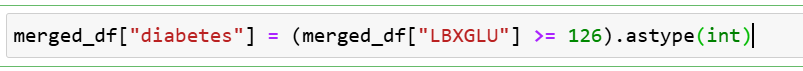
A screen shot of a computer code

AI-generated content may be incorrect.

Since each different dataframes has different number of columns with different numbers of sequence identification numbers. So, for example, merging df1 and df2 can result in a larger dataframe than mergin df1 and df3. So, in order not to lose any data entry, we developed an algorithm to merge the current dataframe with the most effective dataframe.



* Labelling data. **LBXGLU** (fasting glucose in mg/dL). This will give us a **binary label** for our machine learning model.

A graph of blue bars

AI-generated content may be incorrect.

As we can see, although we have done scaling, the features are still not scaling properly. We can scale them further by centering their mean value in 0.5. But we need to be careful since there are also columns which values are only certain numbers such as binary features. We can’t centered their mean to 0.5.