Group Name: PharmaPersist
Team Member Details:
- Name : Yunxin Gan
- Email: xin03040905@gmai.com
- Country: United Kingdom
- College: University of Exeter
- Specialization: Data Science

Github link: https://github.com/YunxinG107112/VC/upload/main/week7


Problem description: To gather insights on the factors impacting the persistency, build a classification model for the given dataset.
Data that will be used in this project involves:
- Target variable (the variable that is used to indicate if the patient was persistent or not, binary variable)
- Demographics (categorical variables)
- Provider attributes (categorical variables)
- Clinical, factors (categorical variables)
- Disease/Treatment factors (categorical variables)

Problems in the data: There are some unknown values in some columns
We can handle the unknown missing values in the categorical columns using different approaches, including imputation with the most frequency category.