Drug persistency and Medical Adherence Code Review

Code written by: Nikola Andrejić
Code reviewed by: Wangu Ndungu
Link to the code: https://github.com/Wangu-ndungu/Drug-persistence-and-Medical-Adherence/blob/main/Drug%20persistency.ipynb
Dataset being analysed: https://github.com/Wangu-ndungu/Drug-persistence-and-Medical-Adherence/blob/main/Data/drug-persistency.csv

The code is a notebook that entails the problem statement, data cleansing and transformation of the dataset. They have began by evaluating for any missing data, of which there was none. They have the proceeded to detect any outliers in both numeric and categoric variables. They have used Log Transformation on the "Dexa_Freq_During_Rx" column to somewhat reduce skewness and then proceeded to chop the outliers on the transformed column. They Chopped the outliers on the "Count_Of_Risks" column.

For the categorical variables, they have used histograms to check for any imbalances then they chopped the outliers. For the binary variables, they dropped the entire column because their presence would not have affected the model.

The code has been well commented with extensive explanations in the markdown making it easy to follow. Numerous visualizations have also been used to supports their code. Their code also follows PEP8 guidelines.

The code also makes use of several functions that automate tasks such as getting outliers and plotting some graphs. This reduced redundancy and made the code much more flexible.

However, chopping variables may cause stronger correlation in our data which may affect our model. We have used different methods of dealing with outliers in order to get the best as they all have their drawbacks.