Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Healthcare: Persistency of a Drug

Date: August 16, 2022
Batch Code: LISUM 10
Team Healthy Bones - Jeeyeon Shon (shon.jeeyeon@gmail.com)
University of Maryland, the United States
Specialization: Data Science
Submitted to: Data Glacier
Link to the Repo: https://github.com/shonjeeyeon/DG_Week_11

Data Glacier
Your Deep Learning Partner

Agenda

Executive Summary

Problem Description

Approach

Exploratory Data Analysis: Overview

Exploratory Data Analysis and Hypothesis Testing

Conclusion

Recommendations

Executive Summary

Data Glacier
Your Deep Learning Partner

- Using a dataset of 3,424 records and 69 features, Exploratory Data Analysis (EDA) was performed to analyze persistency of a certain medication.
- Visual comparisons as well hypothesis tests were used to determine differences in persistency levels among different sub-populations, grouped by features such as specializations of prescribers and comorbidities.
- Summary of the EDA and recommendations for further model development will be included in the presentation.

Problem Description

- Medication persistence refers to completing the medication treatment using the duration set by the prescriber. (Cramer JA et al., 2008)
- Therefore, persistence is important in patients' positive outcomes as well as in pharmaceutical industries' profits.
- Developing a model to automate prediction process will contribute to save time and cost spent by the company, and the prediction results can be used for marketing, patient education, or R&D purposes.

Approach

Data Glacier
Your Deep Learning Partner

Dataset provided by the company
(3,424 patient records)

EDA

Hypothesis Testing

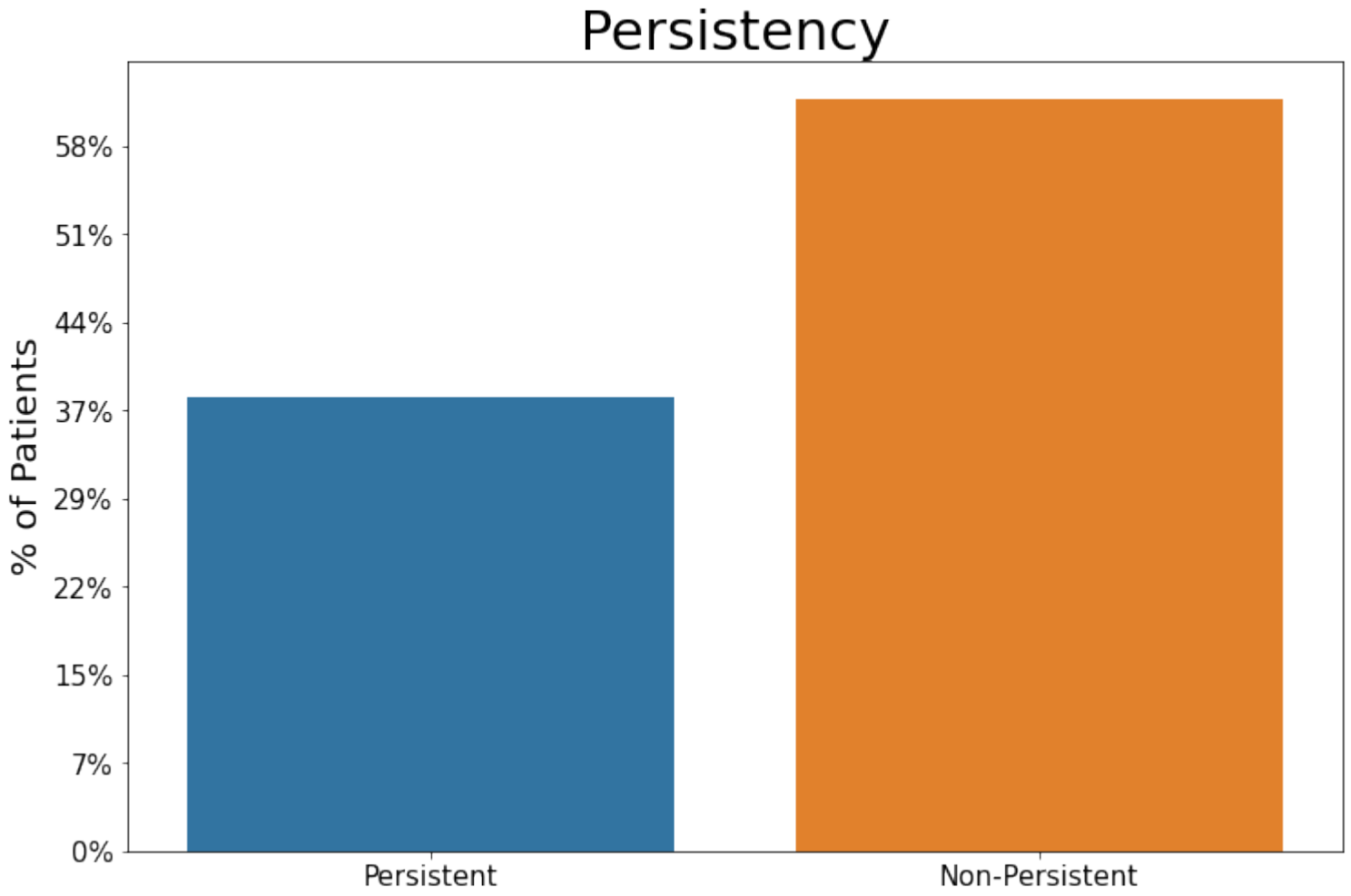
Conclusion and Recommendations for Further Model Development

Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis:

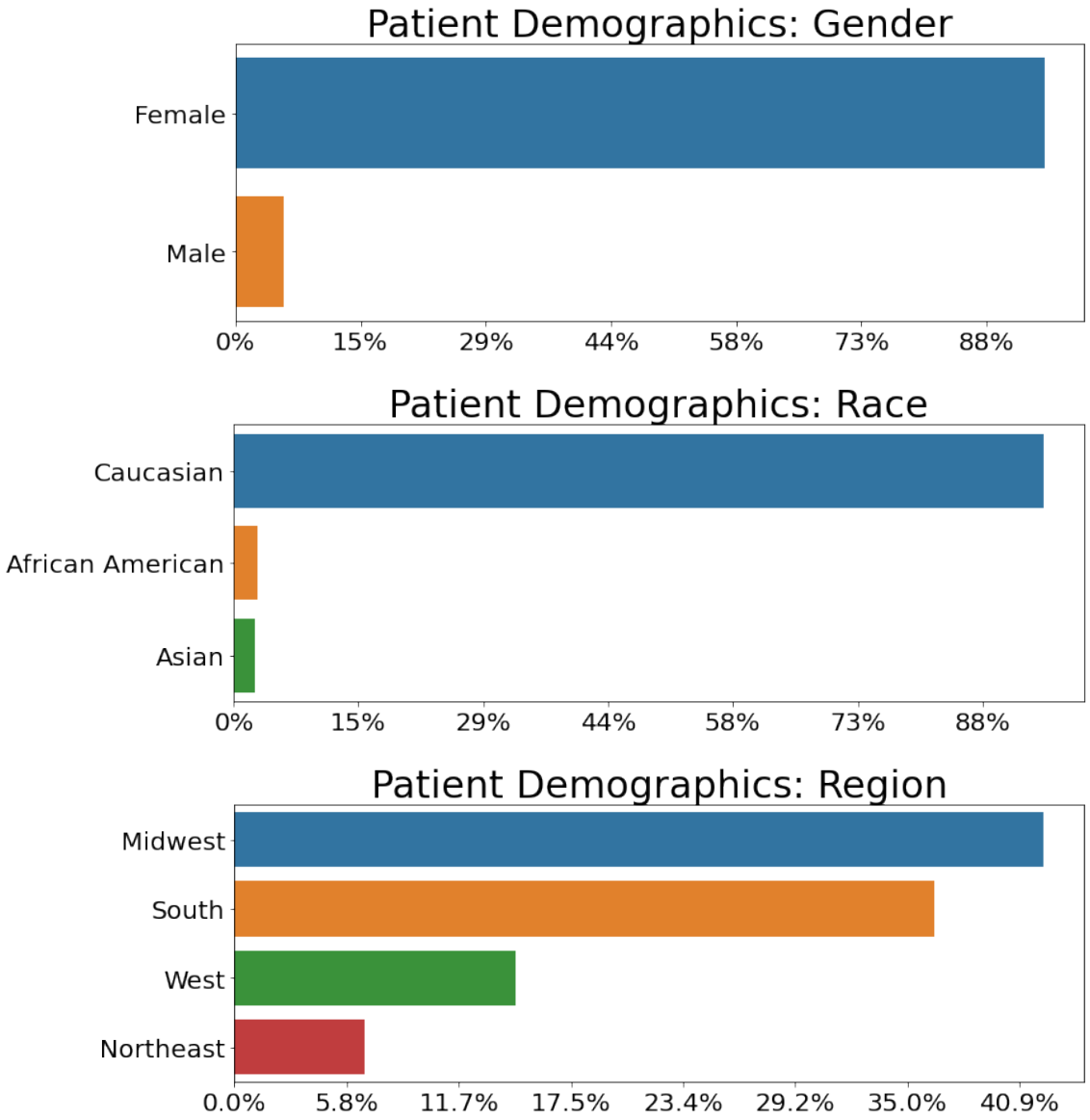
Dataset Overview

Dataset Overview: Persistency



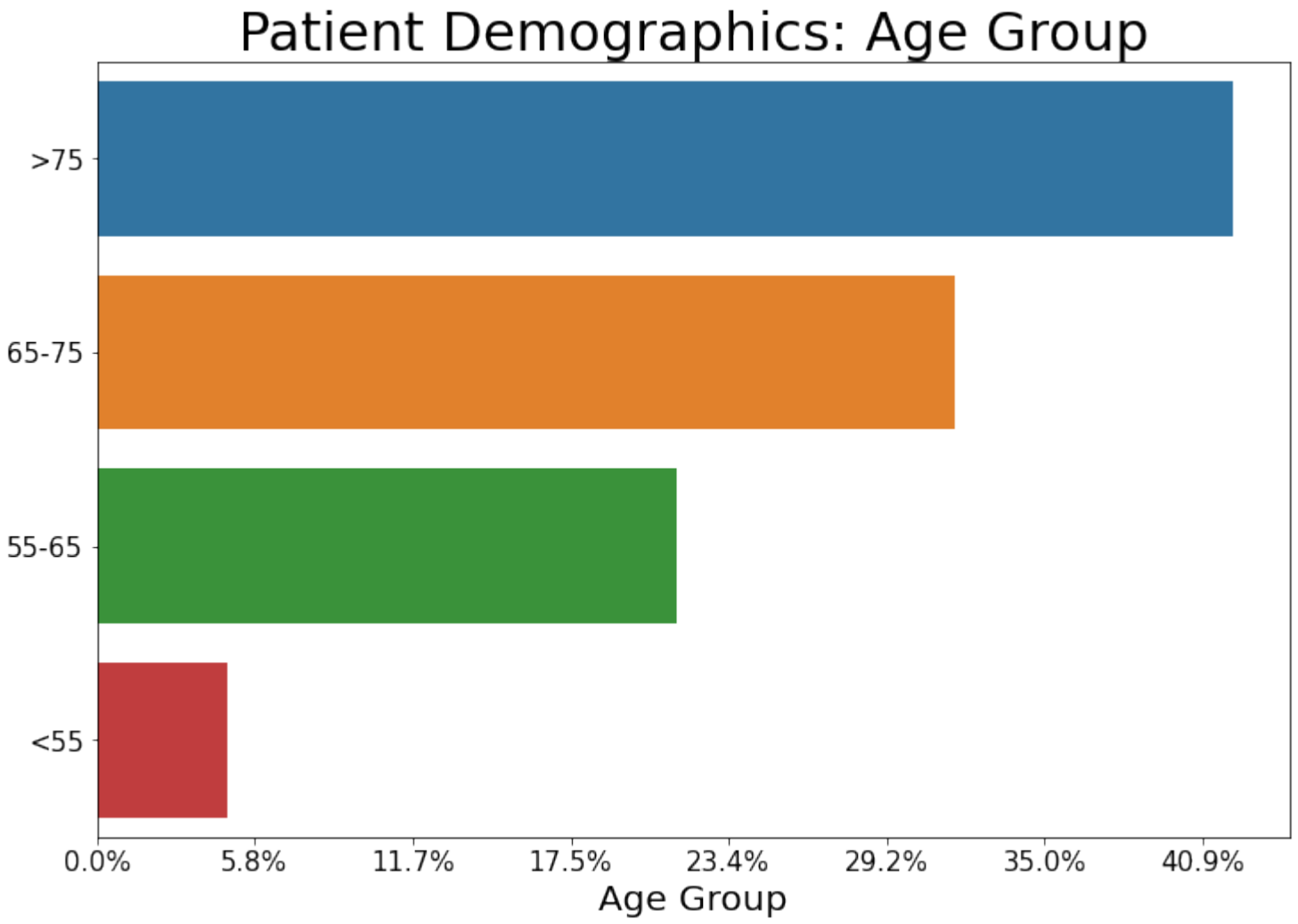
- Among 3,424 patients in the dataset, approximately 40% of the population continued the medication persistently.

Dataset Overview: Patient Demographics



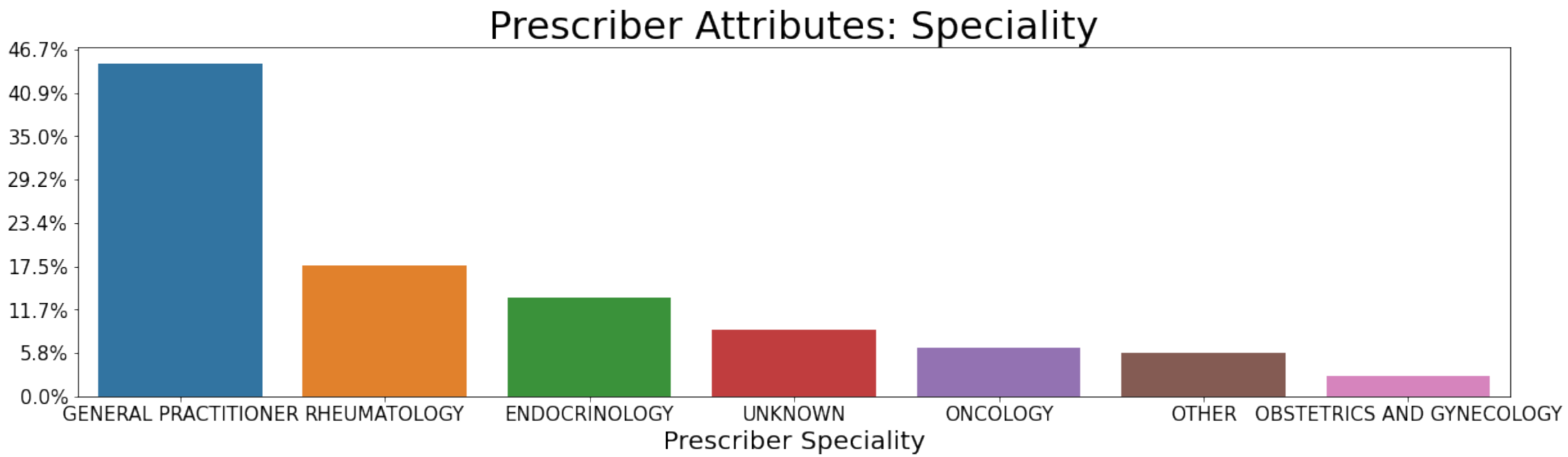
- Most of the patients were Caucasian or Female.
- Midwest was the most popular region, followed by South and West.

Dataset Overview: Patient Demographics



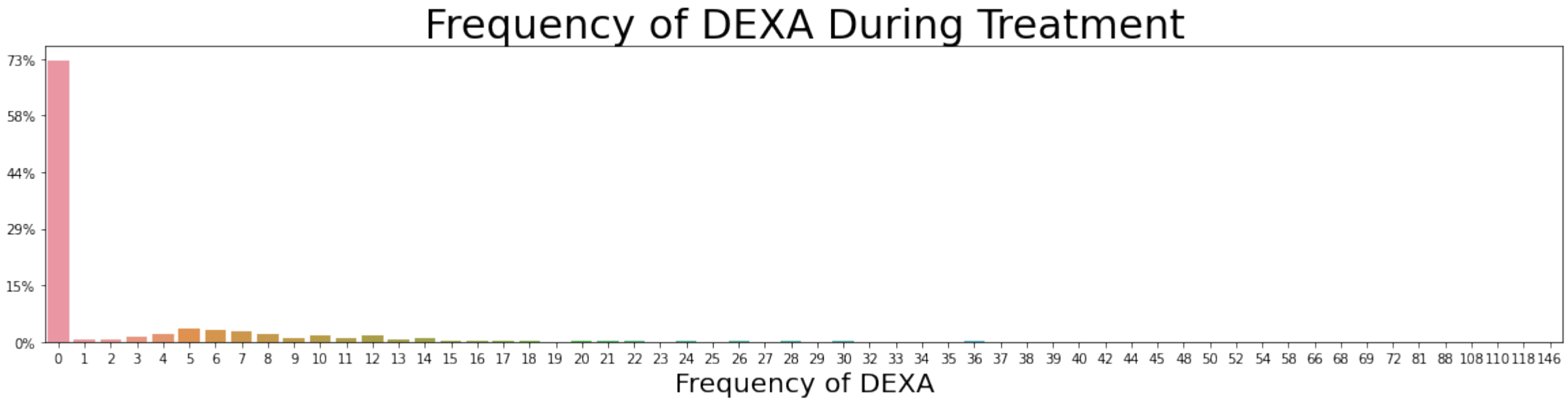
- Most of the patients in the dataset were over 75 years old, followed by patients aged 65 to 75 years old.
- Only about 5% of the population were under 55 years old.

Dataset Overview: Prescriber Attributes



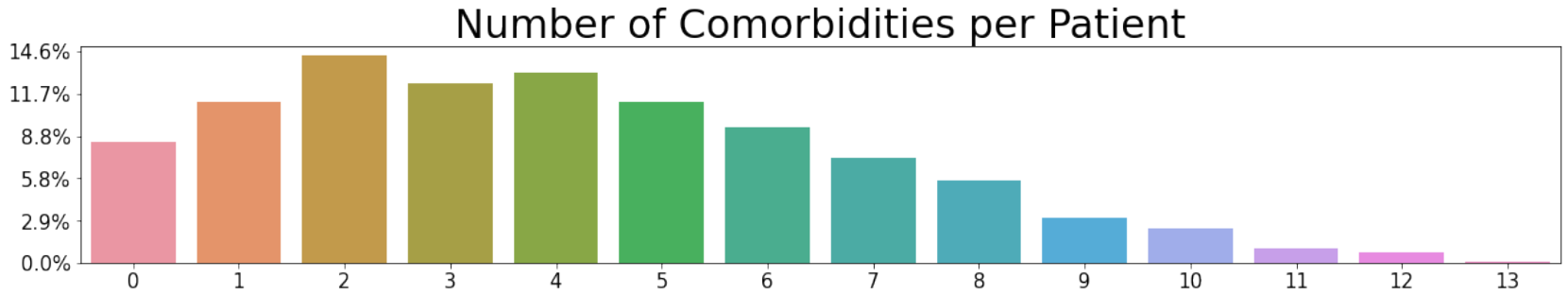
- The majority of the prescribers were General Practitioners.
- The second and third most common specialties were Rheumatology and Endocrinology.

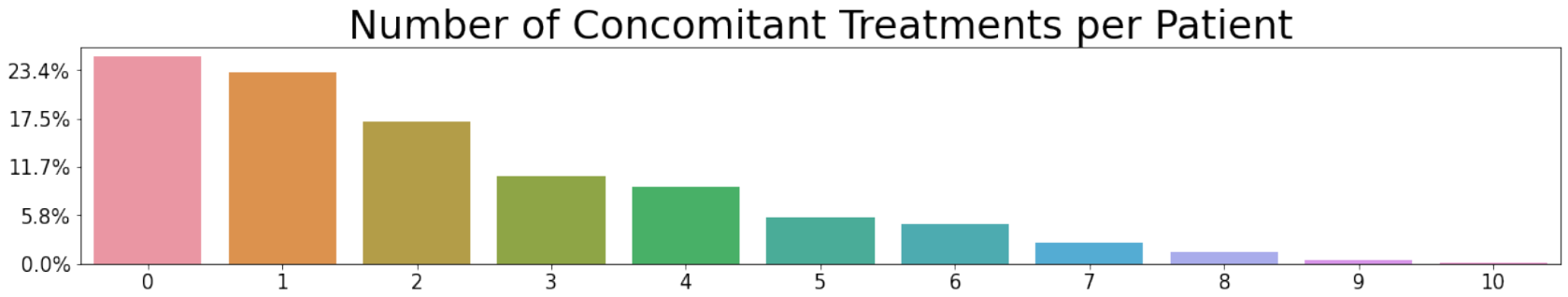
Dataset Overview: Frequency of DEXA

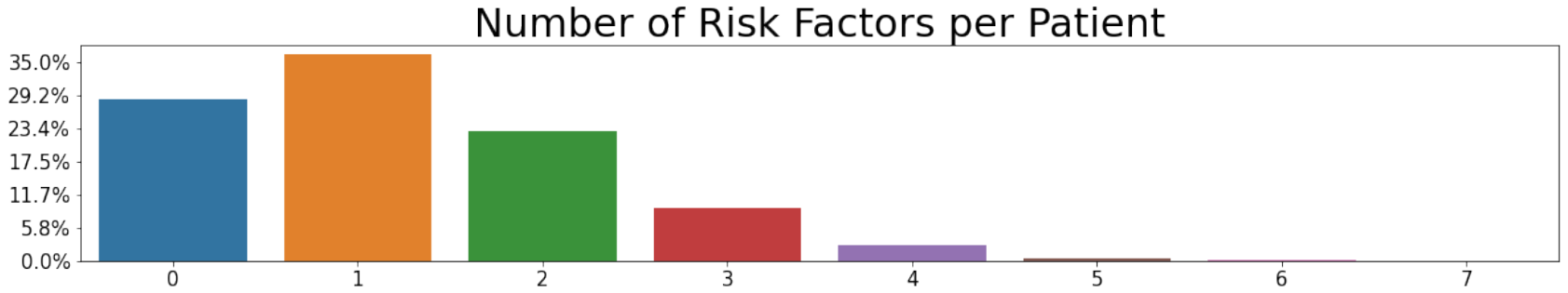


- About 75% of the patient did not take DEXA scan during treatment.
- The frequency DEXA scan during treatment varied substantially among patients, with patients taking the scan more than 140 times.

Dataset Overview:
Comorbidities, Concomitant Treatments, and Risk Factors







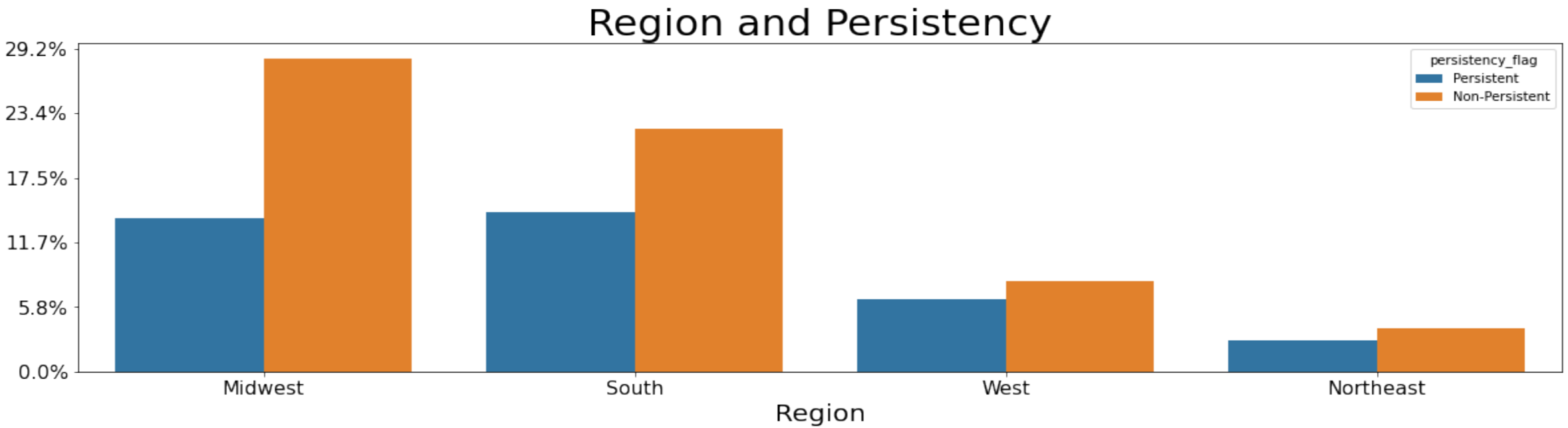
- The majority of the patient had at least one of the followings:
  - Comorbidities
  - Concomitant treatments
  - Risk factors

Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis:

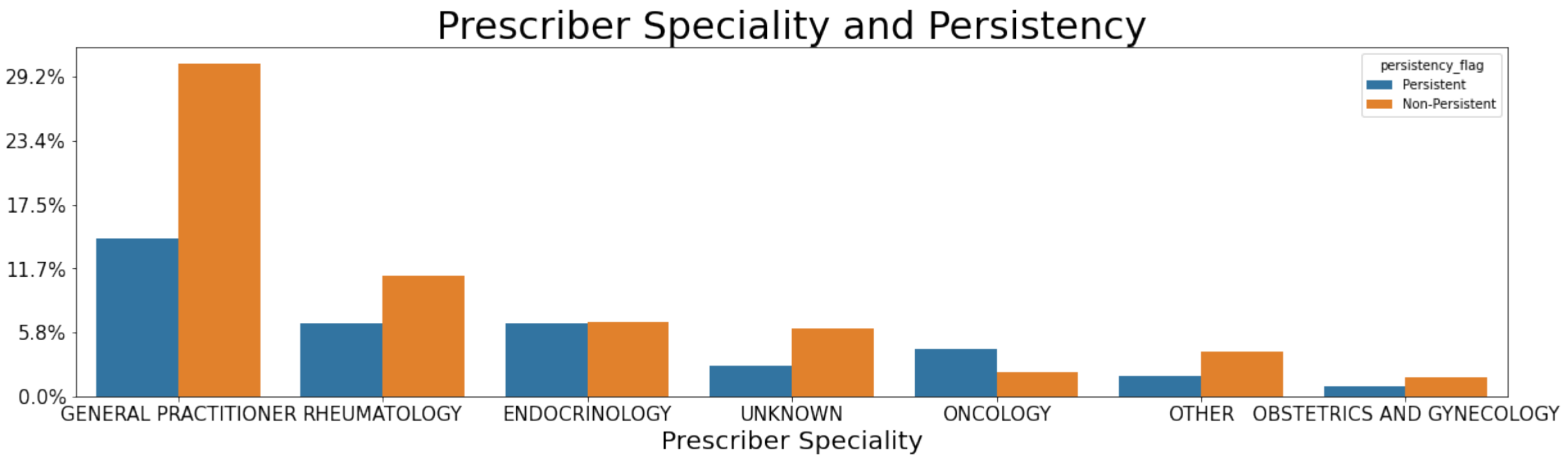
Persistency Comparison and Hypothesis Testing

Region and Persistency



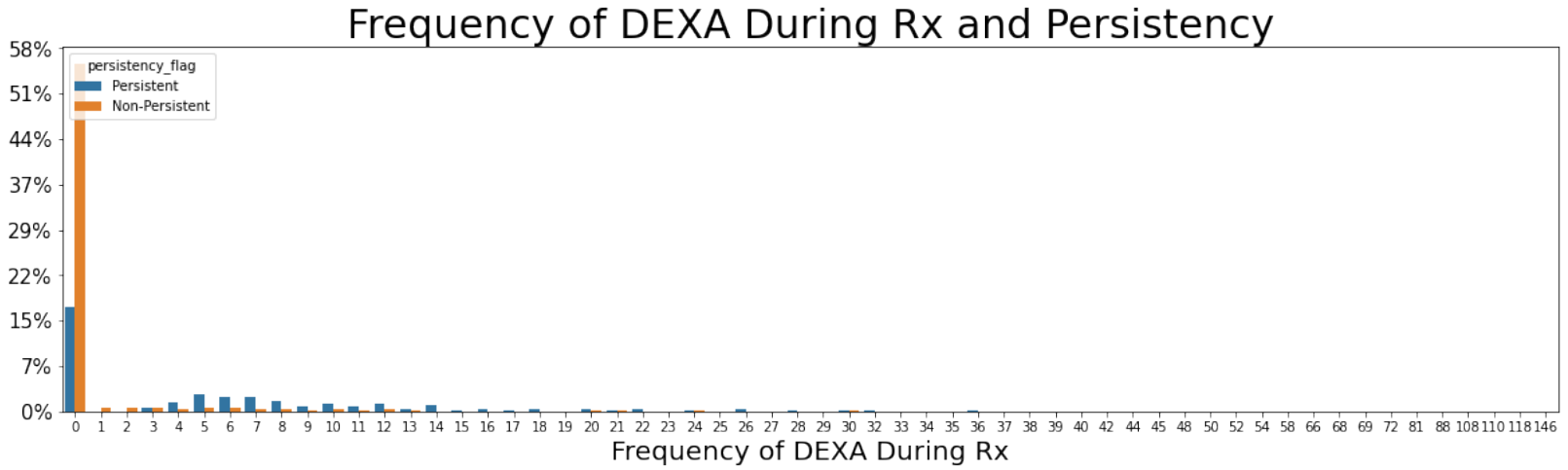
- Midwest has highest proportion of patients, but had the lowest persistency to non-persistency ratio.
- There are significant differences in persistency level among regions (ANOVA, $p < 0.05$)

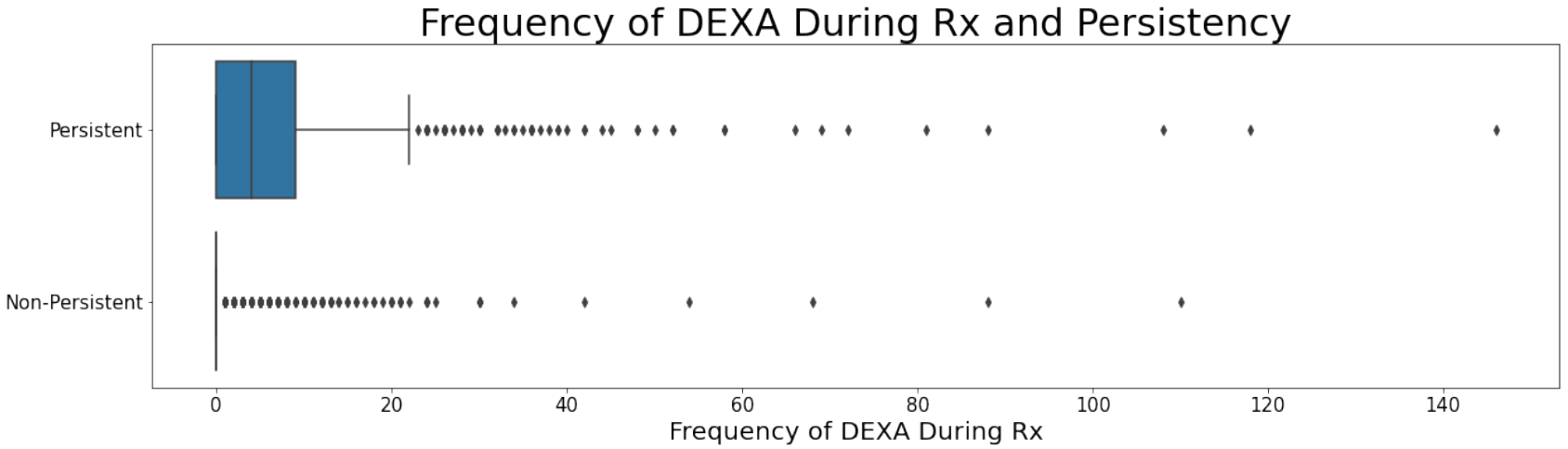
Prescriber Specialization and Persistency



- Endocrinology and Oncology has higher persistent-to-non-persistent ratio
- There are significant differences in persistency level among specializations (ANOVA, $p < 0.05$)

DEXA Frequency and Persistency

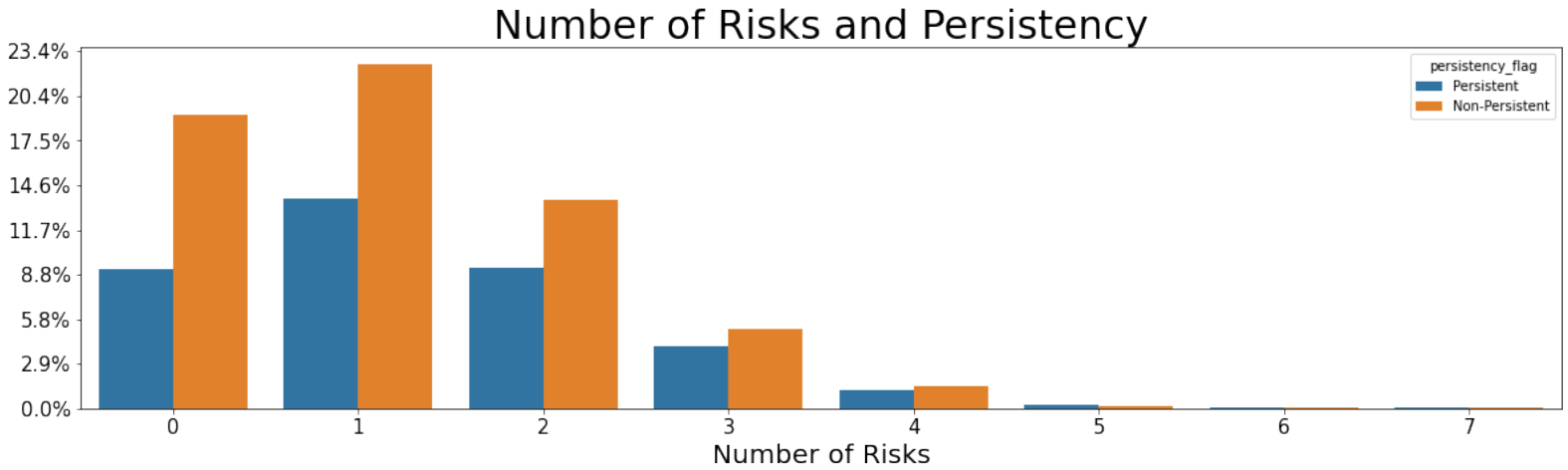


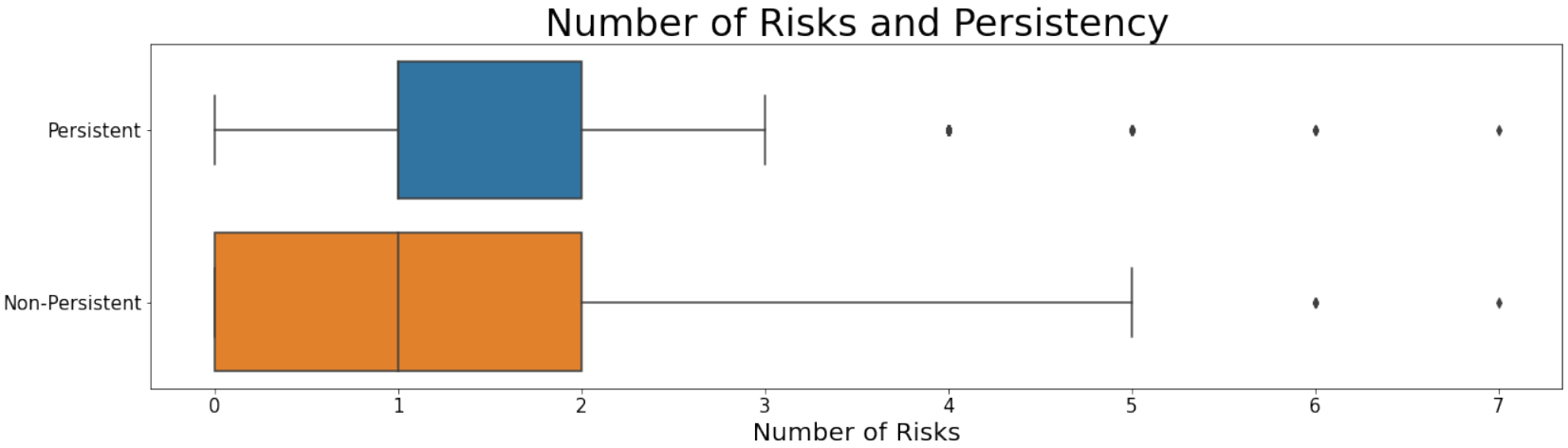


- Persistent group had higher mean number of DEXA scan during therapy
- There are significant differences in persistency level among different DEXA frequency groups

(ANOVA, $p < 0.05$)

Number of Risks and Persistency

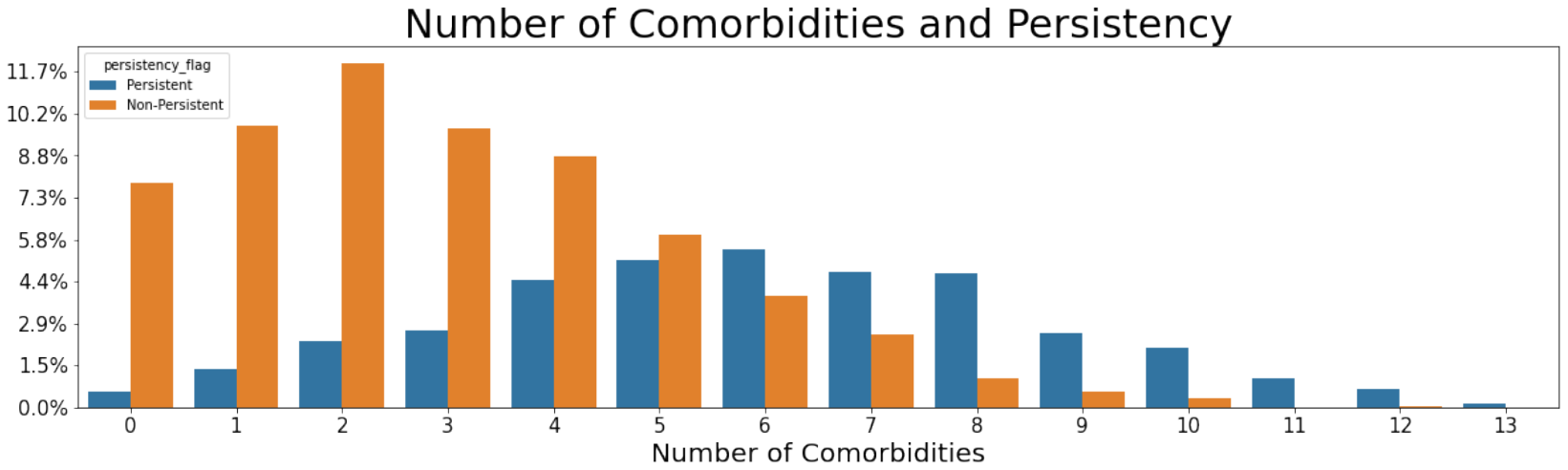


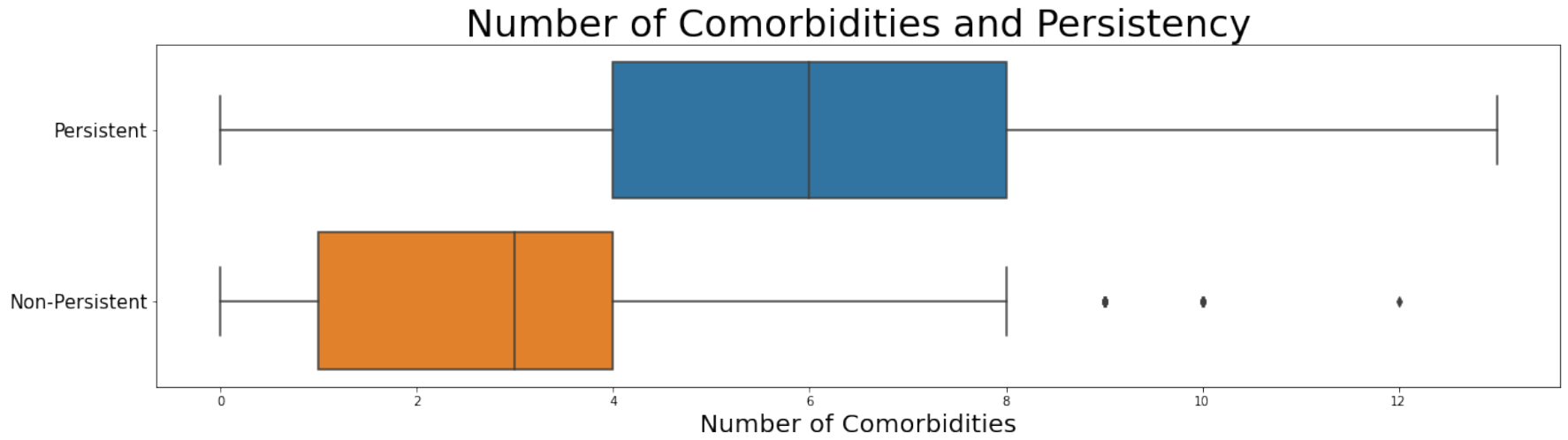


- In general, populations with more risk factors has higher persistent-to-non-persistent ratio
- There are significant differences in persistency level across different risk counts

(ANOVA, $p < 0.05$)

Number of Comorbidities and Persistency

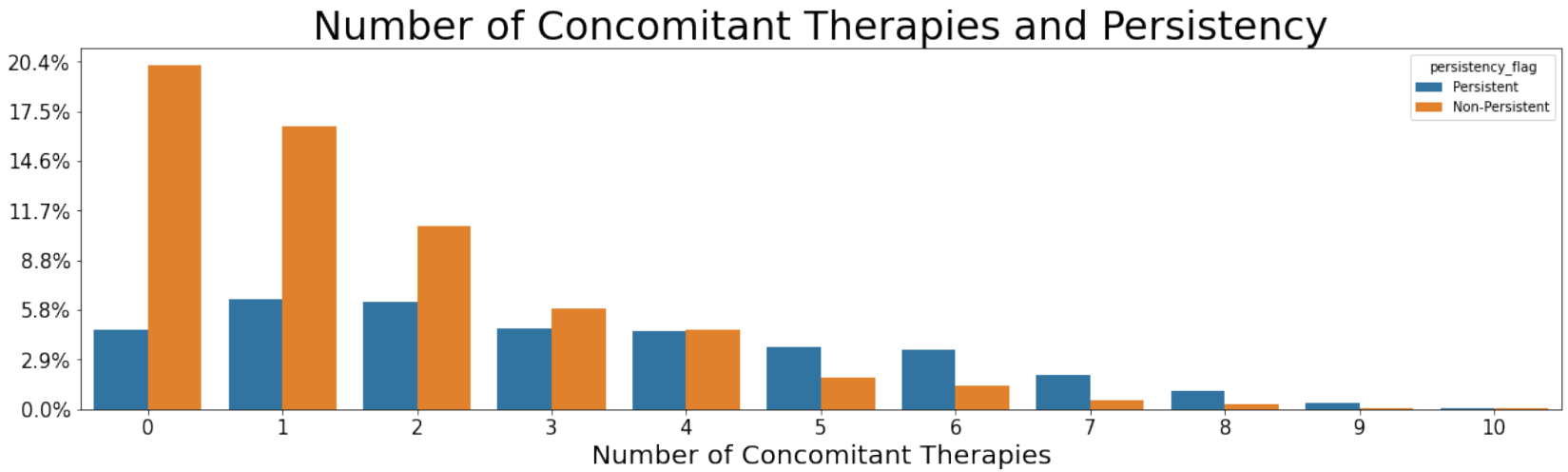


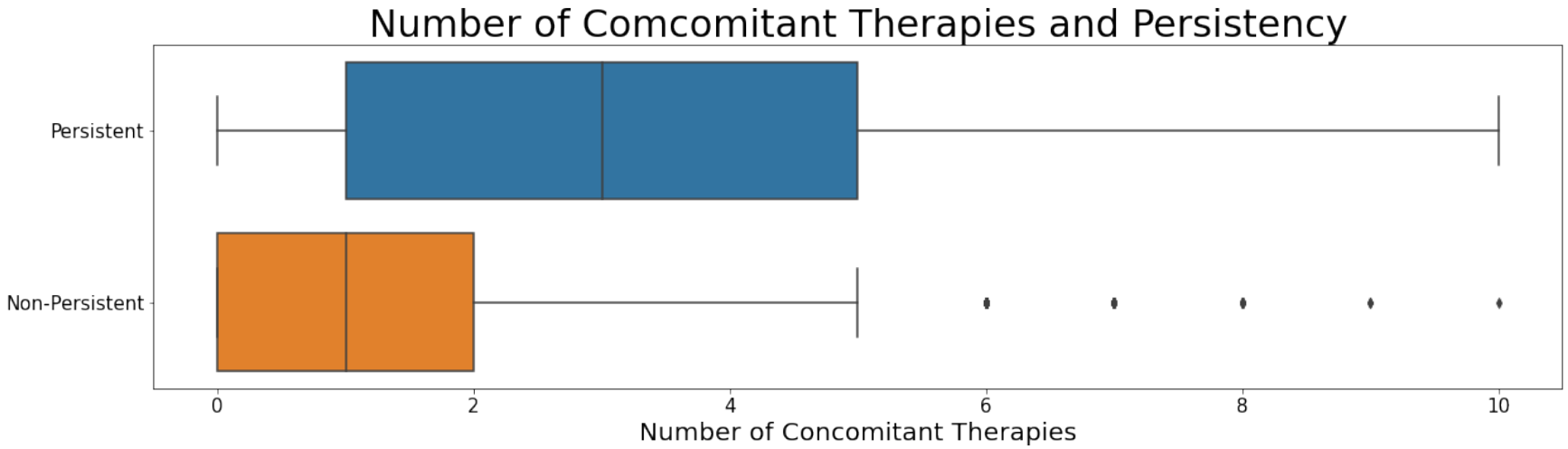


- Persistent group has higher mean in number of comorbidities.
- There are significant differences in persistency level across different numbers of comorbidities

(ANOVA, $p < 0.05$)

Number of Concomitant Therapies and Persistency





- Persistent group has higher mean in number of concomitant therapies.
- There are significant differences in persistency level across different numbers of concomitant therapies

(ANOVA, $p < 0.05$)

Other Features and Hypothesis Testing

- There were significant differences in persistency between:
  - Patients who had used glucocorticoid during therapy and who had not
  - Patients who had experienced fractures during therapy and who had not
  - Patients who had used injectables during therapy and who had not

  (Chi-square tests, p values <0.05)
- Every comorbidity and concomitant therapy was associated with significant difference in persistency (Chi-square tests, p values <0.05)
- Among risk factors, rheumatoid arthritis, chronic hypogonadism, tobacco use, chronic malnutrition, vitamin D insufficiency, and immobilization were associated with significant difference in persistency (Chi-square tests, p values <0.05)

Summary of the EDA

Data Glacier
Your Deep Learning Partner

- Region played a role in differences in persistency.
- Prescribers with certain specializations were associated with higher persistency.
- Taking DEXA scan during therapy were associated with higher persistency.
- Having at least one comorbidities, concomitant therapies, and/or risk factors were associated with higher persistency.
- All of the comorbidities and concomitant therapies, as well as select risk factors, were associated with difference in persistency.

Recommendations

Data Glacier
Your Deep Learning Partner

- Classification should be used for model and application development.
- Simpler models such as decision tree or linear regression should be tried first, before moving on to more complex and time-consuming models.
- The dataset has a substantial number of features. Using PCA or RFE to select important features is highly recommended.

Reference

- Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, Wong PK. Medication compliance and persistence: terminology and definitions. Value Health. 2008 Jan-Feb;11(1):44-7. doi: 10.1111/j.1524-4733.2007.00213.x. PMID: 18237359.

Thank You