

Data Science Project

Healthcare – Persistency of a Drug

Week 10 works

Team member's details:

Group Name:	DS_SS
Name:	SONIYA SUNNY
Email:	soniyasunny1210@gmail.com
Country:	Canada
College/Company:	Data Glacier
Specialization:	Data Science

Problem Description

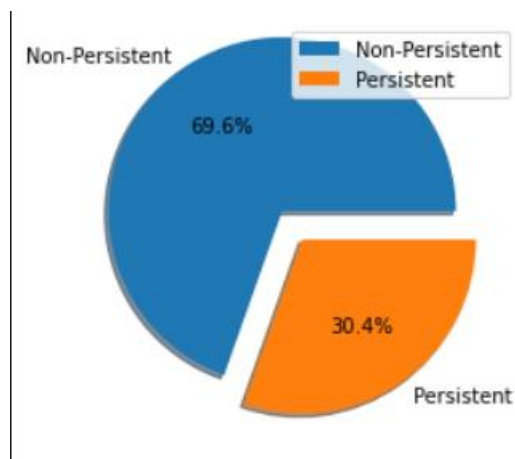
To identify the persistency of a drug, a pharmaceutical company approached to develop a model based on data analysis. Factors that affect the persistence of drugs should be identified, along with data insights with predictive analytics, to help the company for their smooth and efficient functioning, with the help of dataset provided by the company.

Github Repo Link

[Final Project DS SS/week 10 at master · Soniyasunny1/Final Project DS SS \(github.com\)](#)

EDA performed on the data

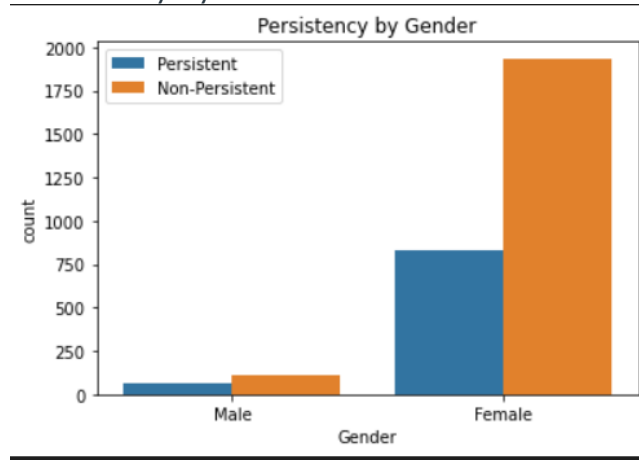
- Basic information
 - 2942 rows and 65 columns for the cleaned dataframe.
 - Ntm_Speciality variable has the maximum number of unique variables
 - No null values
- Target variable count



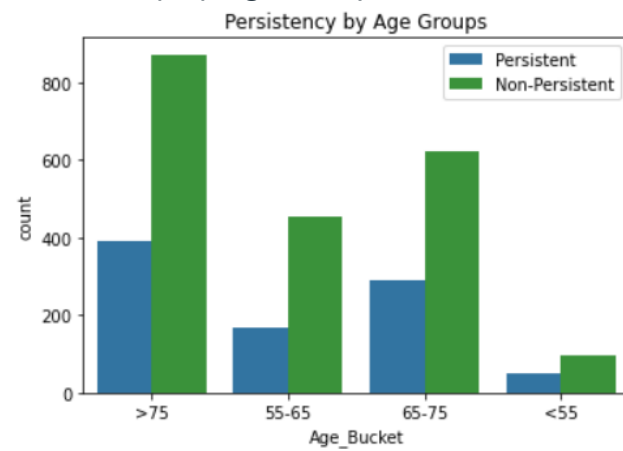
69.6% are non-persistent and 30.4% are persistent.

Demographics Analysis

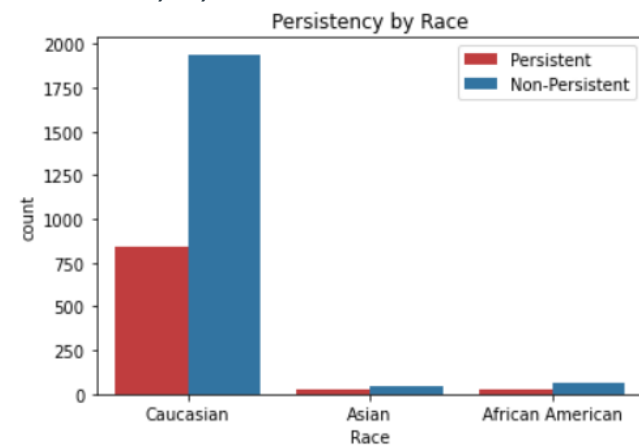
- Persistency by Gender



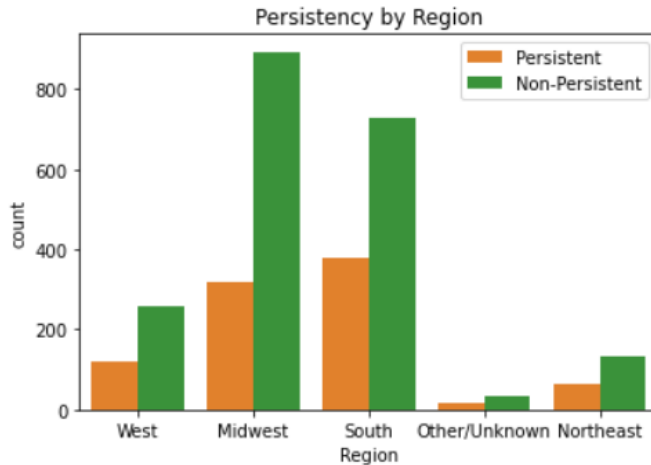
- Persistency by Age Groups



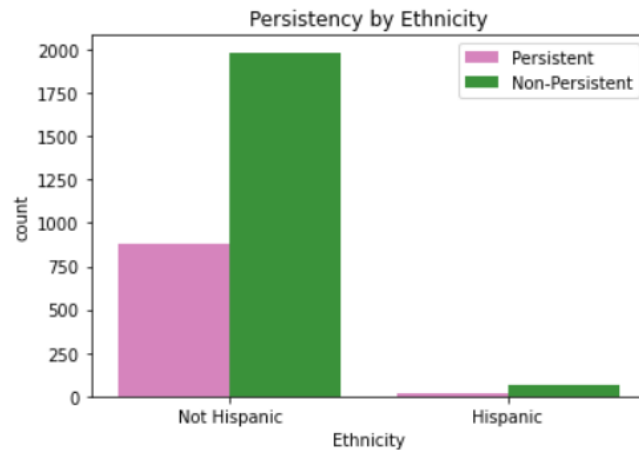
- Persistency by Race



- Persistency by Region

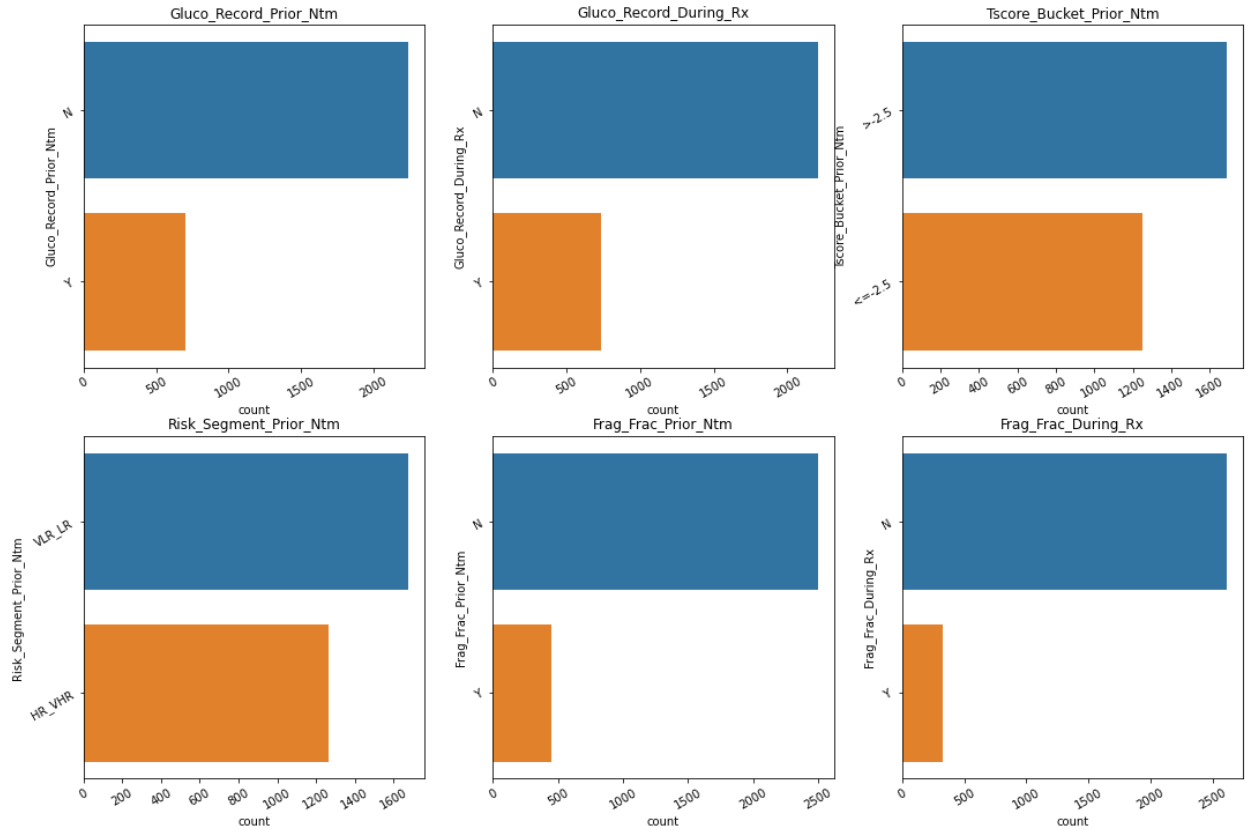


- Persistency by Ethnicity

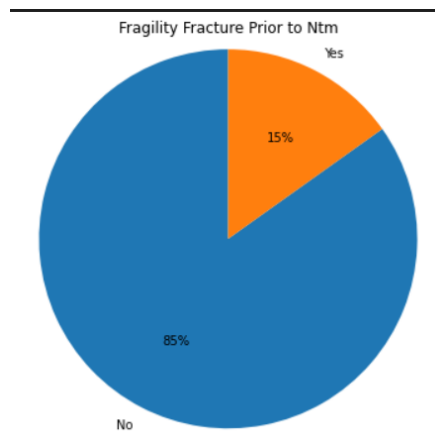


In the demographic analysis, we get an overview of which groups have more counts based on various variables like age, gender, race etc. In gender analysis, we can see that, females are the majority, with more persistency. Old-aged people are more with more persistency as compared to other age buckets. In the case of race, Caucasian is popular and for ethnicity, non-Hispanic are the majority groups.

Clinical Factor Analysis



- Fragility Fracture prior to NTM



Disease/Treatment Factor Analysis

- Influence of different factors
 - NTM-Comorbidity: Disorders of lipoprotein metabolism and other lipidemias has highest influence.
 - NTM-Risk Factors: Vitamin D Insufficiency has highest influence.
 - NTM-Concomitancy: Narcotics has highest influence.

Final Recommendation

Many factors have found to be influential for our modelling. Persistency can be influenced by some demographic features, clinical attributes, and Disease/Treatment factors as well. Better models can be created for finding out the drug persistency.

Data Intake Report

Name: Data Science Final Project – ‘Healthcare – Persistency of a Drug’

Report date: August 06, 2022

Internship Batch: LISUM10: 30

Version:<1.0>

Data intake by: Soniya Sunny

Data intake reviewer:<intern who reviewed the report>

Data storage location: [Healthcare dataset.xlsx - Google Drive](#)

Tabular data details:

Total number of observations	3425
Total number of files	1
Total number of features	69
Base format of the file	.xlsx
Size of the data	899 KB