Chris Donaton, Data Science specialization

Problem description

Given an array of features that describe a patient's medical history, calculate that patient persistency likelihood.

EDA Methodology

First, I checked for correlation within the dataset. There are many mutually correlated fields that will have to be mitigated during the model building process.

Then I looked for the most important features in the dataset. Dexa treatments and comorbidity factors are by far the most predictive factors.

Finally, I assessed skewness and made the necessary transformations to the two numeric columns.

Github Repo: https://github.com/rambles-tech/DG_virtual_internship/week09