EDA PRESENTATION AND PROPOSED MODELING TECHNIQUES

AMNA NAEEM

DATA SCIENCE PROJECT-HEALTHCARE PERSISTENCY OF A DRUG

PROBLEM DESCRIPTION:

- One of the greatest problems faced by Pharmaceutical companies these days is of persistency related to drugs. Persistency is related to the behavior of patient towards the medication advised by doctor. This term reflects the duration of time a drug is taken by a patient from the start of treatment to the end of treatment as maintained by doctor. The persistency is important because it can save not only the health of patient but also the money endured on health care system on illness due to discontinuity of drugs.

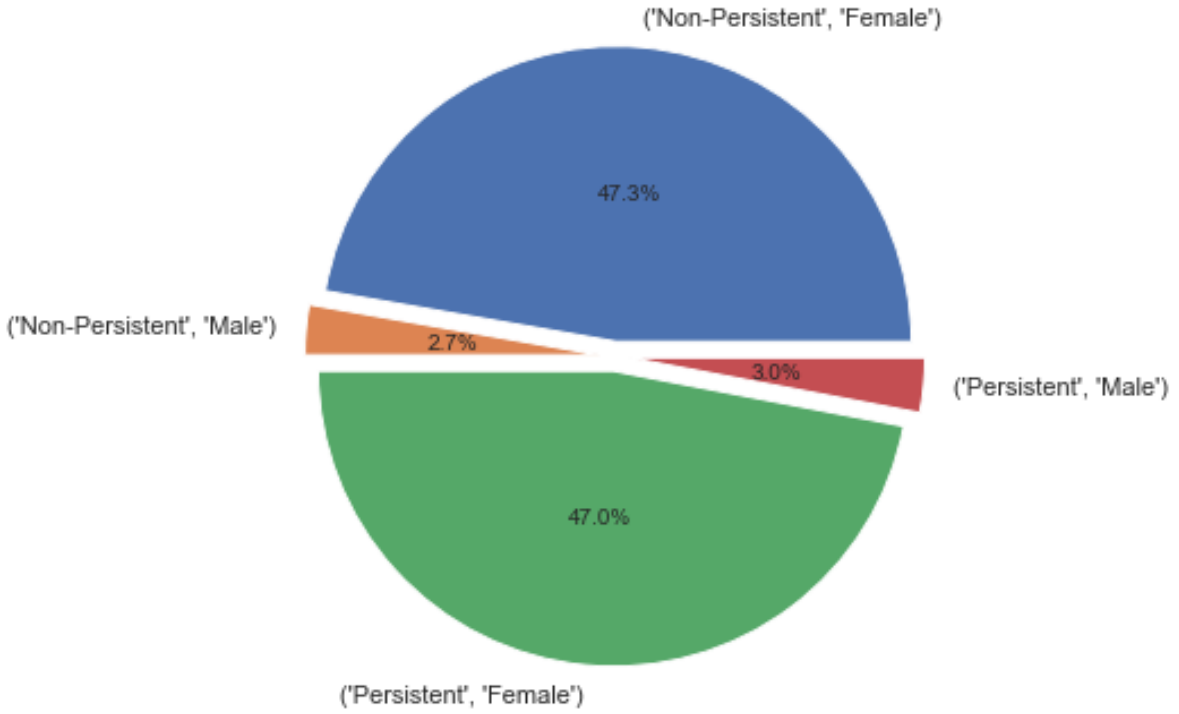
- **OBJECTIVE OF THE STUDY:**

The objective is to gather insights on the factors that are impacting the persistency, build a classification for the given dataset

- **Assumptions:**
  - The analysis is performed after cleaning the data for possible anomalies.
  - NTM –specialty is also categorized for handling the problem related to rare category.

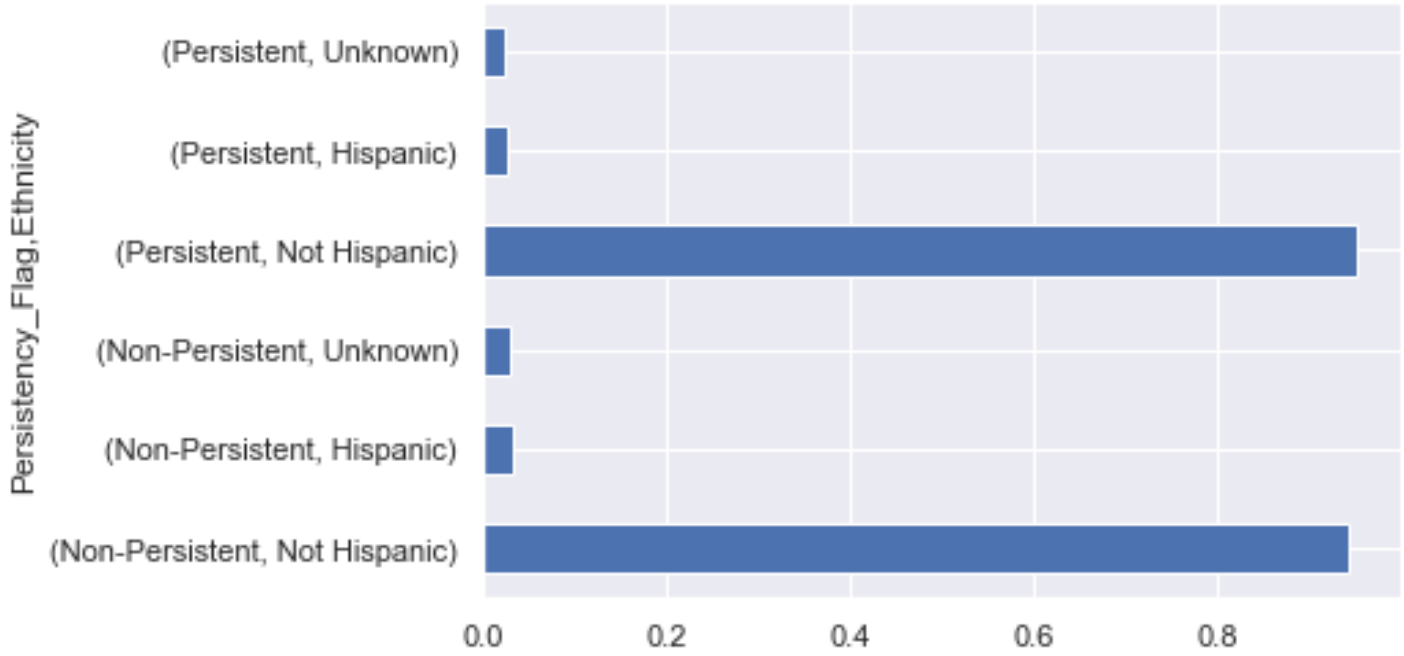
Persistency and Gender:

The persistency in genders are found to be same as **0.3%** difference exist between persistency and non-persistent male and female.



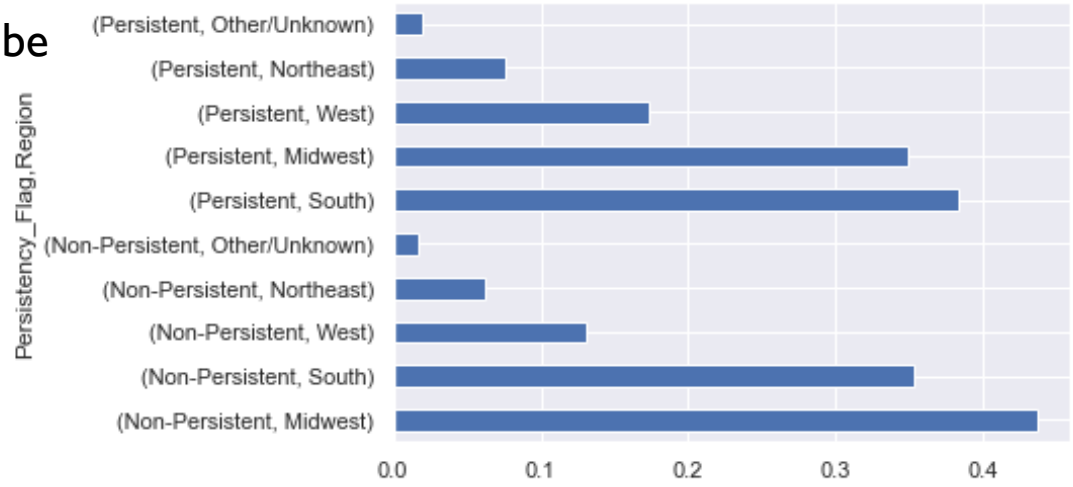
Persistency and Ethnicity:

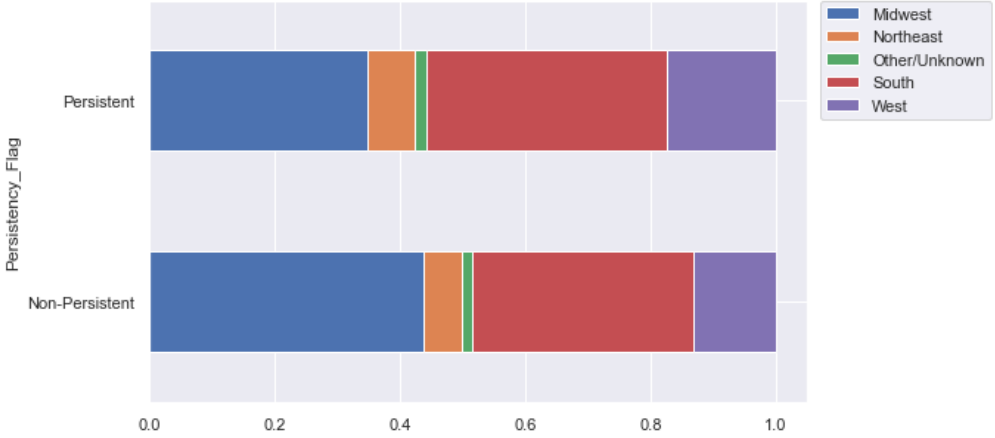
The people belongs to Not Hispanic group are found to be more persistent as compare to Non-Hispanic.



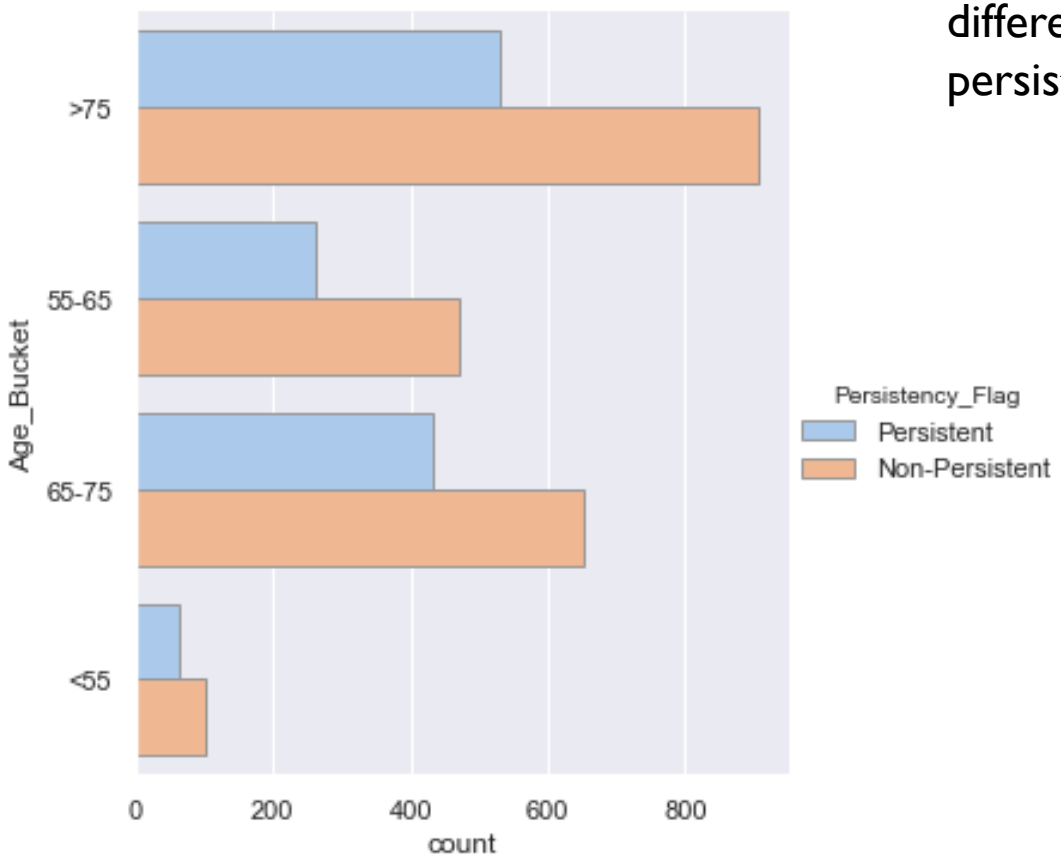
Persistency and Region:

People from all regions are found to be mostly non-persistent however the people from **west regions are appeared to be more persistent** than non persistent people.



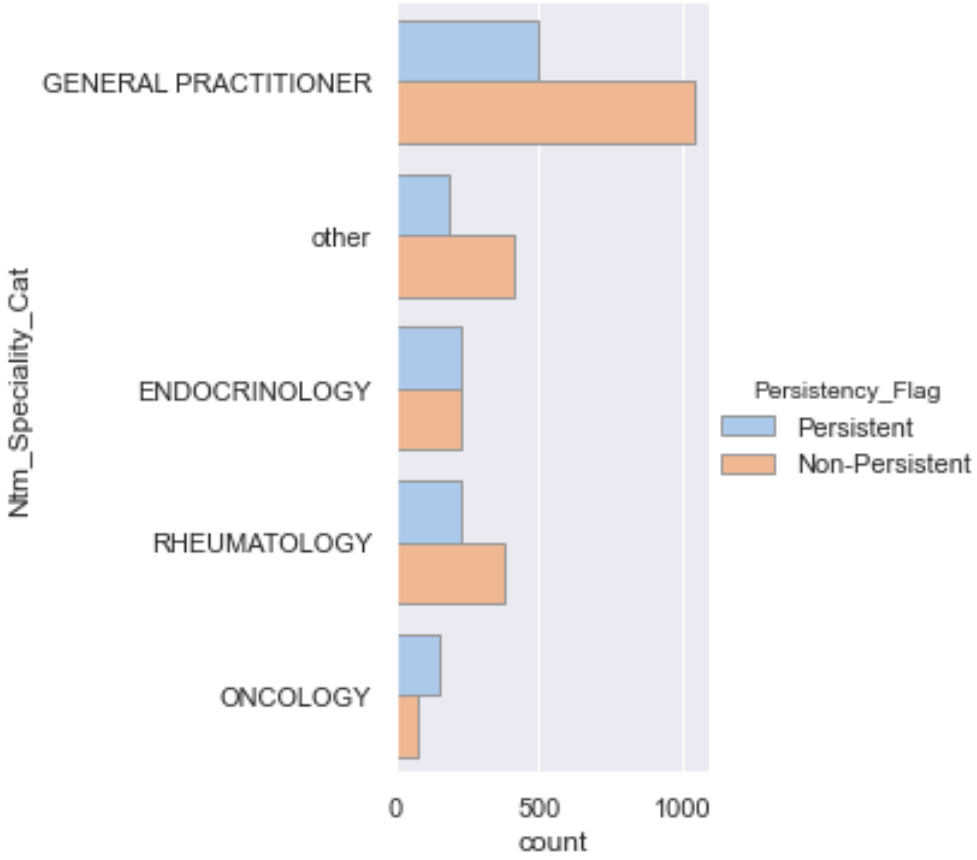


Persistency and Age:



People from all age groups are found to be more non-persistent however the difference between persistent and non-persistent for age group<55 is small.

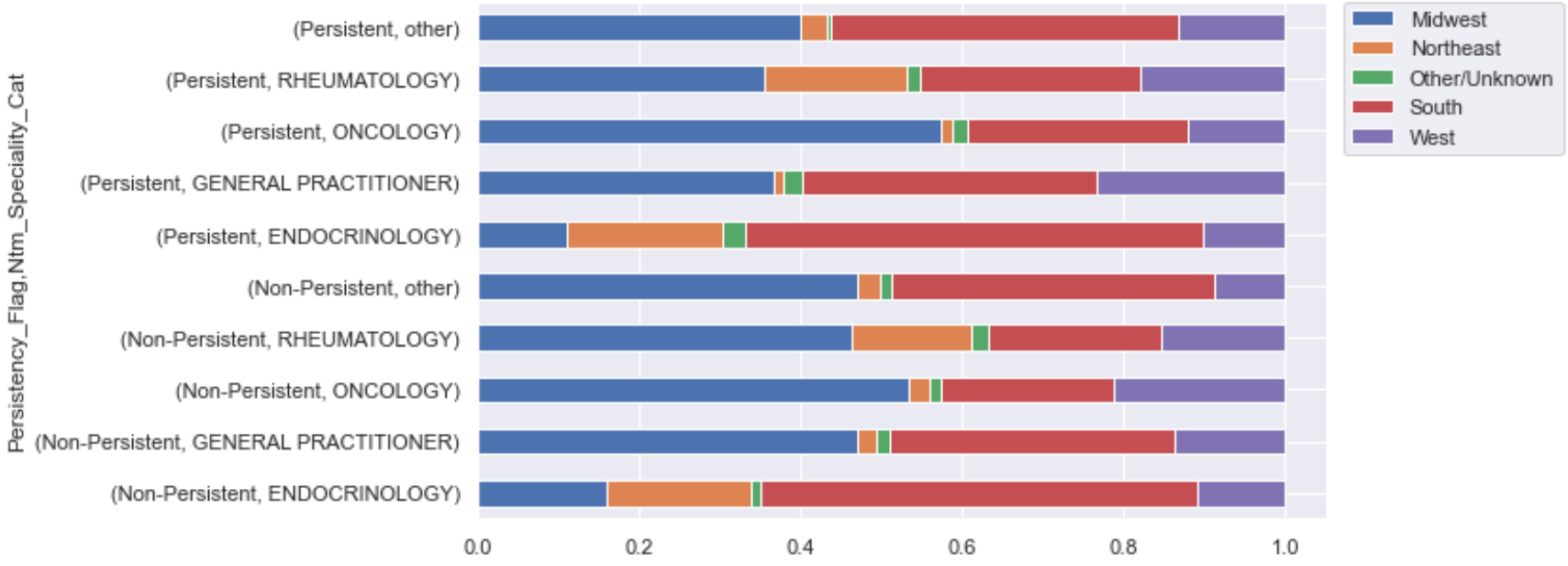
Persistency and Ntm-Speciality:



The medicine prescribed by oncology specialist seems to have more persistency and endocrinology specialists are found to have same number for both categories. Whereas the large number of non persistent in have been found patients in case of general practitioner.

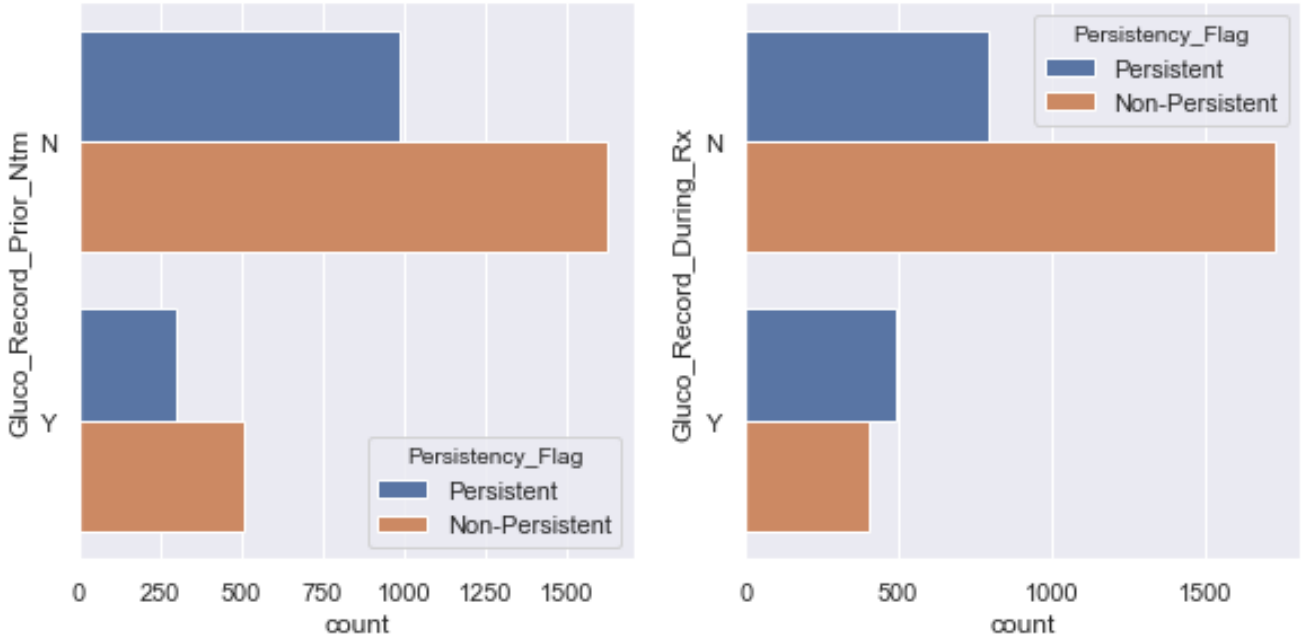
Persistency, Ntm-Speciality and Region:

The west people are found to be more persistent in case of general practitioner where people from Midwest region are found to be more persistent when it comes to Oncology. The people from south are more persistent when it comes to endocrinology.

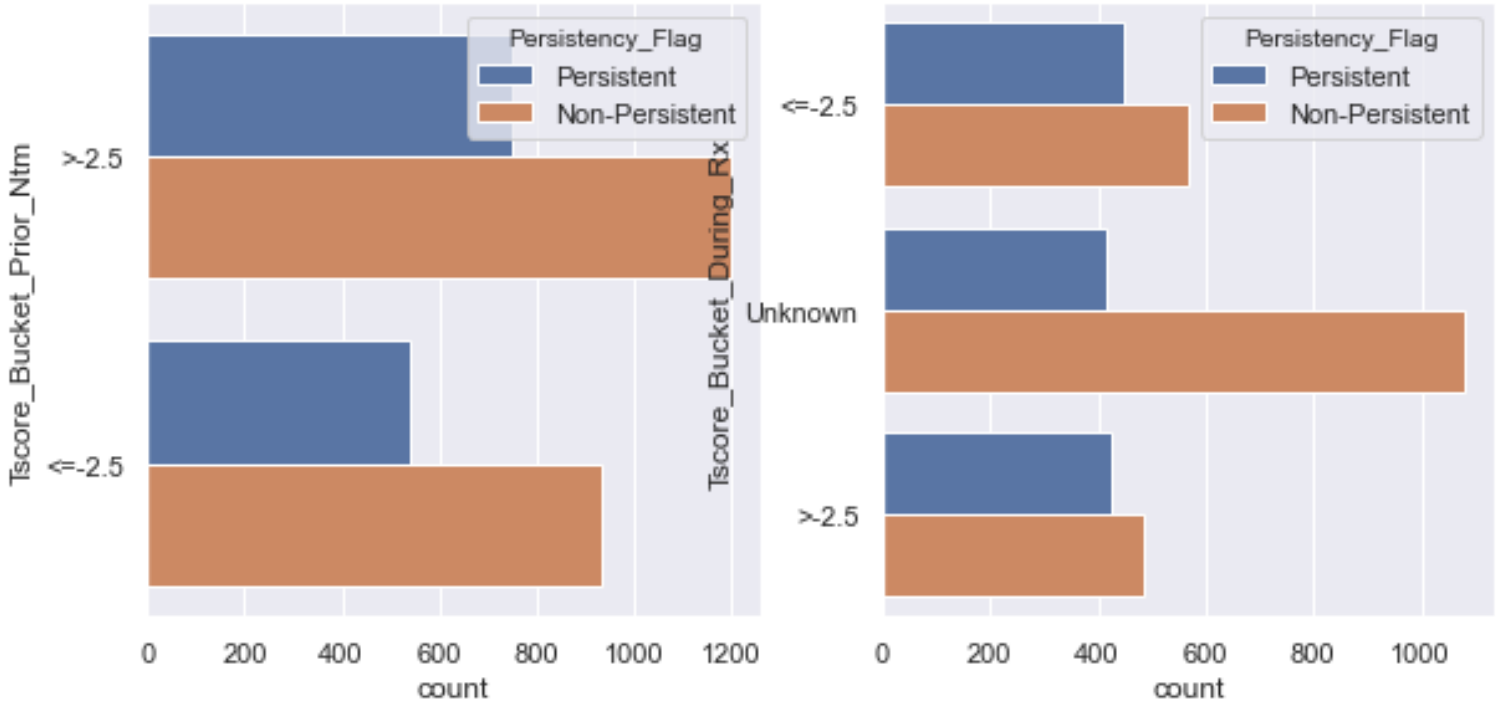


Persistency and Glucocorticoid:

The Glucocorticoid record during Rx showed persistency.

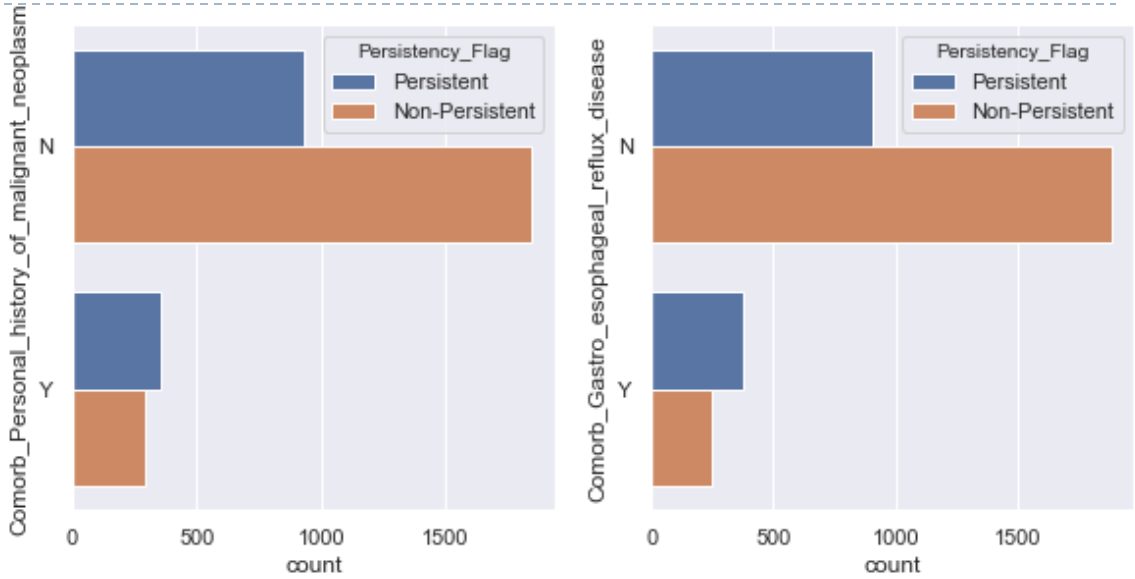


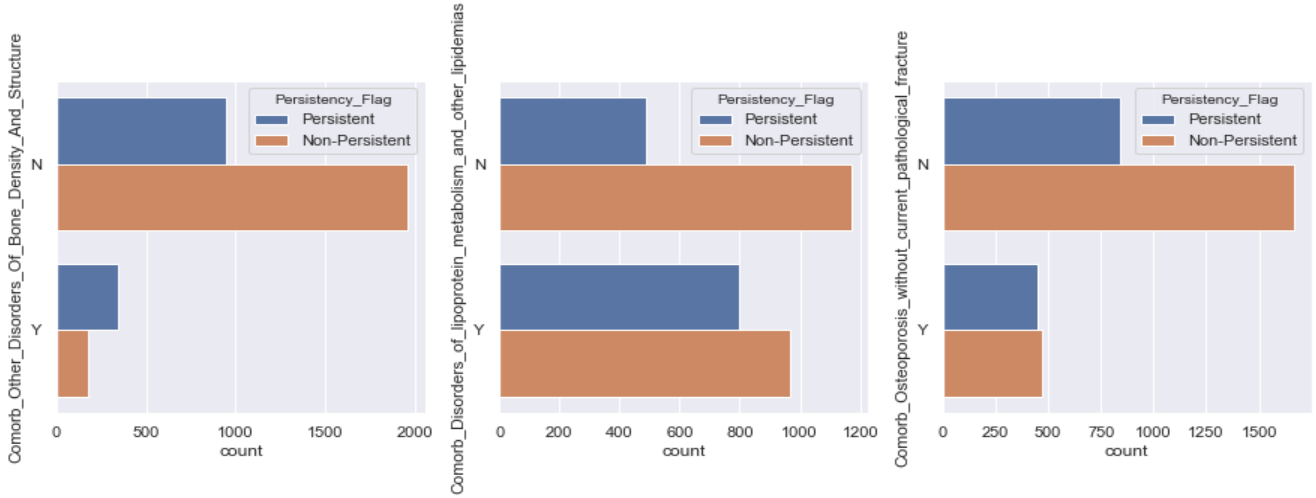
Tscore and Persistency:

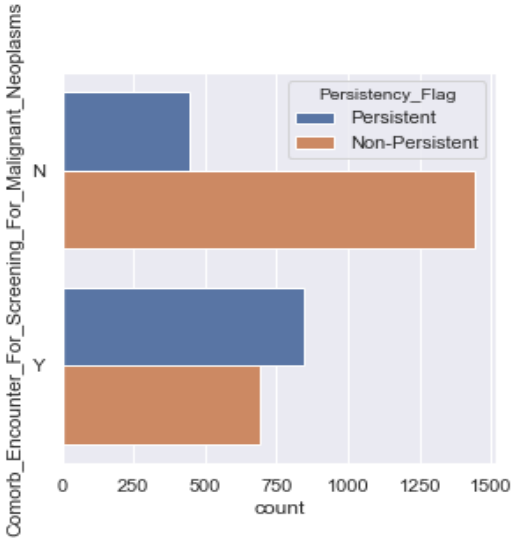


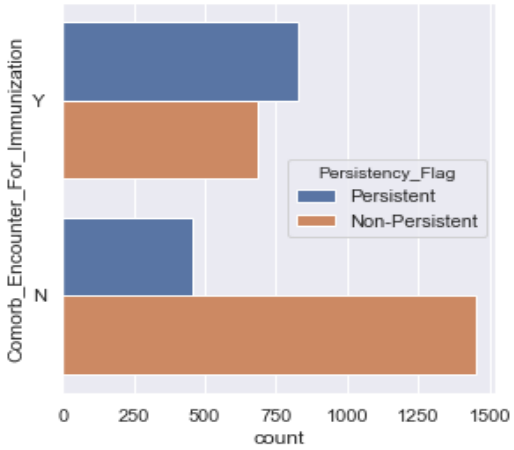
Persistency and Ntm-Comorbidity:

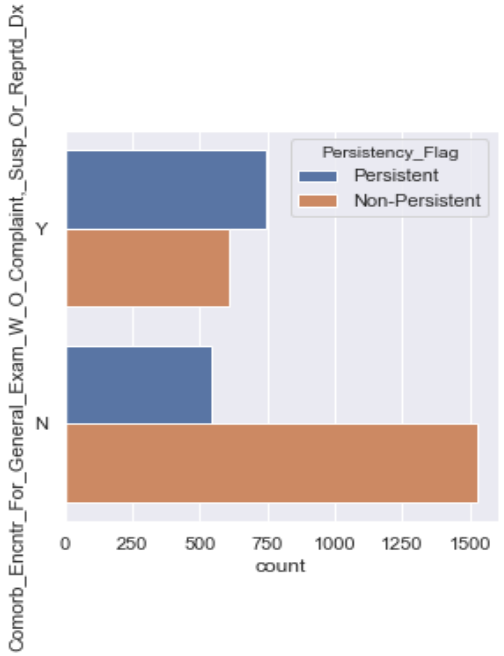
Persistency is examined for different comorbidities and persistency has been found in most people suffering from different diseases.

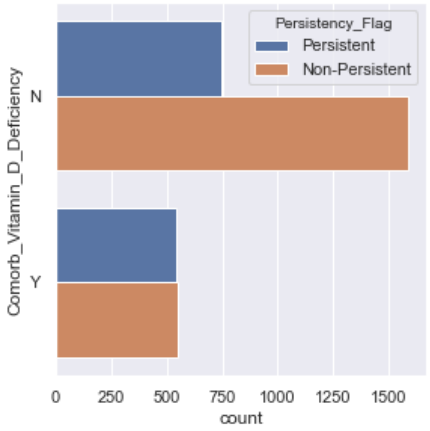


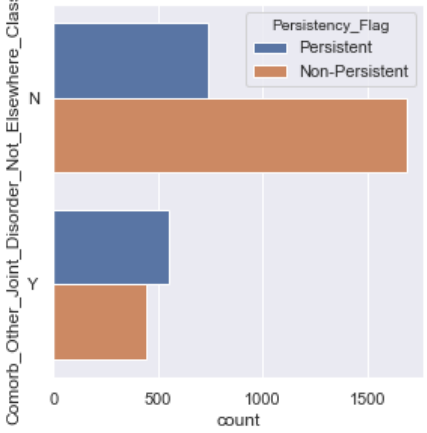


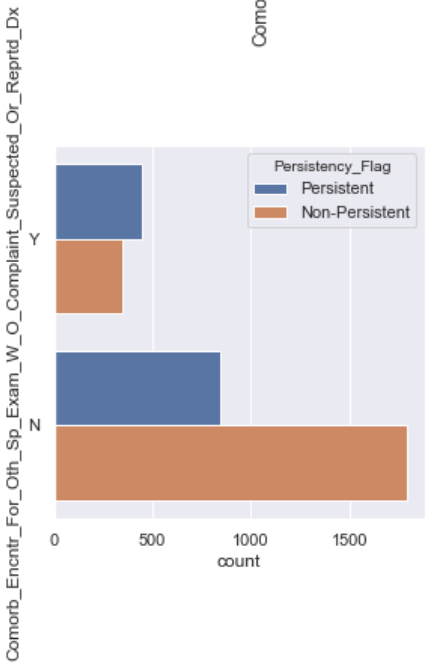

Comorb_Encounter_For_Screening_For_Malignant_Neoplasms
Persistency_Flag
Persistent
Non-Persistent
N
Y
0
250
500
750
1000
1250
1500
count


Comorb_Encounter_For_Immunization
Persistency_Flag
Persistent
Non-Persistent
Y
N
0
250
500
750
1000
1250
1500
count


Comorb_Encntr_For_General_Exam_W_O_Complaint_Susp_Or_Reprtd_Dx
Persistency_Flag
Persistent
Non-Persistent
Y
N
0
250
500
750
1000
1250
1500
count

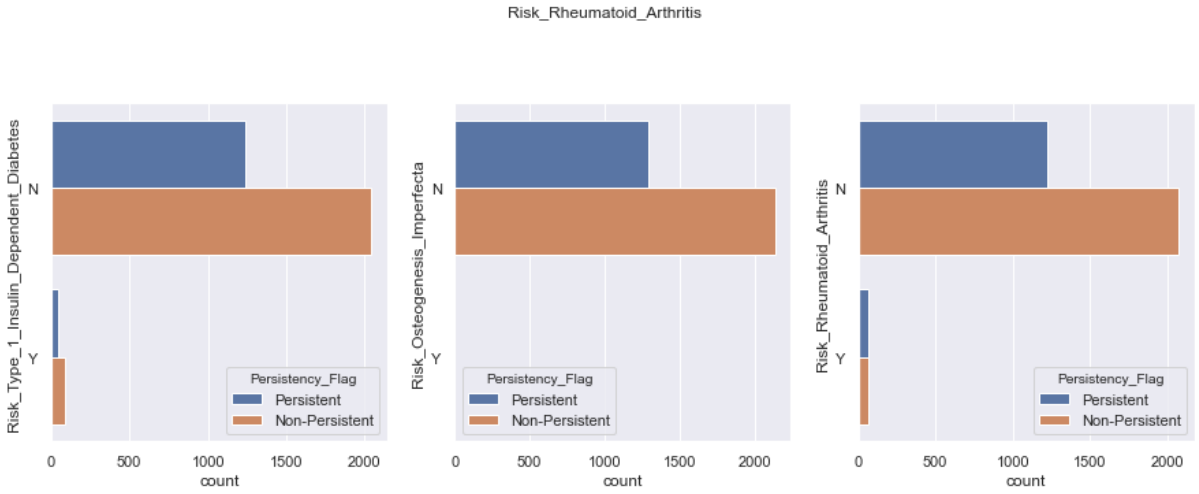

Comorb_Vitamin_D_Deficiency
Persistency_Flag
Persistent
Non-Persistent
N
Y
0
250
500
750
1000
1250
1500
count

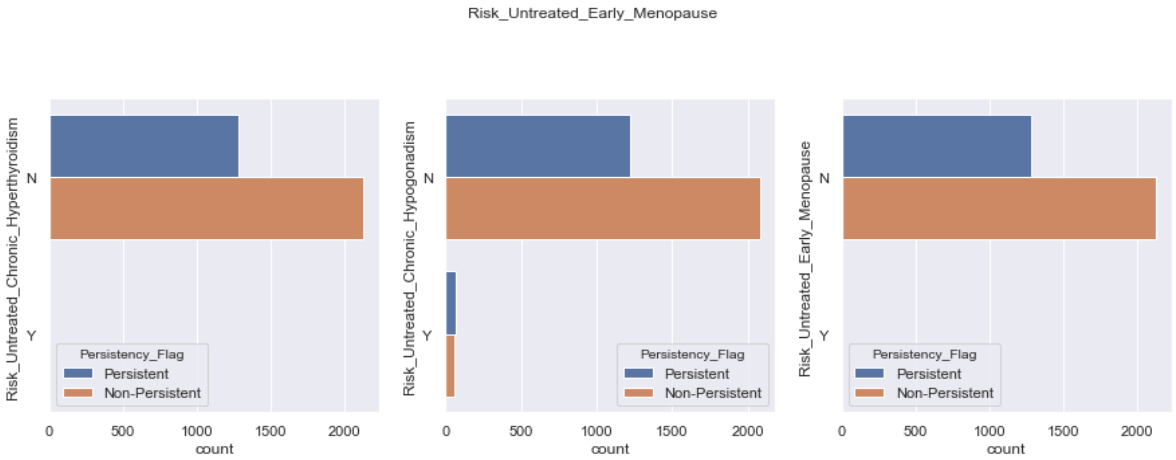

Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified
Persistency_Flag
Persistent
Non-Persistent
N
Y
0
500
1000
1500
count


Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx
Persistency_Flag
Persistent
Non-Persistent
Y
N
0
500
1000
1500
count

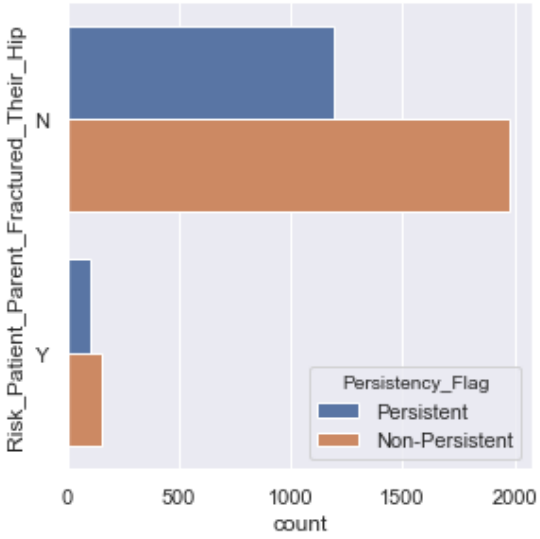
Persistency and Ntm-Risk:

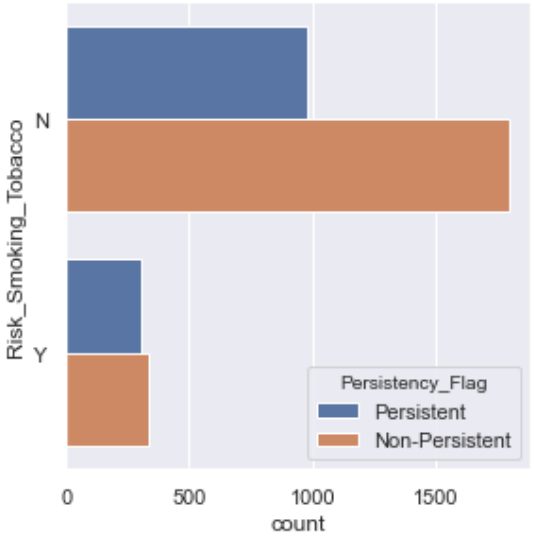
Risk factor has showed less influence on the persistency of people .

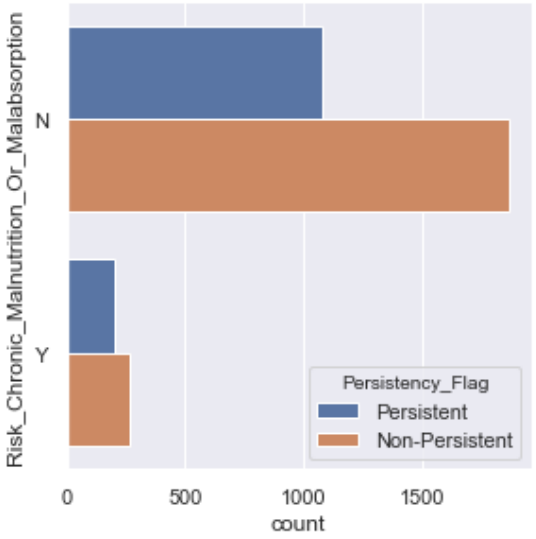




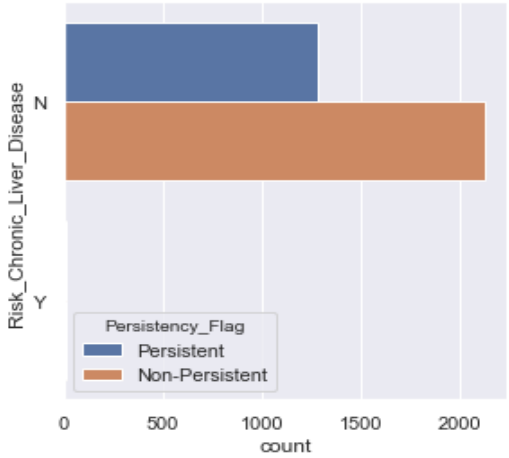
Risk_Chronic_Malnutrition_Or_Malabsorption

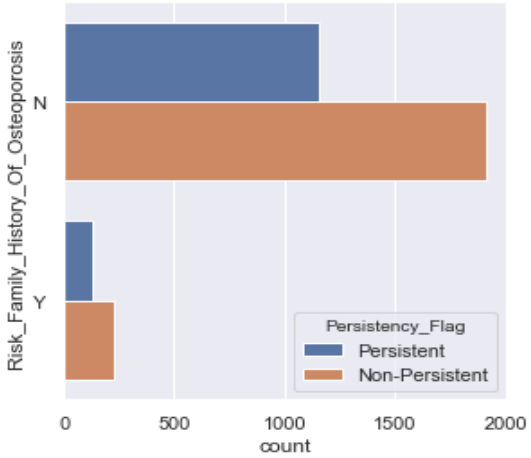


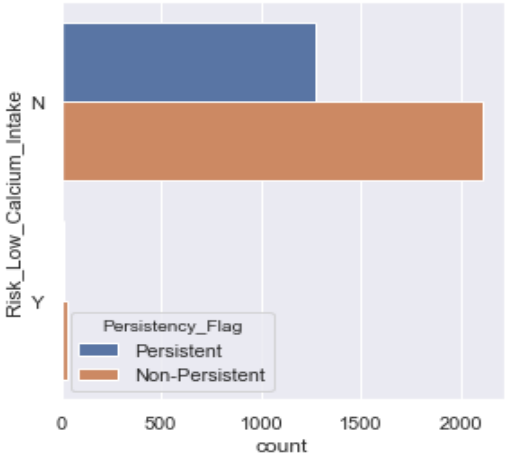




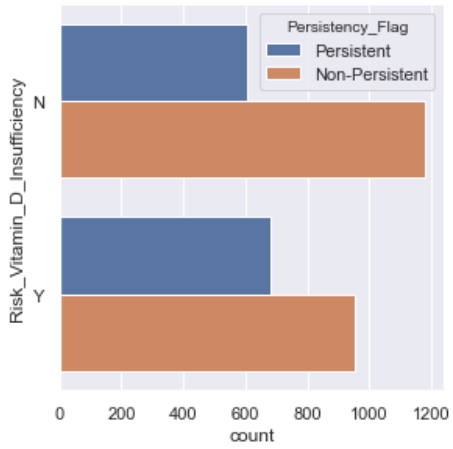
Risk_Low_Calcium_Intake

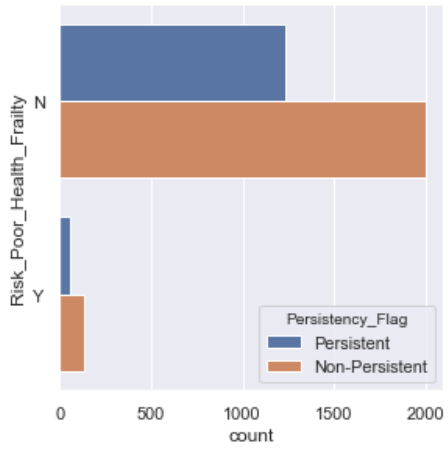


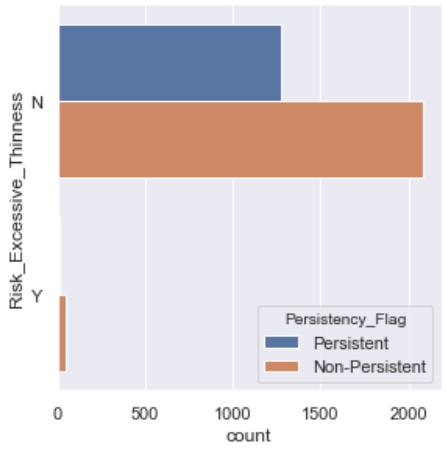




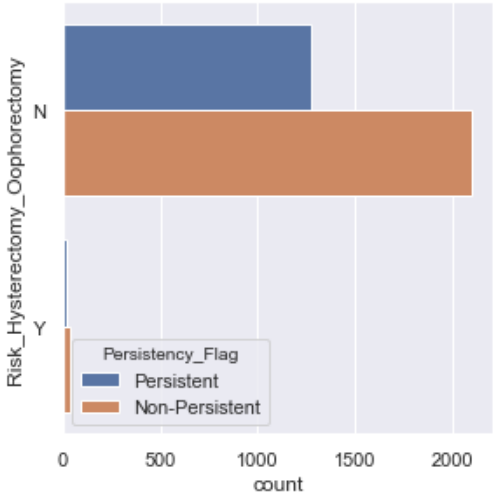
Risk_Excessive_Thinness

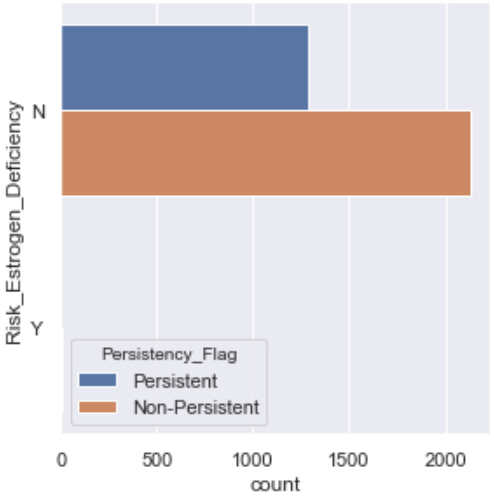


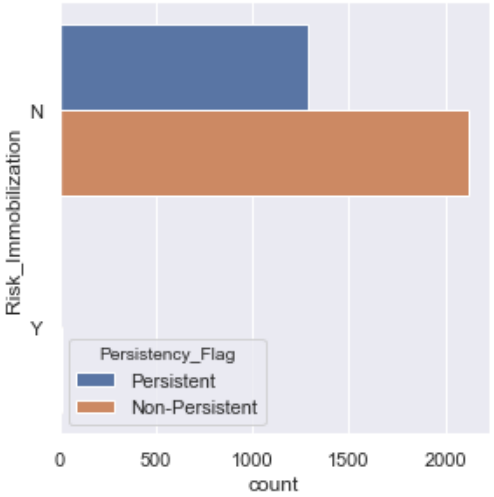


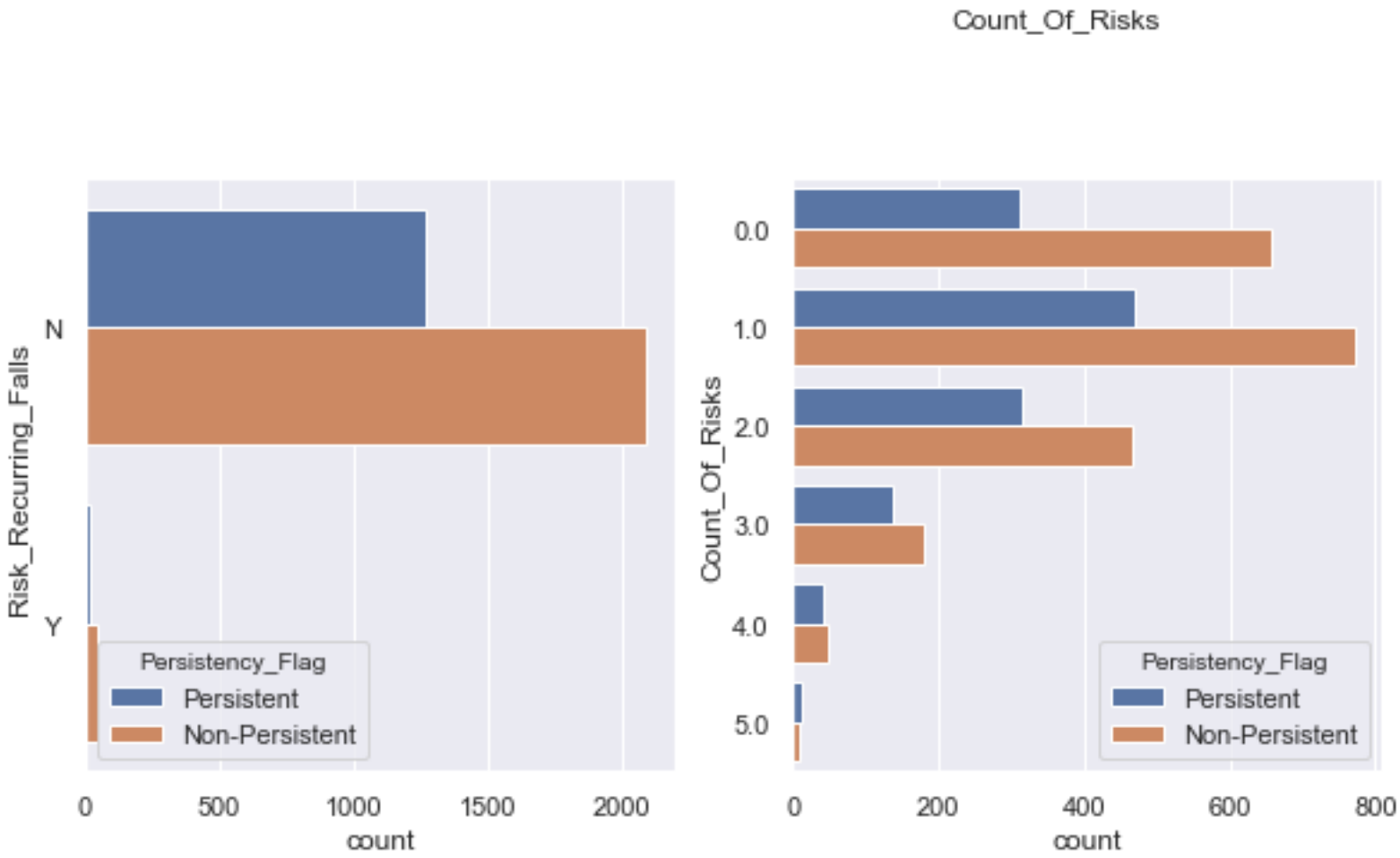


Risk_Immobilization








Count_Of_Risks
Risk_Recurring_Falls
N
Y
Persistency_Flag
Persistent
Non-Persistent
0
500
1000
1500
2000
count
Count_Of_Risks
0.0
1.0
2.0
3.0
4.0
5.0
Persistency_Flag
Persistent
Non-Persistent
0
200
400
600
800
count

Conclusion and Future Recommendations:

- Many factors are found to be related to persistency and also people from different regions has different behavior related to persistency that depends on the type of specialty.

- The number of people included in dataset varies and also the specialty that is found to have different behavior in regions .

- Modeling techniques can also be used and possible algorithms for this study are presented in next slide.

Modeling Techniques:

- Considering the nature of target variable the classification modeling techniques are most suitable for present study. This is a problem of binary classification and models logistic regression , decision tree can be used easily.

- Ensemble modeling techniques like random forest, Ada boost in order to increase the prediction power can also be used.

- These modeling techniques can help us to identify the algorithm that can predict persistency using the behavior of features.