Week 10 EDA

August 9, 2022

1 Drug Persistency and Medical Adherence

Team Name: Feature Transformers Team members: - Name: Wangu Ndungu - Email: nwangu349@gmail.com - Country: Kenya - College/Company: Kenyatta University - Specialization: Data Science

- Name: Nikola Andrejić
- Email: nikola.ing.nl@gmail.com
- Country: Serbia
- College/Company: University of Niš
- Specialization: Data Science

1.1 PROBLEM STATEMENT

According to the World Health Organisation, only 50-70% of patients adhere properly to prescribed drugs during therapy. This is especially true among those with long term medication. This worrying statistic is caused by various factors, for example: patient's condition or disease, their socio-economic status, confusion by the schedule, forgetting, discontinuing because they feel better, just to name a few. Medical non-adherence can lead to devastating consequences on one's health, especially those with chronic illnesses. The purpose of this project is to study trends among patients in a sample and build a model that'll classify a new patient as Persistent or Non-Persistent. This project will givemedical practitioners(especially pharmaceuticals) insight on which patients might require more rigorous follow-ups to ensure they will adhere to their prescriptions.

1.1.1 Importing the required libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, stats
%matplotlib inline
```

1.1.2 Importing the data

[2]:
```
df = pd.read_excel('C:/Users/user/Drug percistency/Healthcare_dataset.xlsx',␣
↪sheet_name='Dataset')
df.head(5)
```

[2]:

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region
0	P1	Persistent	Male	Caucasian	Not Hispanic	West
1	P2	Non-Persistent	Male	Asian	Not Hispanic	West
2	P3	Non-Persistent	Female	Other/Unknown	Hispanic	Midwest
3	P4	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest
4	P5	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest

	Age_Bucket	Ntm_Speciality	Ntm_Specialist_Flag
0	>75	GENERAL PRACTITIONER	Others
1	55-65	GENERAL PRACTITIONER	Others
2	65-75	GENERAL PRACTITIONER	Others
3	>75	GENERAL PRACTITIONER	Others
4	>75	GENERAL PRACTITIONER	Others

	Ntm_Speciality_Bucket	…	Risk_Family_History_Of_Osteoporosis
0	OB/GYN/Others/PCP/Unknown	…	N
1	OB/GYN/Others/PCP/Unknown	…	N
2	OB/GYN/Others/PCP/Unknown	…	N
3	OB/GYN/Others/PCP/Unknown	…	N
4	OB/GYN/Others/PCP/Unknown	…	N

	Risk_Low_Calcium_Intake	Risk_Vitamin_D_Insufficiency
0	N	N
1	N	N
2	Y	N
3	N	N
4	N	N

	Risk_Poor_Health_Frailty	Risk_Excessive_Thinness
0	N	N
1	N	N
2	N	N
3	N	N
4	N	N

	Risk_Hysterectomy_Oophorectomy	Risk_Estrogen_Deficiency	Risk_Immobilization
0	N	N	N
1	N	N	N
2	N	N	N
3	N	N	N
4	N	N	N

	Risk_Recurring_Falls	Count_Of_Risks
0	N	0
1	N	0
2	N	2
3	N	1

```
4                    N              1

[5 rows x 69 columns]
```

1.1.3 Data attributes

[3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3424 entries, 0 to 3423
Data columns (total 69 columns):
 #   Column                                                    Non-
Null Count  Dtype
---  ------
--------------  -----
 0   Ptid                                                      3424
non-null   object
 1   Persistency_Flag                                          3424
non-null   object
 2   Gender                                                    3424
non-null   object
 3   Race                                                      3424
non-null   object
 4   Ethnicity                                                 3424
non-null   object
 5   Region                                                    3424
non-null   object
 6   Age_Bucket                                                3424
non-null   object
 7   Ntm_Speciality                                            3424
non-null   object
 8   Ntm_Specialist_Flag                                       3424
non-null   object
 9   Ntm_Speciality_Bucket                                     3424
non-null   object
 10  Gluco_Record_Prior_Ntm                                    3424
non-null   object
 11  Gluco_Record_During_Rx                                    3424
non-null   object
 12  Dexa_Freq_During_Rx                                       3424
non-null   int64
 13  Dexa_During_Rx                                            3424
non-null   object
 14  Frag_Frac_Prior_Ntm                                       3424
non-null   object
 15  Frag_Frac_During_Rx                                       3424
non-null   object
 16  Risk_Segment_Prior_Ntm                                    3424
```

```
non-null   object
 17  Tscore_Bucket_Prior_Ntm                                    3424
non-null   object
 18  Risk_Segment_During_Rx                                     3424
non-null   object
 19  Tscore_Bucket_During_Rx                                    3424
non-null   object
 20  Change_T_Score                                             3424
non-null   object
 21  Change_Risk_Segment                                        3424
non-null   object
 22  Adherent_Flag                                              3424
non-null   object
 23  Idn_Indicator                                              3424
non-null   object
 24  Injectable_Experience_During_Rx                            3424
non-null   object
 25  Comorb_Encounter_For_Screening_For_Malignant_Neoplasms     3424
non-null   object
 26  Comorb_Encounter_For_Immunization                          3424
non-null   object
 27  Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx  3424
non-null   object
 28  Comorb_Vitamin_D_Deficiency                                3424
non-null   object
 29  Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified       3424
non-null   object
 30  Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx  3424
non-null   object
 31  Comorb_Long_Term_Current_Drug_Therapy                      3424
non-null   object
 32  Comorb_Dorsalgia                                           3424
non-null   object
 33  Comorb_Personal_History_Of_Other_Diseases_And_Conditions   3424
non-null   object
 34  Comorb_Other_Disorders_Of_Bone_Density_And_Structure       3424
non-null   object
 35  Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias  3424
non-null   object
 36  Comorb_Osteoporosis_without_current_pathological_fracture  3424
non-null   object
 37  Comorb_Personal_history_of_malignant_neoplasm              3424
non-null   object
 38  Comorb_Gastro_esophageal_reflux_disease                    3424
non-null   object
 39  Concom_Cholesterol_And_Triglyceride_Regulating_Preparations  3424
non-null   object
 40  Concom_Narcotics                                           3424
```

```
non-null   object
 41  Concom_Systemic_Corticosteroids_Plain                 3424
non-null   object
 42  Concom_Anti_Depressants_And_Mood_Stabilisers          3424
non-null   object
 43  Concom_Fluoroquinolones                               3424
non-null   object
 44  Concom_Cephalosporins                                 3424
non-null   object
 45  Concom_Macrolides_And_Similar_Types                   3424
non-null   object
 46  Concom_Broad_Spectrum_Penicillins                     3424
non-null   object
 47  Concom_Anaesthetics_General                           3424
non-null   object
 48  Concom_Viral_Vaccines                                 3424
non-null   object
 49  Risk_Type_1_Insulin_Dependent_Diabetes                3424
non-null   object
 50  Risk_Osteogenesis_Imperfecta                          3424
non-null   object
 51  Risk_Rheumatoid_Arthritis                             3424
non-null   object
 52  Risk_Untreated_Chronic_Hyperthyroidism                3424
non-null   object
 53  Risk_Untreated_Chronic_Hypogonadism                   3424
non-null   object
 54  Risk_Untreated_Early_Menopause                        3424
non-null   object
 55  Risk_Patient_Parent_Fractured_Their_Hip               3424
non-null   object
 56  Risk_Smoking_Tobacco                                  3424
non-null   object
 57  Risk_Chronic_Malnutrition_Or_Malabsorption            3424
non-null   object
 58  Risk_Chronic_Liver_Disease                            3424
non-null   object
 59  Risk_Family_History_Of_Osteoporosis                   3424
non-null   object
 60  Risk_Low_Calcium_Intake                               3424
non-null   object
 61  Risk_Vitamin_D_Insufficiency                          3424
non-null   object
 62  Risk_Poor_Health_Frailty                              3424
non-null   object
 63  Risk_Excessive_Thinness                               3424
non-null   object
 64  Risk_Hysterectomy_Oophorectomy                        3424
```

```
non-null   object
 65  Risk_Estrogen_Deficiency                                    3424
non-null   object
 66  Risk_Immobilization                                         3424
non-null   object
 67  Risk_Recurring_Falls                                        3424
non-null   object
 68  Count_Of_Risks                                              3424
non-null   int64
dtypes: int64(2), object(67)
memory usage: 1.8+ MB
```

This dataset has a total of 68 variables, Percistency_Flag being our target variable. 67 of these attributes are categorical and 2 of the are continuous.

1.1.4 Checking for missing data

[4]:
```
missing_data = df.isnull()
missing_data.head(5)
```

[4]:

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region	Age_Bucket	\
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	

	Ntm_Speciality	Ntm_Specialist_Flag	Ntm_Speciality_Bucket	…	\
0	False	False	False	…	
1	False	False	False	…	
2	False	False	False	…	
3	False	False	False	…	
4	False	False	False	…	

	Risk_Family_History_Of_Osteoporosis	Risk_Low_Calcium_Intake	\
0	False	False	
1	False	False	
2	False	False	
3	False	False	
4	False	False	

	Risk_Vitamin_D_Insufficiency	Risk_Poor_Health_Frailty	\
0	False	False	
1	False	False	
2	False	False	
3	False	False	
4	False	False	

```
   Risk_Excessive_Thinness  Risk_Hysterectomy_Oophorectomy  \
0                    False                           False
1                    False                           False
2                    False                           False
3                    False                           False
4                    False                           False

   Risk_Estrogen_Deficiency  Risk_Immobilization  Risk_Recurring_Falls  \
0                     False                False                 False
1                     False                False                 False
2                     False                False                 False
3                     False                False                 False
4                     False                False                 False

   Count_Of_Risks
0           False
1           False
2           False
3           False
4           False

[5 rows x 69 columns]
```

[5]:
```
for column in missing_data.columns.values.tolist():
    print(column)
    print (missing_data[column].value_counts())
    print("")
```

```
Ptid
False    3424
Name: Ptid, dtype: int64

Persistency_Flag
False    3424
Name: Persistency_Flag, dtype: int64

Gender
False    3424
Name: Gender, dtype: int64

Race
False    3424
Name: Race, dtype: int64

Ethnicity
False    3424
Name: Ethnicity, dtype: int64
```

```
Region
False    3424
Name: Region, dtype: int64

Age_Bucket
False    3424
Name: Age_Bucket, dtype: int64

Ntm_Speciality
False    3424
Name: Ntm_Speciality, dtype: int64

Ntm_Specialist_Flag
False    3424
Name: Ntm_Specialist_Flag, dtype: int64

Ntm_Speciality_Bucket
False    3424
Name: Ntm_Speciality_Bucket, dtype: int64

Gluco_Record_Prior_Ntm
False    3424
Name: Gluco_Record_Prior_Ntm, dtype: int64

Gluco_Record_During_Rx
False    3424
Name: Gluco_Record_During_Rx, dtype: int64

Dexa_Freq_During_Rx
False    3424
Name: Dexa_Freq_During_Rx, dtype: int64

Dexa_During_Rx
False    3424
Name: Dexa_During_Rx, dtype: int64

Frag_Frac_Prior_Ntm
False    3424
Name: Frag_Frac_Prior_Ntm, dtype: int64

Frag_Frac_During_Rx
False    3424
Name: Frag_Frac_During_Rx, dtype: int64

Risk_Segment_Prior_Ntm
False    3424
Name: Risk_Segment_Prior_Ntm, dtype: int64
```

```
Tscore_Bucket_Prior_Ntm
False    3424
Name: Tscore_Bucket_Prior_Ntm, dtype: int64

Risk_Segment_During_Rx
False    3424
Name: Risk_Segment_During_Rx, dtype: int64

Tscore_Bucket_During_Rx
False    3424
Name: Tscore_Bucket_During_Rx, dtype: int64

Change_T_Score
False    3424
Name: Change_T_Score, dtype: int64

Change_Risk_Segment
False    3424
Name: Change_Risk_Segment, dtype: int64

Adherent_Flag
False    3424
Name: Adherent_Flag, dtype: int64

Idn_Indicator
False    3424
Name: Idn_Indicator, dtype: int64

Injectable_Experience_During_Rx
False    3424
Name: Injectable_Experience_During_Rx, dtype: int64

Comorb_Encounter_For_Screening_For_Malignant_Neoplasms
False    3424
Name: Comorb_Encounter_For_Screening_For_Malignant_Neoplasms, dtype: int64

Comorb_Encounter_For_Immunization
False    3424
Name: Comorb_Encounter_For_Immunization, dtype: int64

Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx
False    3424
Name: Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx, dtype:
int64

Comorb_Vitamin_D_Deficiency
False    3424
Name: Comorb_Vitamin_D_Deficiency, dtype: int64
```

```
Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified
False    3424
Name: Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified, dtype: int64

Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx
False    3424
Name: Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx, dtype:
int64

Comorb_Long_Term_Current_Drug_Therapy
False    3424
Name: Comorb_Long_Term_Current_Drug_Therapy, dtype: int64

Comorb_Dorsalgia
False    3424
Name: Comorb_Dorsalgia, dtype: int64

Comorb_Personal_History_Of_Other_Diseases_And_Conditions
False    3424
Name: Comorb_Personal_History_Of_Other_Diseases_And_Conditions, dtype: int64

Comorb_Other_Disorders_Of_Bone_Density_And_Structure
False    3424
Name: Comorb_Other_Disorders_Of_Bone_Density_And_Structure, dtype: int64

Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias
False    3424
Name: Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias, dtype:
int64

Comorb_Osteoporosis_without_current_pathological_fracture
False    3424
Name: Comorb_Osteoporosis_without_current_pathological_fracture, dtype: int64

Comorb_Personal_history_of_malignant_neoplasm
False    3424
Name: Comorb_Personal_history_of_malignant_neoplasm, dtype: int64

Comorb_Gastro_esophageal_reflux_disease
False    3424
Name: Comorb_Gastro_esophageal_reflux_disease, dtype: int64

Concom_Cholesterol_And_Triglyceride_Regulating_Preparations
False    3424
Name: Concom_Cholesterol_And_Triglyceride_Regulating_Preparations, dtype: int64

Concom_Narcotics
```

```
False    3424
Name: Concom_Narcotics, dtype: int64

Concom_Systemic_Corticosteroids_Plain
False    3424
Name: Concom_Systemic_Corticosteroids_Plain, dtype: int64

Concom_Anti_Depressants_And_Mood_Stabilisers
False    3424
Name: Concom_Anti_Depressants_And_Mood_Stabilisers, dtype: int64

Concom_Fluoroquinolones
False    3424
Name: Concom_Fluoroquinolones, dtype: int64

Concom_Cephalosporins
False    3424
Name: Concom_Cephalosporins, dtype: int64

Concom_Macrolides_And_Similar_Types
False    3424
Name: Concom_Macrolides_And_Similar_Types, dtype: int64

Concom_Broad_Spectrum_Penicillins
False    3424
Name: Concom_Broad_Spectrum_Penicillins, dtype: int64

Concom_Anaesthetics_General
False    3424
Name: Concom_Anaesthetics_General, dtype: int64

Concom_Viral_Vaccines
False    3424
Name: Concom_Viral_Vaccines, dtype: int64

Risk_Type_1_Insulin_Dependent_Diabetes
False    3424
Name: Risk_Type_1_Insulin_Dependent_Diabetes, dtype: int64

Risk_Osteogenesis_Imperfecta
False    3424
Name: Risk_Osteogenesis_Imperfecta, dtype: int64

Risk_Rheumatoid_Arthritis
False    3424
Name: Risk_Rheumatoid_Arthritis, dtype: int64

Risk_Untreated_Chronic_Hyperthyroidism
```

```
False    3424
Name: Risk_Untreated_Chronic_Hyperthyroidism, dtype: int64

Risk_Untreated_Chronic_Hypogonadism
False    3424
Name: Risk_Untreated_Chronic_Hypogonadism, dtype: int64

Risk_Untreated_Early_Menopause
False    3424
Name: Risk_Untreated_Early_Menopause, dtype: int64

Risk_Patient_Parent_Fractured_Their_Hip
False    3424
Name: Risk_Patient_Parent_Fractured_Their_Hip, dtype: int64

Risk_Smoking_Tobacco
False    3424
Name: Risk_Smoking_Tobacco, dtype: int64

Risk_Chronic_Malnutrition_Or_Malabsorption
False    3424
Name: Risk_Chronic_Malnutrition_Or_Malabsorption, dtype: int64

Risk_Chronic_Liver_Disease
False    3424
Name: Risk_Chronic_Liver_Disease, dtype: int64

Risk_Family_History_Of_Osteoporosis
False    3424
Name: Risk_Family_History_Of_Osteoporosis, dtype: int64

Risk_Low_Calcium_Intake
False    3424
Name: Risk_Low_Calcium_Intake, dtype: int64

Risk_Vitamin_D_Insufficiency
False    3424
Name: Risk_Vitamin_D_Insufficiency, dtype: int64

Risk_Poor_Health_Frailty
False    3424
Name: Risk_Poor_Health_Frailty, dtype: int64

Risk_Excessive_Thinness
False    3424
Name: Risk_Excessive_Thinness, dtype: int64

Risk_Hysterectomy_Oophorectomy
```

```
False    3424
Name: Risk_Hysterectomy_Oophorectomy, dtype: int64

Risk_Estrogen_Deficiency
False    3424
Name: Risk_Estrogen_Deficiency, dtype: int64

Risk_Immobilization
False    3424
Name: Risk_Immobilization, dtype: int64

Risk_Recurring_Falls
False    3424
Name: Risk_Recurring_Falls, dtype: int64

Count_Of_Risks
False    3424
Name: Count_Of_Risks, dtype: int64
```

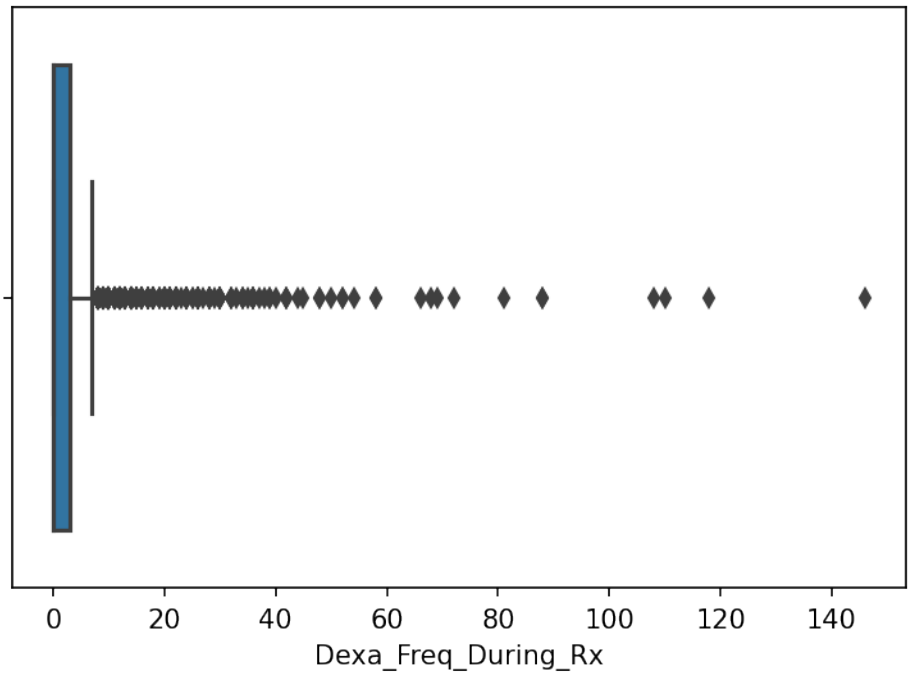
There is no missing data in the dataset

1.1.5 Checking for Outliers

Numerical values Lets visualize the column 'Dexa_Freq_During_Rx' with a box plot

[6]:
```
plt.figure(figsize=(6,4),dpi=150)
sns.boxplot(x=df['Dexa_Freq_During_Rx'])
```

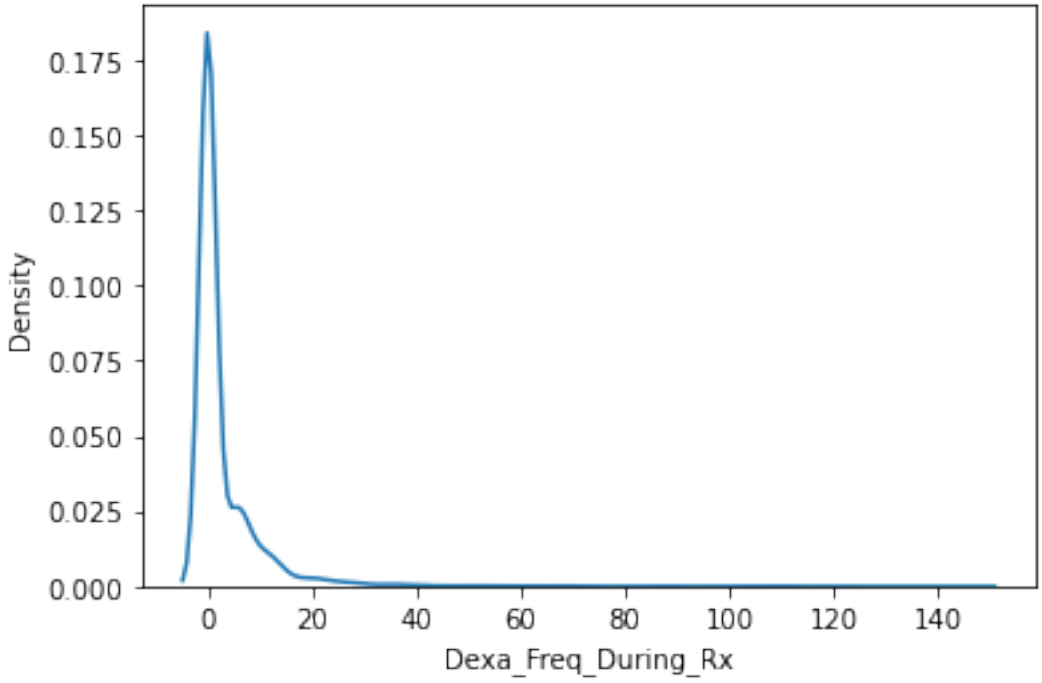
[6]:
```
<AxesSubplot:xlabel='Dexa_Freq_During_Rx'>
```



We can see there's a couple of outliers between 10 and 150

[7]:
```
sns.kdeplot(x=df["Dexa_Freq_During_Rx"])
```

[7]:
```
<AxesSubplot:xlabel='Dexa_Freq_During_Rx', ylabel='Density'>
```



The data is heavily positively skewed. We shall use the .skew() function to find out the exact extent.

[8]:
```
print(skew(df['Dexa_Freq_During_Rx']))
```

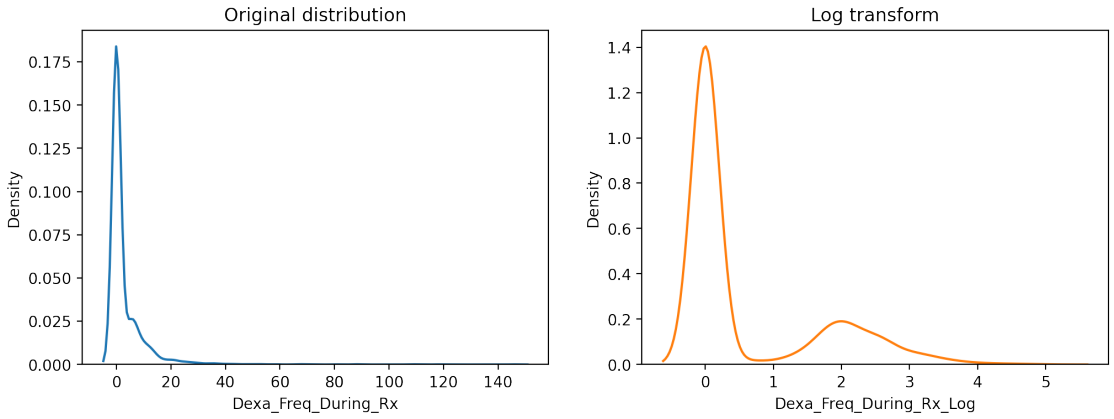
```
6.805747051718919
```

We shall apply log transformation to deal with this and replace the column with the log-transformmed version

[9]:
```
df["Dexa_Freq_During_Rx_Log"] = df['Dexa_Freq_During_Rx'].apply(lambda x: np.
↪log(1+x))
```

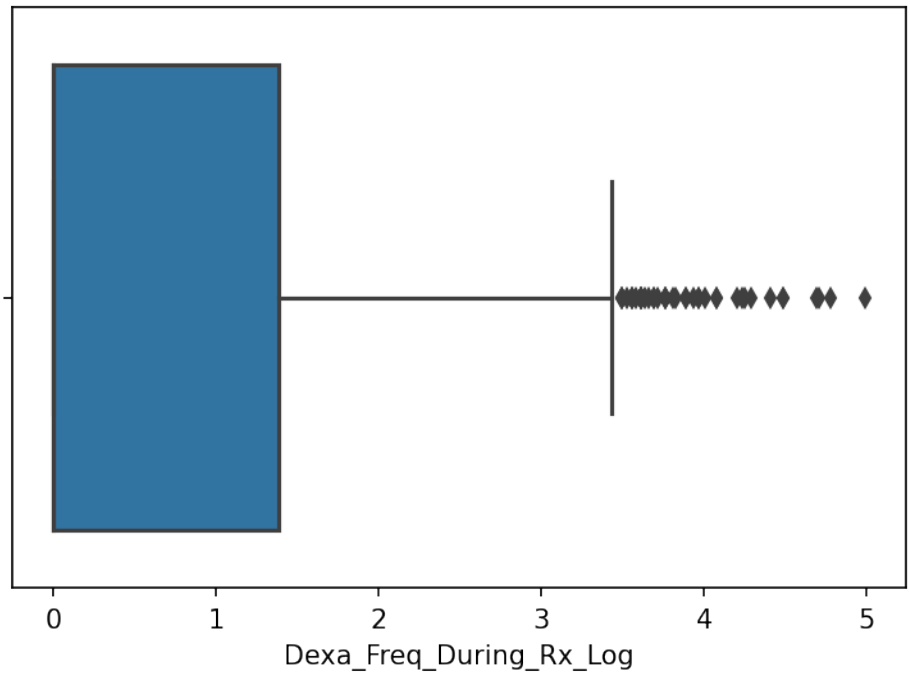
Let's compare these two columns' kde plots side by side

[10]:
```
fig,axes=plt.subplots(nrows=1,ncols=2)
fig.set_size_inches((12,4))
fig.set_dpi(200)
sns.kdeplot(x=df["Dexa_Freq_During_Rx"],ax=axes[0],color="tab:blue")
sns.kdeplot(x=df["Dexa_Freq_During_Rx_Log"],color="tab:orange");
axes[0].set_title("Original distribution")
axes[1].set_title("Log transform");
```



[11]:
```
#Box plot for the log transform data
plt.figure(figsize=(6,4),dpi=150)
sns.boxplot(x=df['Dexa_Freq_During_Rx_Log'])
```

[11]:
```
<AxesSubplot:xlabel='Dexa_Freq_During_Rx_Log'>
```



[12]:
```
print(skew(df['Dexa_Freq_During_Rx_Log']))
```

```
1.4052436284675567
```

The skewness has been greatly improved using log transformation

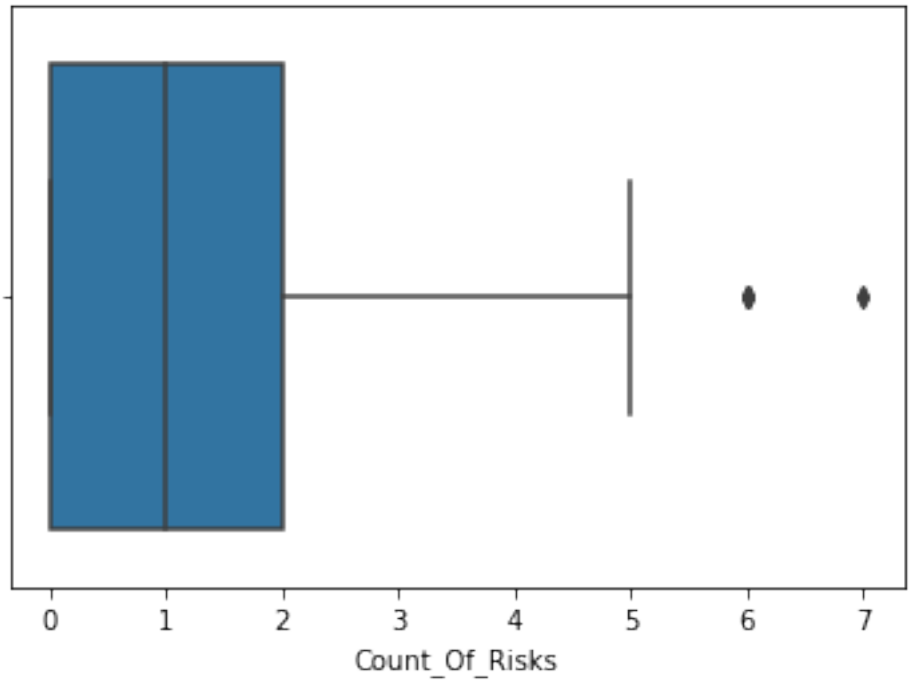
[13]:
```
print(skew(df['Count_Of_Risks']))
```

```
0.8794050541279611
```

A posite skewness is being observed. Lets visualize this column

[14]:
```
sns.boxplot(x=df['Count_Of_Risks'])
```

[14]:
```
<AxesSubplot:xlabel='Count_Of_Risks'>
```



Let's try replacing our outliers with the median

[15]:
```
median = df.loc[df['Count_Of_Risks'] <5, 'Count_Of_Risks'].median()
df.loc[df.Count_Of_Risks > 5, 'Count_Of_Risks'] = np.nan
df.fillna(median,inplace=True)
print(skew(df['Count_Of_Risks']))
```

```
0.7359181096502345
```

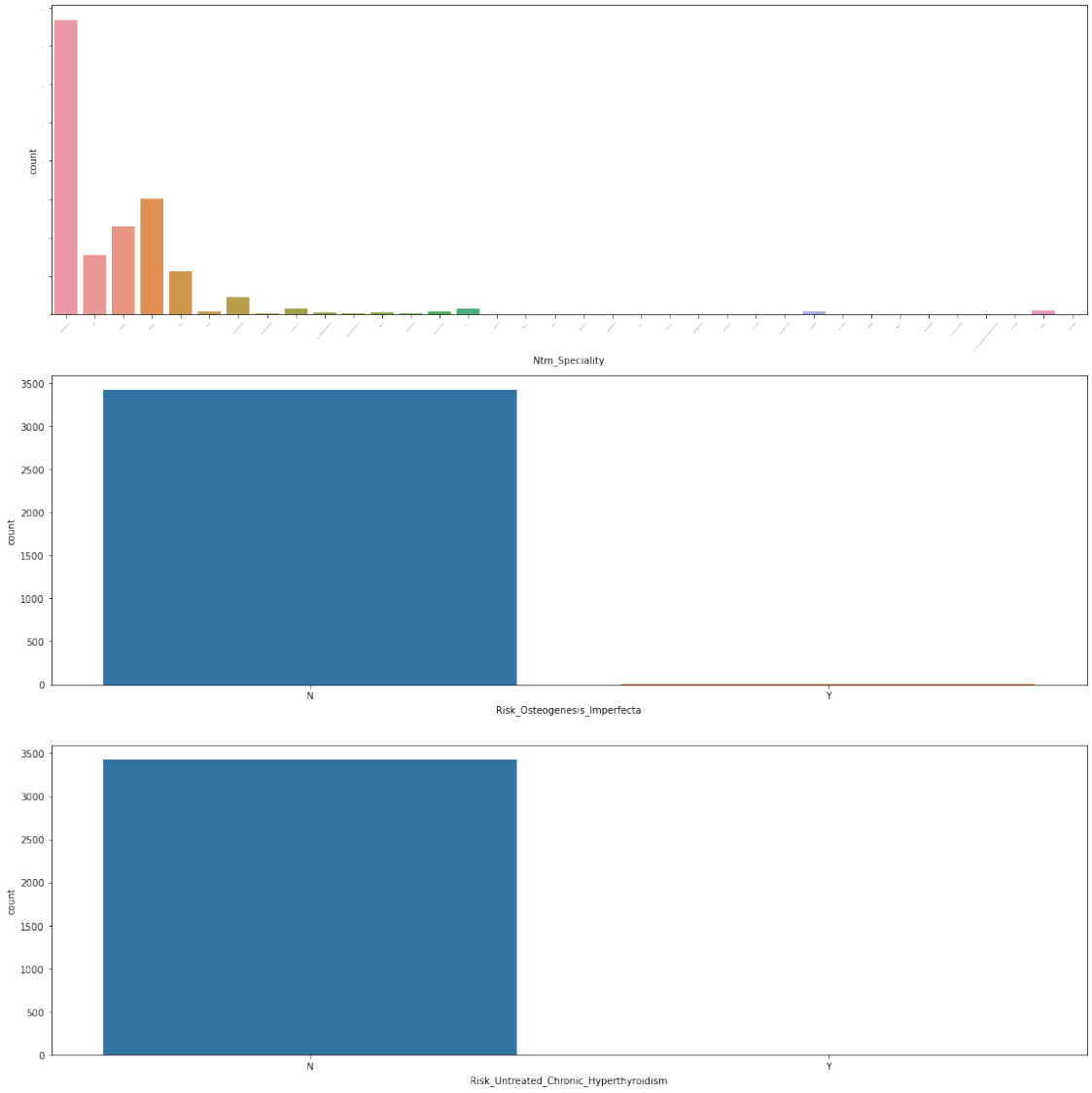
skewness has been reduce a little by replacing the outlier values with the median

Categorical values We will detect outliers by finding categories that have low frequencies with the help of histograms

[16]:
```
#Create a list of categorical columns
cat_cols=df.select_dtypes("object").drop("Ptid",axis=1).columns
#Create a list of categorical columns with outliers
cat_cols_outliers = cat_cols[[any(df[col].value_counts()<=10) for col in␣
 ↪cat_cols]]
```

[17]:
```
#Visualize the imbalance of categorical columns with outliers
fig,axes=plt.subplots(nrows=len(cat_cols_outliers))
fig.set_size_inches((16,4*4))
i=0
for col in cat_cols_outliers:
    sns.countplot(x=df[col],ax=axes[i])
```

```
    i+=1
axes[0].tick_params(rotation=50,labelsize=0)
plt.tight_layout()
```



Now let's retain these categogories and see how it'll affect the model

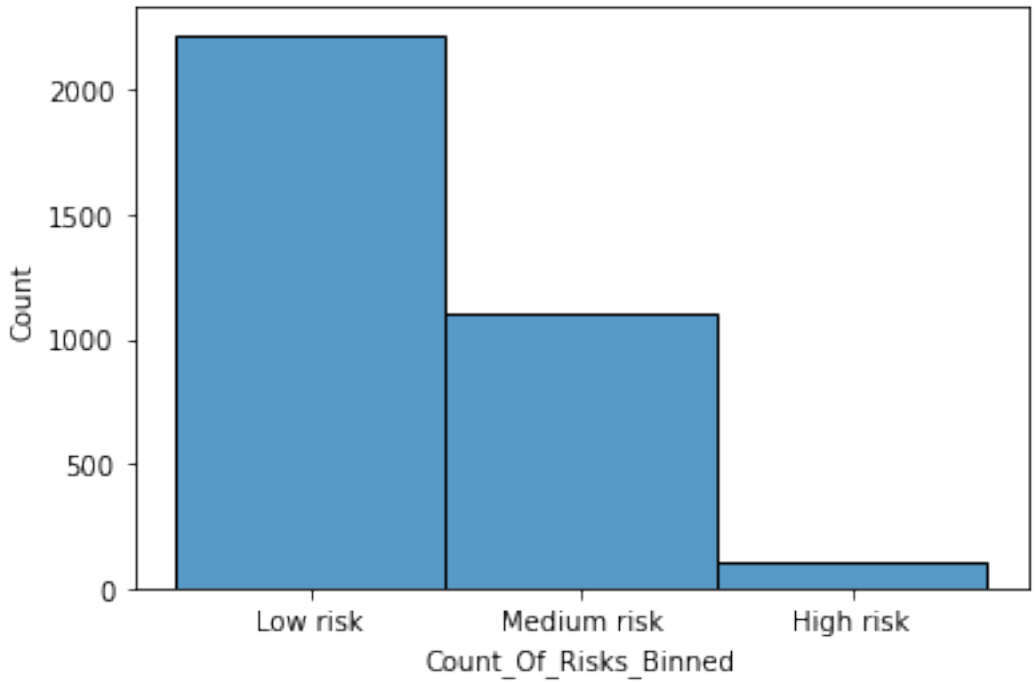
1.2 Feature scaling and transformation

1.2.1 Numerical values

We Have already scaled our 'Dexa_Freq_During_Rx_Log' using log Transformation in order to reduce outliers and minimize skewness. For the attribute "Count_Of_Risks", we will categorize our values in the following bins: Low Risk, Medium Risk and High Risk

[18]:
```
Risk_Bins = np.linspace(min(df['Count_Of_Risks']), max(df['Count_Of_Risks']), 4)
group_names = ["Low risk", "Medium risk", "High risk"]
df['Count_Of_Risks_Binned'] = pd.cut(df['Count_Of_Risks'], Risk_Bins, labels =␣
 ↪group_names, include_lowest = True)
sns.histplot(x=df['Count_Of_Risks_Binned'])
```

[18]:
```
<AxesSubplot:xlabel='Count_Of_Risks_Binned', ylabel='Count'>
```



1.2.2 Categorical values

We have both nominal and ordinal data in our dataset. We will be using ordinal encoding on our ordinal data and frequency encoding on our noiminal data.

[19]:
```
print(cat_cols)
```

```
Index(['Persistency_Flag', 'Gender', 'Race', 'Ethnicity', 'Region',
       'Age_Bucket', 'Ntm_Speciality', 'Ntm_Specialist_Flag',
       'Ntm_Speciality_Bucket', 'Gluco_Record_Prior_Ntm',
       'Gluco_Record_During_Rx', 'Dexa_During_Rx', 'Frag_Frac_Prior_Ntm',
       'Frag_Frac_During_Rx', 'Risk_Segment_Prior_Ntm',
       'Tscore_Bucket_Prior_Ntm', 'Risk_Segment_During_Rx',
       'Tscore_Bucket_During_Rx', 'Change_T_Score', 'Change_Risk_Segment',
       'Adherent_Flag', 'Idn_Indicator', 'Injectable_Experience_During_Rx',
       'Comorb_Encounter_For_Screening_For_Malignant_Neoplasms',
       'Comorb_Encounter_For_Immunization',
```

```
       'Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx',
       'Comorb_Vitamin_D_Deficiency',
       'Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified',
       'Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx',
       'Comorb_Long_Term_Current_Drug_Therapy', 'Comorb_Dorsalgia',
       'Comorb_Personal_History_Of_Other_Diseases_And_Conditions',
       'Comorb_Other_Disorders_Of_Bone_Density_And_Structure',
       'Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias',
       'Comorb_Osteoporosis_without_current_pathological_fracture',
       'Comorb_Personal_history_of_malignant_neoplasm',
       'Comorb_Gastro_esophageal_reflux_disease',
       'Concom_Cholesterol_And_Triglyceride_Regulating_Preparations',
       'Concom_Narcotics', 'Concom_Systemic_Corticosteroids_Plain',
       'Concom_Anti_Depressants_And_Mood_Stabilisers',
       'Concom_Fluoroquinolones', 'Concom_Cephalosporins',
       'Concom_Macrolides_And_Similar_Types',
       'Concom_Broad_Spectrum_Penicillins', 'Concom_Anaesthetics_General',
       'Concom_Viral_Vaccines', 'Risk_Type_1_Insulin_Dependent_Diabetes',
       'Risk_Osteogenesis_Imperfecta', 'Risk_Rheumatoid_Arthritis',
       'Risk_Untreated_Chronic_Hyperthyroidism',
       'Risk_Untreated_Chronic_Hypogonadism', 'Risk_Untreated_Early_Menopause',
       'Risk_Patient_Parent_Fractured_Their_Hip', 'Risk_Smoking_Tobacco',
       'Risk_Chronic_Malnutrition_Or_Malabsorption',
       'Risk_Chronic_Liver_Disease', 'Risk_Family_History_Of_Osteoporosis',
       'Risk_Low_Calcium_Intake', 'Risk_Vitamin_D_Insufficiency',
       'Risk_Poor_Health_Frailty', 'Risk_Excessive_Thinness',
       'Risk_Hysterectomy_Oophorectomy', 'Risk_Estrogen_Deficiency',
       'Risk_Immobilization', 'Risk_Recurring_Falls'],
      dtype='object')
```

After going through the dataset, I have found the following ordinal columns: - 'Age_Bucket' - 'Tscore_Bucket_Prior_Ntm' - 'Tscore_Bucket_During_Rx'

[20]:
```
# Importing ordinal encoder
data_categorical = df[cat_cols]
from sklearn.preprocessing import OrdinalEncoder
Age_column = data_categorical[["Age_Bucket"]]
encoder = OrdinalEncoder()
Age_encoded = encoder.fit_transform(Age_column)
Age_encoded
```

[20]:
```
array([[3.],
       [0.],
       [1.],
       …,
       [3.],
       [0.],
       [1.]])
```

[21]:
```
Tscore_Bucket_Prior_Ntm_column = data_categorical[["Tscore_Bucket_Prior_Ntm"]]
encoder = OrdinalEncoder()
Tscore_Bucket_Prior_Ntm_encoded = encoder.
 ↪fit_transform(Tscore_Bucket_Prior_Ntm_column)
Tscore_Bucket_Prior_Ntm_encoded
```

[21]:
```
array([[1.],
       [1.],
       [0.],
       …,
       [1.],
       [1.],
       [1.]])
```

[22]:
```
Tscore_Bucket_During_Rx_column = data_categorical[["Tscore_Bucket_During_Rx"]]
encoder = OrdinalEncoder()
Tscore_Bucket_During_Rx_encoded = encoder.
 ↪fit_transform(Tscore_Bucket_During_Rx_column)
Tscore_Bucket_During_Rx_encoded
```

[22]:
```
array([[0.],
       [2.],
       [0.],
       …,
       [0.],
       [2.],
       [2.]])
```

[23]:
```
df["Age_encoded"] = Age_encoded
df["Tscore_Bucket_Prior_Ntm_encoded"] = Tscore_Bucket_Prior_Ntm_encoded
df["Tscore_Bucket_During_Rx_encoded"] = Tscore_Bucket_During_Rx_encoded
```

Now for the Nominal colums

[24]:
```
df_nominal = data_categorical.drop(["Tscore_Bucket_During_Rx",␣
 ↪"Tscore_Bucket_Prior_Ntm", "Age_Bucket"], axis = 1)
```

[25]:
```
# Using frequency encoding
for column in df_nominal:
    Freq_enc = (df_nominal.groupby(column).size()) / len(df)
    print(Freq_enc)
    df[column + 'freq_enc'] = df_nominal[column].apply(lambda x : Freq_enc[x])
```

```
Persistency_Flag
Non-Persistent    0.62354
Persistent        0.37646
dtype: float64
Gender
```

```
Female    0.943341
Male      0.056659
dtype: float64
Race
African American    0.027745
Asian               0.024533
Caucasian           0.919393
Other/Unknown       0.028329
dtype: float64
Ethnicity
Hispanic        0.028621
Not Hispanic    0.944801
Unknown         0.026577
dtype: float64
Region
Midwest          0.403914
Northeast        0.067757
Other/Unknown    0.017523
South            0.364194
West             0.146612
dtype: float64
Ntm_Speciality
CARDIOLOGY                                                          0.006425
CLINICAL NURSE SPECIALIST                                           0.000292
EMERGENCY MEDICINE                                                  0.000292
ENDOCRINOLOGY                                                       0.133762
GASTROENTEROLOGY                                                    0.000584
GENERAL PRACTITIONER                                                0.448306
GERIATRIC MEDICINE                                                  0.000584
HEMATOLOGY & ONCOLOGY                                               0.004089
HOSPICE AND PALLIATIVE MEDICINE                                     0.000584
HOSPITAL MEDICINE                                                   0.000292
NEPHROLOGY                                                          0.000876
NEUROLOGY                                                           0.000292
NUCLEAR MEDICINE                                                    0.000292
OBSTETRICS & OBSTETRICS & GYNECOLOGY & OBSTETRICS & GYNECOLOGY      0.000292
OBSTETRICS AND GYNECOLOGY                                           0.026285
OCCUPATIONAL MEDICINE                                               0.000292
ONCOLOGY                                                            0.065713
OPHTHALMOLOGY                                                       0.000292
ORTHOPEDIC SURGERY                                                  0.008762
ORTHOPEDICS                                                         0.000876
OTOLARYNGOLOGY                                                      0.004089
PAIN MEDICINE                                                       0.000292
PATHOLOGY                                                           0.004673
PEDIATRICS                                                          0.003797
PHYSICAL MEDICINE AND REHABILITATION                                0.003213
PLASTIC SURGERY                                                     0.000584
```

```
PODIATRY                                                    0.000292
PSYCHIATRY AND NEUROLOGY                                    0.001168
PULMONARY MEDICINE                                          0.002336
RADIOLOGY                                                   0.000292
RHEUMATOLOGY                                                0.176402
SURGERY AND SURGICAL SPECIALTIES                            0.002336
TRANSPLANT SURGERY                                          0.000584
UROLOGY                                                     0.009638
Unknown                                                     0.090537
VASCULAR SURGERY                                            0.000584
dtype: float64
Ntm_Specialist_Flag
Others        0.587909
Specialist    0.412091
dtype: float64
Ntm_Speciality_Bucket
Endo/Onc/Uro                 0.209112
OB/GYN/Others/PCP/Unknown    0.614486
Rheum                        0.176402
dtype: float64
Gluco_Record_Prior_Ntm
N    0.764895
Y    0.235105
dtype: float64
Gluco_Record_During_Rx
N    0.736565
Y    0.263435
dtype: float64
Dexa_During_Rx
N    0.726636
Y    0.273364
dtype: float64
Frag_Frac_Prior_Ntm
N    0.838785
Y    0.161215
dtype: float64
Frag_Frac_During_Rx
N    0.878213
Y    0.121787
dtype: float64
Risk_Segment_Prior_Ntm
HR_VHR    0.43604
VLR_LR    0.56396
dtype: float64
Risk_Segment_During_Rx
HR_VHR     0.281834
Unknown    0.437208
VLR_LR     0.280958
```

```
dtype: float64
Change_T_Score
Improved     0.027453
No change    0.484813
Unknown      0.437208
Worsened     0.050526
dtype: float64
Change_Risk_Segment
Improved     0.006425
No change    0.307243
Unknown      0.650993
Worsened     0.035339
dtype: float64
Adherent_Flag
Adherent        0.949474
Non-Adherent    0.050526
dtype: float64
Idn_Indicator
N    0.253213
Y    0.746787
dtype: float64
Injectable_Experience_During_Rx
N    0.107477
Y    0.892523
dtype: float64
Comorb_Encounter_For_Screening_For_Malignant_Neoplasms
N    0.552278
Y    0.447722
dtype: float64
Comorb_Encounter_For_Immunization
N    0.558119
Y    0.441881
dtype: float64
Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx
N    0.60514
Y    0.39486
dtype: float64
Comorb_Vitamin_D_Deficiency
N    0.680783
Y    0.319217
dtype: float64
Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified
N    0.708236
Y    0.291764
dtype: float64
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx
N    0.768984
Y    0.231016
```

```
dtype: float64
Comorb_Long_Term_Current_Drug_Therapy
N    0.76139
Y    0.23861
dtype: float64
Comorb_Dorsalgia
N    0.772488
Y    0.227512
dtype: float64
Comorb_Personal_History_Of_Other_Diseases_And_Conditions
N    0.802278
Y    0.197722
dtype: float64
Comorb_Other_Disorders_Of_Bone_Density_And_Structure
N    0.848715
Y    0.151285
dtype: float64
Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias
N    0.484521
Y    0.515479
dtype: float64
Comorb_Osteoporosis_without_current_pathological_fracture
N    0.732185
Y    0.267815
dtype: float64
Comorb_Personal_history_of_malignant_neoplasm
N    0.810456
Y    0.189544
dtype: float64
Comorb_Gastro_esophageal_reflux_disease
N    0.816005
Y    0.183995
dtype: float64
Concom_Cholesterol_And_Triglyceride_Regulating_Preparations
N    0.65479
Y    0.34521
dtype: float64
Concom_Narcotics
N    0.639895
Y    0.360105
dtype: float64
Concom_Systemic_Corticosteroids_Plain
N    0.715829
Y    0.284171
dtype: float64
Concom_Anti_Depressants_And_Mood_Stabilisers
N    0.719918
Y    0.280082
```

```
dtype: float64
Concom_Fluoroquinolones
N    0.81396
Y    0.18604
dtype: float64
Concom_Cephalosporins
N    0.82389
Y    0.17611
dtype: float64
Concom_Macrolides_And_Similar_Types
N    0.833236
Y    0.166764
dtype: float64
Concom_Broad_Spectrum_Penicillins
N    0.871787
Y    0.128213
dtype: float64
Concom_Anaesthetics_General
N    0.854848
Y    0.145152
dtype: float64
Concom_Viral_Vaccines
N    0.896904
Y    0.103096
dtype: float64
Risk_Type_1_Insulin_Dependent_Diabetes
N    0.959404
Y    0.040596
dtype: float64
Risk_Osteogenesis_Imperfecta
N    0.999124
Y    0.000876
dtype: float64
Risk_Rheumatoid_Arthritis
N    0.962033
Y    0.037967
dtype: float64
Risk_Untreated_Chronic_Hyperthyroidism
N    0.999416
Y    0.000584
dtype: float64
Risk_Untreated_Chronic_Hypogonadism
N    0.962909
Y    0.037091
dtype: float64
Risk_Untreated_Early_Menopause
N    0.996495
Y    0.003505
```

```
dtype: float64
Risk_Patient_Parent_Fractured_Their_Hip
N    0.925234
Y    0.074766
dtype: float64
Risk_Smoking_Tobacco
N    0.811916
Y    0.188084
dtype: float64
Risk_Chronic_Malnutrition_Or_Malabsorption
N    0.862734
Y    0.137266
dtype: float64
Risk_Chronic_Liver_Disease
N    0.994743
Y    0.005257
dtype: float64
Risk_Family_History_Of_Osteoporosis
N    0.895444
Y    0.104556
dtype: float64
Risk_Low_Calcium_Intake
N    0.987734
Y    0.012266
dtype: float64
Risk_Vitamin_D_Insufficiency
N    0.522196
Y    0.477804
dtype: float64
Risk_Poor_Health_Frailty
N    0.943925
Y    0.056075
dtype: float64
Risk_Excessive_Thinness
N    0.980432
Y    0.019568
dtype: float64
Risk_Hysterectomy_Oophorectomy
N    0.984229
Y    0.015771
dtype: float64
Risk_Estrogen_Deficiency
N    0.996787
Y    0.003213
dtype: float64
Risk_Immobilization
N    0.995911
Y    0.004089
```

```
dtype: float64
Risk_Recurring_Falls
N    0.979848
Y    0.020152
dtype: float64
```

[26]:
```
# dropping dulicate rows
df = df.drop_duplicates()
```

[27]:
```
df.head(5)
```

[27]:
```
  Ptid Persistency_Flag  Gender           Race     Ethnicity   Region  \
0   P1       Persistent    Male      Caucasian  Not Hispanic     West
1   P2   Non-Persistent    Male          Asian  Not Hispanic     West
2   P3   Non-Persistent  Female  Other/Unknown      Hispanic  Midwest
3   P4   Non-Persistent  Female      Caucasian  Not Hispanic  Midwest
4   P5   Non-Persistent  Female      Caucasian  Not Hispanic  Midwest

  Age_Bucket        Ntm_Speciality Ntm_Specialist_Flag  \
0        >75  GENERAL PRACTITIONER              Others
1      55-65  GENERAL PRACTITIONER              Others
2      65-75  GENERAL PRACTITIONER              Others
3        >75  GENERAL PRACTITIONER              Others
4        >75  GENERAL PRACTITIONER              Others

       Ntm_Speciality_Bucket  … Risk_Chronic_Liver_Diseasefreq_enc  \
0  OB/GYN/Others/PCP/Unknown  …                           0.994743
1  OB/GYN/Others/PCP/Unknown  …                           0.994743
2  OB/GYN/Others/PCP/Unknown  …                           0.994743
3  OB/GYN/Others/PCP/Unknown  …                           0.994743
4  OB/GYN/Others/PCP/Unknown  …                           0.994743

  Risk_Family_History_Of_Osteoporosisfreq_enc  \
0                                    0.895444
1                                    0.895444
2                                    0.895444
3                                    0.895444
4                                    0.895444

   Risk_Low_Calcium_Intakefreq_enc Risk_Vitamin_D_Insufficiencyfreq_enc  \
0                         0.987734                              0.522196
1                         0.987734                              0.522196
2                         0.012266                              0.522196
3                         0.987734                              0.522196
4                         0.987734                              0.522196

  Risk_Poor_Health_Frailtyfreq_enc Risk_Excessive_Thinnessfreq_enc  \
```

```
0                          0.943925                          0.980432
1                          0.943925                          0.980432
2                          0.943925                          0.980432
3                          0.943925                          0.980432
4                          0.943925                          0.980432

  Risk_Hysterectomy_Oophorectomyfreq_enc Risk_Estrogen_Deficiencyfreq_enc  \
0                                0.984229                          0.996787
1                                0.984229                          0.996787
2                                0.984229                          0.996787
3                                0.984229                          0.996787
4                                0.984229                          0.996787

  Risk_Immobilizationfreq_enc Risk_Recurring_Fallsfreq_enc
0                    0.995911                     0.979848
1                    0.995911                     0.979848
2                    0.995911                     0.979848
3                    0.995911                     0.979848
4                    0.995911                     0.979848

[5 rows x 137 columns]
```

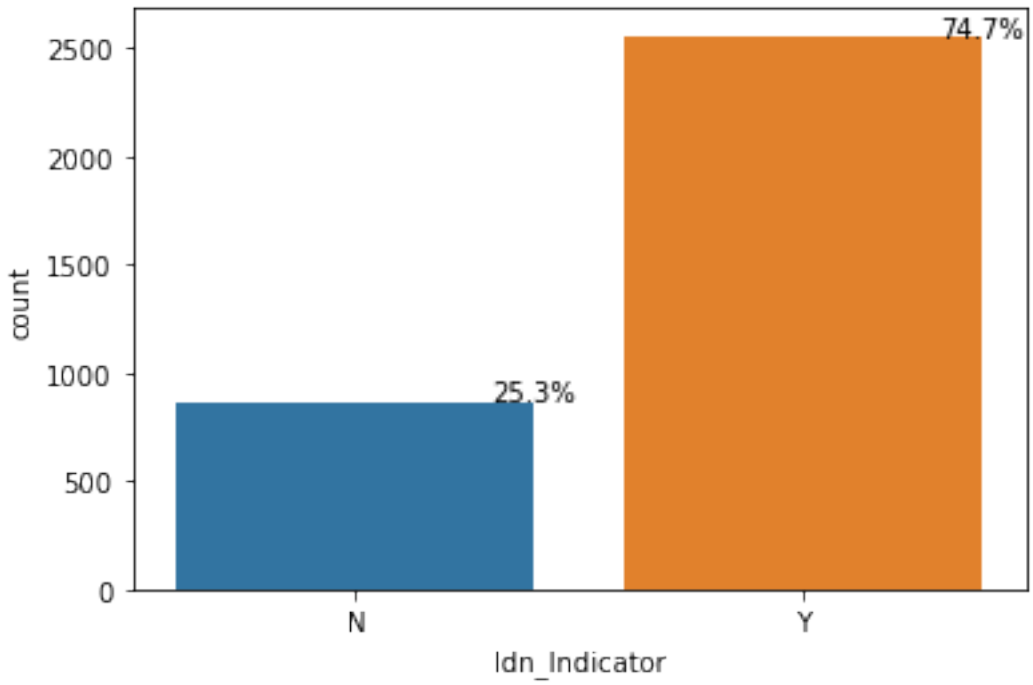
1.3 Exploratory data Analysis

Let's take a deep dive into our dataframe and try to formulate a few hypotheses that we will be testing using exploratory data analysis. - Patients mapped to IDNs belong to a particular race or ethnicity - Patients mapped to IDNs get more dexa scans - Patients of a certain age group have lower T-scores prior to any prescriptions - A certain group of patients are more prone to fragility fractures - Drug persistency have an effect on a patients T-score - Patients on concomitant drugs have a lower rate of drug percistency - Patients with Cormobiditant illness have lower bone density and lower rates of drugs persistency - Patients with injectable drugs have a lower rate of drug persistency

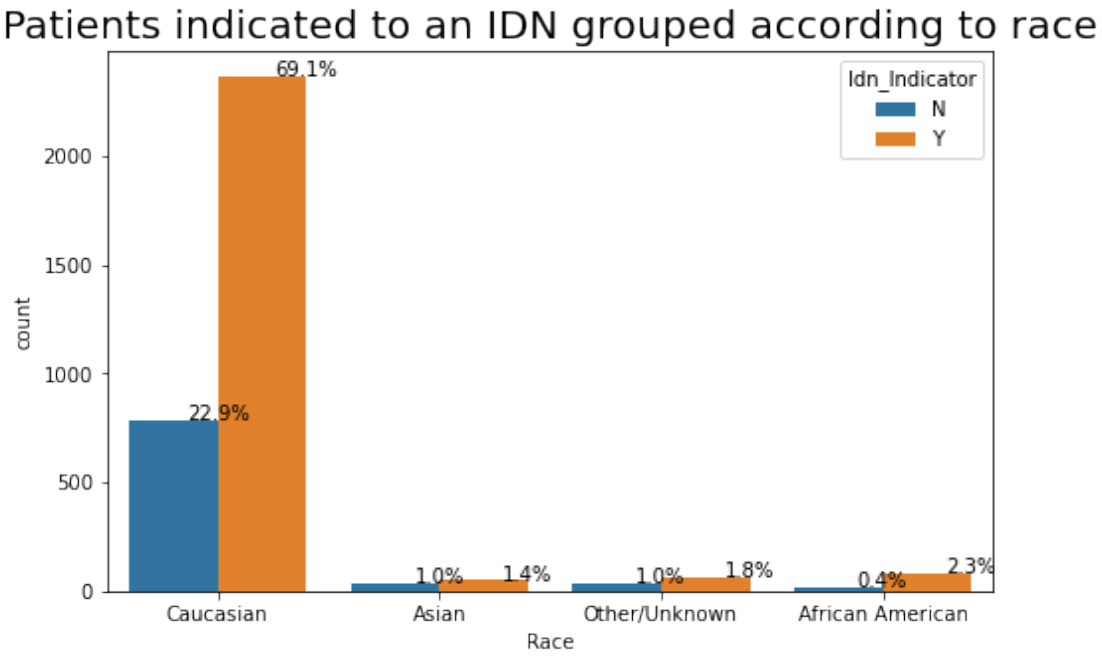
1.3.1 Patients' analysis

Background information

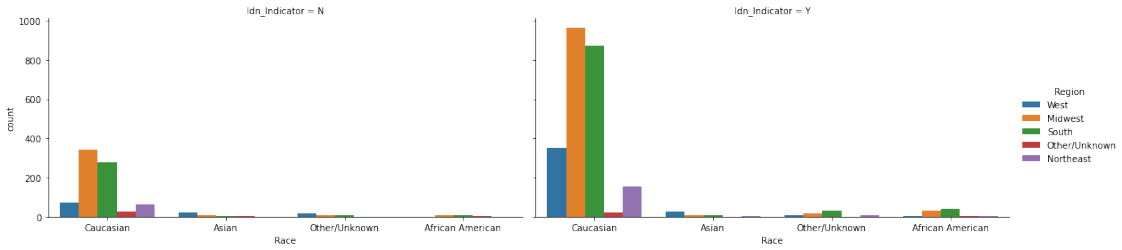
[28]:
```
total = float(len(df))
ax = sns.countplot(x= 'Idn_Indicator', data = df)
for p in ax.patches:
    percentage = '{:.1f}%'.format(100 * p.get_height()/total)
    x = p.get_x() + p.get_width()
    y = p.get_height()
    ax.annotate(percentage, (x, y),ha='center')
plt.show()
```



[29]:
```
plt.figure(figsize=(8,5))
total = float(len(df))
ax = sns.countplot(x="Race", hue="Idn_Indicator", data=df)
plt.title('Patients indicated to an IDN grouped according to race', fontsize=20)
for p in ax.patches:
    percentage = '{:.1f}%'.format(100 * p.get_height()/total)
    x = p.get_x() + p.get_width()
    y = p.get_height()
    ax.annotate(percentage, (x, y),ha='center')
plt.show()
```



[30]:
```
sns.catplot(x="Race", hue="Region", col="Idn_Indicator",
                data=df, kind="count",
                height=4, aspect=2);
```



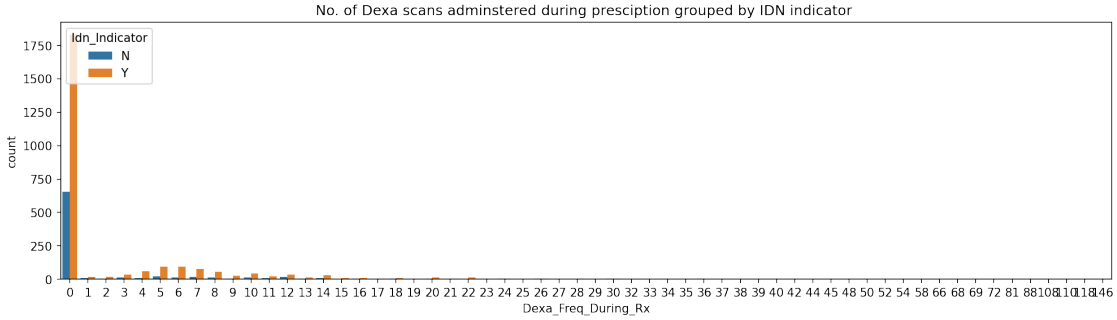
Findings

- 75% of the patients belong to IDNs. This may be influenced by the fact that with IDNs, one gets a full range of quality(and sometimes more affordable) healthcare without having to visit other facilities
- More patients are mapped to IDNs across each race, with caucasian patients being the vast majority of study subjects within our dataset
- Majority of patients mapped to IDNs belonging to the West, Midwest and South regions. This may be attributed by most of our patients coming from these regions

Who gets more Dexa scans?

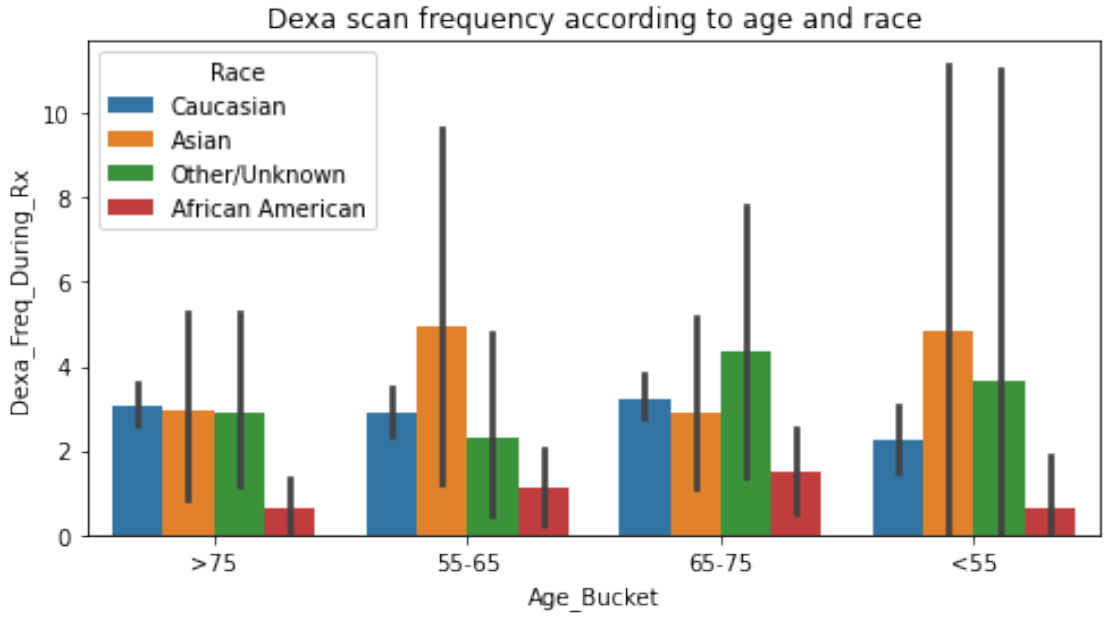
[31]:
```
plt.figure(figsize=(16,4),dpi=150)
plt.title('No. of Dexa scans adminstered during presciption grouped by IDN␣
 ↪indicator ')
sns.countplot(x= 'Dexa_Freq_During_Rx', hue = 'Idn_Indicator',  data = df )
```

[31]:
```
<AxesSubplot:title={'center':'No. of Dexa scans adminstered during presciption
grouped by IDN indicator '}, xlabel='Dexa_Freq_During_Rx', ylabel='count'>
```



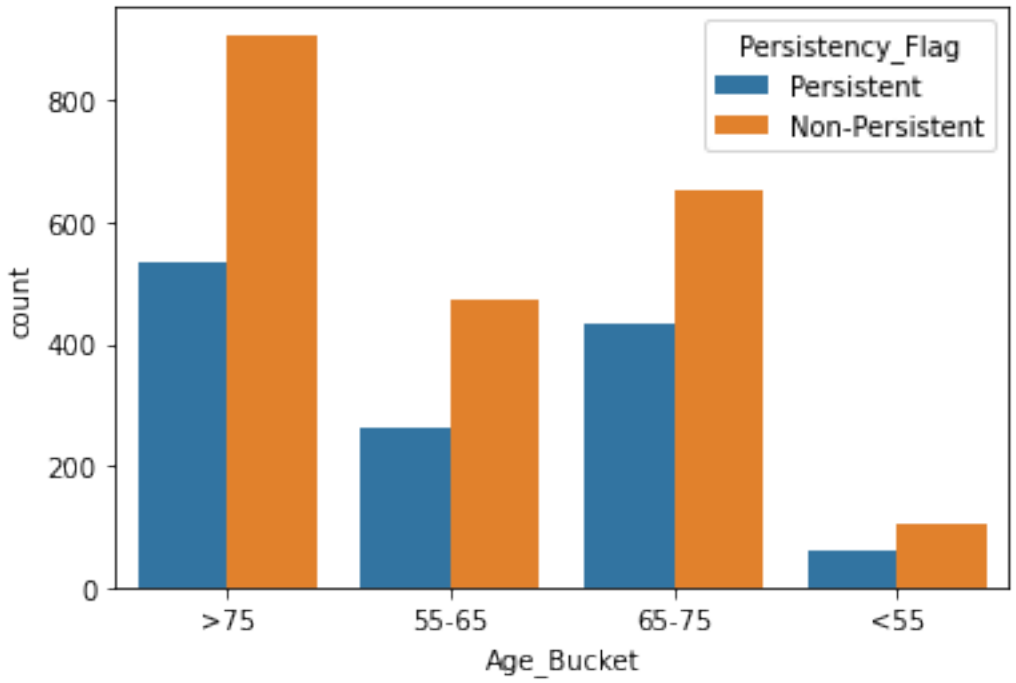
[32]:
```
plt.figure(figsize=(8,4))
plt.title('Dexa scan frequency according to age and race')
sns.barplot(x = 'Age_Bucket', y= 'Dexa_Freq_During_Rx', hue= 'Race', data = df)
```

[32]:
```
<AxesSubplot:title={'center':'Dexa scan frequency according to age and race'},
xlabel='Age_Bucket', ylabel='Dexa_Freq_During_Rx'>
```



[34]:
```
sns.countplot(x="Age_Bucket", hue="Persistency_Flag", data=df)
```

[34]:
```
<AxesSubplot:xlabel='Age_Bucket', ylabel='count'>
```

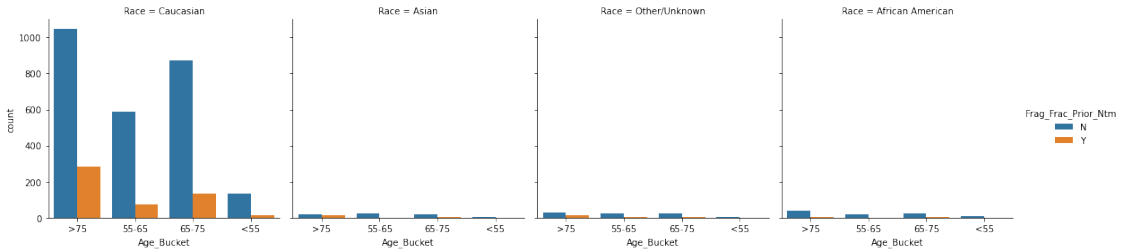


Findings

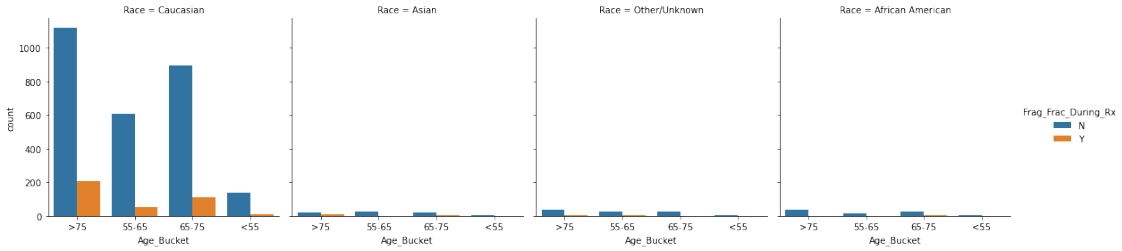
- Most patients did not have Dexa scans admistered to them during their prescriptions
- Those that did get dexa scans vastly got them from IDNs
- Older patients are more likely to be non-persistent with their medication

Fragility fractures At what age do most fragility fractures occur?

[35]:
```
sns.catplot(x="Age_Bucket", hue="Frag_Frac_Prior_Ntm", col="Race",
                data=df, kind="count",
                height=4, aspect=1);
```

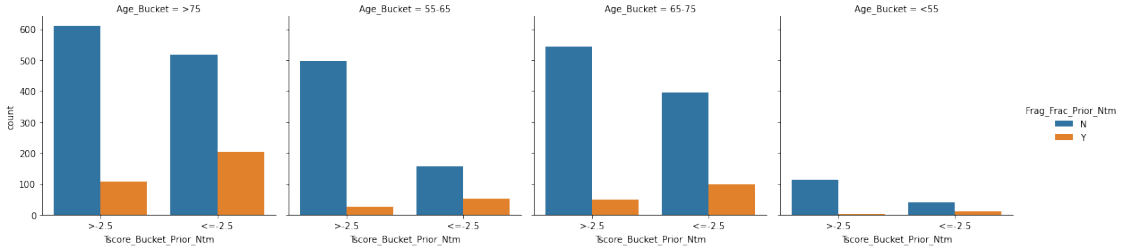


[36]:
```
sns.catplot(x="Age_Bucket", hue="Frag_Frac_During_Rx", col="Race",
                data=df, kind="count",
                height=4, aspect=1);
```

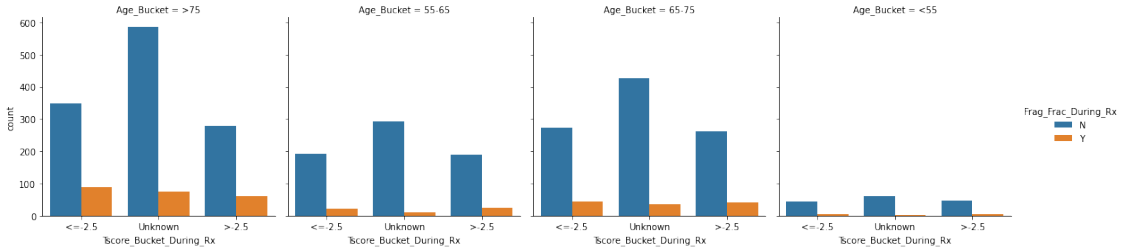


Is a patients T-score directly related to whether or not they are at risk of getting fragility fractures?

[38]:
```
sns.catplot(x="Tscore_Bucket_Prior_Ntm", hue="Frag_Frac_Prior_Ntm",␣
↪col="Age_Bucket",
                data=df, kind="count",
                height=4, aspect=1);
```



[39]:
```
sns.catplot(x="Tscore_Bucket_During_Rx", hue="Frag_Frac_During_Rx",␣
↪col="Age_Bucket",
                data=df, kind="count",
                height=4, aspect=1);
```



Patients with cormobid conditions Cormobidity is the presence of two or more medical conditions. Prescription of various kinds of treatments is required for these patients. Are these patiemts more likely to be non-persistent with their medication? Does Cormobidity have an effect on a patient's T-score?

[40]:
```
# creating a dataframe with the columns we want to use
PatientInfo = df[["Age_encoded", "Persistency_Flagfreq_enc",␣
 ↪"Tscore_Bucket_Prior_Ntm_encoded"]]
Cormobid = df.iloc[:, 94:108]
PatientCormobidity = [PatientInfo, Cormobid]
Cormobid_df = pd.concat(PatientCormobidity, axis=1, join="inner")
Cormobid_df.head(5)
```

[40]:

	Age_encoded	Persistency_Flagfreq_enc	Tscore_Bucket_Prior_Ntm_encoded	\
0	3.0	0.37646	1.0	
1	0.0	0.62354	1.0	
2	1.0	0.62354	0.0	
3	3.0	0.62354	1.0	
4	3.0	0.62354	0.0	

	Comorb_Encounter_For_Screening_For_Malignant_Neoplasmsfreq_enc	\
0	0.552278	
1	0.552278	
2	0.447722	
3	0.552278	
4	0.447722	

	Comorb_Encounter_For_Immunizationfreq_enc	\
0	0.441881	
1	0.558119	
2	0.558119	
3	0.441881	
4	0.441881	

	Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dxfreq_enc	\
0	0.39486	
1	0.39486	
2	0.39486	
3	0.39486	
4	0.39486	

	Comorb_Vitamin_D_Deficiencyfreq_enc	\
0	0.680783	
1	0.680783	
2	0.680783	
3	0.680783	
4	0.680783	

```
   Comorb_Other_Joint_Disorder_Not_Elsewhere_Classifiedfreq_enc  \
0                                           0.708236
1                                           0.708236
2                                           0.708236
3                                           0.291764
4                                           0.708236

   Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dxfreq_enc  \
0                                           0.231016
1                                           0.768984
2                                           0.768984
3                                           0.768984
4                                           0.768984

   Comorb_Long_Term_Current_Drug_Therapyfreq_enc  Comorb_Dorsalgiafreq_enc  \
0                                        0.76139                  0.227512
1                                        0.76139                  0.772488
2                                        0.76139                  0.772488
3                                        0.76139                  0.227512
4                                        0.76139                  0.227512

   Comorb_Personal_History_Of_Other_Diseases_And_Conditionsfreq_enc  \
0                                           0.197722
1                                           0.802278
2                                           0.802278
3                                           0.802278
4                                           0.197722

   Comorb_Other_Disorders_Of_Bone_Density_And_Structurefreq_enc  \
0                                           0.848715
1                                           0.848715
2                                           0.848715
3                                           0.848715
4                                           0.848715

   Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemiasfreq_enc  \
0                                           0.484521
1                                           0.484521
2                                           0.484521
3                                           0.515479
4                                           0.484521

   Comorb_Osteoporosis_without_current_pathological_fracturefreq_enc  \
0                                           0.732185
1                                           0.732185
2                                           0.732185
```

```
3                                                0.732185
4                                                0.732185

   Comorb_Personal_history_of_malignant_neoplasmfreq_enc  \
0                                                0.810456
1                                                0.810456
2                                                0.810456
3                                                0.810456
4                                                0.810456

   Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0                                         0.816005
1                                         0.816005
2                                         0.816005
3                                         0.183995
4                                         0.816005
```

[41]:
```
Cormobid_df.corr()
```

[41]:
```
                                                    Age_encoded  \
Age_encoded                                            1.000000
Persistency_Flagfreq_enc                               0.001976
Tscore_Bucket_Prior_Ntm_encoded                       -0.139659
Comorb_Encounter_For_Screening_For_Malignant_Ne…       0.153437
Comorb_Encounter_For_Immunizationfreq_enc             -0.074153
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…       0.031601
Comorb_Vitamin_D_Deficiencyfreq_enc                    0.019368
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…      -0.013027
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…       0.118082
Comorb_Long_Term_Current_Drug_Therapyfreq_enc         -0.036832
Comorb_Dorsalgiafreq_enc                              -0.061233
Comorb_Personal_History_Of_Other_Diseases_And_C…      -0.029494
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…       0.055870
Comorb_Disorders_of_lipoprotein_metabolism_and_…       0.132569
Comorb_Osteoporosis_without_current_pathologica…       0.042129
Comorb_Personal_history_of_malignant_neoplasmfr…      -0.025022
Comorb_Gastro_esophageal_reflux_diseasefreq_enc       -0.051770

                                                    Persistency_Flagfreq_enc  \
Age_encoded                                                         0.001976
Persistency_Flagfreq_enc                                            1.000000
Tscore_Bucket_Prior_Ntm_encoded                                    -0.018903
Comorb_Encounter_For_Screening_For_Malignant_Ne…                    0.322320
Comorb_Encounter_For_Immunizationfreq_enc                           0.314887
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…                    0.289828
Comorb_Vitamin_D_Deficiencyfreq_enc                                 0.172664
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…                    0.233279
```

```
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…                  0.213413
Comorb_Long_Term_Current_Drug_Therapyfreq_enc                      0.352760
Comorb_Dorsalgiafreq_enc                                           0.215307
Comorb_Personal_History_Of_Other_Diseases_And_C…                 0.219665
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…                 0.247283
Comorb_Disorders_of_lipoprotein_metabolism_and_…                -0.163495
Comorb_Osteoporosis_without_current_pathologica…                 0.139920
Comorb_Personal_history_of_malignant_neoplasmfr…                 0.174835
Comorb_Gastro_esophageal_reflux_diseasefreq_enc                    0.220644

Tscore_Bucket_Prior_Ntm_encoded  \
Age_encoded
-0.139659
Persistency_Flagfreq_enc
-0.018903
Tscore_Bucket_Prior_Ntm_encoded
1.000000
Comorb_Encounter_For_Screening_For_Malignant_Ne…
-0.049224
Comorb_Encounter_For_Immunizationfreq_enc
0.017948
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
-0.018860
Comorb_Vitamin_D_Deficiencyfreq_enc
0.045292
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.013277
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
-0.029792
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
-0.029717
Comorb_Dorsalgiafreq_enc
0.047666
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.044071
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
-0.202228
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.006727
Comorb_Osteoporosis_without_current_pathologica…
-0.168502
Comorb_Personal_history_of_malignant_neoplasmfr…
-0.057491
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
-0.006128

Comorb_Encounter_For_Screening_For_Malignant_Neoplasmsfreq_enc  \
```

```
Age_encoded
0.153437
Persistency_Flagfreq_enc
0.322320
Tscore_Bucket_Prior_Ntm_encoded
-0.049224
Comorb_Encounter_For_Screening_For_Malignant_Ne…
1.000000
Comorb_Encounter_For_Immunizationfreq_enc
0.227780
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.295191
Comorb_Vitamin_D_Deficiencyfreq_enc
0.072620
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.136603
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.279885
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.146354
Comorb_Dorsalgiafreq_enc
0.137612
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.147308
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.144372
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.158386
Comorb_Osteoporosis_without_current_pathologica…
0.092100
Comorb_Personal_history_of_malignant_neoplasmfr…
0.113031
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.151479

Comorb_Encounter_For_Immunizationfreq_enc  \
Age_encoded
-0.074153
Persistency_Flagfreq_enc
0.314887
Tscore_Bucket_Prior_Ntm_encoded
0.017948
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.227780
Comorb_Encounter_For_Immunizationfreq_enc
1.000000
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
```

```
0.348389
Comorb_Vitamin_D_Deficiencyfreq_enc
0.100954
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.196080
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.144376
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.202801
Comorb_Dorsalgiafreq_enc
0.175037
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.150386
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.141320
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.235447
Comorb_Osteoporosis_without_current_pathologica…
0.092693
Comorb_Personal_history_of_malignant_neoplasmfr…
0.111365
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.161815

Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dxfreq_enc  \
Age_encoded
0.031601
Persistency_Flagfreq_enc
0.289828
Tscore_Bucket_Prior_Ntm_encoded
-0.018860
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.295191
Comorb_Encounter_For_Immunizationfreq_enc
0.348389
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
1.000000
Comorb_Vitamin_D_Deficiencyfreq_enc
0.050520
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.113738
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.205070
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.112700
Comorb_Dorsalgiafreq_enc
0.127417
```

```
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.139031
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.129161
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.208106
Comorb_Osteoporosis_without_current_pathologica…
0.071393
Comorb_Personal_history_of_malignant_neoplasmfr…
0.127646
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.126806

Comorb_Vitamin_D_Deficiencyfreq_enc  \
Age_encoded
0.019368
Persistency_Flagfreq_enc
0.172664
Tscore_Bucket_Prior_Ntm_encoded
0.045292
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.072620
Comorb_Encounter_For_Immunizationfreq_enc
0.100954
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.050520
Comorb_Vitamin_D_Deficiencyfreq_enc
1.000000
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.113153
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.045334
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.116411
Comorb_Dorsalgiafreq_enc
0.070730
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.086342
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.027354
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.148653
Comorb_Osteoporosis_without_current_pathologica…
0.040007
Comorb_Personal_history_of_malignant_neoplasmfr…
0.012513
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
```

```
0.062884

Comorb_Other_Joint_Disorder_Not_Elsewhere_Classifiedfreq_enc  \
Age_encoded
-0.013027
Persistency_Flagfreq_enc
0.233279
Tscore_Bucket_Prior_Ntm_encoded
0.013277
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.136603
Comorb_Encounter_For_Immunizationfreq_enc
0.196080
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.113738
Comorb_Vitamin_D_Deficiencyfreq_enc
0.113153
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
1.000000
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.128372
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.196903
Comorb_Dorsalgiafreq_enc
0.273890
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.155629
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.096583
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.110607
Comorb_Osteoporosis_without_current_pathologica…
0.142843
Comorb_Personal_history_of_malignant_neoplasmfr…
0.096133
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.154516

Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dxfreq_enc  \
Age_encoded
0.118082
Persistency_Flagfreq_enc
0.213413
Tscore_Bucket_Prior_Ntm_encoded
-0.029792
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.279885
```

```
Comorb_Encounter_For_Immunizationfreq_enc
0.144376
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.205070
Comorb_Vitamin_D_Deficiencyfreq_enc
0.045334
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.128372
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
1.000000
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.135355
Comorb_Dorsalgiafreq_enc
0.107500
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.168067
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.122473
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.089092
Comorb_Osteoporosis_without_current_pathologica…
0.086311
Comorb_Personal_history_of_malignant_neoplasmfr…
0.107968
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.106331

Comorb_Long_Term_Current_Drug_Therapyfreq_enc  \
Age_encoded
-0.036832
Persistency_Flagfreq_enc
0.352760
Tscore_Bucket_Prior_Ntm_encoded
-0.029717
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.146354
Comorb_Encounter_For_Immunizationfreq_enc
0.202801
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.112700
Comorb_Vitamin_D_Deficiencyfreq_enc
0.116411
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.196903
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.135355
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
```

```
1.000000
Comorb_Dorsalgiafreq_enc
0.204507
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.236488
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.214935
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.138277
Comorb_Osteoporosis_without_current_pathologica…
0.155037
Comorb_Personal_history_of_malignant_neoplasmfr…
0.173323
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.236386
```

	Comorb_Dorsalgiafreq_enc
Age_encoded	-0.061233
Persistency_Flagfreq_enc	0.215307
Tscore_Bucket_Prior_Ntm_encoded	0.047666
Comorb_Encounter_For_Screening_For_Malignant_Ne…	0.137612
Comorb_Encounter_For_Immunizationfreq_enc	0.175037
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…	0.127417
Comorb_Vitamin_D_Deficiencyfreq_enc	0.070730
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…	0.273890
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…	0.107500
Comorb_Long_Term_Current_Drug_Therapyfreq_enc	0.204507
Comorb_Dorsalgiafreq_enc	1.000000
Comorb_Personal_History_Of_Other_Diseases_And_C…	0.209854
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…	0.064448
Comorb_Disorders_of_lipoprotein_metabolism_and_…	-0.113528
Comorb_Osteoporosis_without_current_pathologica…	0.090259
Comorb_Personal_history_of_malignant_neoplasmfr…	0.103705
Comorb_Gastro_esophageal_reflux_diseasefreq_enc	0.189981

	Comorb_Personal_History_Of_Other_Diseases_And_Conditionsfreq_enc
Age_encoded	-0.029494
Persistency_Flagfreq_enc	0.219665
Tscore_Bucket_Prior_Ntm_encoded	0.044071
Comorb_Encounter_For_Screening_For_Malignant_Ne…	0.147308
Comorb_Encounter_For_Immunizationfreq_enc	0.150386
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…	

```
0.139031
Comorb_Vitamin_D_Deficiencyfreq_enc
0.086342
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.155629
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.168067
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.236488
Comorb_Dorsalgiafreq_enc
0.209854
Comorb_Personal_History_Of_Other_Diseases_And_C…
1.000000
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.107601
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.121815
Comorb_Osteoporosis_without_current_pathologica…
0.100498
Comorb_Personal_history_of_malignant_neoplasmfr…
0.137847
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.174930

Comorb_Other_Disorders_Of_Bone_Density_And_Structurefreq_enc  \
Age_encoded
0.055870
Persistency_Flagfreq_enc
0.247283
Tscore_Bucket_Prior_Ntm_encoded
-0.202228
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.144372
Comorb_Encounter_For_Immunizationfreq_enc
0.141320
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.129161
Comorb_Vitamin_D_Deficiencyfreq_enc
0.027354
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.096583
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.122473
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.214935
Comorb_Dorsalgiafreq_enc
0.064448
```

```
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.107601
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
1.000000
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.097824
Comorb_Osteoporosis_without_current_pathologica…
0.480896
Comorb_Personal_history_of_malignant_neoplasmfr…
0.143106
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.125557

Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemiasfreq_enc  \
Age_encoded
0.132569
Persistency_Flagfreq_enc
-0.163495
Tscore_Bucket_Prior_Ntm_encoded
-0.006727
Comorb_Encounter_For_Screening_For_Malignant_Ne…
-0.158386
Comorb_Encounter_For_Immunizationfreq_enc
-0.235447
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
-0.208106
Comorb_Vitamin_D_Deficiencyfreq_enc
-0.148653
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
-0.110607
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
-0.089092
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
-0.138277
Comorb_Dorsalgiafreq_enc
-0.113528
Comorb_Personal_History_Of_Other_Diseases_And_C…
-0.121815
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
-0.097824
Comorb_Disorders_of_lipoprotein_metabolism_and_…
1.000000
Comorb_Osteoporosis_without_current_pathologica…
-0.108619
Comorb_Personal_history_of_malignant_neoplasmfr…
-0.082683
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
```

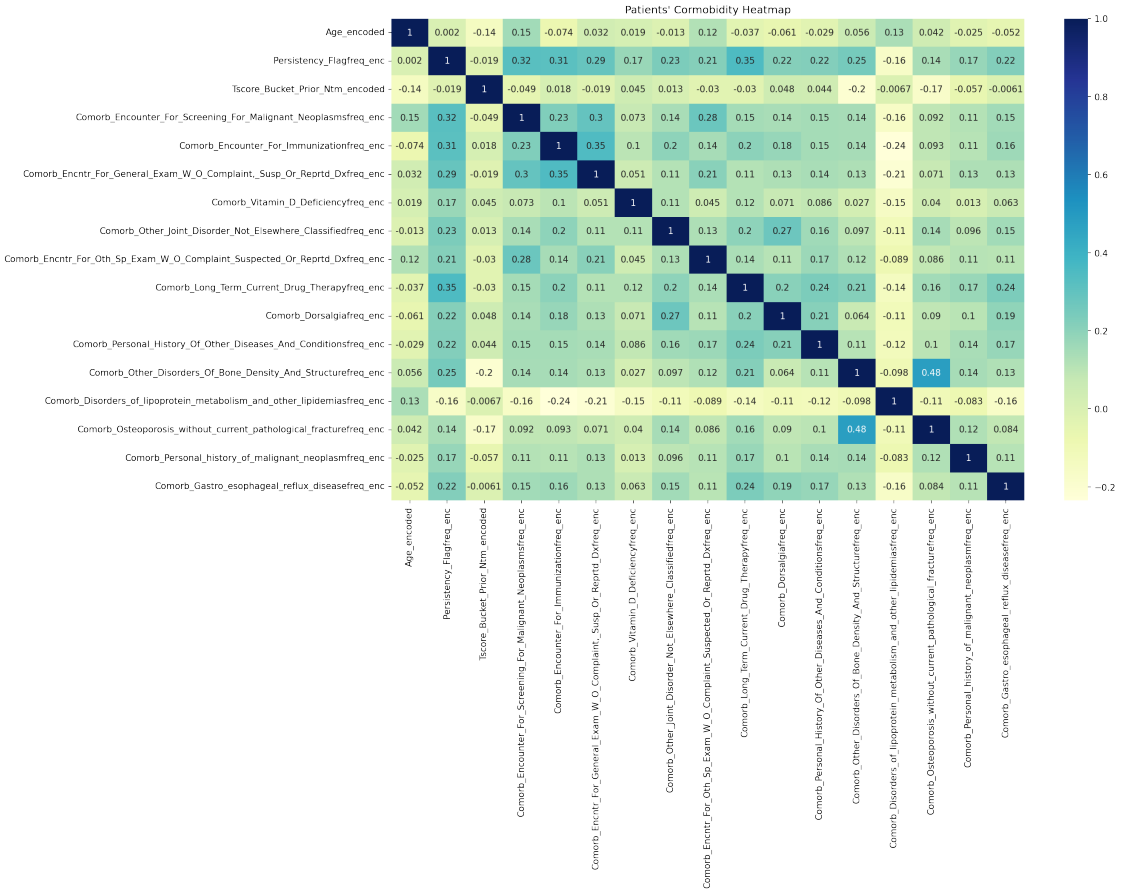
```
-0.164767

Comorb_Osteoporosis_without_current_pathological_fracturefreq_enc  \
Age_encoded
0.042129
Persistency_Flagfreq_enc
0.139920
Tscore_Bucket_Prior_Ntm_encoded
-0.168502
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.092100
Comorb_Encounter_For_Immunizationfreq_enc
0.092693
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.071393
Comorb_Vitamin_D_Deficiencyfreq_enc
0.040007
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.142843
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.086311
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.155037
Comorb_Dorsalgiafreq_enc
0.090259
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.100498
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.480896
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.108619
Comorb_Osteoporosis_without_current_pathologica…
1.000000
Comorb_Personal_history_of_malignant_neoplasmfr…
0.123156
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.083874

Comorb_Personal_history_of_malignant_neoplasmfreq_enc  \
Age_encoded
-0.025022
Persistency_Flagfreq_enc
0.174835
Tscore_Bucket_Prior_Ntm_encoded
-0.057491
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.113031
```

```
Comorb_Encounter_For_Immunizationfreq_enc
0.111365
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.127646
Comorb_Vitamin_D_Deficiencyfreq_enc
0.012513
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.096133
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.107968
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
0.173323
Comorb_Dorsalgiafreq_enc
0.103705
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.137847
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.143106
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.082683
Comorb_Osteoporosis_without_current_pathologica…
0.123156
Comorb_Personal_history_of_malignant_neoplasmfr…
1.000000
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
0.110744

Comorb_Gastro_esophageal_reflux_diseasefreq_enc
Age_encoded
-0.051770
Persistency_Flagfreq_enc
0.220644
Tscore_Bucket_Prior_Ntm_encoded
-0.006128
Comorb_Encounter_For_Screening_For_Malignant_Ne…
0.151479
Comorb_Encounter_For_Immunizationfreq_enc
0.161815
Comorb_Encntr_For_General_Exam_W_O_Complaint,_S…
0.126806
Comorb_Vitamin_D_Deficiencyfreq_enc
0.062884
Comorb_Other_Joint_Disorder_Not_Elsewhere_Class…
0.154516
Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus…
0.106331
Comorb_Long_Term_Current_Drug_Therapyfreq_enc
```

```
0.236386
Comorb_Dorsalgiafreq_enc
0.189981
Comorb_Personal_History_Of_Other_Diseases_And_C…
0.174930
Comorb_Other_Disorders_Of_Bone_Density_And_Stru…
0.125557
Comorb_Disorders_of_lipoprotein_metabolism_and_…
-0.164767
Comorb_Osteoporosis_without_current_pathologica…
0.083874
Comorb_Personal_history_of_malignant_neoplasmfr…
0.110744
Comorb_Gastro_esophageal_reflux_diseasefreq_enc
1.000000
```

Let's create a heatmap that'll better visualize correlation between our intended variable in a manner that is more easy on the eye

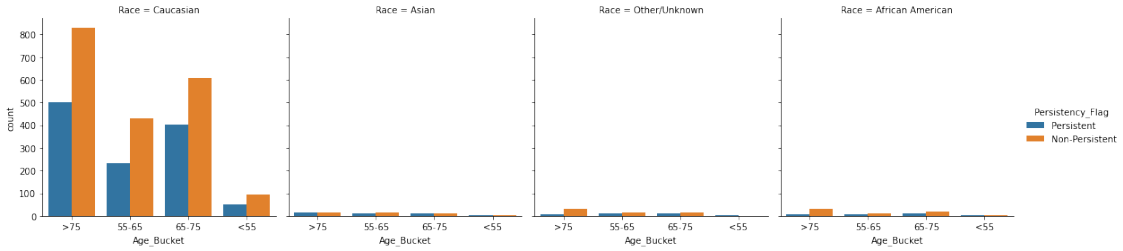
[42]:
```
plt.figure(figsize=(16,10),dpi=150)
plt.title("Patients' Cormobidity Heatmap")
dataplot1 = sns.heatmap(Cormobid_df.corr(), cmap="YlGnBu", annot=True)
```



Findings

- The Risk of a patient getting a fragility fracture increases with age
- Patients above 75 years of age are the most vulnerable to fragility fractures
- Patients with lower T-scores get more fragility fractures, this is probably attributed by their lower bone density
- Patients above 75 years of age with lower T-scores is our most vulnarable group
- There is some correlation between a patient's persistency flag and cormobidity with current long term therapy and cormobid encounters for screening for malignant neoplasms
- There is high correlation between Comorbidity of Osteoporosis without current pathological fracture and cormobidity with other disoders of bone density and structures. A patient diagnosed with osteoporsis is likely to have other disorders of bone density and stucture.

[43]:
```
sns.catplot(x="Age_Bucket", hue="Persistency_Flag", col="Race",
                data=df, kind="count",
                height=4, aspect=1);
```



From the above plot, it is difficult to tell which group is more likely to be non-persistent due to the dataset having a majority of caucasian people to begin with

1.3.2 Administered drugs

Here we shall be investigating whether the types of drugs adminstered have any effect on the patients' persistency #### Drug Concormity This is the prescription of two or more drugs give at the same time or almost the same time to a patient. Let's see if this has any effect on a patient's persitency

[44]:
```
# creating a dataframe with the specific columns we would like to use
PatientInfo = df[["Age_encoded", "Persistency_Flagfreq_enc",␣
 ↪"Tscore_Bucket_Prior_Ntm_encoded"]]
Concorm = df.iloc[:, 108:118]
PatientConcormity = [PatientInfo, Concorm]
Concorm_df = pd.concat(PatientConcormity, axis=1, join="inner")
Concorm_df.head(5)
```

[44]:

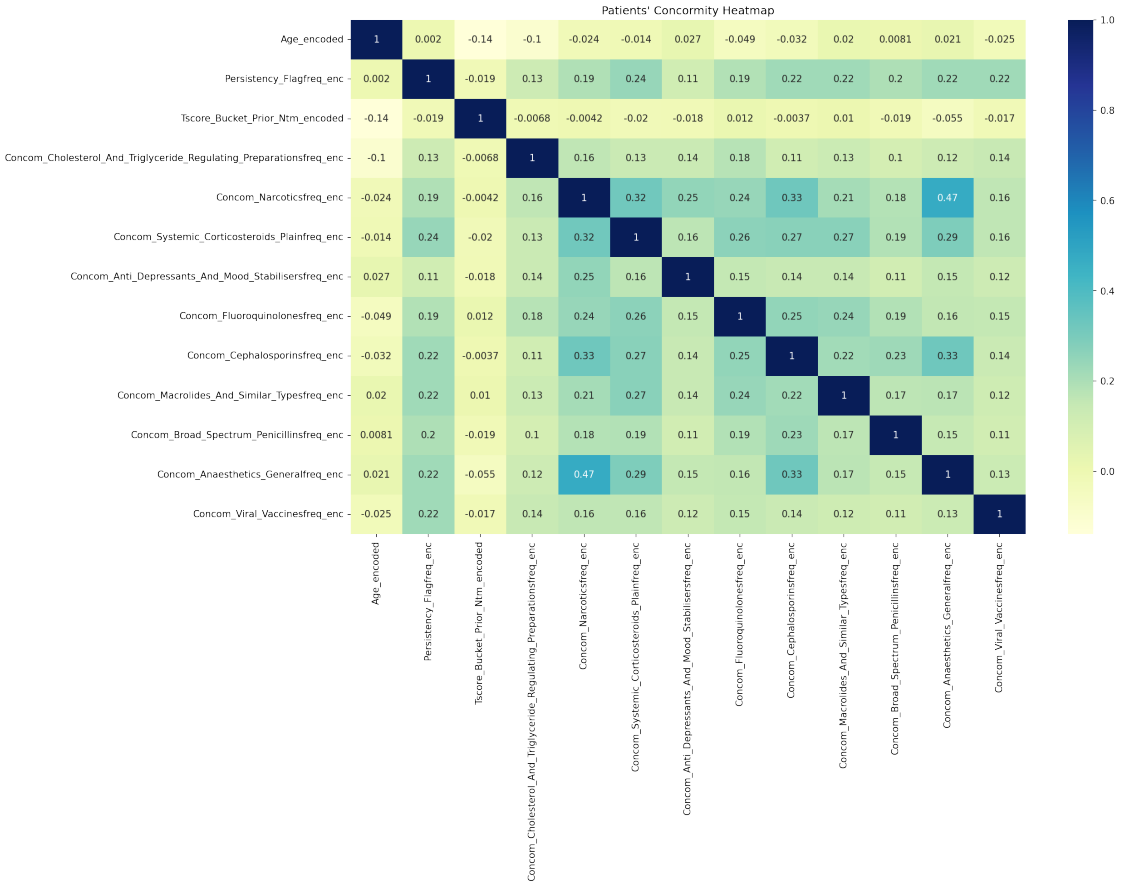
	Age_encoded	Persistency_Flagfreq_enc	Tscore_Bucket_Prior_Ntm_encoded	\
0	3.0	0.37646	1.0	
1	0.0	0.62354	1.0	
2	1.0	0.62354	0.0	
3	3.0	0.62354	1.0	
4	3.0	0.62354	0.0	

	Concom_Cholesterol_And_Triglyceride_Regulating_Preparationsfreq_enc	\
0	0.65479	
1	0.65479	
2	0.34521	
3	0.65479	
4	0.65479	

	Concom_Narcoticsfreq_enc	Concom_Systemic_Corticosteroids_Plainfreq_enc	\
0	0.639895	0.715829	
1	0.639895	0.715829	
2	0.639895	0.715829	
3	0.360105	0.284171	
4	0.360105	0.284171	

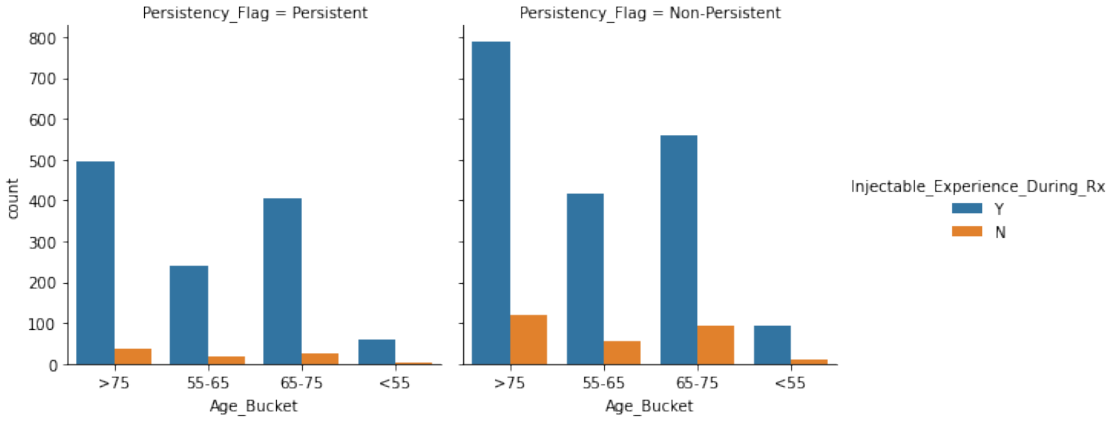
```
   Concom_Anti_Depressants_And_Mood_Stabilisersfreq_enc  \
0                                           0.719918
1                                           0.719918
2                                           0.719918
3                                           0.719918
4                                           0.280082

   Concom_Fluoroquinolonesfreq_enc  Concom_Cephalosporinsfreq_enc  \
0                          0.81396                        0.82389
1                          0.81396                        0.82389
2                          0.81396                        0.82389
3                          0.81396                        0.82389
4                          0.81396                        0.82389

   Concom_Macrolides_And_Similar_Typesfreq_enc  \
0                                     0.833236
1                                     0.833236
2                                     0.833236
3                                     0.833236
4                                     0.833236

   Concom_Broad_Spectrum_Penicillinsfreq_enc  \
0                                   0.871787
1                                   0.871787
2                                   0.871787
3                                   0.871787
4                                   0.871787

   Concom_Anaesthetics_Generalfreq_enc  Concom_Viral_Vaccinesfreq_enc
0                             0.854848                       0.896904
1                             0.854848                       0.896904
2                             0.854848                       0.896904
3                             0.854848                       0.103096
4                             0.854848                       0.896904
```

[45]:
```
plt.figure(figsize=(16,10),dpi=150)
plt.title("Patients' Concormity Heatmap")
dataplot2 = sns.heatmap(Concorm_df.corr(), cmap="YlGnBu", annot=True)
```



Let's take a look at patients who have Injectectable medication prescribed to them during the prescription.

[46]:
```
sns.catplot(x="Age_Bucket", hue="Injectable_Experience_During_Rx",␣
↪col="Persistency_Flag",
                data=df, kind="count",
                height=4, aspect=1);
```



Findings

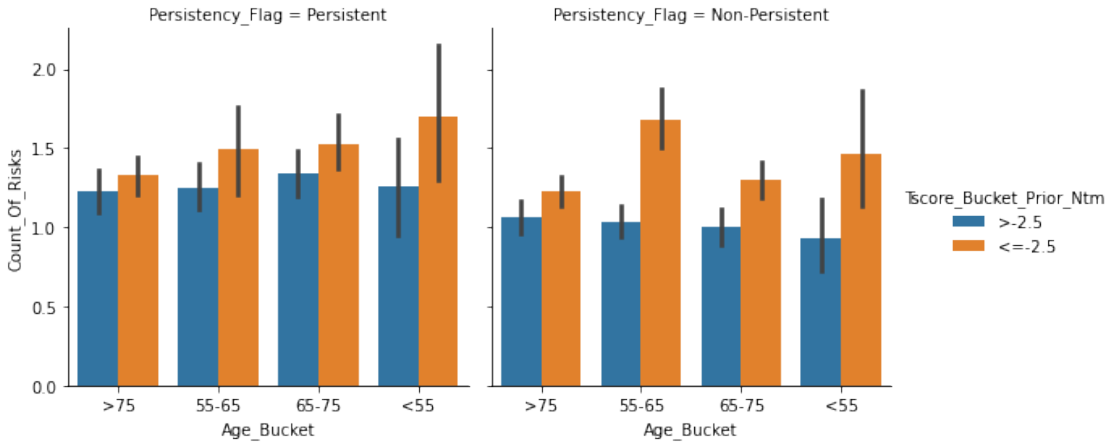
- A large number of peopke who have injectable medication in their prescription are not persistent with their treatments

1.3.3 Patient's Risk

In this section, we will investigate what may put patients at different kinds of risks

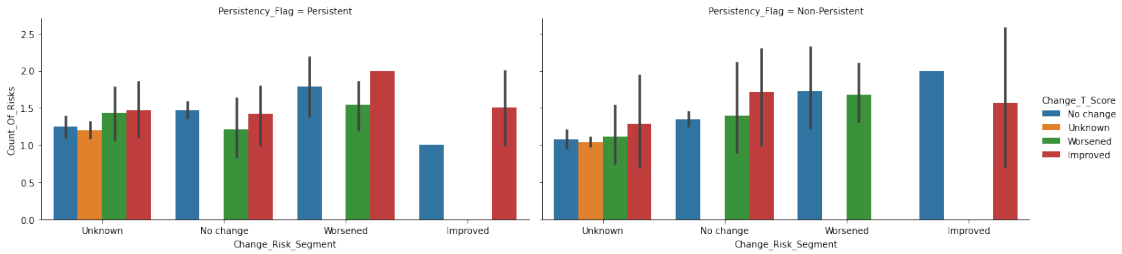
Does a patient's age, race or t-score affect their risk count?

[47]:
```
g = sns.catplot(x="Age_Bucket", y="Count_Of_Risks",
                hue="Tscore_Bucket_Prior_Ntm", col="Persistency_Flag",
                data=df, kind="bar",
                height=4, aspect=1);
```



Factors that could cause a possible change in a patients risk segment

[48]:
```
g = sns.catplot(x="Change_Risk_Segment", y="Count_Of_Risks",
                hue="Change_T_Score", col="Persistency_Flag",
                data=df, kind="bar",
                height=4, aspect=2);
```



[49]:
```
# creating a new dataframe with a count of risks column
Risk = df["Count_Of_Risks"]
Risk2 = [Cormobid_df, Risk]
Cormobid_Risk = pd.concat(Risk2, axis=1, join="inner")
Cormobid_Risk.head()
```

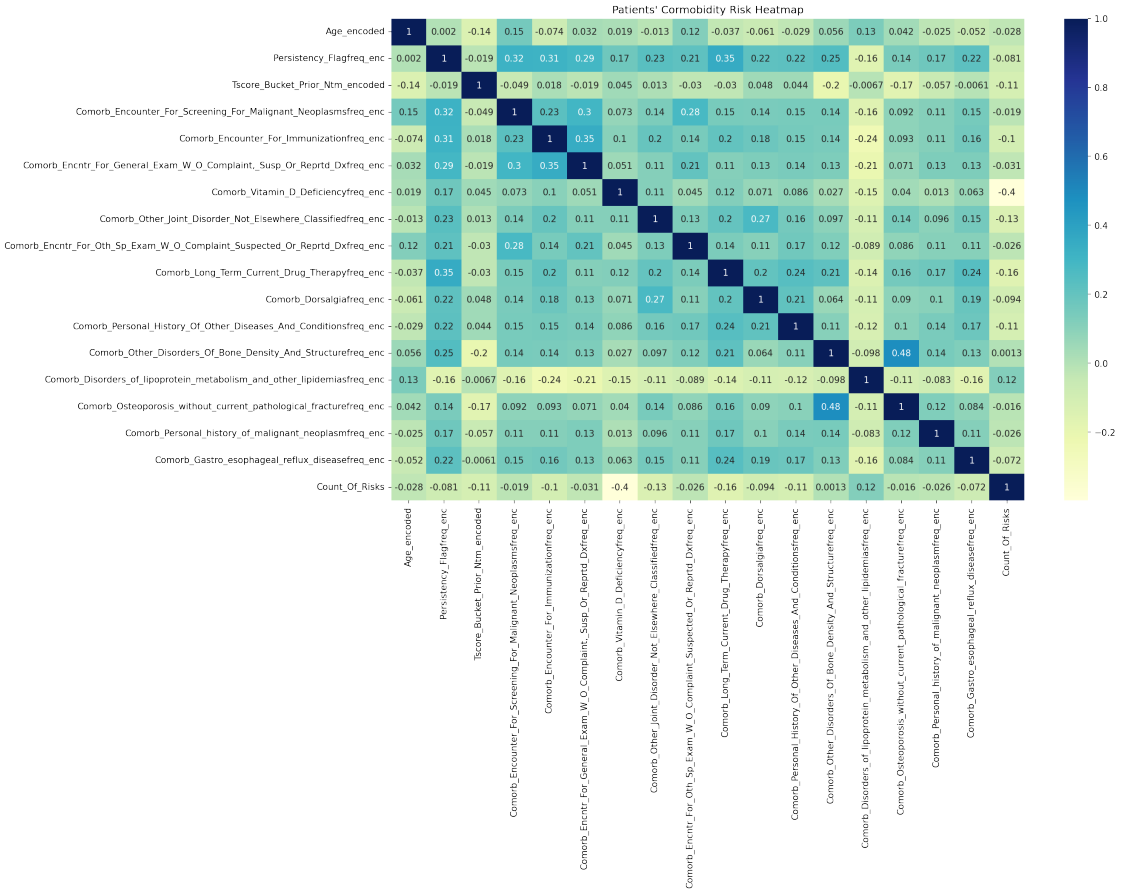
[49]:

	Age_encoded	Persistency_Flagfreq_enc	Tscore_Bucket_Prior_Ntm_encoded	Comorb_Encounter_For_Screening_For_Malignant_Neoplasmsfreq_enc	Comorb_Encounter_For_Immunizationfreq_enc	Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dxfreq_enc	Comorb_Vitamin_D_Deficiencyfreq_enc	Comorb_Other_Joint_Disorder_Not_Elsewhere_Classifiedfreq_enc
0	3.0	0.37646	1.0	0.552278	0.441881	0.39486	0.680783	0.708236
1	0.0	0.62354	1.0	0.552278	0.558119	0.39486	0.680783	0.708236
2	1.0	0.62354	0.0	0.447722	0.558119	0.39486	0.680783	
3	3.0	0.62354	1.0	0.552278	0.441881	0.39486	0.680783	
4	3.0	0.62354	0.0	0.447722	0.441881	0.39486	0.680783	

```
2                                                0.708236
3                                                0.291764
4                                                0.708236

   Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dxfreq_enc  \
0                                                0.231016
1                                                0.768984
2                                                0.768984
3                                                0.768984
4                                                0.768984

   Comorb_Long_Term_Current_Drug_Therapyfreq_enc  Comorb_Dorsalgiafreq_enc  \
0                                        0.76139                  0.227512
1                                        0.76139                  0.772488
2                                        0.76139                  0.772488
3                                        0.76139                  0.227512
4                                        0.76139                  0.227512

   Comorb_Personal_History_Of_Other_Diseases_And_Conditionsfreq_enc  \
0                                                0.197722
1                                                0.802278
2                                                0.802278
3                                                0.802278
4                                                0.197722

   Comorb_Other_Disorders_Of_Bone_Density_And_Structurefreq_enc  \
0                                                0.848715
1                                                0.848715
2                                                0.848715
3                                                0.848715
4                                                0.848715

   Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemiasfreq_enc  \
0                                                0.484521
1                                                0.484521
2                                                0.484521
3                                                0.515479
4                                                0.484521

   Comorb_Osteoporosis_without_current_pathological_fracturefreq_enc  \
0                                                0.732185
1                                                0.732185
2                                                0.732185
3                                                0.732185
4                                                0.732185

   Comorb_Personal_history_of_malignant_neoplasmfreq_enc  \
```

```
0                                                0.810456
1                                                0.810456
2                                                0.810456
3                                                0.810456
4                                                0.810456

   Comorb_Gastro_esophageal_reflux_diseasefreq_enc  Count_Of_Risks
0                                         0.816005             0.0
1                                         0.816005             0.0
2                                         0.816005             2.0
3                                         0.183995             1.0
4                                         0.816005             1.0
```

[50]:
```
plt.figure(figsize=(16,10),dpi=150)
plt.title("Patients' Cormobidity Risk Heatmap")
dataplot1 = sns.heatmap(Cormobid_Risk.corr(), cmap="YlGnBu", annot=True)
```



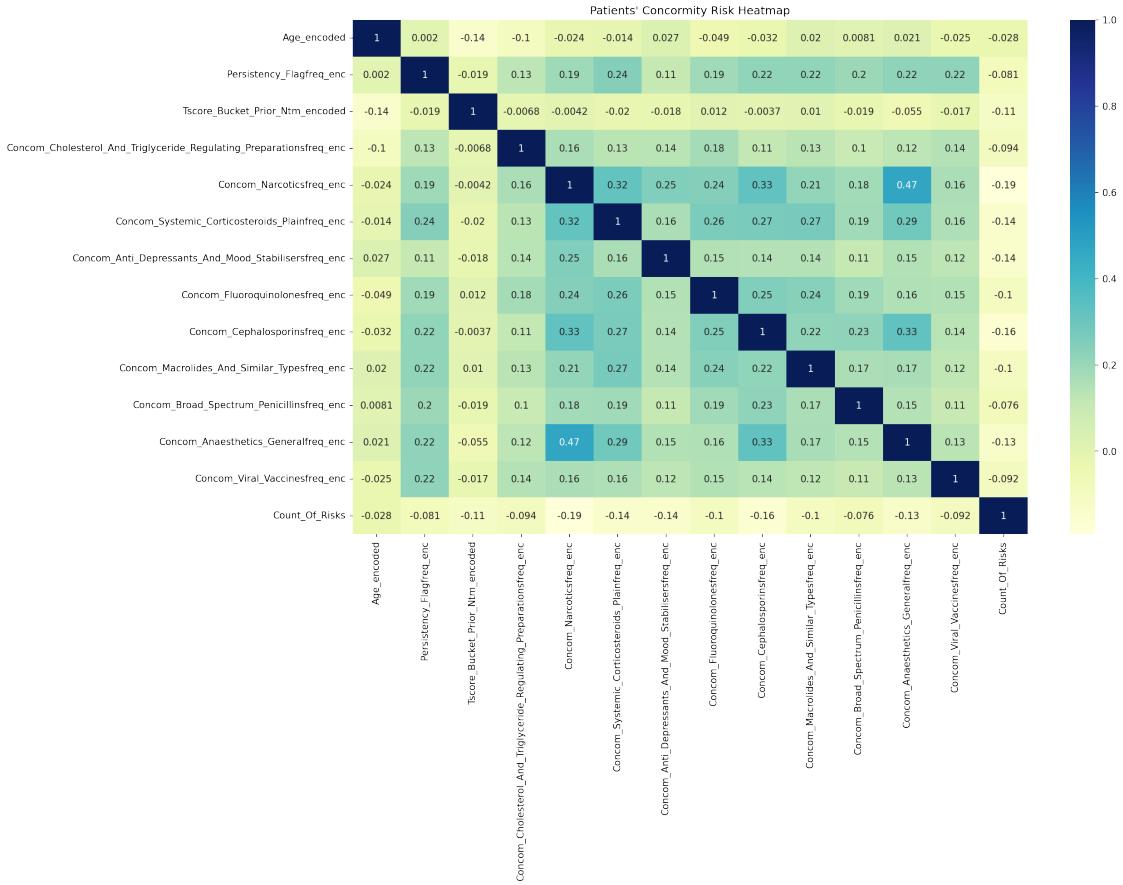
[51]:
```
Risk3 = [Concorm_df, Risk]
Concorm_Risk = pd.concat(Risk3, axis=1, join="inner")
Concorm_Risk.head()
```

[51]:

```
   Age_encoded  Persistency_Flagfreq_enc  Tscore_Bucket_Prior_Ntm_encoded  \
0          3.0                   0.37646                              1.0
1          0.0                   0.62354                              1.0
2          1.0                   0.62354                              0.0
3          3.0                   0.62354                              1.0
4          3.0                   0.62354                              0.0

   Concom_Cholesterol_And_Triglyceride_Regulating_Preparationsfreq_enc  \
0                                            0.65479
1                                            0.65479
2                                            0.34521
3                                            0.65479
4                                            0.65479

   Concom_Narcoticsfreq_enc  Concom_Systemic_Corticosteroids_Plainfreq_enc  \
0                  0.639895                                       0.715829
1                  0.639895                                       0.715829
2                  0.639895                                       0.715829
3                  0.360105                                       0.284171
4                  0.360105                                       0.284171

   Concom_Anti_Depressants_And_Mood_Stabilisersfreq_enc  \
0                                           0.719918
1                                           0.719918
2                                           0.719918
3                                           0.719918
4                                           0.280082

   Concom_Fluoroquinolonesfreq_enc  Concom_Cephalosporinsfreq_enc  \
0                          0.81396                        0.82389
1                          0.81396                        0.82389
2                          0.81396                        0.82389
3                          0.81396                        0.82389
4                          0.81396                        0.82389

   Concom_Macrolides_And_Similar_Typesfreq_enc  \
0                                     0.833236
1                                     0.833236
2                                     0.833236
3                                     0.833236
4                                     0.833236

   Concom_Broad_Spectrum_Penicillinsfreq_enc  \
```

```
0                                    0.871787
1                                    0.871787
2                                    0.871787
3                                    0.871787
4                                    0.871787

   Concom_Anaesthetics_Generalfreq_enc  Concom_Viral_Vaccinesfreq_enc  \
0                             0.854848                       0.896904
1                             0.854848                       0.896904
2                             0.854848                       0.896904
3                             0.854848                       0.103096
4                             0.854848                       0.896904

   Count_Of_Risks
0             0.0
1             0.0
2             2.0
3             1.0
4             1.0
```

[52]:
```
plt.figure(figsize=(16,10),dpi=150)
plt.title("Patients' Concormity Risk Heatmap")
dataplot1 = sns.heatmap(Concorm_Risk.corr(), cmap="YlGnBu", annot=True)
```



1.3.4 Findings

- Patients with lower Tscores are generally more at risk compared to those with higher Tscores
- Concormity and cormobidity generally have no effect on a patient's risk count

1.3.5 Recommendations

- Due to the fact that patients who belong to IDNs get more Dexa svans and are able to access a wide range of healthcare services in the same facility, we recommend patients seeking medical care from these facilities.
- Patients should be administered with dexa scans at least once and those with lower Tscores should be closely monitored as they at a high risk of getting fragility fractures. We recommend frequent check-ins with their healthcare provider.
- Patients with injectable medication should be given frequent reminders by their local pharmacists to take their medication as they have a high rate of non-persistency.
- Pharmacists should closely monitor their older patients as the rate of non-persistency is increasing with age. They should set remainders for them or keep in touch with a patient's loved one or care giver to ensure they are taking their medication