Week 10

Group Name: Ignotus

Team Member Details:

- Corey Hamren, hamhams86@gmail.com, USA, Self-taught, Data Science
- Motamen MohammedAhmed, Motamen.salih@hotmail.com, UAE, Zayed University, Data Science
- Francis Kim, fkim39@gmail.com, Fkim39@gmail.com, USA, University of Maryland, Data Science
- Inna Soltsman-Groysman, innasol90@gmail.com, USA, Data Science

Problem description:

There are different diseases that affect people around the world. Pharmaceutical companies started to manufacture cures to defeat them. One of the challenges for these companies is to understand the persistence of drugs as per the physician prescription and the related factors. To solve this problem ABC pharma company approached an analytics company to automate this process of identification.

EDA performed:

Corey:
- Graphed numerical columns with histograms
- Graphed categorical columns with bar graphs
- Checked for null values
- Printed value counts for columns

Motamen:
- Changing the target variable to numerical using the label encoder.
- Changing all the binary and non binary categorical variables that have values other than Y and N into numerical using the ordinal encoder.
- Applying the WOE with Inna as an extra step for handling variables with so many values.

Francis:
- Checked for duplicate values
- Transformed binary categorical
- Pie plots for EDA

Inna:

- Correlation plot
- Pie plot for age feature

Model Recommendation:

Corey:

- Gradient Boosting

Motamen:

- Random Forest

Inna:

- Gradient Boosting Model

Francis:

- Gradient Boosting Model

Final Recommendation:

- Gradient Boosting