Team Member Details:

Group Name: Data Hacks

Name: Suhas Yogeshwara

Email: suhas.gys1996@gmail.com

Country: Germany

College: SRH University of Applied Sciences Berlin

Specialization: Data Science

Problem Description:

The pharmaceutical industry is currently having trouble keeping track of whether a prescription remains to be applied in practice advised by a physician. Classification must be required in order to automate the procedure in to address this problem.

Methods:

For NA values: as there are no Null values in the Dataset for the methods such as mean/median/model approach to be applied for.

For Outliers: For outliers to be shown in graphical format the numerical values should be either in the decimal or fraction. Main focus in onto Persistency_Flag so the Outliers can be avoided in such cases.

Ipynb Notebook:

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv('Healthcare_dataset.csv')
df.head()
```

Out[2]:

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region	Age_Bucket	Ntm_Speciality	Ntm_Specialist_Flag	Ntm_Speciality_Bucket	...	Risk_F
0	P1	Persistent	Male	Caucasian	Not Hispanic	West	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
1	P2	Non-Persistent	Male	Asian	Not Hispanic	West	55-65	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
2	P3	Non-Persistent	Female	Other/Unknown	Hispanic	Midwest	65-75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
3	P4	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
4	P5	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	

5 rows × 69 columns

In [4]:
```
df.shape
```

Out[4]:
```
(3424, 69)
```

In [6]:
```
df.isnull()
```

Out[6]:

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region	Age_Bucket	Ntm_Speciality	Ntm_Specialist_Flag	Ntm_Speciality_Bucket	...	Risk_Family_Histo
0	False	False	False	False	False	False	False	False	False	False	...	
1	False	False	False	False	False	False	False	False	False	False	...	
2	False	False	False	False	False	False	False	False	False	False	...	
3	False	False	False	False	False	False	False	False	False	False	...	
4	False	False	False	False	False	False	False	False	False	False	...	
...	...	...	...	...	...	...	...	...	...	...	...	...
3419	False	False	False	False	False	False	False	False	False	False	...	
3420	False	False	False	False	False	False	False	False	False	False	...	
3421	False	False	False	False	False	False	False	False	False	False	...	
3422	False	False	False	False	False	False	False	False	False	False	...	
3423	False	False	False	False	False	False	False	False	False	False	...	

3424 rows × 69 columns

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
Ptid                                0
Persistency_Flag                    0
Gender                              0
Race                                0
Ethnicity                           0
                                   ..
Risk_Hysterectomy_Oophorectomy      0
Risk_Estrogen_Deficiency            0
Risk_Immobilization                 0
Risk_Recurring_Falls                0
Count_Of_Risks                      0
Length: 69, dtype: int64
```