

Practice2B (Pima Datasets)

Dataset

Data used: Pima CSV

Preprocessing

Dropped 2 columns with high number of missing values (otherwise imputation can cause strong bias)

Outliers

Cast to NaN with IQR (for future imputation)

Missing Values

Were held with Scaling + Linear imputer (Used RobustScaler as there were plenty of strongly skewed features)

Visualizations

Different visualizations (boxplots, histplots, heatmap)

Result

Ready to proceed dataset with clean data