Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis and Proposed Modeling Technique

Healthcare – Persistency of a drug

12/11/2022

Data Glacier

Your Deep Learning Partner

Group FTR

Name	Email	Country	College/Company	Specialization
Fabian Umeh	Fabianumeh335@gmail.com	UK	Teesside University	Data Science
Rukevwe Ovuowo	rukevwe10@gmail.com	Nigeria	GBG Data science Academy	Data Science
Olutayo Oladeinbo	oladeinboolutayo@yahoo.com	UK	Teesside University	Data Science

Agenda

- Executive Summary
  - Problem Statement
  - Objectives
  - Approach
- EDA
- EDA Summary
- Proposed modeling technique

Executive Summary

Problem statement

- One of the challenge for all Pharmaceutical companies is to understand the persistency of drugs as per the physician prescription. To solve this problem ABC pharma company would like the process Automated.

Objectives

- The overall aim of the analysis part of the project is to provide insights into factors that impact the persistency of drugs, which afterwards will lay the foundation on building a suitable classification model and also propose some modelling technique to be used.

Approach

- Understanding the dataset
- Identifying the most impactful factors
- Making recommendations.
- Proposed modeling technique

Data Glacier
Your Deep Learning Partner

Data Understanding

```
data.head()
```

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region	Age_Bucket	Ntm_Speciality	Ntm_Specialist_Flag	Ntm_Speciality_Bucket	...	Risk_Family_History_Of
0	P1	Persistent	Male	Caucasian	Not Hispanic	West	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
1	P2	Non-Persistent	Male	Asian	Not Hispanic	West	55-65	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
2	P3	Non-Persistent	Female	Other/Unknown	Hispanic	Midwest	65-75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
3	P4	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
4	P5	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	

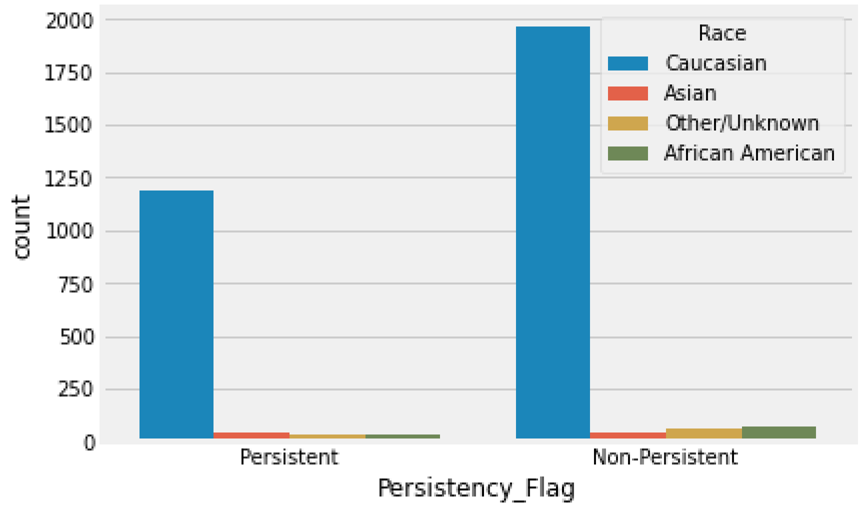
5 rows × 69 columns

- The dataset contains 3424 rows and 69 columns.

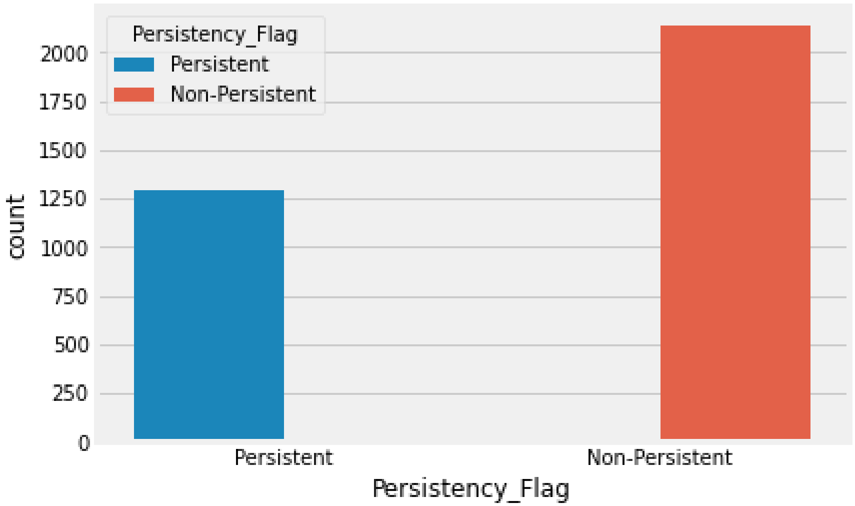
Bucket	Variable	Variable Description
Unique Row Id	Patient ID	Unique ID of each patient
Target Variable	Persistency_Flag	Flag indicating if a patient was persistent or not
Demographics	Age	Age of the patient during their therapy
	Race	Race of the patient from the patient table
	Region	Region of the patient 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
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

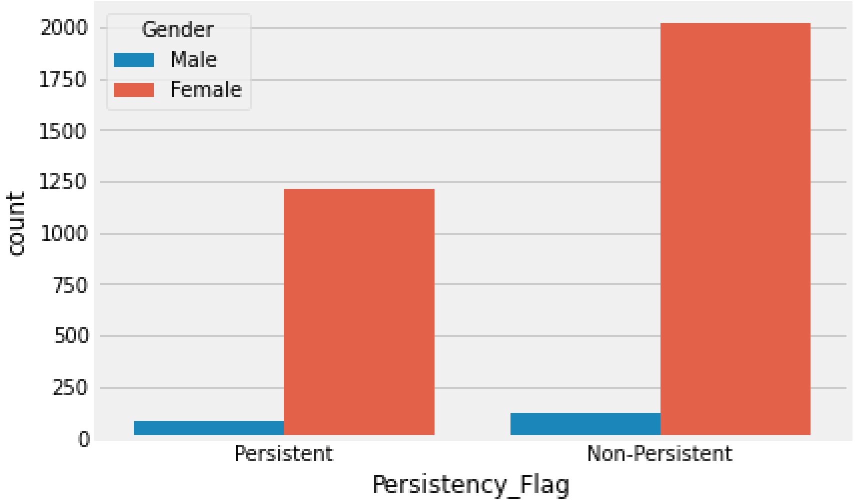
Data Glacier

Exploratory data analysis

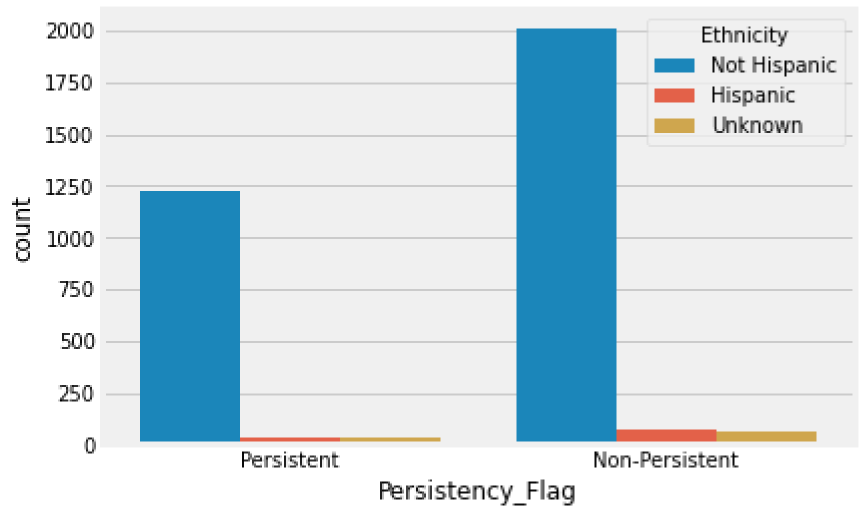


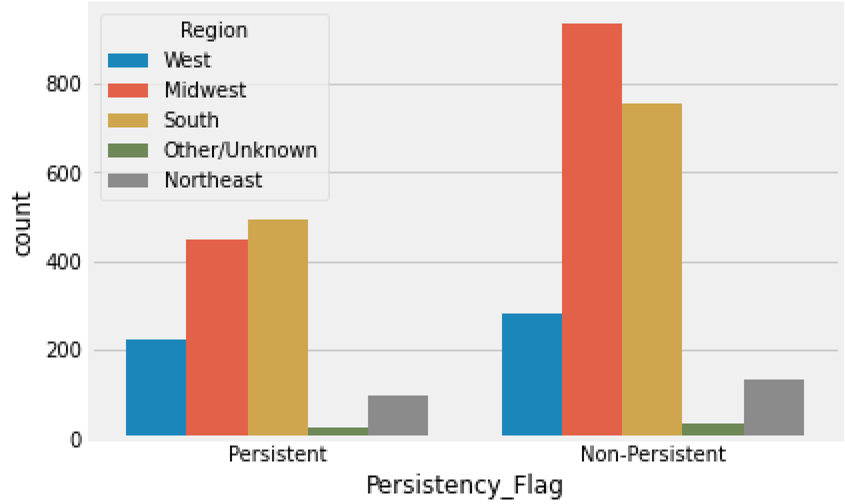
- The number of cases where the drugs proved to be non-persistent were higher compared to number od persistency cases.
- The dataset reveal that more females partook in this analysis than male.
- People of Caucasian race when compered to other races were the most common in the study.





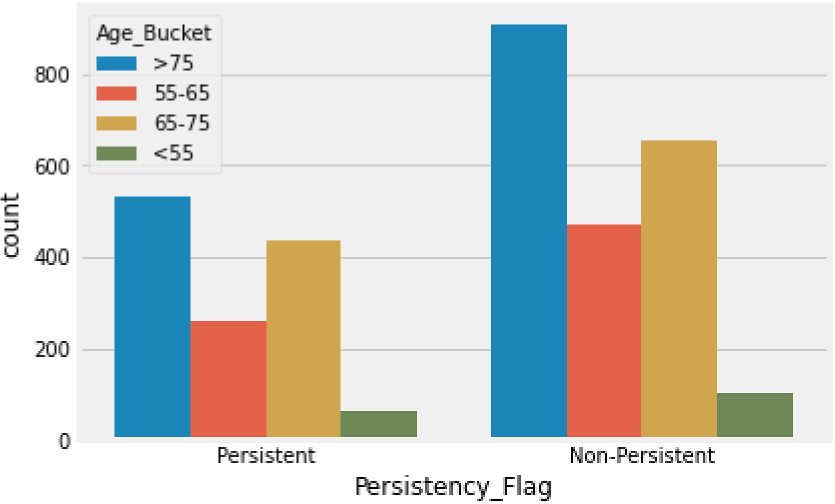
Exploratory data analysis

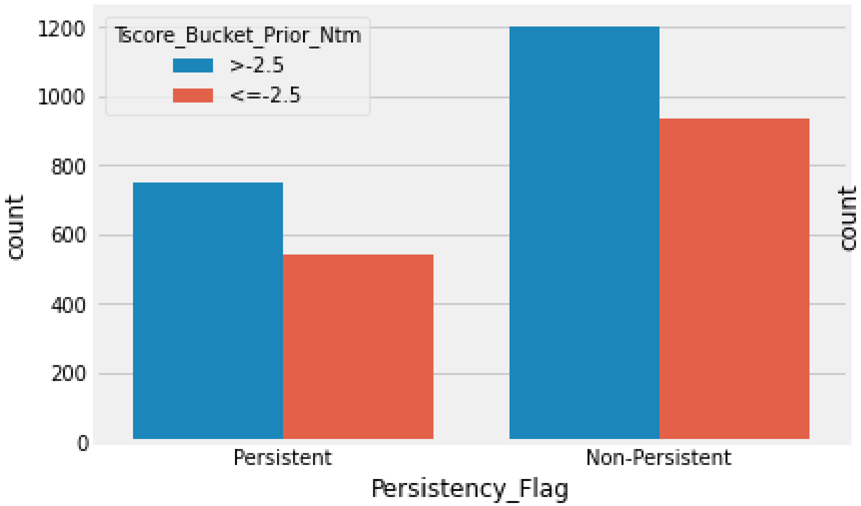


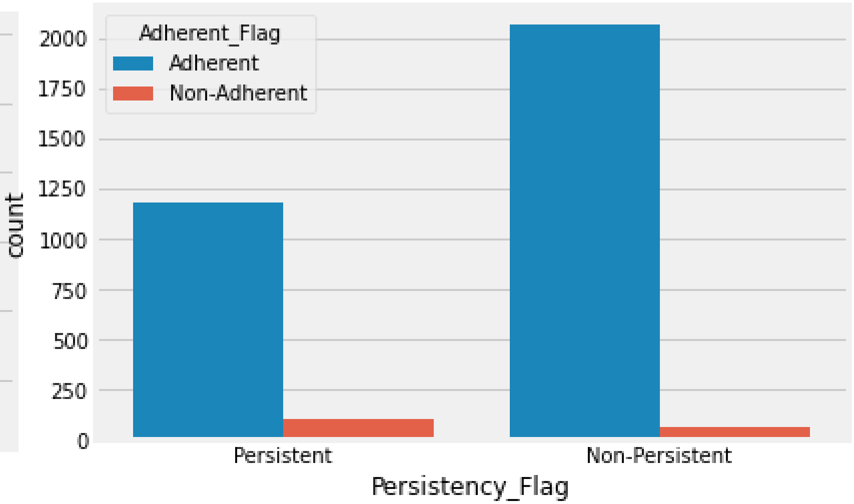


- The non-Hispanic ethnic group were the most common in the study.
- There were more people from the Midwest and South region compared to other regions.

Exploratory data analysis



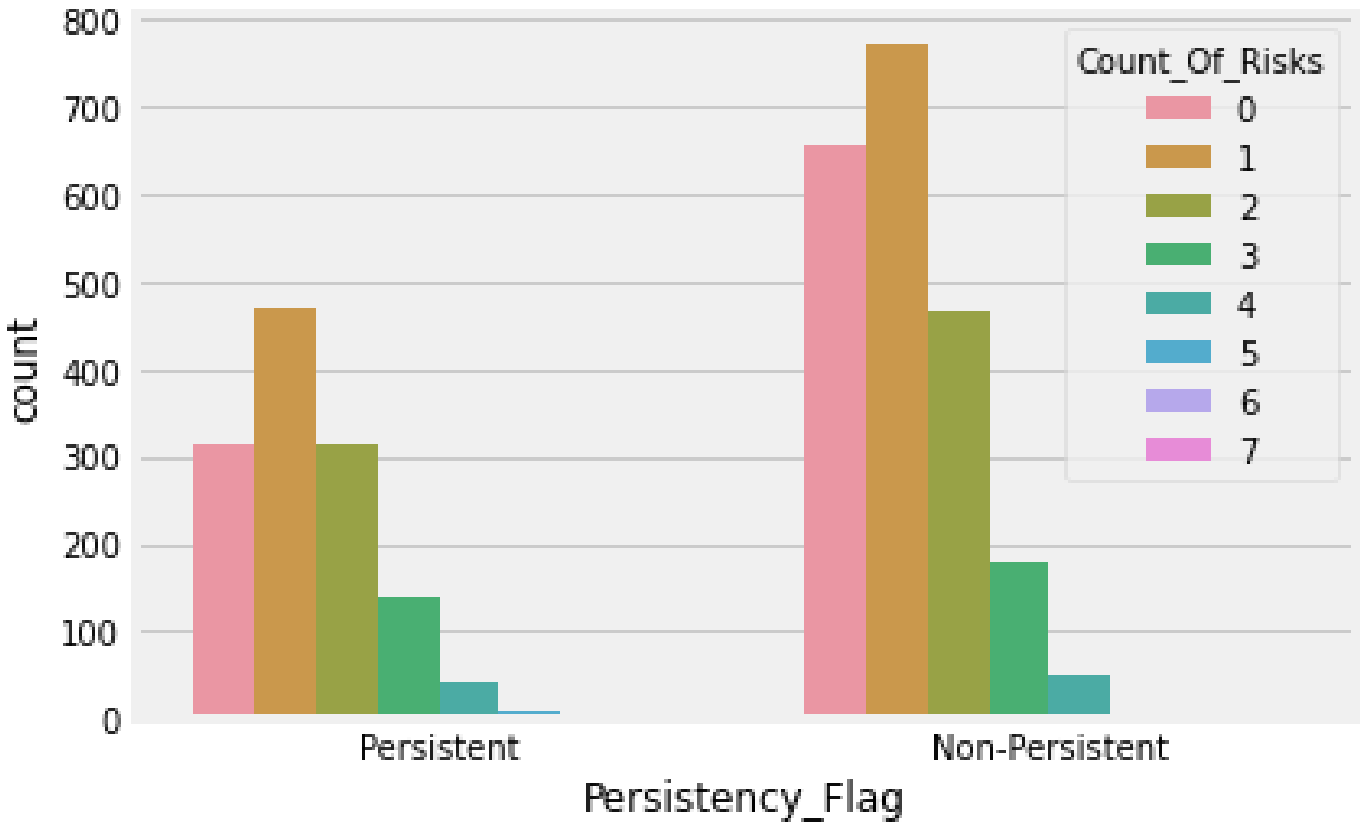




- **For this study, the majority of people selected are greater than 75 years of age.**
- **People with a Tscore of >-2.5 have a higher chance of drug being non-persistent.**

Data Glacier
Your Deep Learning Partner

Exploratory data analysis



- The chart reveals people with a lower count of risk have a higher chance of drug being non-persistent.

Data Glacier

Summary and recommendation

EDA SUMMARY

- The dataset contains 3424 rows and 69 columns.
- The number of cases where the drugs proved to be non-persistent were higher compared to number od persistency cases.
- The dataset reveal that more females partook in this analysis than male.
- People of Caucasian race when compered to other races were the most common in the study.
- The non-Hispanic ethnic group were the most common in the study.
- There were more people from the Midwest and South region compared to other regions.
- For this study, most people selected are greater than 75 years of age.
- People with a Tscore of >-2.5 have a higher chance of drug being non-persistent.

Proposed modeling technique

The project is aimed at using certain factors relative to a patient in classifying successfully whether a drug is persistent or not. From the machine learning aspect of things, the task is a classification task and a binary classification task to be specific. For this project, we will focus on state-of-the-art machine learning classification models to build our drug persistency classifier. They include:

1. Logistic regression model
2. Support vector machines (SVM)
3. K-nearest neighbours (KNN)
4. Gradient Boost model

Repository details

- **Github Repo link**: [Ovuowo-Rukevwe/-Healthcare-Persistency-of-a-Drug (github.com)]

Data Glacier

Your Deep Learning Partner

Thank You

Data Glacier