



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

HEALTHCARE - PERSISTENCY OF A DRUG

Group Name: DS_SS

Name: SONIYA SUNNY

Email: soniyasunny1210@gmail.com

Country: Canada

Company: Data Glacier

Specialization: Data Science

Agenda

Executive Summary

Problem Statement

Approach

Data Cleaning

EDA

EDA Summary

Recommendations

Recommended Models

Problem Statement

- 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.

Approach

- Collect data
- Analyze the data
- Detect outliers
- Combine datasets
- Curate the data
- Feature engineering
- Detect correlations
- Analyze patterns
- Provide insights
- Tools used – Microsoft Excel, Python, Microsoft Power BI

Data Cleaning

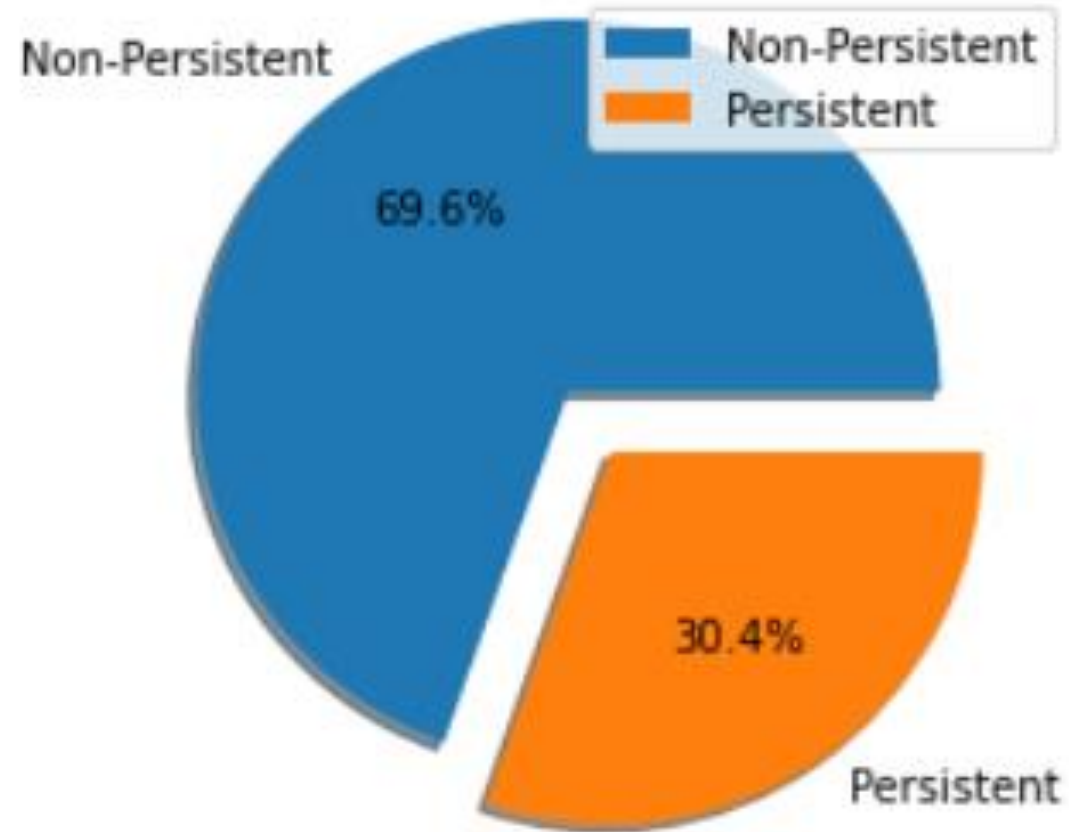
- Dataset details:
 - File name: Healthcare_dataset.xlsx
 - No. of rows = 3424
 - No. of columns = 69
 - Target variable = **'Persistency_Flag'**
- There were no null values in the dataset.
- Some variables have large number of value counts, which is reduced by grouping same names, treating "Unknown" values, etc.
- A new dataset was created after data curation steps, with 2942 rows & 66 columns.



Exploratory Data Analysis

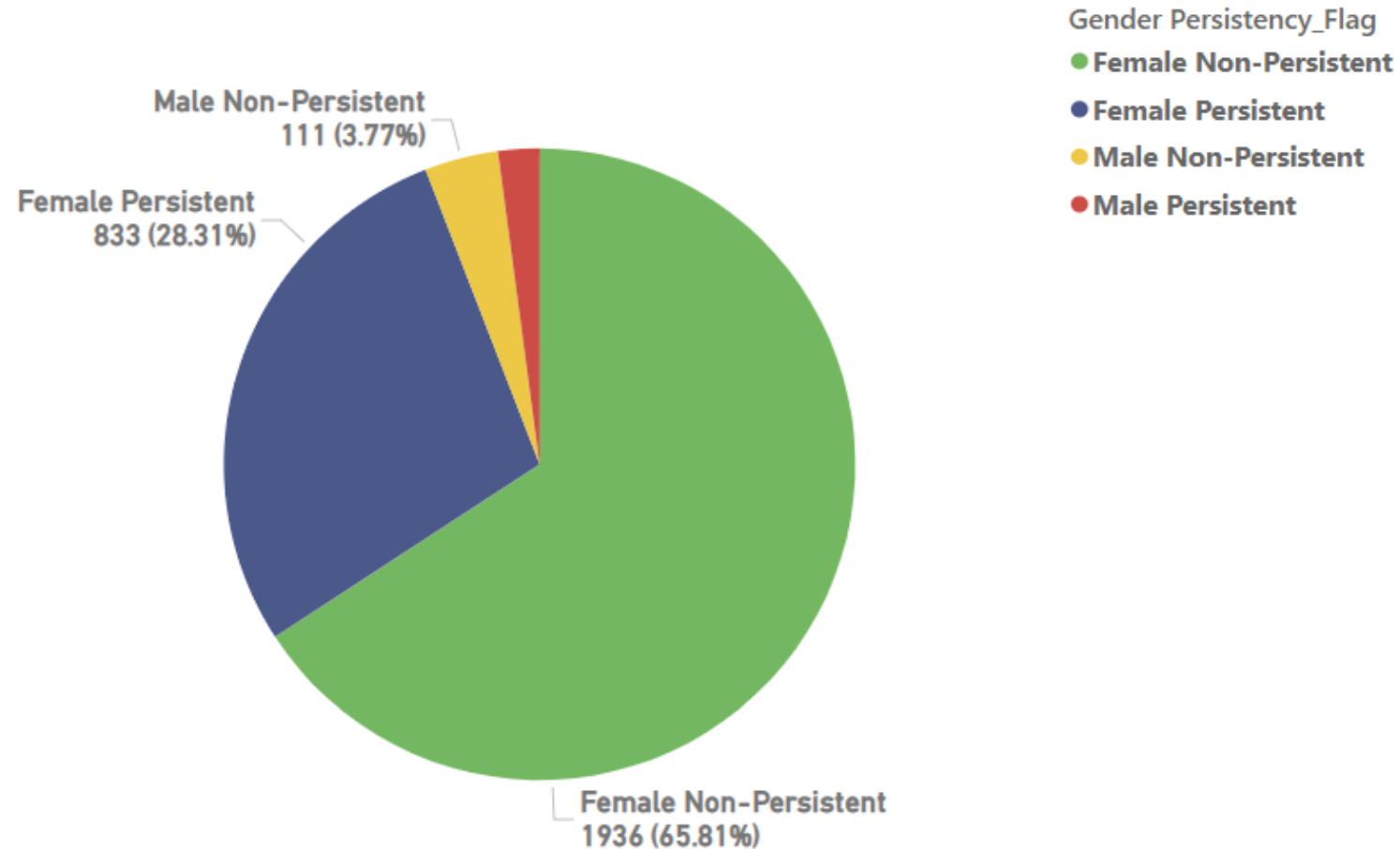
1. Basic Data Exploration
2. Demographic Analysis
 - Gender
 - Age
 - Race
 - Region
 - Ethnicity
3. Clinical Factors' Analysis
 - T score
 - Risk Segment
 - Glucocorticoid Recency
 - Fragility Fracture Recency
4. Disease/Treatment Factors' Analysis
 - Comorbidity
 - Risk factors
 - Concomitancy

Target Variable Analysis



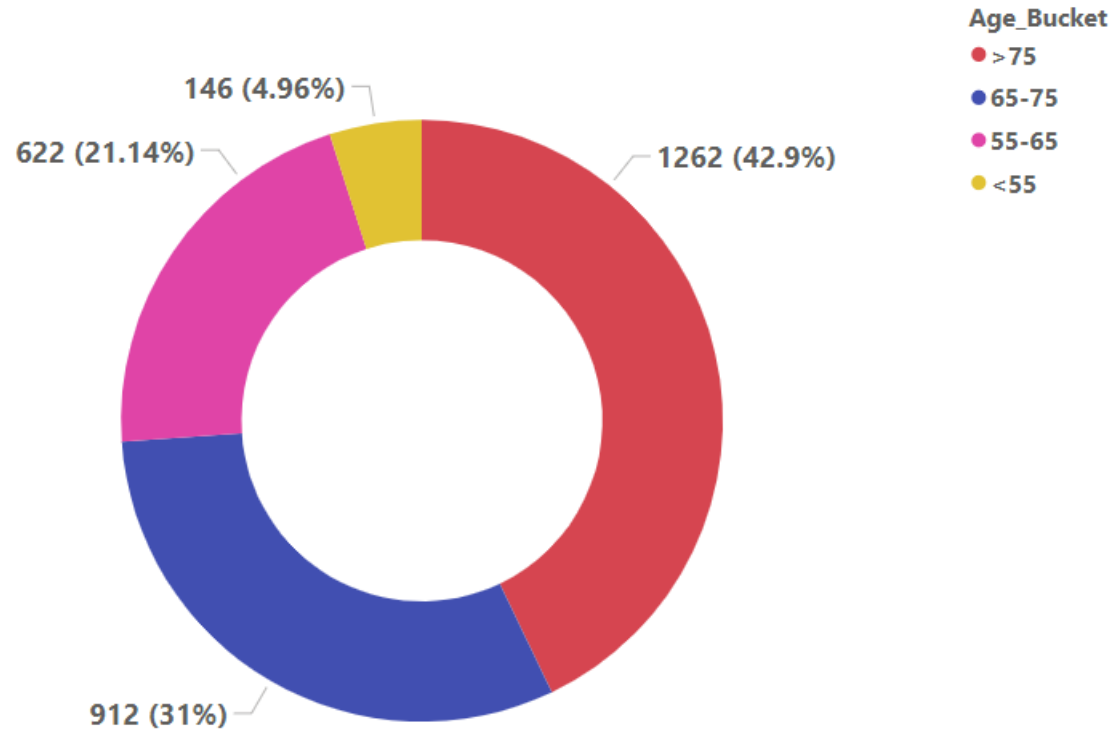
Gender Analysis

Patients' Count by Gender and Persistency_Flag

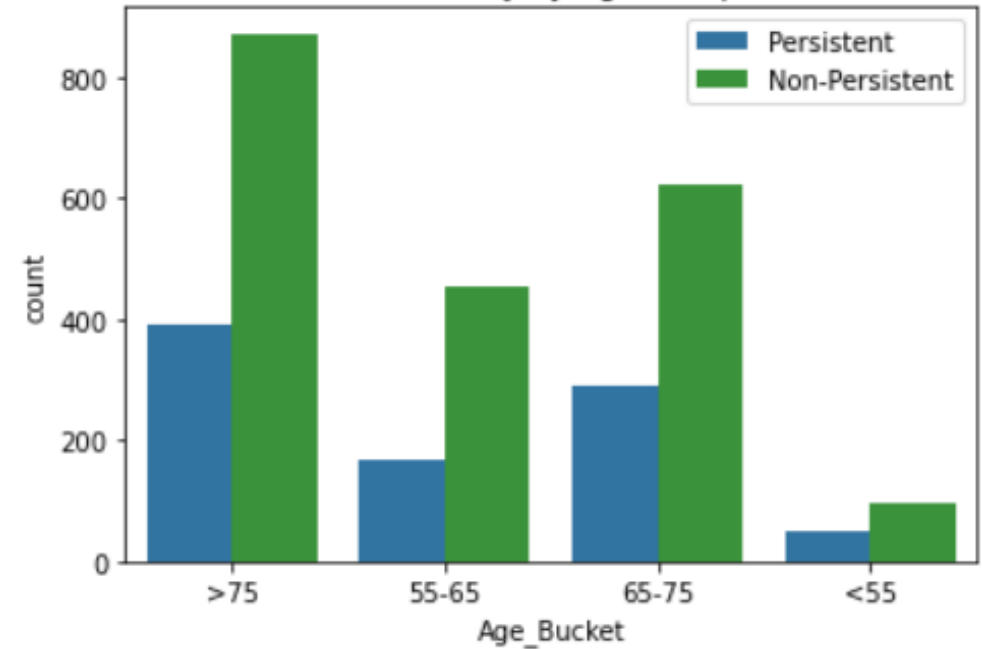


Age Group Analysis

Age Group Count

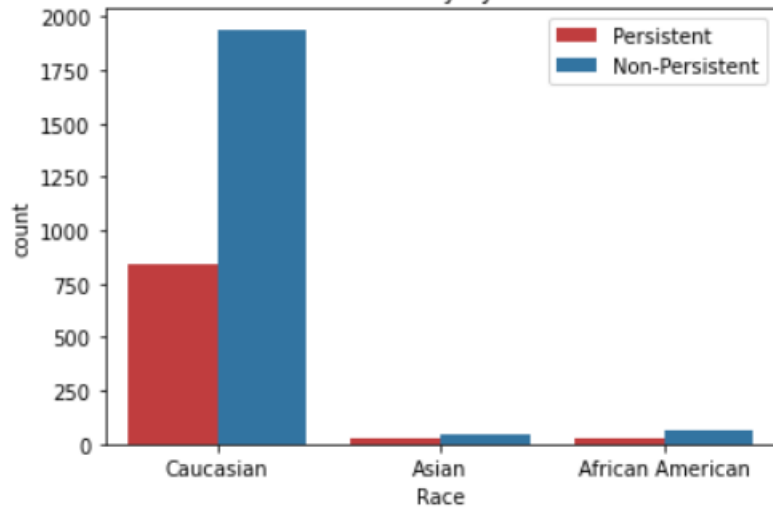


Persistency by Age Groups

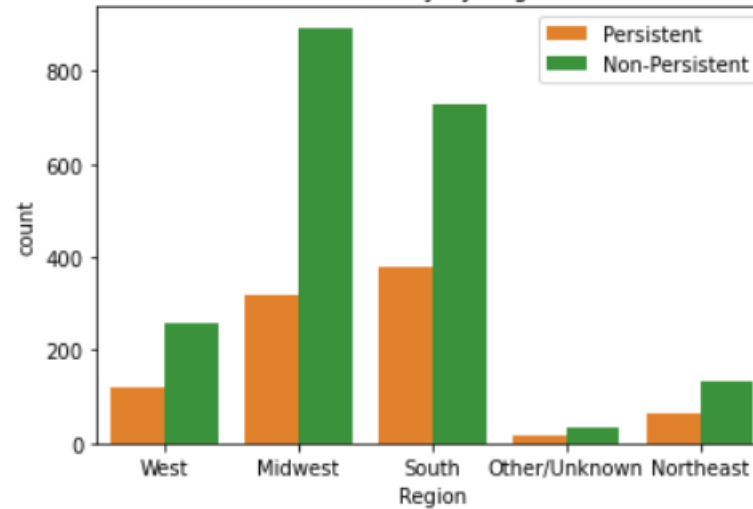


Race, Region & Ethnicity

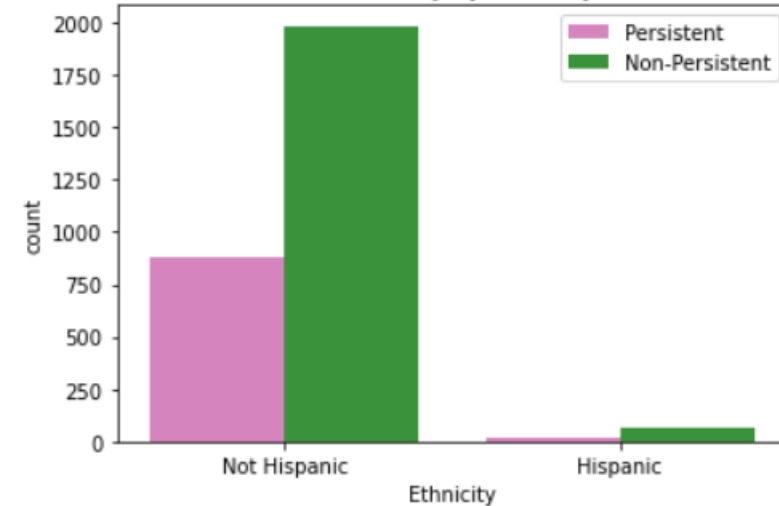
Persistence by Race



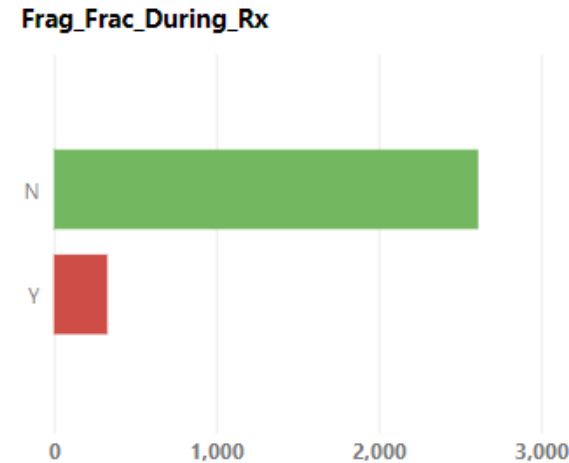
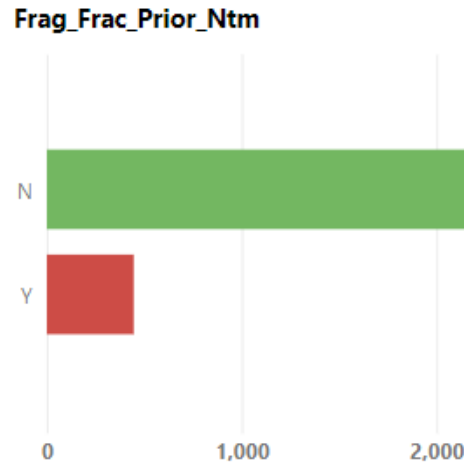
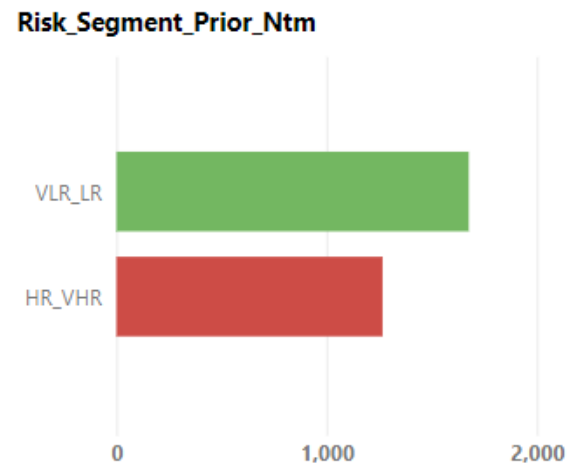
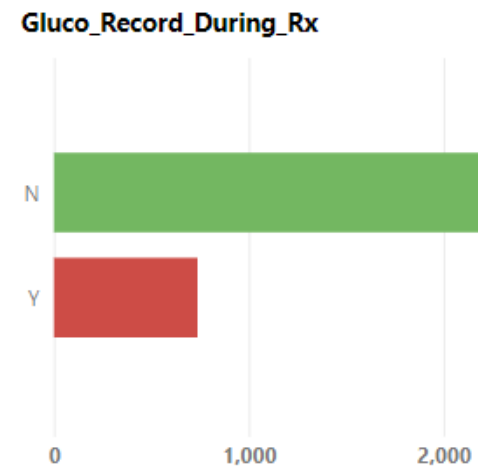
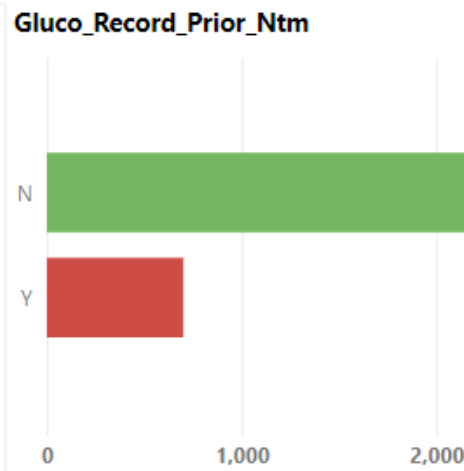
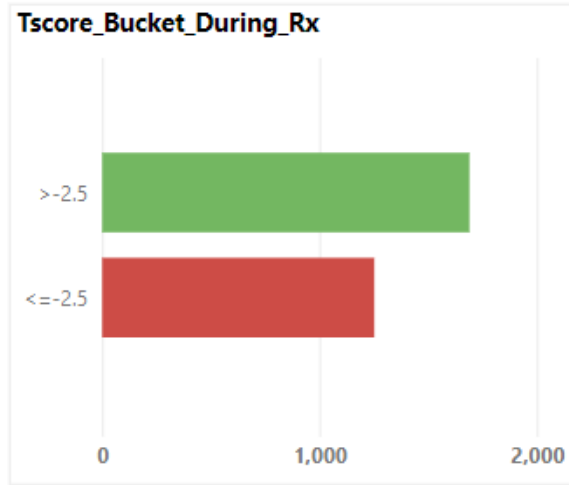
Persistence by Region



Persistence by Ethnicity



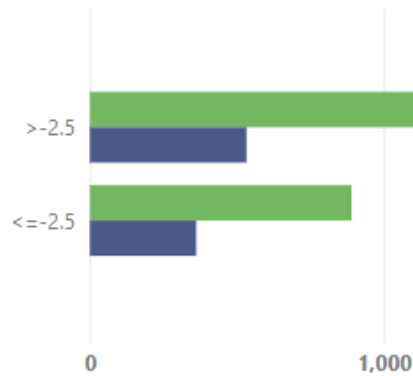
Clinical Factors



Clinical Factors - Persistency

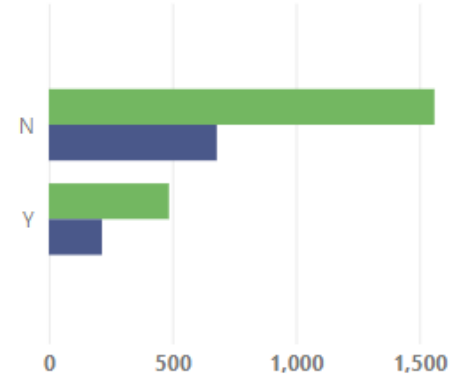
Tscore_Bucket_During_Rx

Persistenc... ● Non-Persistent ● Persistent



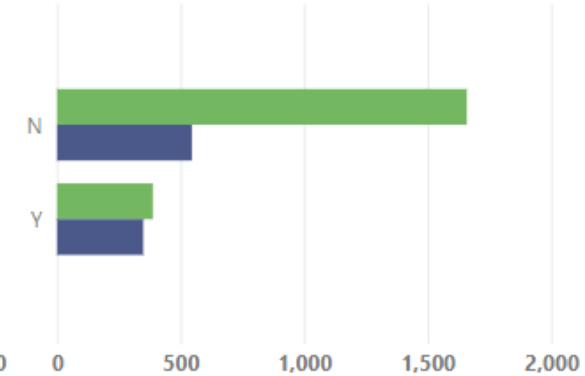
Gluco_Record_Prior_Ntm

Persistency_Flag ● Non-Persistent ● Persistent



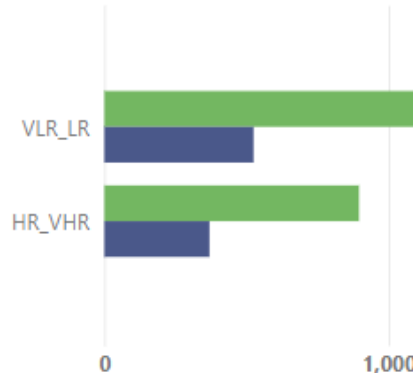
Gluco_Record_During_Rx

Persistency_Flag ● Non-Persistent ● Persistent



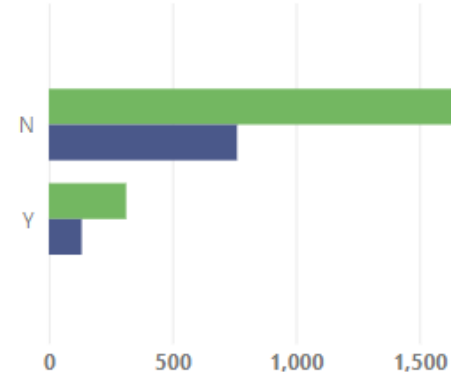
Risk_Segment_Prior_Ntm

Persistency_Flag ● Non-Persistent ● Persistent



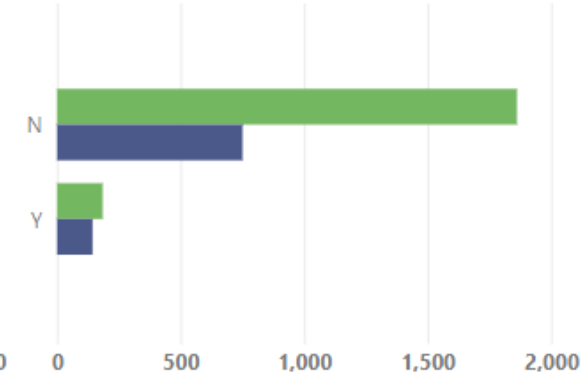
Frag_Frac_Prior_Ntm

Persistency_Flag ● Non-Persistent ● Persistent



Frag_Frac_During_Rx

Persistency_Flag ● Non-Persistent ● Persistent



Disease/Treatment Factors

- NTM-Comorbidity: Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias has highest influence.
- NTM-Risk Factors: Risk_Vitamin_D_Insufficiency has highest influence.
- NTM-Concomitancy: Concom_Narcotics has highest influence.

EDA Summary

- Drug Persistency is affected by many factors
- Patients older than 65 years show more persistency
- T-square also has influence in persistency
- Disease/Treatment factors:
 - Comorbidity
 - Risk factors
 - Concomitancy

Recommendations

- Proposed Models:
 - Random Forest Classifier
 - Decision Tree Classifier
 - Logistic Regression
- Classifier algorithms tend to produce more effective results
- These models help to categorize patients based on certain factors

Thank You