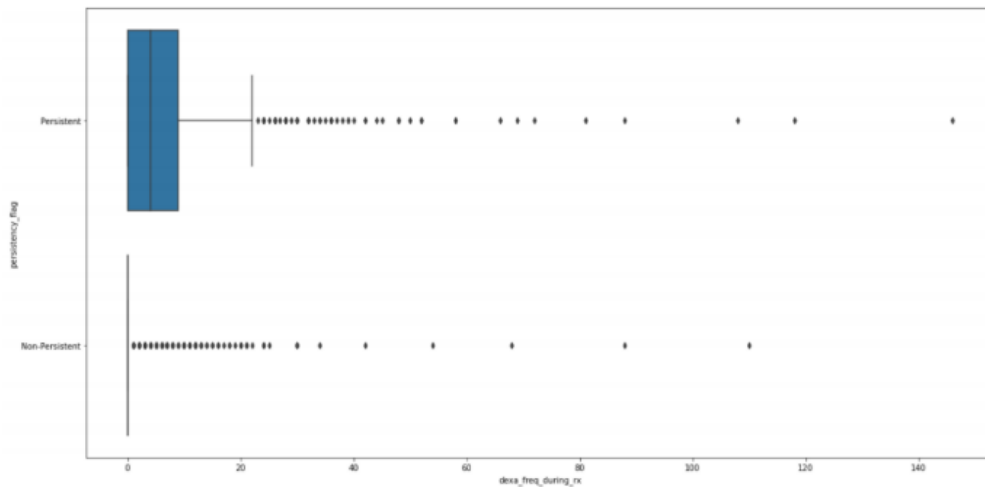
Data Problems

- Null Values: The dataset has no Null values.
- Outliers: There are only two numerical columns, both of which have some outliers.
  1. count_of_risks:

  2. dexa_freq_during_rx:

- Skewness and Kurtosis: There are only two numerical columns, both of which have some outliers.
1. count_of_risks:
   a. Count of risks skewness: 0.8797905232898707
   b. Count of risks Kurtosis: 0.9004859968892842
2. dexa_freq_during_rx:
   a. dexa_freq_during_rx skewness: 6.8087302112992285
   b. dexa_freq_during_rx Kurtosis: 74.75837754795428