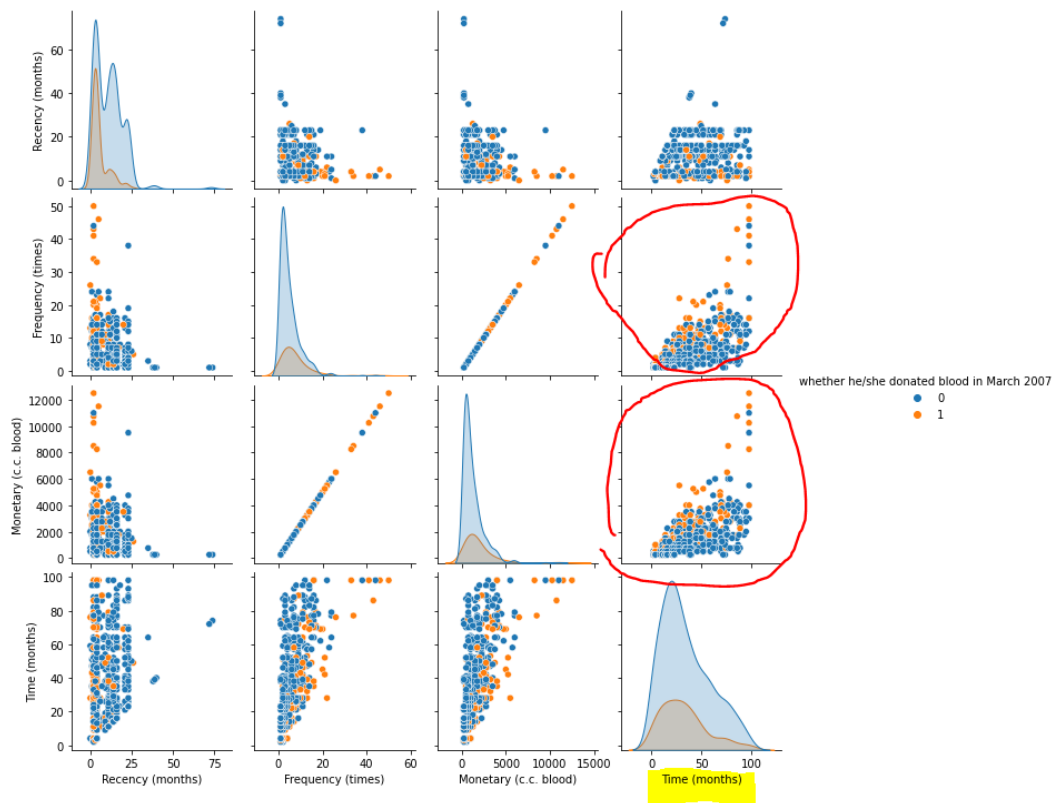


According to the documentation this is all we know about the features:

R (Recency - months since last donation),
 F (Frequency - total number of donation),
 M (Monetary - total blood donated in c.c.),
 T (Time - months since first donation), and
 a binary variable representing whether he/she donated blood in March 2007 (1 stand for donating blood; 0 stands for not donating blood)

As we can see in the count-plot our data is obviously imbalanced so, that would possibly make classification difficult. We have 570 records with 1 as their label and 178 labeled with 0. I plot a pair-plot in which we can see all of scatter plots with different features as their x and y axis. Expectedly, our data is not well classified in any of them. If we take a look at the Time(months) column in the pair-plot, we can observe that it's a bit better in comparison to the others.



In next section, I displayed our binary label (“whether he/she donated blood in March 2007”) in box-plots with different features as their y axes. We can’t actually make a strong statement with these box-plots, but we can say that probably Recency (month) would a better feature compare to others. Because it separates the boxes (not completely though). By separating, I mean that orange box and blue box have less overlap with Recency as y axis.

