DATA SCIENCE HEALTHCARE PROJECT

INDIVIDUAL PROJECT NAME: HEALTH and CARE

Member Name: ABINA AZEES

Submision Date: 16/7/2025

Drug Persistency case study

Project Objective:

- To explore, analyze, and identify the factors affecting drug persistency.

- To build a classification model that can predict whether a patient is likely to be persistent with a prescribed drug.

- **Approach:**

The analysis is divided into the following three key stages:

Data Understanding:
- Explore the dataset to understand its structure, variables, and data quality.
- Identify missing values, data types, and potential anomalies.

Data Insights and Visualization:
- Perform exploratory data analysis (EDA) to uncover patterns, correlations, and trends.
- Visualize relationships between key variables and persistency behavior.

Modeling and Recommendations:
- Build and evaluate classification models to predict drug persistency.
- Derive business recommendations based on model insights and data patterns.

Data Exploration

The objective of this phase is to gain an initial understanding of the dataset's structure, quality, and key characteristics.
**Dataset Overview:**

- The dataset consists of **68 features**, categorized into the following groups:

- **General Features**
  Includes demographic information (e.g., age, gender, region) and provider-related attributes (e.g., specialty, experience).

- **Disease/Drug Factors**
  Captures the type and nature of diseases diagnosed and drugs prescribed, including treatment history and drug categories.

- **Clinical Factors**
  Involves lab results, diagnostic codes, comorbidities, and other clinical measurements relevant to treatment decisions.

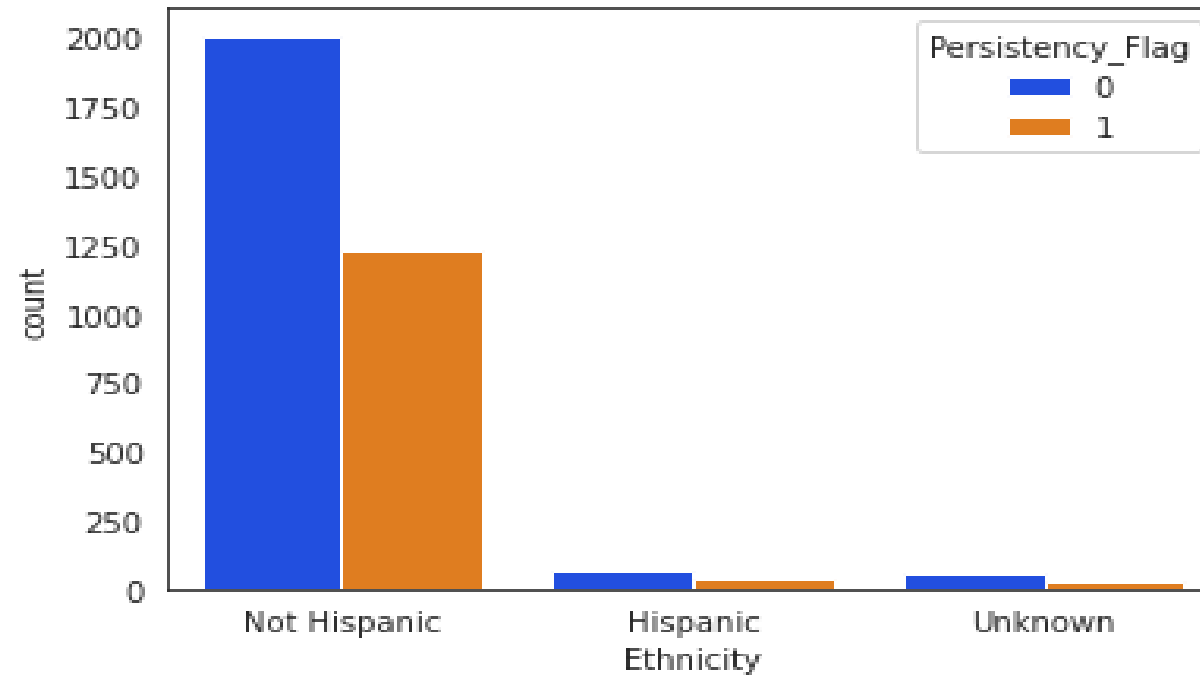- **Total Number of Patients: 3,424**

ANALYSIS

GENDER PROPORTION IN THE DATASET
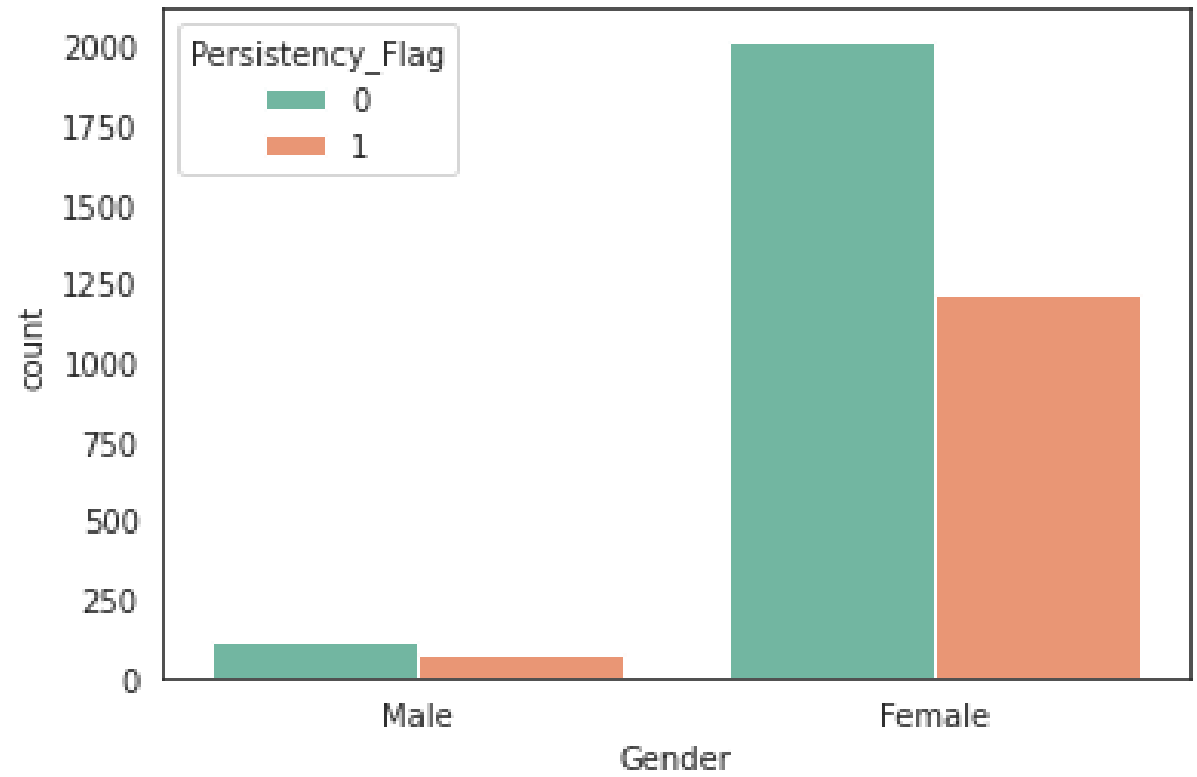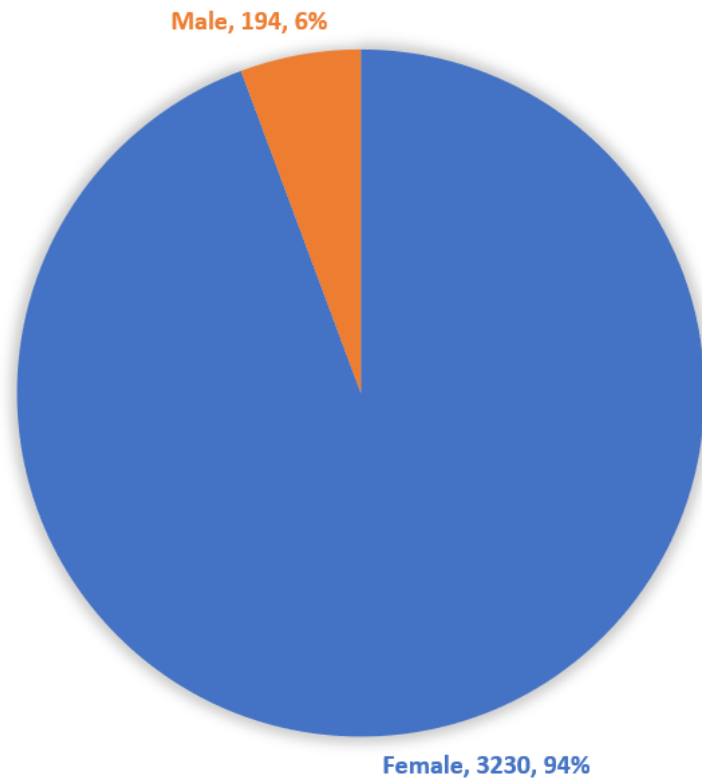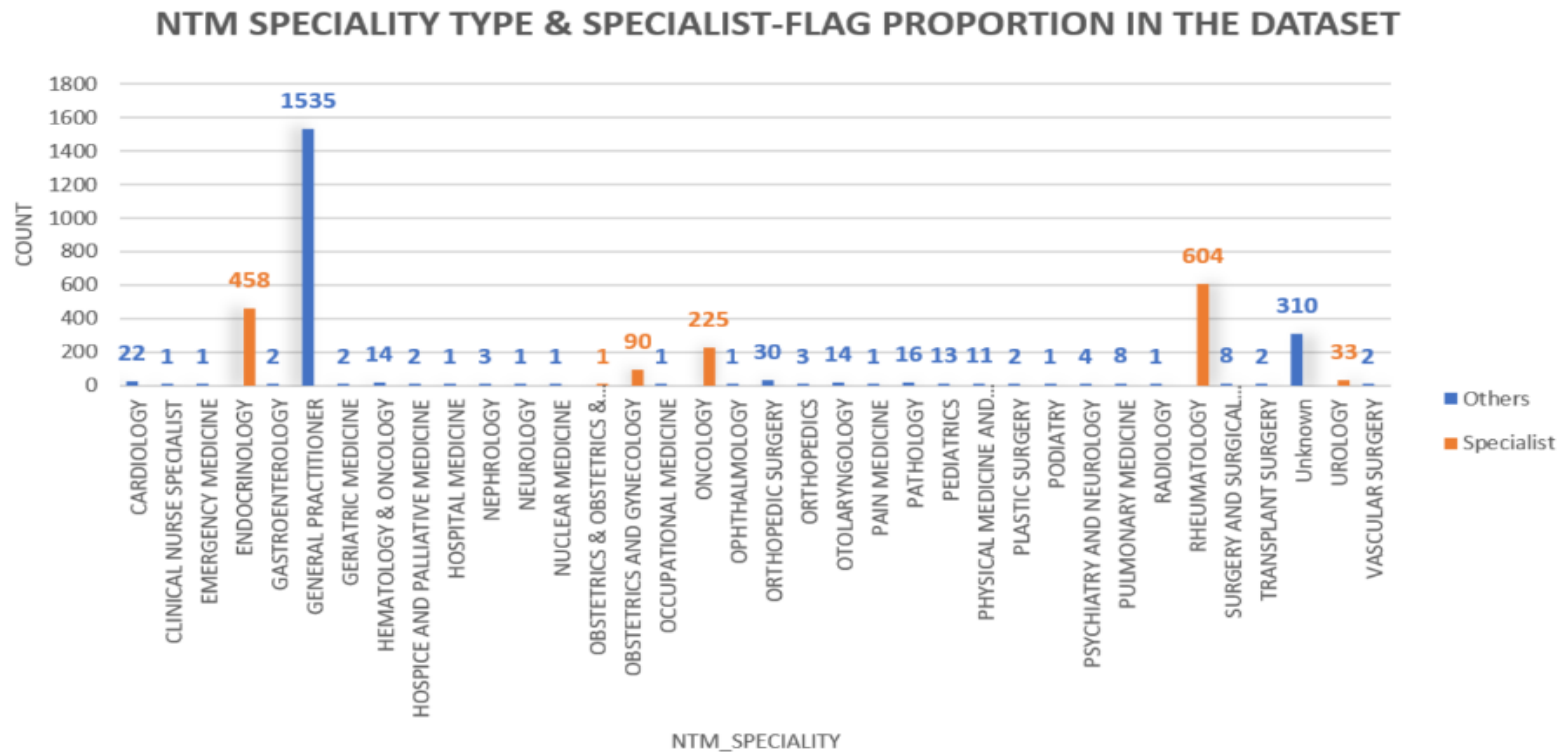
Age Bucket vs. Persistency Flag
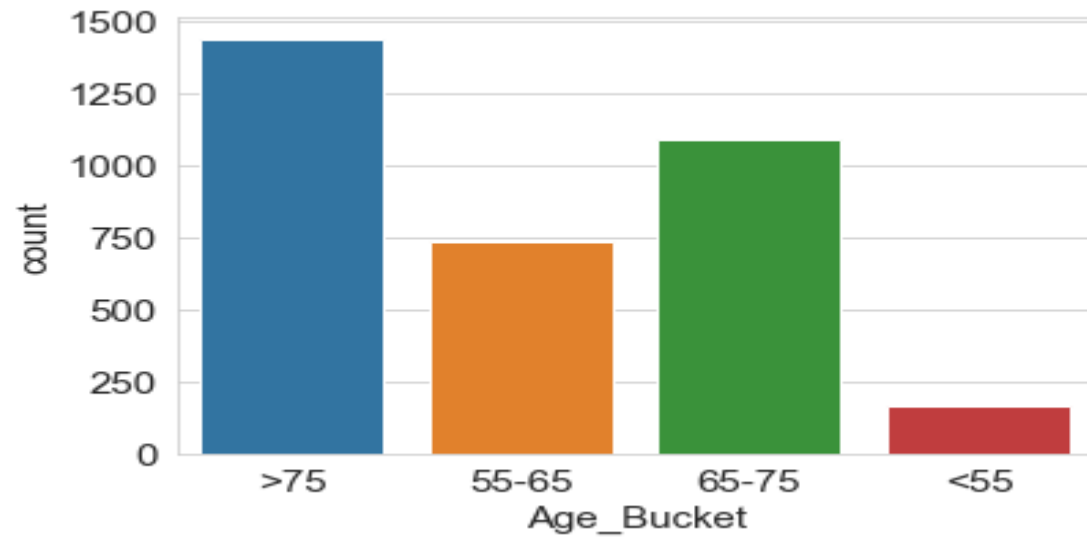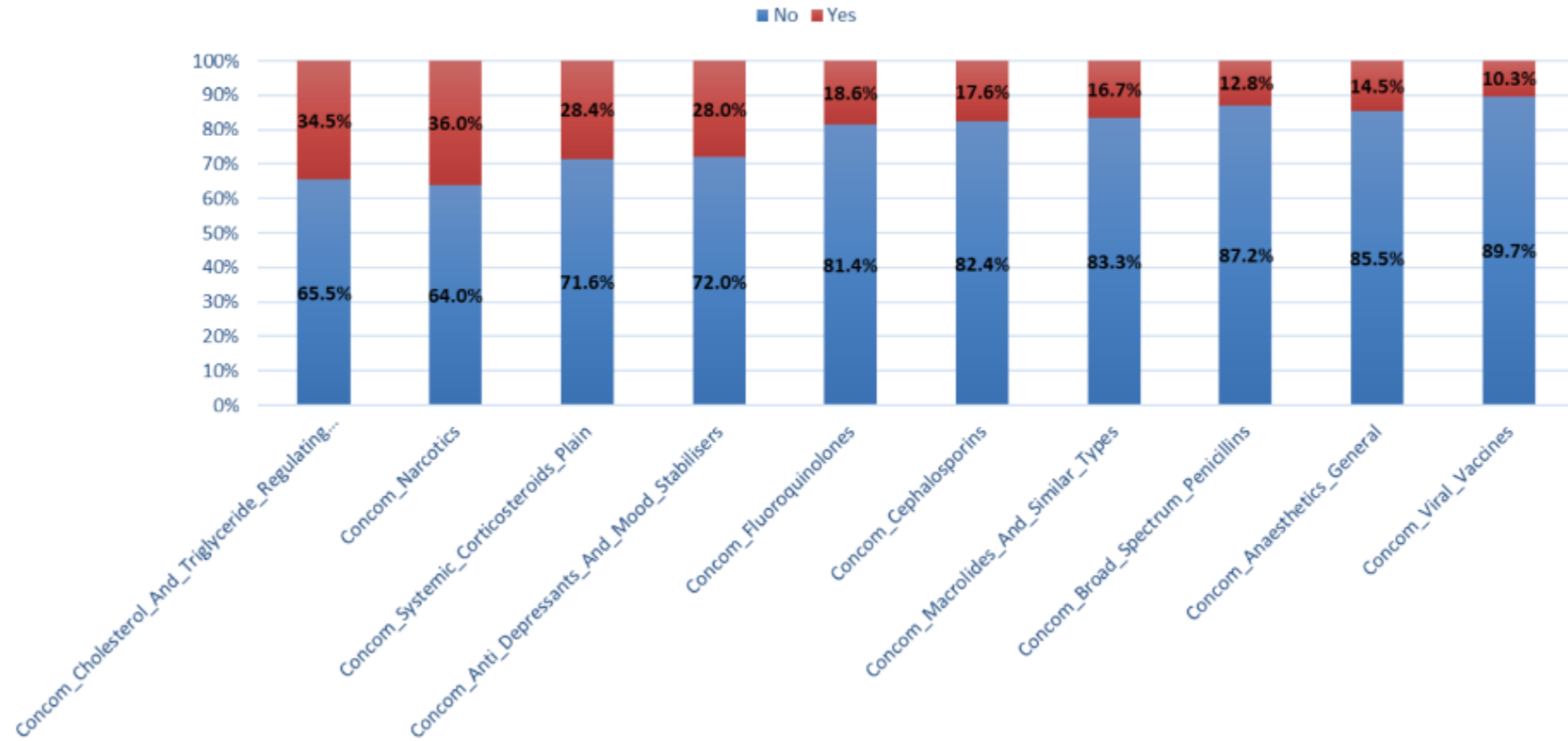
Ethnicity vs. Persistency Flag

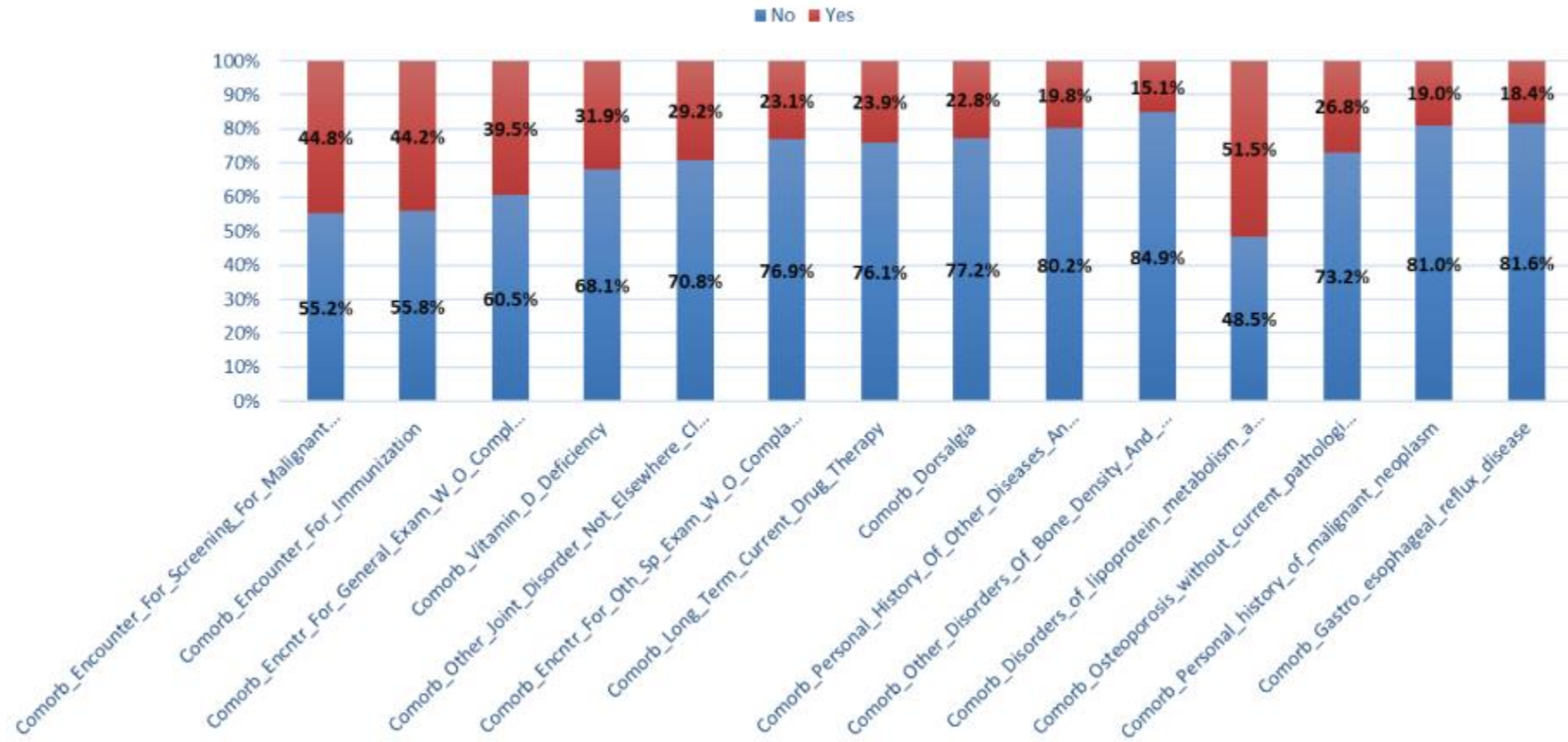# Analysis of Disease Types and Responsible Physician Specialties

This section focuses on understanding how various disease types and physician specialties are related to drug persistency outcomes

NTM SPECIALITY TYPE & SPECIALIST-FLAG PROPORTION IN THE DATASET

# Drug-related Feature Evaluation

# Disease-related Feature Evaluation

Clinical Risk Evaluation

- A significant portion of non-persistent patients had fewer than 3 recorded risk factors.

- However, patients with more than 3 risk factors showed the highest non-persistency rate relative to their total group size, indicating that increased clinical risk is strongly associated with treatment discontinuation.
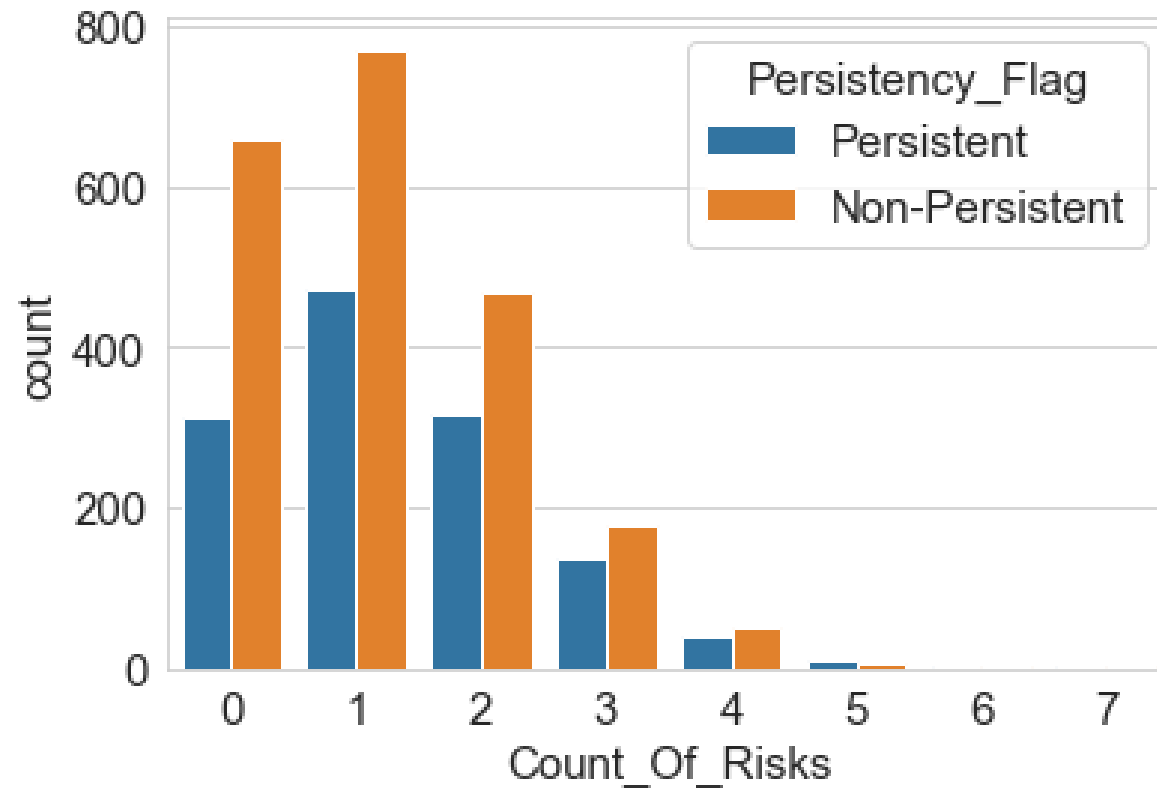
Feature Importance Analysis

The dominance analysis identified the top 15 most influential features impacting drug persistency. Notably, clinical parameters emerged as the most influential contributors, suggesting that a patient's medical and diagnostic history plays a critical role in determining treatment adherence.

Percentage Relative Importance Waterfall
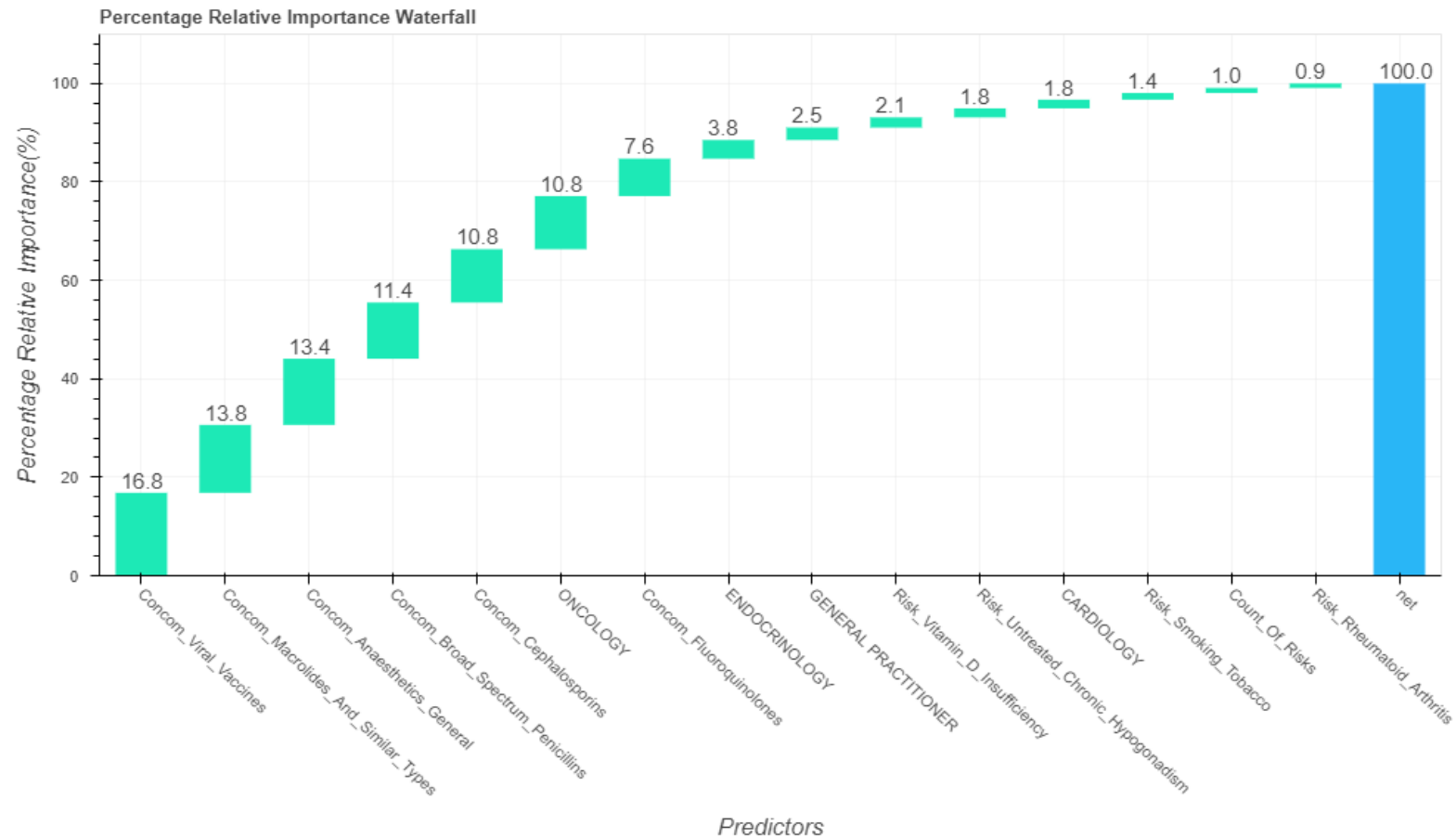
Recommendations

- From the EDA conducted on the dataset, the following recommendations are provided to ABC Pharma's technical team:

- Demographic factors included in the dataset do not exhibit a strong relationship with patient persistency levels and may have limited predictive value.

- The NTM Specialist type and Specialist Flag variables show no significant correlation with the target variable and could be considered for exclusion in further modeling.

- Through feature importance analysis, several key parameters were identified. These influential features can be used to create a refined subset of the dataset for more focused quantitative analysis.

- Clinical factors such as drug concomitancy, comorbidities, and risk factors display meaningful correlations with persistency levels, warranting further investigation using quantitative methods like machine learning to enhance predictive accuracy.