

Data Science Internship

Data Intake Report

Name: Predicting the persistency of a drug (Healthcare)

Report date: 14th of April 2023

Internship Batch: LISUM19

Version: 1.0

Data intake by: Mario Rodrigues Peres

Data intake reviewer: Data Glacier

Data storage location: GitHub

Tabular data details:

Dataset - Healthcare

Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	xlsx
Size of the data	920 KB
Feature - data type (unique values)	<ol style="list-style-type: none">1. Ptid - object (3424)2. Persistency_Flag - object (2)3. Gender - object (2)4. Race - object (4)5. Ethnicity - object (3)6. Region - object (5)7. Age_Bucket - object (4)8. Ntm_Speciality - object (36)9. Ntm_Specialist_Flag - object (2)10. Ntm_Speciality_Bucket - object (3)11. Gluco_Record_Prior_Ntm - object (2)12. Gluco_Record_During_Rx - object (2)13. Dexa_Freq_During_Rx - integer (58)14. Dexa_During_Rx - object (2)15. Frag_Frac_Prior_Ntm - object (2)

Feature - data type (unique values)	<p> 16. Frag_Frac_During_Rx - object (2) 17. Risk_Segment_Prior_Ntm - object (2) 18. Tscore_Bucket_Prior_Ntm - object (2) 19. Risk_Segment_During_Rx - object (3) 20. Tscore_Bucket_During_Rx - object (3) 21. Change_T_Score - object (4) 22. Change_Risk_Segment - object (4) 23. Adherent_Flag - object (2) 24. Idn_Indicator - object (2) 25. Injectable_Experience_During_Rx - object (2) 26. Comorb_Encounter_For_Screening_For_Malignant Neoplasms - object (2) 27. Comorb_Encounter_For_Immunization - object (2) 28. Comorb_Encntr_For_General_Exam_W_O_Complaint, Susp_Or_Reprtd_Dx - object (2) 29. Comorb_Vitamin_D_Deficiency - object (2) 30. Comorb_Other_Joint_Disorder_Not_Elsewhere Classified - object (2) 31. Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint Suspected_Or_Reprtd_Dx - object (2) 32. Comorb_Long_Term_Current_Drug_Therapy - object (2) 33. Comorb_Dorsalgia - object (2) 34. Comorb_Personal_History_Of_Other_Diseases_And Conditions - object (2) 35. Comorb_Other_Disorders_Of_Bone_Density_And Structure - object (2) 36. Comorb_Disorders_of_lipoprotein_metabolism_and other_lipidemias - object (2) 37. Comorb_Osteoporosis_without_current_pathological fracture - object (2) 38. Comorb_Personal_history_of_malignant_neoplasm - object (2) 39. Comorb_Gastro_esophageal_reflux_disease - object (2) 40. Concom_Cholesterol_And_Triglyceride_Regulating Preparations - object (2) 41. Concom_Narcotics - object (2) 42. Concom_Systemic_Corticosteroids_Plain - object (2) 43. Concom_Anti_Depressants_And_Mood_Stabilisers - object (2) 44. Concom_Fluoroquinolones - object (2) 45. Concom Cephalosporins - object (2) </p>
--	--

Feature - data type (unique values)	<p>46. Concom_Macrolides_And_Similar_Types - object (2)</p> <p>47. Concom_Broad_Spectrum_Penicillins - object (2)</p> <p>48. Concom_Anaesthetics_General - object (2)</p> <p>49. Concom_Viral_Vaccines - object (2)</p> <p>50. Risk_Type_1_Insulin_Dependent_Diabetes - object (2)</p> <p>51. Risk_Osteogenesis_Imperfecta - object (2)</p> <p>52. Risk_Rheumatoid_Arthritis - object (2)</p> <p>53. Risk_Untreated_Chronic_Hyperthyroidism - object (2)</p> <p>54. Risk_Untreated_Chronic_Hypogonadism - object (2)</p> <p>55. Risk_Untreated_Early_Menopause - object (2)</p> <p>56. Risk_Patient_Parent_Fractured_Their_Hip - object (2)</p> <p>57. Risk_Smoking_Tobacco - object (2)</p> <p>58. Risk_Chronic_Malnutrition_Or_Malabsorption - object (2)</p> <p>59. Risk_Chronic_Liver_Disease - object (2)</p> <p>60. Risk_Family_History_Of_Osteoporosis - object (2)</p> <p>61. Risk_Low_Calcium_Intake - object (2)</p> <p>62. Risk_Vitamin_D_Insufficiency - object (2)</p> <p>63. Risk_Poor_Health_Frailty - object (2)</p> <p>64. Risk_Excessive_Thinness - object (2)</p> <p>65. Risk_Hysterectomy_Oophorectomy - object (2)</p> <p>66. Risk_Estrogen_Deficiency - object (2)</p> <p>67. Risk_Immobilization - object (2)</p> <p>68. Risk_Recurring_Falls - object (2)</p> <p>69. Count_Of_Risks - integer (8)</p>
--	---

Proposed Approach:

- There is no duplicated registers, and therefore, all registers are unique.
- There is no missing values in the dataset.
- There are only two integer variables (Dexa_Freq_During_Rx and Count_Of_Risks), and most of the variables present only two categories (Y or N).