Week 8: Data Science Healthcare Project

Name: Serhat Uğur

E-mail: ugur.serhat@outlook.com

Country: Turkey

University: Anadolu University

Specialization: Data Science

Problem Description

One of the challenges for all pharmaceutical companies is to understand the persistence of drugs as per the physician's prescription. To solve this problem, ABC Pharma Company approached an analytics company to automate the identification process.

Data Understanding
Columns:

1- Unique Row Id

• **Patient ID:** Unique ID of each patient.

2- Target Variable

• **Persistency_Flag:** Flag indicating if a patient was persistent or not.

3- 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.

4- Provider Attributes

• **NTM - Physician Specialty:** Specialty of the HCP that prescribed the NTM Rx.

5- 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 Facture 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

6- Disease/Treatment Factor

- **NTM - Injectable Experience:** Flag indicating any injectable drug usage in the recent 12 months before the NTM OP Rx

- **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

The original dataset contains 3424 rows and 69 columns. Most of the columns have an object datatype. There are no null data in sight, but we will look at them deeply next week. We will also look at outliers and skewness in the data.

```
RangeIndex: 3424 entries, 0 to 3423
Data columns (total 69 columns):
 #   Column                                                            Non-Null Count  Dtype
---  ------                                                            --------------  -----
 0   Ptid                                                              3424 non-null   object
 1   Persistency_Flag                                                  3424 non-null   object
 2   Gender                                                            3424 non-null   object
 3   Race                                                              3424 non-null   object
 4   Ethnicity                                                         3424 non-null   object
 5   Region                                                            3424 non-null   object
 6   Age_Bucket                                                        3424 non-null   object
 7   Ntm_Speciality                                                    3424 non-null   object
 8   Ntm_Specialist_Flag                                               3424 non-null   object
 9   Ntm_Speciality_Bucket                                             3424 non-null   object
 10  Gluco_Record_Prior_Ntm                                            3424 non-null   object
 11  Gluco_Record_During_Rx                                            3424 non-null   object
 12  Dexa_Freq_During_Rx                                               3424 non-null   int64
 13  Dexa_During_Rx                                                    3424 non-null   object
 14  Frag_Frac_Prior_Ntm                                               3424 non-null   object
 15  Frag_Frac_During_Rx                                               3424 non-null   object
 16  Risk_Segment_Prior_Ntm                                            3424 non-null   object
 17  Tscore_Bucket_Prior_Ntm                                           3424 non-null   object
 18  Risk_Segment_During_Rx                                            3424 non-null   object
 19  Tscore_Bucket_During_Rx                                           3424 non-null   object
 20  Change_T_Score                                                    3424 non-null   object
 21  Change_Risk_Segment                                               3424 non-null   object
 22  Adherent_Flag                                                     3424 non-null   object
 23  Idn_Indicator                                                     3424 non-null   object
 24  Injectable_Experience_During_Rx                                   3424 non-null   object
 25  Comorb_Encounter_For_Screening_For_Malignant_Neoplasms            3424 non-null   object
 26  Comorb_Encounter_For_Immunization                                 3424 non-null   object
 27  Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx   3424 non-null   object
 28  Comorb_Vitamin_D_Deficiency                                       3424 non-null   object
 29  Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified              3424 non-null   object
 30  Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx 3424 non-null  object
 31  Comorb_Long_Term_Current_Drug_Therapy                             3424 non-null   object
 32  Comorb_Dorsalgia                                                  3424 non-null   object
 33  Comorb_Personal_History_Of_Other_Diseases_And_Conditions          3424 non-null   object
 34  Comorb_Other_Disorders_Of_Bone_Density_And_Structure              3424 non-null   object
 35  Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias   3424 non-null   object
 36  Comorb_Osteoporosis_without_current_pathological_fracture         3424 non-null   object
 37  Comorb_Personal_history_of_malignant_neoplasm                     3424 non-null   object
 38  Comorb_Gastro_esophageal_reflux_disease                           3424 non-null   object
 39  Concom_Cholesterol_And_Triglyceride_Regulating_Preparations       3424 non-null   object
 40  Concom_Narcotics                                                  3424 non-null   object
 41  Concom_Systemic_Corticosteroids_Plain                             3424 non-null   object
 42  Concom_Anti_Depressants_And_Mood_Stabilisers                      3424 non-null   object
 43  Concom_Fluoroquinolones                                           3424 non-null   object
 44  Concom_Cephalosporins                                             3424 non-null   object
 45  Concom_Macrolides_And_Similar_Types                               3424 non-null   object
 46  Concom_Broad_Spectrum_Penicillins                                 3424 non-null   object
 47  Concom_Anaesthetics_General                                       3424 non-null   object
 48  Concom_Viral_Vaccines                                             3424 non-null   object
 49  Risk_Type_1_Insulin_Dependent_Diabetes                            3424 non-null   object
 50  Risk_Osteogenesis_Imperfecta                                      3424 non-null   object
 51  Risk_Rheumatoid_Arthritis                                         3424 non-null   object
 52  Risk_Untreated_Chronic_Hyperthyroidism                            3424 non-null   object
 53  Risk_Untreated_Chronic_Hypogonadism                               3424 non-null   object
 54  Risk_Untreated_Early_Menopause                                    3424 non-null   object
 55  Risk_Patient_Parent_Fractured_Their_Hip                           3424 non-null   object
 56  Risk_Smoking_Tobacco                                              3424 non-null   object
 57  Risk_Chronic_Malnutrition_Or_Malabsorption                        3424 non-null   object
 58  Risk_Chronic_Liver_Disease                                        3424 non-null   object
 59  Risk_Family_History_Of_Osteoporosis                               3424 non-null   object
 60  Risk_Low_Calcium_Intake                                           3424 non-null   object
 61  Risk_Vitamin_D_Insufficiency                                      3424 non-null   object
 62  Risk_Poor_Health_Frailty                                          3424 non-null   object
 63  Risk_Excessive_Thinness                                           3424 non-null   object
 64  Risk_Hysterectomy_Oophorectomy                                    3424 non-null   object
 65  Risk_Estrogen_Deficiency                                          3424 non-null   object
 66  Risk_Immobilization                                               3424 non-null   object
 67  Risk_Recurring_Falls                                              3424 non-null   object
 68  Count_Of_Risks                                                    3424 non-null   int64
dtypes: int64(2), object(67)
memory usage: 1.8+ MB
```

For the missing values, I will try to replace them with a meanable value. If it is not important, I will delete them.