



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Individual Project – Healthcare Persistency

12/20/2024

Agenda

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Project Approach

Data Understanding

- Analyze demographic and regional disparities in persistency rates.
- Focus on features like Gender, Age_Bucket, Race, and Region.

Exploratory Data Analysis (EDA)

- Perform univariate and bivariate analyses to identify key trends.
- Use combined analysis to uncover deeper insights into persistency patterns.

Modeling

- Develop predictive models to forecast persistency based on demographic and behavioral features.
- Test multiple machine learning algorithms (Logistic Regression, Random Forest, Gradient Boosting).

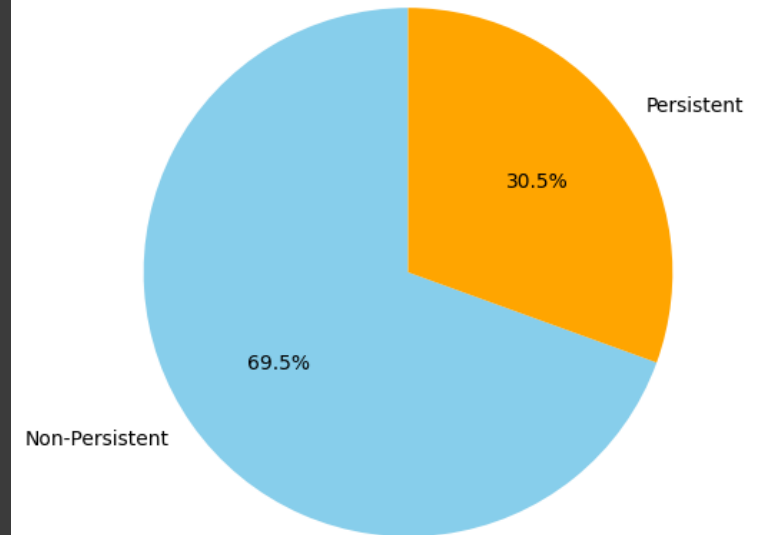
Recommendations

- Provide data-driven strategies for improving perscription adherence, tailored to high-impact demographics.
- Suggest tools and interventions based on insights from the analysis.

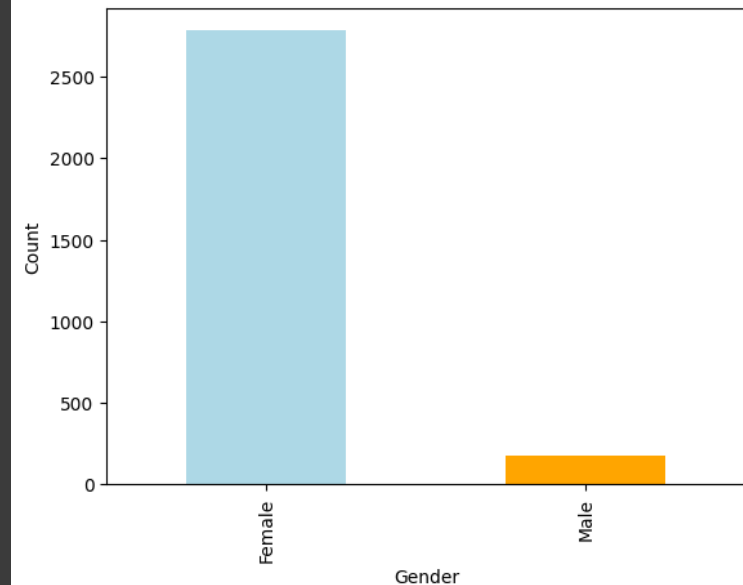
Key EDA Insights

- Key demographic factors like age, gender, race, and region impact persistency rates.
- Majority of patients are "Non-Persistent" (69.5%).
- Improving persistency rates will directly impact business revenue and patient health outcomes.

Distribution of Persistency_Flag

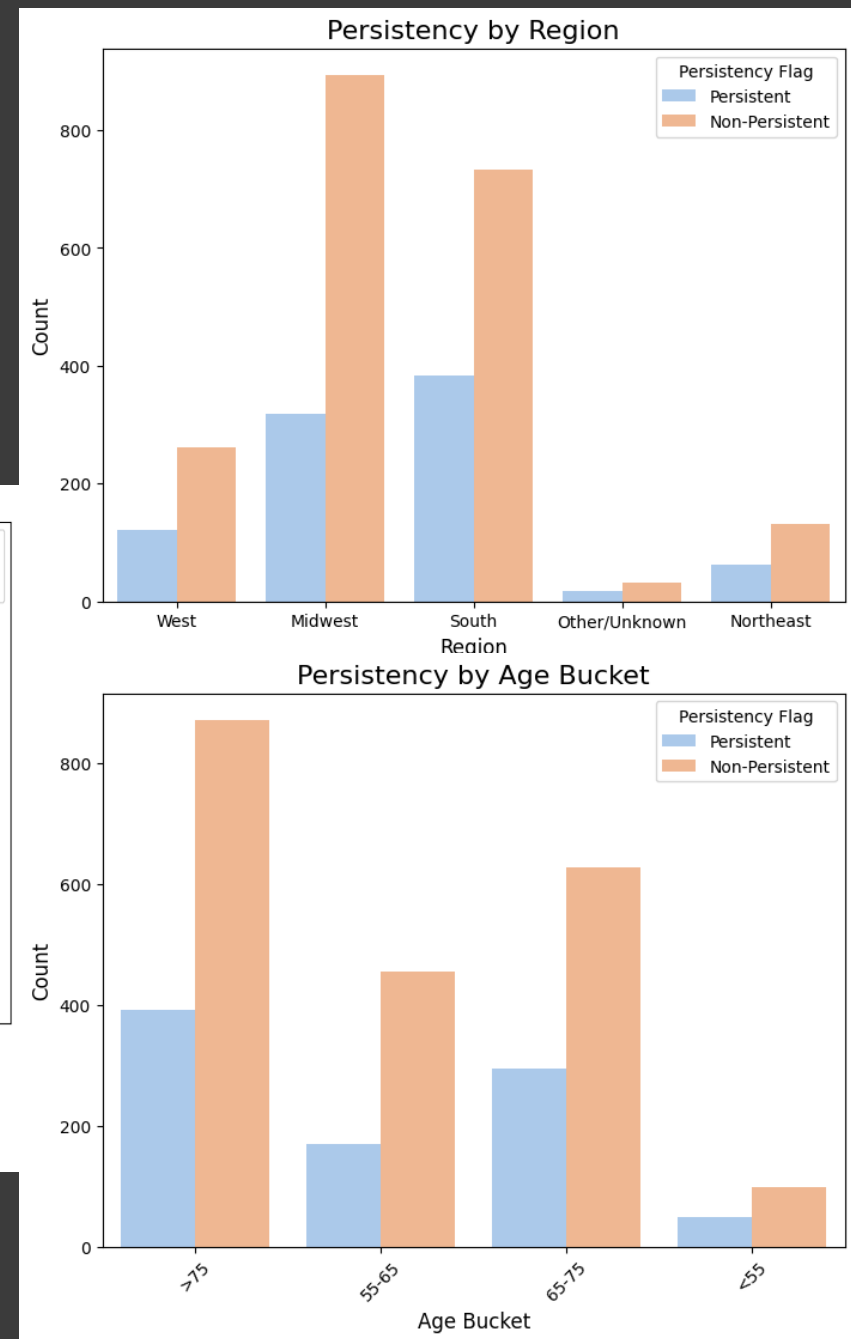
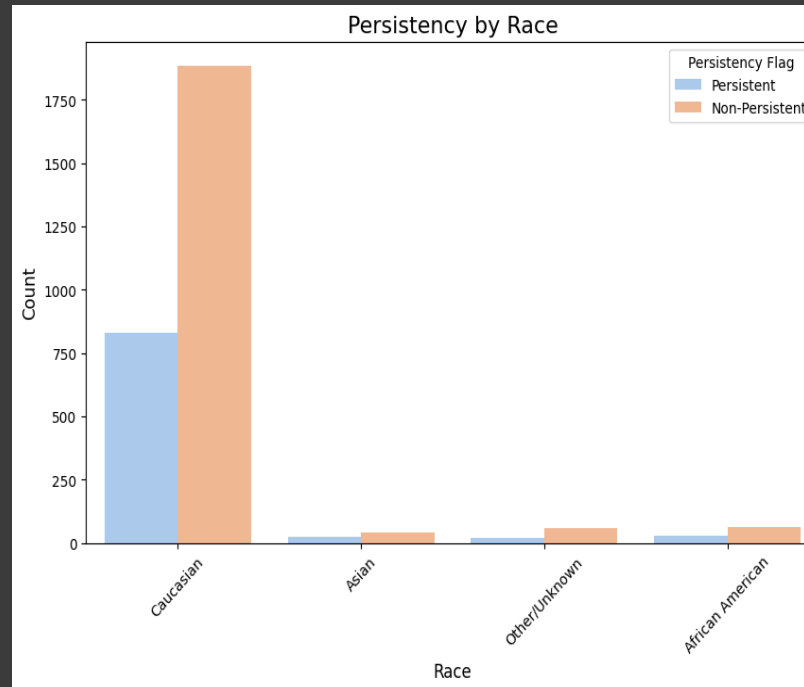


Distribution of Gender



EDA Key Insights

- **Persistency Distribution:** Majority are "Non-Persistent."
- **Gender:** Females show higher persistency rates.
- **Age Groups:** Patients >75 years are more persistent.
- **Regions:** Non-persistent rates are highest in the Midwest and South



Recommendations to Business

- Prioritize targeting males, non-Caucasians, and younger patients (<55 years) with educational programs.
- Implement digital tools for adherence reminders, especially for younger patients.
- Focus on demographic and regional variations to design customized strategies.

Recommended Models for Technical User

These are types of machine learning techniques that can be used to make predictions. Each has its own strengths:

- **Logistic Regression:** Helps identify which factors (like age or gender) influence whether a patient is persistent.
- **Random Forest:** Handles complex datasets well and can identify patterns in the data.
- **Gradient Boosting:** Combines multiple smaller models to make smarter predictions.

Focus on Key Features:

These are the most important factors identified during the data analysis, like Age, Gender, and Region. The recommendation is to build a model that uses these features to predict persistency.

Fixing Class Imbalance:

Since there are more "Non-Persistent" than "Persistent" patients, the model might be biased. This step makes sure both groups are treated fairly by either balancing the data or changing how the model learns from it.

How to Check Model Accuracy:

After building the model, technical users will need to measure its precision and success.

This involves checking things like:

- **Accuracy:** How often the model predicts correctly.
- **Precision and Recall:** How well the model predicts each group.
- **ROC-AUC:** A score that shows how reliable the model is overall.

Deploying the Model:

This step makes the model usable in real-world settings, like integrating it into an app or a dashboard.

Regular updates ensure the model stays accurate over time.

Thank You