Exploratory Data Analysis and Modelling Proposal

Persistency of a Drug

Team: Quantic

Members: Ansel Vallejo

Date: November 16th , 2023

Data Glacier

Agenda

Business Problem

Approach

EDA

EDA Recap

Recommendations

Data Glacier

Your Deep Learning Partner

Business Problem

ABC Pharma Company, in its relentless pursuit of pharmaceutical excellence, grappled with a persistent challenge: understanding the duration of drug persistence aligned with physician prescriptions. To surmount this obstacle, ABC Pharma astutely partnered with Quantic, a trailblazer in streamlining and automating processes. Through the strategic application of data analytics, the pharmaceutical titan aimed to unravel intricate drug persistency patterns, empowering informed decision-making and ushering in a new era of superior patient care. This synergistic collaboration between ABC Pharma and Quantic underscores a steadfast commitment to harnessing the power of data-driven solutions, effectively tackling pivotal industry challenges head-on. The ultimate objective remains crystal clear: to accurately predict patient adherence to physician-prescribed treatments, ensuring uninterrupted medication adherence throughout the course of therapy.

Approach

The dataset presented contains **3,424** datapoints and **69** variables.

For better understanding, the data is grouped in buckets:
- Demographics
- Provider Attributes
- Clinical Factors
- Disease and Treatment Factors

- The exploratory data analysis (**EDA**) was conducted using a similar framework, encompassing the entirety of the dataset, thereby extracting potential insights from the comprehensive analysis.

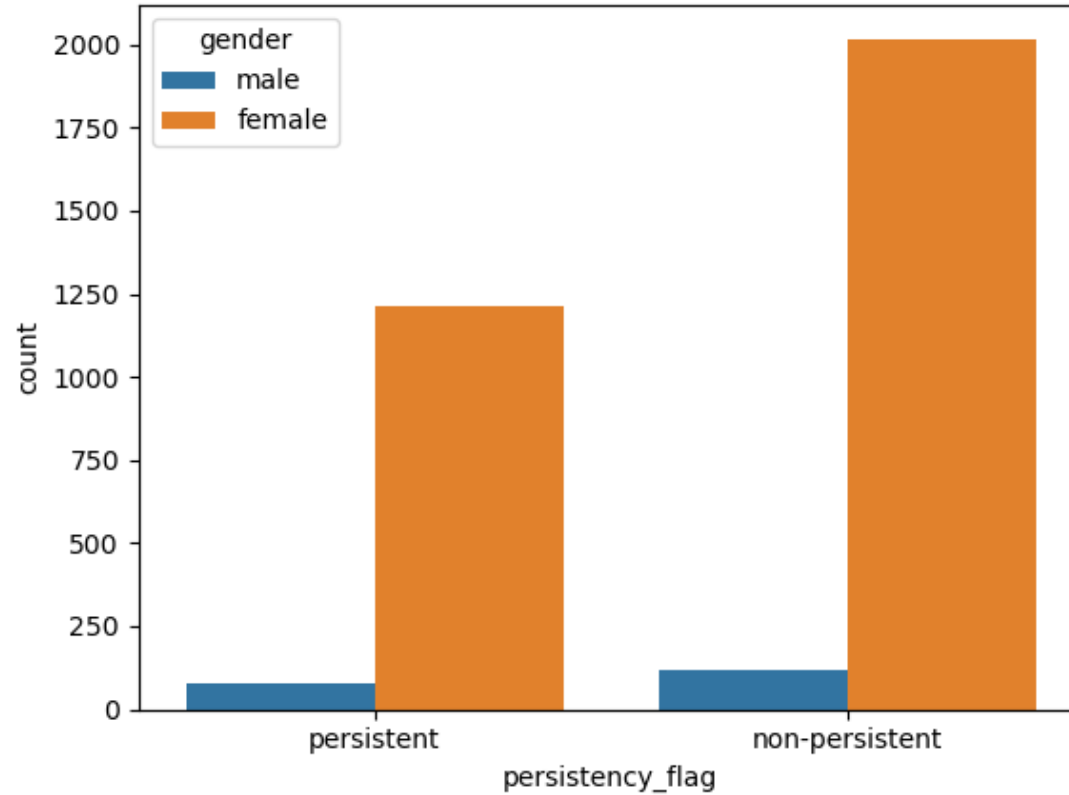- Two **model** proposals have been formulated.

Data Glacier
Your Deep Learning Partner
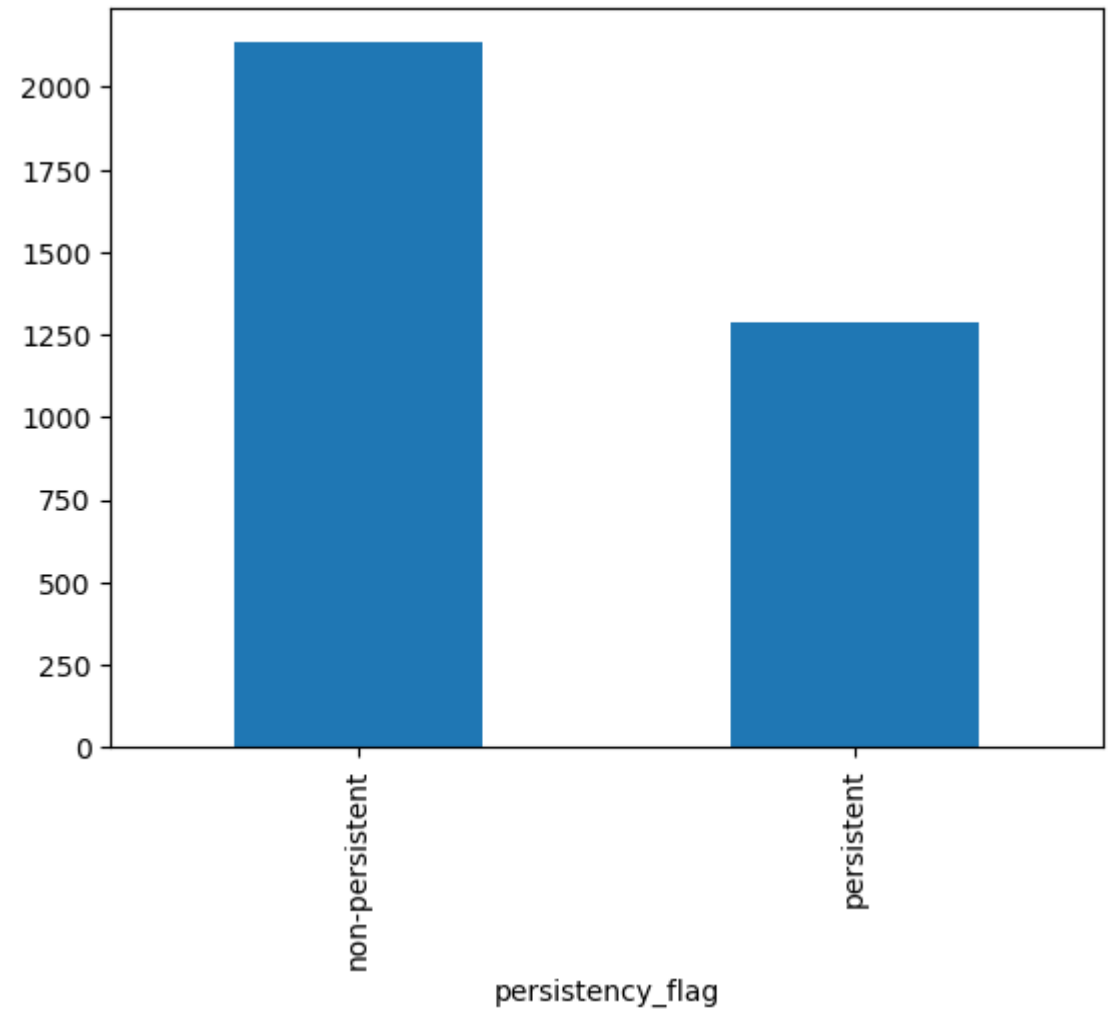
Gender

**What gender is most persistent of taking the drug?**

Males have the lowest sample size of persistent and non-persistent of a drug totaling 194, whereas Females have the highest sample size totaling 3230.

Males tend to have higher persistency than Females by 2.17%, whereas Females have higher non-persistency than Males by 2.17%.

Age

## Is age a big contribution to drug persistency?
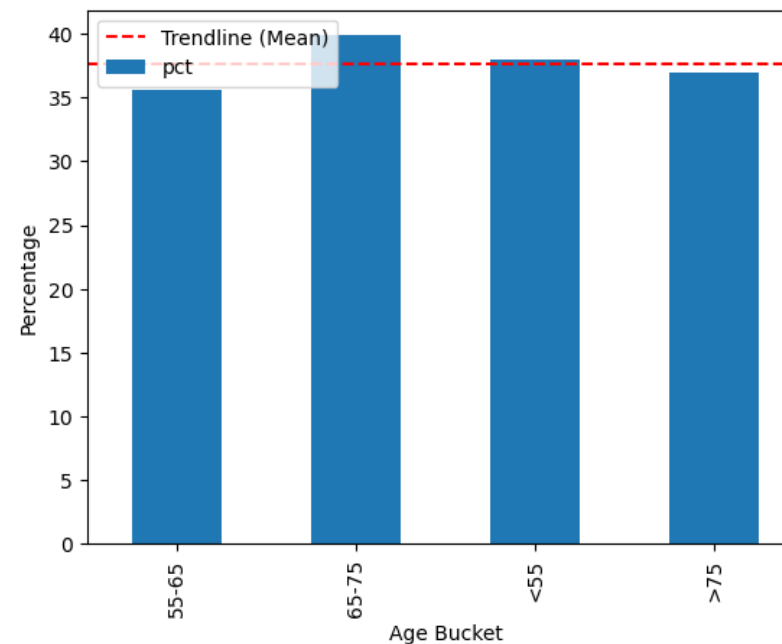
persistency_flag	non-persistent	persistent	pct_of_persistency	pct_of_non_persistency	total
<55	103	63	37.95	62.05	166
55-65	472	261	35.61	64.39	733
65-75	653	433	39.87	60.13	1086
>75	907	532	36.97	63.03	1439

Age group of 65-75 has the highest percentage of drug persistency at 39.87% out of the others, exceeding the mean persistency by 2.27% from all age groups.

Data Glacier
Your Deep Learning Partner

Race

What is the percentage of persistency based on race?
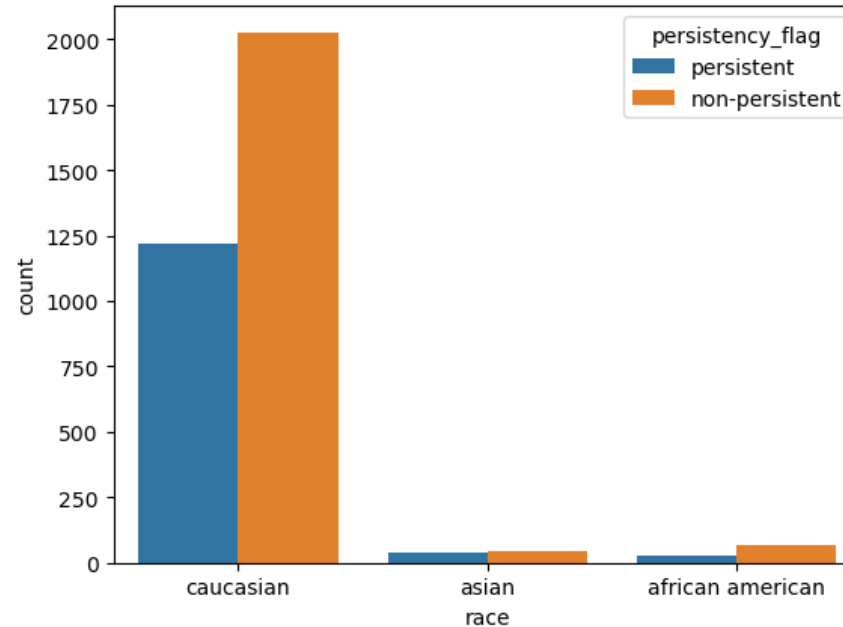
persistency_flag	race	non-persistent	persistent	persistent_pct	non_persistent_pct	total
0	african american	65	30	31.58	68.42	95
1	asian	43	41	48.81	51.19	84
2	caucasian	2027	1218	37.53	62.47	3245

Therefore, based on the calculated differences, African Americans exhibit the most substantial contrast in drug persistency percentages between persistent and non-persistent cases among the provided racial groups.

Data Glacier
Your Deep Learning Partner

Risk Factor

Risk of Vitamin D Insufficiency
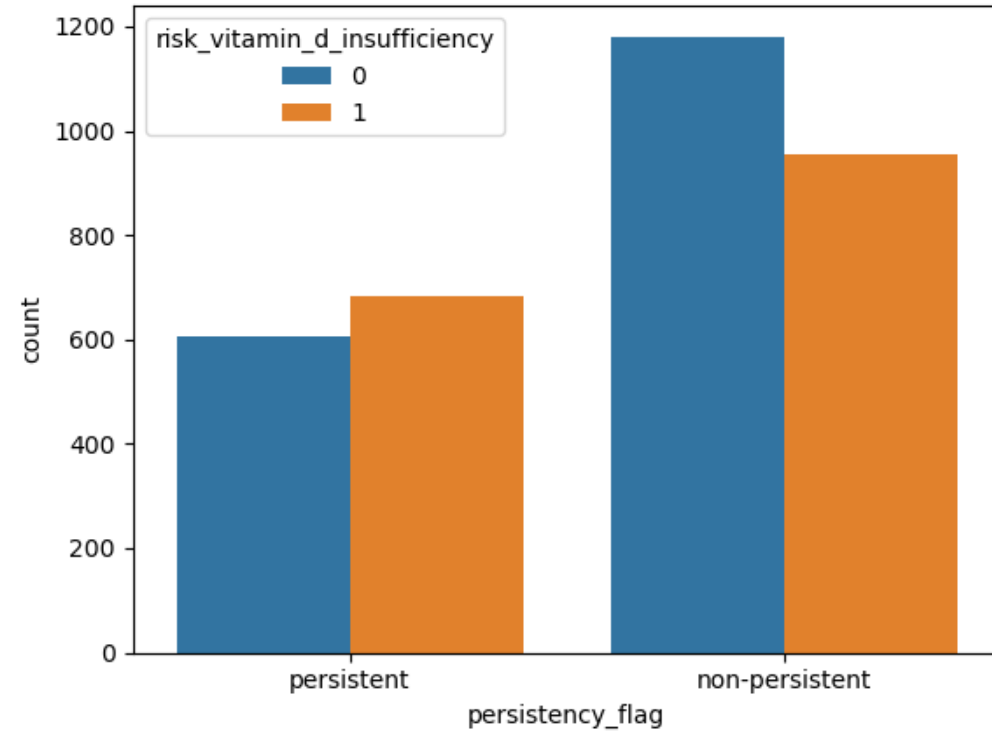
There is an interesting relationship between persistency and the risk of Vitamin D insufficiency.

If the patient is persistent, the chances of Vitamin D insufficiency is greater than non-persistent patients.

EDA Recap

The dataset comprises details of 3,424 patients, encompassing demographic data, clinical records, information on other diseases as risk factors, and details about their physicians' specialties.

Age and Gender play a key role to help further explain the disparities between the categories in terms of persistency.

Distinct differences emerge between individuals of Asian descent and other racial groups within the dataset, suggesting noteworthy disparities in various aspects.

Notable variations are evident across patients concerning factors like vitamin D deficiencies.

Data Glacier
Your Deep Learning Partner

Model proposals (technical review)

Logistic Regression Model:

Advantages:

- Straightforward binary classification.
- Provides probability estimates.

Suitability:

Ideal for handling binary outcomes and offering insights into class probabilities.

Decision Tree Model:

Advantages:

- Offers interpretability with visual decision pathways.
- Accommodates diverse data types effectively.

Suitability:

Useful for transparent decision-making processes and handling both categorical and numerical data.

Thanks! ☺