Group Name: PharmaPersist
Team Member Details:

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

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)

Github link: https://github.com/YunxinG107112/VC/tree/main/week10

The EDA that was done:

- Data Summary and Distribution:
  - Count the number of unique values for categorical features like Race, Region, Ethnicity, Gender, and others.
  - Visualize the distribution of categorical features using bar charts.
- Target Variable Analysis
  - Visualize the distribution of the target variable using a bar chart or pie chart to understand the class balance
- Correlation Analysis
  - Visualize the correlations between numerical features and the target variable using a heatmap to Identify which numerical features have the highest correlation with persistency.
- Relationship analysis
  - Visualize the relationship between age, gender and race, dexa scan, risk segment, comorbidity, etc with persistency of the medicine.

Final recommendation:
Based on the exploratory data analysis (EDA) findings, several factors appear to influence drug persistency. Demographic factors such as age, gender, region, ethnicity, and clinical factors like Gluco_Record_During_Rx, Dexa_During_Rx, Frag_Frac_During_Rx, Change_T_Score, Adherent_Flag, and Idn_Indicator show strong associations with drug persistency. For example, Patients aged over 75 years tend to have a higher rate of drug persistency which suggests that age is a significant factor influencing drug adherence, patients with glucose records during treatment are more likely to be persistent suggesting that ensures regular monitoring of relevant clinical parameters could improve adherence.